Ditton, Elizabeth Rose (2022) ***Artificial Intelligence for autonomous persona generation to shape tailored communications and products and incentivise disaster preparation behaviours.*** **PhD Thesis, James Cook University.**

Artificial Intelligence for Autonomous Persona Generation to Shape Tailored Communications and Products and Incentivise Disaster Preparation Behaviours

By

Elizabeth Rose Ditton

BIT (Honours)

Thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

College of Science and Engineering

James Cook University

December 2022

ACKNOWLEDGEMENTS

I would like to start by thanking everyone who made this thesis possible.

From our first discussion in a café to our most recent meetings, my supervisors Professor Trina Myers and Dr Anne Swinbourne have provided consistent support, guidance, and enthusiasm.

I would like to thank the IT writing/research group who took me in at the beginning of my journey and provided invaluable support during the early days. So, thank you to Mangalam Sankupellay, Randy Zhu, Tim Hart, Kranthi Addanki, Fahad Nazir, and Alzayat Saleh. Also, thank you to the Psychology research group for providing support and friendly faces when I was just getting into the world of psychology.

I would also like to thank my friends and family for their support. Thank you to my Mum, Merrilee, and my sisters, Juliet and Maddy, who all provided unwavering sanity and support. Thank you to my friends Jono, Jaimee, and Tori who were always available to chat and provide a much-needed distraction.

Finally, a special thanks goes to my husband, Dean, for his unconditional support. You were there for me every step of the way.

STATEMENT OF CONTRIBUTION OF OTHERS

This thesis was completed by Elizabeth Ditton under the supervision of Professor Trina Myers and Dr Anne Swinbourne.

Nature of Assistance	Contribution	Co-Contributors
Intellectual Support	Supervision, Editorial Assistance, Survey design, Conceptual guidance	Prof Trina Myers (QUT) Dr Anne Swinbourne (JCU)
Financial Assistance	Stipend and Fee offset/waiver	JCU Research Scholarship
Infrastructure Assistance	Use of High-Performance Computers	JCU

COPYRIGHT DECLARATION

Every reasonable effort has been made to gain permission and acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

ABSTRACT

Targeted messaging is imperative to effective communication. In Australia, millions of dollars are spent each year on recovery from natural disasters such as tropical cyclones. As such, effective communication and education surrounding measures to protect the population and reduce damage during severe weather events is essential. In *North Queensland* (NQ), Australia nearly all residents (92%) have experienced at least one cyclone, and almost a third (29%) have experienced more than five cyclones. As the perceptions of cyclones and risk mitigation behaviours can differ greatly between individuals, segmenting the audience based on their perceptions and attitudes allows for the most effective communication.

One common way to communicate audience segments with a variety of stakeholders is to create a persona to represent each segment. A persona is a description of a fictitious person that allows for information to be communicated without the need for domain-specific jargon and emphasises the human attributes to elicit empathy. However, manual persona development is time and resource intensive, requires a high level of specialisation, and often requires being completely repeated for additional use-cases or data sets. Additionally, manual persona development methods are critiqued due to the potential for bias during creation and the inability for an individual to effectively analyse large datasets.

As such, there has been a push towards more automated methods of persona development. To automate persona development, machine learning algorithms, specifically clustering algorithms, are applied to automatically identify groups within the data. Automated persona development methods are often criticised as unable to capture the complex concepts and nuance that are essential for many persona use-cases, resulting in many current persona development methods taking a semi-automated approach. The primary aim of this project is to determine whether machine learning techniques, specifically clustering algorithms, could be employed to facilitate the development of deep, nuanced personas based on behavioural models.

One key issue with semi-automated and automated persona development methods, alongside any other problem requiring the application of a clustering algorithm, is algorithm and parameter selection, also known as hyperparameter tuning. Each approach to clustering targets a specific type of cluster, and the performance of a specific algorithm can vary significantly depending on the nature of the clusters present within a data set. Despite the clustering algorithm selected having a significant impact on the clusters developed, minimal documented research and evaluation has gone into the selection of a clustering algorithm for persona development.

The hyperparameter tuning of clustering algorithms is difficult due to the lack of ground truth values. To address these challenges and facilitate persona development HyPersona, a semi-automated hyperparameter tuning and persona development framework was developed. HyPersona allows for a series of clustering algorithms and parameters to be compared for persona development and uses naïve internal metric thresholds to rule out poor quality results. To assist in ruling out poor quality results, and to provide additional insights into the cluster sets developed, *Average Feature Significance* (AFS), a novel internal evaluation metric, was developed. AFS focuses on how distinct the clusters are from one another and the general population.

To determine whether clustering algorithms could be applied to persona development where the persona's basis on behavioural theory is important, a range of algorithms were tested with HyPersona. Over 20 clustering algorithms, each with a variety of parameters, were run over two complementary data sets. The first data set was from a survey of NQ residents on their perceptions and attitudes around cyclones, cyclone risk, and protective behaviours, as well as gathering some general demographic information. The second data set was a repeat of the first survey, with some additional questions added, such as questions about insurance status. A behavioural theory expert had previously used the first data set to develop a set of personas based on their attitudes and perceptions surrounding cyclone risk and cyclone shutters. The persona set developed by a behavioural theory were used as a gold standard for a set of behavioural theory driven personas developed for the given use case.

A total of 3,404 algorithm and parameter combinations were applied to each data set using HyPersona, which automatically ruled out over 60% of combinations. The cluster sets developed by each algorithm were evaluated in terms of overall performance and consistency. Overall, most algorithms were found to perform consistently across the two data sets. While the algorithm and parameter variations differed more greatly, reinforcing the importance of algorithm parameter selection. The internal metrics were found to be good indicators of cluster set quality, however, could not be used to identify the best cluster set for the given use case. Primarily because the clusters developed were found to differ significantly between algorithm and parameter combinations even when the cluster quality, according to internal metrics, was comparable.

The best cluster sets for the current use case were identified with domain-specific evaluation and then used to develop a set of personas. The persona set developed with HyPersona was compared to a set of personas developed by a behavioural expert on the same data, then evaluated based on how well the persona set related to behavioural theory and could be used to target communication. The persona set developed with HyPersona was found to align with behavioural theory, provide insight into the data set, and could be used to effectively target communication. As a result, HyPersona was found to have developed a persona set of a comparable quality to those developed by a behavioural expert. Thus, clustering algorithms were found to be able to effectively mimic expert decision making. Further, as the entire data set was used, the personas developed with HyPersona were more robust and could be used to target a wider range of behaviours.

This study found that hyperparameter tuning and some manual intervention was required to mimic expert decision making with clustering algorithms. However, with HyPersona and clustering algorithms the effort and time required for persona development was greatly minimised, even in comparison to other semi-automated persona development methodologies. AFS was also found to provide unique and useful insight into cluster quality. Most importantly, clustering algorithms were found to be able to develop personas that achieve the same level of depth and nuance as manually developed personas, without the required resources.

PUBLICATIONS

Publication	Input of each author (CRediT statement)
Ditton, E., Swinbourne, A., Myers, T., & Scovell, M. (2021). *Applying Semi-Automated Hyperparameter Tuning for Clustering Algorithms* [Late Breaking Results Poster Session]. The 36th IEEE/ACM International Conference on Automated Software Engineering. arXiv:2108.11053 [cs], Aug. 2021. Available: http://arxiv.org/abs/2108.11053	**Elizabeth Ditton:** Writing - Original Draft, Conceptualization, Methodology, Software, Validation, Formal Analysis. **Dr Anne Swinbourne:** Supervision, Writing - Review & Editing, Conceptualization. **Prof Trina Myers:** Supervision, Writing - Review & Editing, Conceptualization. **Dr Mitchell Scovell:** Data Curation, Investigation.
Ditton, E., Swinbourne, A., & Myers, T. (2022). Selecting a clustering algorithm: A semi-automated hyperparameter tuning framework for effective persona development. *Array*, 14, 100186. https://doi.org/10.1016/j.array.2022.100186	**Elizabeth Ditton:** Writing – Original Draft, Conceptualization, Methodology, Software, Validation, Formal Analysis. **Dr Anne Swinbourne:** Supervision, Writing – Review & Editing, Conceptualization. **Prof Trina Myers:** Supervision, Writing – Review & Editing, Conceptualization.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

General terms	
NQ	North Queensland
SRQ	Secondary Research Question
ML	Machine Learning
Behavioural Models	
EV	Expectation-Value
PMT	Protective Motivation Theory
PADM	Protective Action Decision Model
Internal evaluation metrics	
AFS	Average Feature Significance
SC	Silhouette Coefficient
CHI	Calinski-Harabasz Index
DBI	Davies-Bouldin Index
Other	
RBF	Radial Basis Function
Clustering Algorithms	
AHC	Agglomerative Hierarchical Clustering
BIRCH	Balanced Iterative Reducing and Clustering using Hierarchies
CURE	Clustering Using Representatives
MST	Minimum Spanning Tree
DBSCAN	Density Based Spatial Clustering of Applications with Noise
OPTICS	Ordering Points To Identify the Clustering Structure
SVC	Support Vector Clustering
EM	Expectation-Maximisation
SOM	Self-Organising Map
GA	Genetic Algorithm
SFLA	Shuffled Frog-Leaping Algorithm
ABC	Artificial Bee Colony
AP	Affinity Propagation
NMF	Non-negative Matrix Factorization

Chapter 1: INTRODUCTION AND BACKGROUND

Millions of dollars are spent each year on recovery from natural disasters in Australia [1]. Northern Australia frequently experiences tropical cyclones during the summer months as they form over the warm waters surrounding the region [2]. The Australian Bureau of Meteorology reports that, on average, 9 to 11 tropical cyclones form off the Australian coast every cyclone season, four of which will typically make landfall [3]. In *North Queensland* (NQ), the most densely populated region of northern Australia, nearly all residents (92%) have experienced at least one cyclone, and almost a third (29%) have experienced more than five cyclones [4]–[6].

1.1 Tropical Cyclones

A tropical cyclone is made up of heavy thunderstorms rotating around the eye of the cyclone, accompanied by severe wind [2]. A tropical cyclone occurs when a low-pressure system forms over warm water, and can continue to strengthen for days [2]. Similar systems formed in other areas of the world are often referred to as hurricanes or typhoons. The severity of a cyclone is defined by its sustained wind speed [2]. There are five categories of cyclone, given in Table 1-1 [2].

TABLE 1-1: CYCLONE CATEGORY DEFINITIONS [2]

Category	Max Mean Wind (km/h)	Typical Strongest Gust (km/h)	Typical Effects
1	63 - 88	< 125	Damaging winds. Negligible house damage. Damage to some crops, trees, and caravans. Craft may drag moorings.
2	89 - 117	125 - 164	Destructive winds. Minor house damage. Significant damage to signs, trees, and caravans. Heavy damage to some crops. Risk of power failure. Small craft may break moorings.
3	118 - 159	165 - 224	Very destructive winds. Some roof and structural damage. Some caravans destroyed. Power failures likely.
4	160 - 199	225 - 279	Significant roofing loss and structural damage. Many caravans destroyed and blown away. Dangerous airborne debris. Widespread power failures
5	> 200	> 279	Extremely dangerous with widespread destruction.

Cyclones can cause significant damage to property and infrastructure through destructive winds, heavy rainfall, storm surge, and subsequent flooding. The effects of property damage extend beyond the obvious economic impacts to the personal costs of damage to irreplaceable sentimental items and being displaced from home or work while a property is repaired. Cyclones that do not make landfall, or downgrade to a tropical low before making landfall, can still cause significant damage as an effect of increased rainfall and storm surge. In February 2021, Tropical Cyclone Niran formed off the Queensland coast, and while Tropical Cyclone Niran never made landfall, gale force winds in the area caused minor property damage, primarily from falling trees, and banana plantations in the area reported a significant loss of crops [7].

As tropical cyclones typically develop when the sea-surface temperature is above 26.5°C, they frequently form over the oceans around the northern areas of Australia during the summer months [2]. This period, between November and April, is referred to as the cyclone season [2]. The Australian Bureau of Meteorology reports that on average 9 to 11 tropical cyclones form off the Australian coast every cyclone season, four of which would typically make landfall [3].

Cyclones are expected to cause more damage in the future. As the oceans warm fewer low-category cyclones and more high-category cyclone are expected, as the increased temperatures provide more energy to the system [8], [9]. As large, high-category cyclones require large amounts of energy the likelihood for multiple cyclones to occur at once decreases [8], [9]. Due to the different damage potential, multiple category 1 or category 2 cyclones require far less to protect against compared to a single category 5 cyclone. As sea levels rise storm surge is expected to become a bigger issue when cyclones occur [9]. As these risks increase, the importance of effective education around cyclones and damage mitigation strategies increases.

The areas that are most commonly affected by tropical cyclones sit above the Tropic of Capricorn [2]. Figure 1-1 shows the location of the Tropic of Capricorn over Australia, and the population density of the areas [4], [5]. The eastern coast of Queensland is the most densely populated region of Australia above the Tropic of Capricorn [4], [5]. Although there is no strict border, the region of Queensland above the city of Rockhampton is known as *North Queensland* (NQ). Due to the regions high susceptibility to cyclones, 92% of NQ residents report having experienced at least one cyclone and almost a third (29%) of residents report having experienced more than five cyclones [6].

The most severe example of a cyclone to affect NQ in recent history was Tropical Cyclone Debbie. Tropical Cyclone Debbie was a category 4 cyclone that made landfall on the NQ coast near Airlie Beach, a coastal town between Townsville and Mackay, during March 2017 [10]. Tropical Cyclone Debbie had peak wind gusts of 263 km/h, caused a 2.6m storm surge, and resulted in torrential rain in central to southeast Queensland over the following days [10].

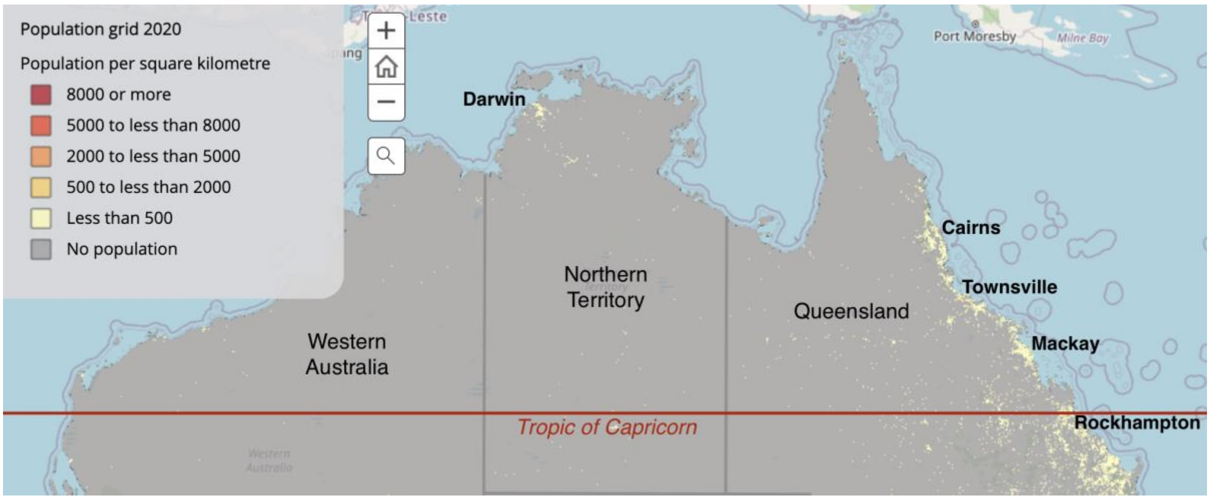


FIGURE 1-1: MAP OF NORTHERN AUSTRALIA WITH POPULATION CENTRES [5]

Tropical Cyclone Debbie demonstrates the impacts a cyclone can have on a community. Numerous properties were damaged, resulting in over AU$1.7 billion in residential insurance claims, including 944 properties left uninhabitable and a further 2,360 damaged, and billions more in losses from industry estimated [11]. There was widespread power loss, with an estimated 65,000 households that lost power which took 16 days to get 95% restored [11]. On top of voluntary evacuations, 4,357 households were evacuated due to storm surge risk [11], and returning to a premises once evacuated is often difficult due to roads being cut by flooding, fallen trees, or landslides, and airports are often closed due to the weather conditions. The financial impacts of cyclones can be severe, beyond the costs of repairing and replacing damaged items/property, higher insurance premiums, loss of work, and increased costs of regular produce can all place significant financial strain on residents.

In the short term, increased prices of produce are often due to damaged warehouses and trucks carrying goods unable to access the area. However, damage to farmlands/industry can have a longer lasting impact. For example, during Cyclone Yasi in February 2011, an estimated AU$350million of bananas were destroyed, which drove prices from their usual AU$2/kg to over AU$14/kg, and prices did not return to normal until later in the year [12]. On top of the monetary loss, there is the potential loss of irreplaceable sentimental items and the mental distress associated with being displaced from home or work while repairs take place.

Due to a widening tropical zone and warmer ocean temperatures cyclonic activity is expected to affect a much larger area [8], [9]. When cyclones affect areas further south, they put new, currently unprepared communities at risk, including the many densely populated zones along the south-east and south-west coast [8], [9].

Most recently, Tropical Cyclone Seroja, a category 3 cyclone, crossed the WA border unusually far south during April 2021, just south of Kalbarri, bringing the strongest winds to the area in more than

50 years [13]. In the affected areas, 70% of the properties reported damage, most consisting of lost roofs, and an estimated 40% of those properties were completely destroyed [14]. Which is a much higher percentage of properties seriously damaged in comparison to what a higher category cyclone, Cyclone Debbie, caused in NQ. Additionally, many of the damaged properties were older, revealing asbestos which created further danger for individuals returning to the area and cleaning up [15]. Cyclone Seroja was estimated to have caused AU$200 million in damages, which is not as high as the costs associated with Cyclone Debbie, as Cyclone Seroja did not impact any densely populated areas [15]. However, at a similar latitude on the east coast of Australia is Brisbane, the capital city of Queensland, which has a population close to 2.5 million.

There are two primary factors that cause cyclones to do significantly more damage in unprepared areas; 1) the properties and buildings have not been built to mitigate cyclone damage and are less likely to have had structural upgrades to reduce potential damage; and 2) the population are not familiar with cyclones and how to appropriately prepare once a cyclone warning has been issued. As such, there is a large population unaware of the risks associated with cyclones due to lack of experience and exposure, so effective and immediate messaging is required.

1.2 Damage Mitigation Strategies

Damage mitigation strategies range from simple, low-cost actions such as tidying a yard or securing loose outdoor items, to more difficult, costly actions such as structural upgrades to the home. The performance of damage mitigation behaviours is critical to minimise damage from cyclones, as seen by the large amount of damage done by Tropical Cyclone Seroja.

For example, bringing in outdoor furniture and other items is a simple action that should be performed immediately prior to a cyclone [16], [17]. Which is an example of a ‘low-cost’ action which helps reduce the quantity of dangerous debris which may act as projectiles during the cyclone. Conversely, garage door reinforcement is a relatively expensive structural upgrade that needs to be planned for and completed before a cyclone occurs. The cost and planning required, alongside the need to organise installation before cyclone season act as barriers to uptake of garage door reinforcement. However on the benefit side, garage door reinforcement is one of the only ways to protect a car from getting damaged from the garage door collapsing inwards when hit by debris [18].

As cyclones cause a range of different weather hazards, such as wind, rain or storm surge, a variety of behaviours should be undertaken to mitigate the different hazards [16], [17]. Insurance claims after Tropical Cyclone Yasi in 2011, showed that approximately 29% of the claims costs were for minor damage, most of which could have been prevented through the performance of simple preparatory behaviours [19]. The best protection against minor damage is to inspect and prepare the premises at the beginning of cyclone season so any weak areas can be addressed [16], [17]. Areas of the property that should be checked are [16], [17]:

- The entire house for rust, rotten timber, and termite infections
- The structural integrity of the walls
- Look for loose fittings
- The integrity of the fence
- Check the roof, including any roof tiles or sheets, and the eaves

Other behaviours that can be performed at the beginning of cyclone season to minimise damage is to trim treetops and branches, especially those close to the property, and to clean the gutters and down pipes [16], [17]. Installing netting to avoid clogged gutters can also help. Before the cyclone season begins, there are other structural upgrades that can be made to help avoid damage. Structural upgrades help to avoid the large-scale damage, such as those seen during the aftermath of Tropical Cyclone Seroja [18]. Structural upgrades that can mitigate cyclone damage can include [18]:

- Cyclone shutters: A structural upgrade that allow for a roller door to cover windows or doors during a cyclone, which can mitigate damage from flying debris.
- Metal screens on glass areas (windows, doors, etc.)
- Complete roof replacement or upgrades: Most important on houses built before 1982, the upgrade depends on the existing roof and can include replacing the external cladding, upgrading the batten to rafter attachments, and upgrading tie-downs from rafter.
- Shed upgrade: Can include anchoring the shed to a concrete slab or reinforcing the shed using a cyclone kit.
- Dead locks on external doors.
- Roller door bracing

Some of the structural upgrades offer additional uses, other than just protecting against cyclone damage, which cause them to be more popular. Dead locks and metal screens on glass areas can act as security, protecting against break-ins, which can make them more popular.

In addition to behaviours and upgrades that can be performed in advance, there are some behaviours that are usually performed in the lead up to a cyclone [16], [17]. These are usually activities that performing in advance would potentially obstruct regular day-to-day life or could require repeating in the lead up to a cyclone anyway. However, the unpredictability of cyclones can lead to residents putting off these tasks until the last possible moment, not wanting to perform the behaviour if the cyclone does not hit, and then not having the appropriate amount of time to perform the tasks. These activities include [16], [17]:

- Clearing the yard of any loose items –including fallen branches or palm leaves
- Secure or bring inside outdoor furniture and garden items
- Put plywood up on glass windows and doors
- Take down shade sails

1.3 Encouraging Performance of Damage Mitigation Behaviours

Understandably, people are more likely to undertake the simple, low-cost options rather than the more difficult, high-cost options, despite the more expensive methods being highly effective and often the only method to protect against certain types of damage [6], [20]. Misinformation, incorrect perceptions, and a lack of awareness can all impact whether a risk mitigation behaviour is performed. Additionally, the continual exposure to cyclones, especially cyclones that do not make landfall or only cause minimal damage, can cause complacency in residents. To ensure that those living in NQ and the surrounding regions undertake all possible risk mitigation strategies, effective communication and education is required.

As the perceptions of cyclones and risk mitigation behaviours can differ greatly between individuals targeted messaging is required for effective communication. Behavioural models attempt to describe why an individual preforms a given behaviour and the various reasons they may not perform a desired behaviour [21], [22]. Communication can be targeted to an individual by employing behavioural models to identify and address the primary perceptions or attitudes that are stopping the individual from performing the desired protective behaviours. Audience segmentation based on behavioural theory allows for communication to be effectively customised for a large percentage of the population at a time through social marketing campaigns.

The creation of personas is a technique often employed to effectively communicate audience segments and analytical data between stakeholders while emphasising human attributes to elicit empathy [23], [24]. A persona is a description of a fictitious person, usually including a name, photo, and a description of the persona's traits, attitudes, or behaviours [23]. Using personas allows for information to be communicated without the need for domain-specific jargon, meaning a set of personas could be used by a wide range of stakeholders to customise various forms of communication and incentives.

To this end, Scovell et al. [6], [25] developed a set of three personas representing the attitudes of NQ residents towards installing cyclone shutters, a structural upgrade to mitigate cyclone damage. The personas were created from survey data assessing the psychological characteristics and cyclone-related attitudes of over 500 NQ residents [6], [25]. These personas did not have a name or photo, as is often seen with traditional personas, but described the behaviours and attitudes of the individuals that made up each segment in an empathetic and easy to understand manner. However, manual expert driven persona development is time and resource intensive, requires a high level of specialisation, and often requires being completely repeated for additional use-cases or data sets [26], [27].

The high costs associated with manual persona development, alongside critiques of the objectivity and the ability for an individual to effectively analyse large datasets having been driving forces behind a push towards more automated methods of persona development [26], [27]. However, one of the major criticisms of automated persona development methods is that they cannot capture the complex concepts and nuance that are essential for many persona use-cases [26], [27]. As a result, most current approaches to persona development are semi-automated and rely heavily on manual guidance before or after the clustering algorithm is applied [26]–[28]. However, semi-automated persona development methods can become victim to the pitfalls of both automated and manual persona development, requiring significant resources whilst failing to capture significant nuance and depth.

Automated and semi-automated methods of persona development are usually based around the application of a clustering algorithm [27], [28]. A clustering algorithm is an unsupervised machine learning algorithm that attempts to identify distinct groups of data points within a data set [29]. For an automated or semi-automated persona development method to create a set of personas that can be effectively applied to target communication around mitigation behaviours, the clustering algorithm employed needs to effectively mimic expert decision making.

Despite the impact of algorithm choice, minimal documented research and evaluation goes into the selection of a clustering algorithm for persona development [27], [28]. Each approach to clustering targets a specific type of cluster, and the performance of a specific algorithm can vary significantly depending on the nature of the clusters present within a data set [29]. Evaluating and selecting a best preforming clustering algorithm is a complicated process, there are no ground truth values available during clustering so objective metrics, such as accuracy, are not available [29]. Furthermore, the performance of an algorithm on one use-case is not generalizable to different use-cases or data sets [30].

1.4 Thesis Overview

The primary aim of this thesis was to determine whether machine learning techniques, specifically clustering algorithms, could be employed to facilitate the development of deep, nuanced personas based on behavioural models. An effective clustering algorithm would be able to mimic expert decision making to develop a set of distinct clusters that represent different states within a behavioural model. The primary research question developed for this thesis was:

Can clustering algorithms facilitate the development of deep, nuanced personas based on behavioural models, replicating the decision making of experts, for the automation of persona development?

Primary research question was broken into four *Secondary Research Questions* (SRQs) where addressing each SRQ would allow the primary research question to be confidently answered. The four SRQ are:

SRQ1. How can a range of clustering algorithms and the clusters they develop be efficiently evaluated and compared?

SRQ2. Are the performances of clustering algorithms and approaches to clustering for persona development consistent?

SRQ3. Does the selection of a clustering algorithm and parameters significantly impact the set of personas developed?

SRQ4. How do personas created by clustering algorithms compare to behavioural theory and personas created through the application of behavioural theory?

HyPersona, a framework for testing and evaluating a series of clustering algorithms was developed, to identify whether a clustering algorithm could achieve sufficient depth. As part of the HyPersona framework, a new metric specific to the priorities during persona development, *Average Feature Significance* (AFS), was proposed. AFS is a metric that evaluates a set of clusters based on their statistical significance, an element that is indicative of cluster quality, and important to persona development, which is not captured in existing metrics. The HyPersona framework was then used to apply a range of clustering algorithms representing different clustering approaches to two data sets, the data set used by Scovell et al. [6], [25] and a subsequent confirmatory study. Clustering algorithms were found to be able to effectively replicate expert decision making, with the meta-heuristic approach to clustering, specifically the *Shuffled Frog-Leaping Algorithm* (SFLA) [31], [32], performing particularly well. A set of personas was able to be created from each data set that reflected behavioural theory.

Through identifying that clustering algorithms can effectively replicate expert decision making, the persona development process can be significantly automated. Meta-heuristic algorithms have been proven to be a strong starting point when deep, nuanced personas are required and the metric proposed as part of the HyPersona framework, AFS, was found to be a strong indicator of algorithm performance for persona development. Both HyPersona and AFS could be applied across a variety of persona development problems. The findings of this thesis can be applied to adjacent fields, allowing for the customisation of communication around a range of threats and preparatory behaviours and facilitating the adoption of risk mitigation strategies that would lead to a safer population during the next natural disaster.

1.5 Thesis structure

This thesis has been broken into three main sections: 1) context chapters, 2) the preparatory phase, and 3) the persona development phase. A graphical overview of the thesis is given in Figure 1-2. The current chapter provides the necessary background around the problem domain and the previous study performed by Scovell et al. [6], [25].

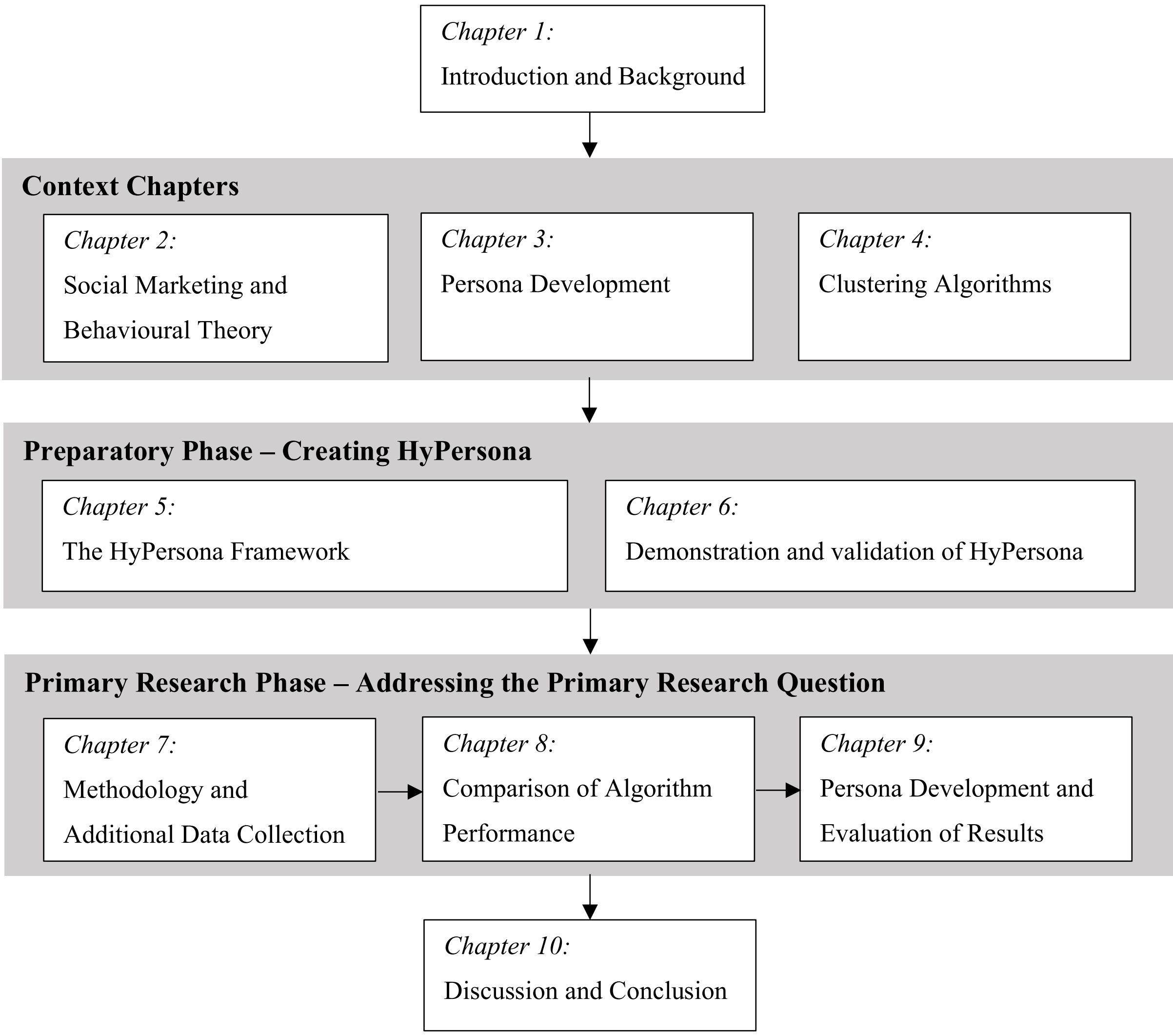


FIGURE 1-2: THESIS OVERVIEW DIAGRAM

Chapters 2-4 make up the context chapters. Chapter 2 will detail the psychological basis of the personas, discussing social marketing and behavioural theory. Chapter 3 gives an overview of the current state of the literature around personas, focussing on persona development methodologies. Chapter 4 presents the literature surrounding clustering, detailing each of the clustering algorithms that will be used during this thesis.

The following two chapters make up the preparatory phase, detailing the steps taken to begin to answer the research question and address SRQ1. Chapter 5 presents HyPersona and AFS, then chapter 6 contains a proof-of-concept study, validating both HyPersona and the idea that clustering algorithms can develop clusters that create effective personas based in behavioural theory.

Chapters 7-9 give the details of the primary phase of the research, determining whether clustering algorithms can effectively mimic expert decision making to develop behavioural theory driven personas. Chapter 7 details the methodology followed, including preforming the survey which found a confirmatory sample for the data collected by Scovell et al. [6], [25]. The additional data was imperative to determining the stability and consistency of the performances of the clustering algorithms and HyPersona. Chapter 8 gives a comprehensive comparison of the performance of each clustering algorithm for persona development. Chapter 9 develops a set of personas from the best performing algorithms on each data set, compares the personas to those created by Scovell et al. [6], [25], and evaluates how well the personas mimic expert decision making. Finally, Chapter 10 discusses the findings and concludes the thesis.

1.6 Research Significance

The findings presented in this thesis have implications across multiple research fields and areas of industry. The HyPersona framework and associated AFS metric will allow for more efficient persona development and clustering algorithm selection, which lowers the barrier of entry to persona use and simplifies persona maintenance. As a result, more initiatives, such as those surrounding cyclone damage mitigation in NQ, will be able to utilise personas to create an avenue for effective education and incentives that individuals are more likely to respond to.

The HyPersona framework and the AFS metric begins to address the challenges associated with the evaluation of clusters, personas, and their respective development methods. HyPersona proposes a semi-automated approach to persona development and the hyperparameter tuning of clustering algorithms to minimise manual intervention whilst not relying completely on metrics that cannot comment on the usefulness of a set of clusters or personas. As such, HyPersona begins to bridge the disconnect between the requirements of automation for quantitative, data-driven metrics and the requirements of the use-cases for domain-specific evaluation.

Previous approaches to automated and semi-automated persona development have only used a small subset of clustering algorithms with minimal, if any, evaluation performed prior to algorithm selection. However, existing automated and semi-automated persona development methods are criticised as being shallow and unable to capture the complex details that make personas realistic and able to provoke empathy, prompting the move towards more complex and convoluted methods. The current research found that more complex approaches to persona development are not needed when more sophisticated clustering algorithms are applied to the problem. Deep personas can be developed that mimic those created through the expert decision making used in manual approaches by identifying and applying the most appropriate clustering algorithm. The HyPersona framework facilitates the informed selection of a clustering algorithm for persona development reducing the required complexity for the creation of effective personas.

Chapter 2: SOCIAL MARKETING AND BEHAVIOURAL THEORY

Cyclone damage can have many effects on an individual. There are obvious economic impacts, such as the cost of repairing damage, as well as more personal costs, such as damage to irreplaceable sentimental items or being displaced from home or work during repairs. Thus, it is optimal that individuals living in areas with a high cyclone risk are encouraged to perform a multitude of protective behaviours to mitigate damage. Protective behaviours range from simple, low-cost actions such as securing outdoor furniture to more complicated, high-cost actions, such as reinforcing garage doors. Understandably, people are more likely to undertake the simple, low-cost actions than the more difficult, high-cost actions [6], [20]. However, the high-cost actions are often the most effective or the only method of mitigating certain risks, such as the garage door blowing in and damaging a car.

To promote the undertaking of high-cost behaviours, communication educating the population of their risk is required. One of the most effective methods for influencing behaviour change is social marketing. Social marketing is a form of marketing designed for selling a message rather than a product, such as campaigns for health initiatives [33]. One successful example of social marketing is The Road Crew, an anti drink driving campaign run in rural Wisconsin, America [34]. Anti drink driving campaigns are a common use of social marketing, often supported by policy, such as fines for drink driving. At the end of the first year, the campaign evaluation suggested that the program had decreased alcohol-related crashes by 17% [34].

As individuals have different perceptions surrounding mitigation behaviours, segmenting the audience based on underlying motivators can allow for more effective targeting of communication [35]. Effective segmentation needs to reflect both the individual's perceptions and the reasoning behind those perceptions, so communication can be customised based on those factors [36]. For example, an individual may not get their garage door reinforced because they do not realise that garage door reinforcement is an option, or because they do not believe the upgrade will effectively mitigate damage, or they are unable to afford the upgrade. Behavioural theory attempts to explain the processes that determine intent to perform a particular behaviour. Thus, segmentation informed by behavioural theory should be based on the underlying motivations to perform behaviours to allow for communication to be effectively targeted.

2.1 Social Marketing

As described by Grier and Bryant [33], social marketing is the field of selling behaviour change or lifestyle. Social marketing aims to alter core perceptions to make the desired behaviour more advantageous and to reinforce negative consequences associated with not performing the behaviour. Social marketing can include education and policy elements, which are both common approaches to behaviour change, but in collaboration with marketing techniques.

- **Policy:** Policy or legislation is most required when society is not willing to pay the costs of the undesirable behaviour, however individuals are unlikely to find changing within their self-interest [33]. For example, individuals may be unwilling to stop drinking and driving, believing they are unlikely to get into an accident or that they are not too drunk to drive. Thus, to avoid the unnecessary risk of life, strict policies around drink driving is enforced.
- **Education:** Education is a strong tool in social media campaigns. As a standalone method education is most effective when societal and individual values align with the desired behaviours, the results of the behaviour are attractive and immediate, and the costs of the behaviour are relatively low [33]. Learning what to do during a natural disaster, such as a cyclone, is an example of an initiative that may only require education. Individuals are already invested in the problem, surviving a natural disaster, with the benefits of knowing what to do far outweighing the costs of learning.
- **Social Marketing:** Social marketing as the primary activity is most effective when the desired behaviour is not immediately consistent with the individual's self-interest, but the behaviour aligns with societal goals and individuals can be influenced to perform the behaviour [33]. Social marketing can also be used to reinforce policy or education-based initiatives.

An example of where social marketing has been used alongside education and policy is in anti-speeding campaigns. Many countries have laws and policies against speeding whilst driving. In Queensland, Australia, as of 2021, the maximum penalty of being caught speeding is a 6 months driver's licence suspension and a fine of AU$1245 [37]. There are also many educational initiatives, aimed at students and young drivers, and demonstrating the ability to follow the speed limit is required for an individual to receive their licence. These elements are used in conjunction with social marketing campaigns, such as the "Slow Down Stupid" campaign in Australia which focuses on the consequences of speeding [38]. The "Slow Down Stupid" campaign was found to be very effective, with 80% of individuals surveyed following the campaign saying "Since seeing the advertising I'm less inclined to speed" [38].

Social marketing shares many similarities to commercial marketing but faces unique challenges specific to the task of attempting to change behaviours, lifestyles, or social norms. Aspects such as competition and exchange are harder to define in many social marketing campaigns than in their commercial counterparts. Social campaigns, such as those promoting exercise, do not have a singular product or exchange that marketing can easily be focussed on, so research needs to be done into identifying which aspects garner the most response. However, the use and understanding of commercial marketing principles can be beneficial to social marketing campaigns once their differences are understood. A core principle of both commercial and social marketing is the marketing mix, also known as the four P’s [33]:

1. **Product:** What is being sold, in social marketing the product may be the benefits of the behaviour change rather than the behaviour change itself
2. **Price:** The cost of the behaviour change, includes financial costs, time commitment, and personal discomfort
3. **Place:** Where the behaviour change takes place, could be a general location such as in the kitchen or at the dinner table
4. **Promotion:** Communication strategies designed to influence behaviour

Both product and price are related to the notion of exchange. Generally, in commercial marketing, the exchange is giving money for goods or services. While social marketing requires convincing the individual to give up behaviours or lifestyles for results that are often intangible or long-term. An effective social marketing campaign requires acknowledging all the costs associated with the desired behaviour and identifying what the target audience truly values. For example, when promoting exercise, focussing on the exercise as a way to have fun and form closer relationships within the individual’s family may be effective than focussing on avoiding adverse health conditions [33].

An important aspect of program uptake is audience segmentation, a method that is growing in popularity in social marketing. Research shows that a one-size-fits-all approach is not as effective as targeted messaging that is more relevant to an individual [35]. How the population is segmented significantly affects the efficacy of the targeting. Traditionally segments are based on demographic factors, however, these segments are only effective when the demographic factors are reliable predictors of behaviour [33].

Often social marketing uses factors such as the individual’s wants, needs, lifestyle, and behaviours to develop better audience segments [33]. When looking at disaster preparatory behaviours, demographic or lifestyle factors do not provide the insights needed to determine motivation [36]. Factors relating to psychological perceptions and behaviours are suggested to be stronger predictors of behaviours, thus similar factors are required to effectively segment an audience based on an individual’s reported motivation to perform preparatory behaviours. That is, segmentation based on behavioural models can, in principle, lead to more effective audience segmentation.

2.2 Behavioural Models

A behavioural model attempts to reflect the processes that determine the intent of an individual to perform a particular behaviour. Generally, these models are based on the *Expectancy-Value* (EV) theory. EV theory suggests that an individual’s behaviour is dependent on the expected outcome of that behaviour and the values associated with the expected outcome [21]. Both expectancies and values are believed to be influenced by task-specific beliefs, perceptions of other people's attitudes, previous experience, social norms, cultural environment, and historical events [22]. An outcome is believed to have four general categories of values associated [22]:

1. **Attainment Value:** The importance of the outcome to the individual's sense of self or identity.
2. **Intrinsic Value:** The enjoyment an individual gets from performing the activity or their subjective interest in it.
3. **Utility Value:** The perception of the usefulness of the outcome or activity.
4. **Cost:** The cost of the activity, including any potential negative consequences.

EV theory proposes that, for an individual to perform an activity, the costs associated with the activity needs to be outweighed by the positive values associated with the outcome. Between the various behavioural models based on EV theory, the primary differences are the cognitive factors that are proposed to be predictors of behaviour [39]. A behavioural model tends to be targeted towards a certain behaviour or class of behaviours, with different theories better at explaining different behaviours [39].

Two primary models have been applied to the motivation to perform disaster mitigation behaviours: *Protective Motivation Theory* (PMT) [40], [41] and *Protective Action Decision Model* (PADM) [42]–[44]. PMT was originally proposed by Rogers [40] and applied to disaster mitigation by Grothmann and Reusswig [41]. PADM was originally proposed by Lindell and Perry [42] and applied to disaster mitigation by Lindell and Perry [43] and Terpstra and Lindell [44].

2.2.1 Protective Motivation Theory

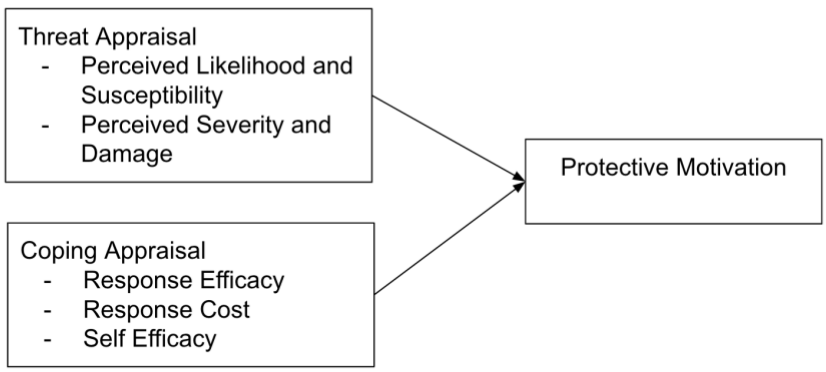


FIGURE 2-1: PROTECTIVE MOTIVATION THEORY

The *Protective Motivation Theory* (PMT) defines an individual's level of protective motivation, that is, their motivation to perform a protective behaviour. Grothmann and Reusswig [41] conceptualised the motivation to perform a protective behaviour as the combination of threat appraisal and coping appraisal. Threat appraisal, or risk perception, is reflective of the individual's beliefs about the hazard or risk. Coping appraisal reflects the individual's beliefs about the protective behaviour. Figure 2-1 gives a visual representation of PMT.

Threat appraisal is defined in terms of two components: how likely the threat is to occur, and the severity of the threat or what potential damage the threat could do [41]. With the example of cyclones as the threat, the likelihood that a cyclone of a particular category would hit, the damage that a cyclone of that category could do, and the likelihood that the individual will receive that damage, would all be factors in threat appraisal. Coping appraisal is defined in terms of three components regarding the protective behaviour itself, such as bringing outdoor furniture inside leading up to a cyclone [41]:

1. **Perceived efficacy of the action:** How effective the individual believes that bringing in furniture is at reducing the damage the individual may receive.
2. **Perceived self-efficacy:** Whether the individual believes they could either bring the furniture inside themselves or could organise for someone else to do it.
3. **Cost:** The cost may be in terms of: time, how long bringing the furniture in will take; money, do they believe they will need to pay someone to bring the furniture in or are they likely to break anything in the process; and, knowledge, do they know how to get the furniture inside i.e. will they need to break the furniture down or take a door off its hinges to get everything inside.

Grothmann and Reusswig [41] suggested that the combination of threat appraisal and coping appraisal leads to the level of protective motivation. For an individual to perform an action, PMT proposes that both the threat appraisal and coping appraisal need to be relatively high. If the threat appraisal is too low, the individual may not see the threat as being worth mitigating. However, if the threat appraisal is overwhelmingly high, the individual may feel there is nothing that could be done to mitigate the risk and instead turn to maladaptive behaviours, such as avoidance or fatalism. Each element of the coping appraisal can act as a barrier to performing the activity. If the perceived efficacy of the action is too low the individual will not believe the mitigation behaviour is worth performing. If the perceived self-efficacy is too low or the cost too high, the individual may not believe they can perform the behaviour, leading them towards maladaptive behaviours. However, when the coping appraisal is significant the individual may experience a reduction of threat appraisal even if the behaviour is not performed.

2.2.2 Protective Action Decision Model

The *Protective Action Decision Model* (PADM) proposes that there are three primary phases in determining whether an individual will perform a protective behaviour [43]. The three phases are: 1) the pre-decisional process; 2) core perceptions; and, 3) protective action decision making [43]. The primary difference between PMT and PADM is the additional factors, such as the pre-decisional processes and stakeholder perceptions, that PADM accounts for. Figure 2-2 gives a visual overview of PADM, including the elements going into the model and the results.

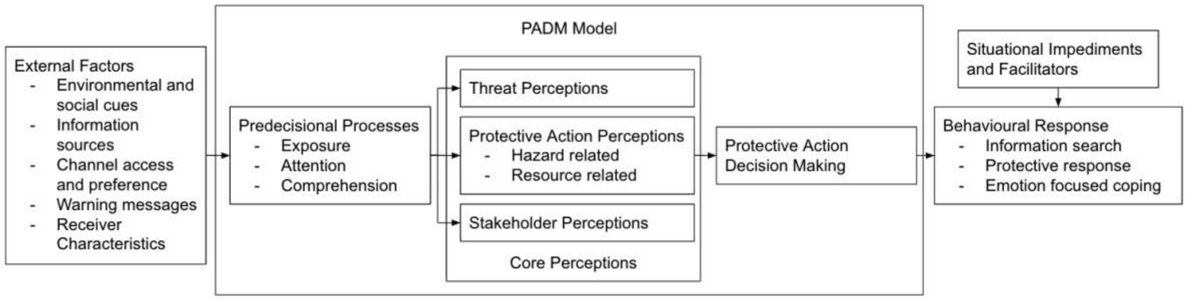


FIGURE 2-2: PROTECTIVE ACTION DECISION MODEL

The pre-decisional processes pertain to the information available to the individual surrounding both the threat and protective behaviour. There are three key parts of the pre-decisional process, as described by as defined by Lindell and Perry [43]: 1) exposure; 2) attention; and, 3) comprehension. Exposure is about the information that is available to the individual, the information can come from a variety of sources, such as environmental and social cues, warning messages, and information channels such as TV, radio, or the internet. Attention is an important element, as the information available to an individual only has an impact when the information provided is notable, interpretable, and relevant to the individual, capturing the individual's attention. Lastly, only the information that is effectively comprehended and internalised by the individual will become the foundations for the core perceptions phase.

The core perceptions phase focuses on the individual's perceptions of the threat and protective actions, similar to the threat and coping appraisal, and the individual's perceptions of key stakeholders [43]. Threat perceptions encompass the perceived likelihood that the threat would occur, the expected impacts and outcomes, and the severity of the given outcomes [43].

The perceptions of the protective actions, also referred to as hazard adjustment perceptions, is the combination of two attributes, hazard related attributes and resource related attributes [44]. Hazard related perceptions include elements such as the efficacy of the protective behaviour, and whether the mitigation is suitable for other purposes [43], [44]. Resource related perceptions include the evaluation of financial costs, skill requirements, time requirements, effort requirements, and whether performing the action requires cooperation with others [43], [44].

Stakeholders can include the authorities, watchdogs, employers, and households [43]. The importance the individual places on performing the protective action can depend on how responsible and trustworthy they find the stakeholders to be [43]. For example, an individual who believes the government will provide adequate financial aid after a cyclone may not be as motivated to perform protective action as those who do not believe adequate financial aid will be available.

The protective action decision making phase is often reflective. During the protective action decision making phase, the individual reassesses the results of the previous phases and the information available to determine a behavioural response [43]. Taking the previous elements into account one of three general types of outcomes is reached:

1. **Information Search:** The individual decides to look for additional information about the threat, the protective behaviour, or both feeding back into the model.
2. **Protective Response:** The individual performs the desired protective behaviour.
3. **Emotion Focussed Searching:** Responding with emotion, often through maladaptive behaviours such as fear or denial.

2.2.3 Key Features for Disaster Mitigation Behaviour

Across behavioural models there are several common factors that are believed to determine an individual's motivation to perform mitigation behaviours. Generally, these factors relate to the threat itself and the behaviour to be performed. Factors around the threat generally surround the probability of the threat occurring and the likely severity and outcome of the threat [41], [43].

Risk perception is informed by a wide variety of elements. Factors relating to risk perception include the believed likelihood that a cyclone will occur, the likely severity of the cyclone, the likely damage each category of cyclone would cause. The perceived likelihood and severity of cyclones are generally informed by education and experience. Prior experience can increase risk perception, as the individual is more aware of the risks, however prior experience can also have the opposite effect [45], [46]. If the individual is only minimally impacted or overestimates the severity of the cyclone they have experienced, they can tend to underestimate the likely severity of a future cyclone [45], [46]. People also have the tendency to believe that if they have lived through one unlikely event, such as a cyclone, the chance of having a second occur is exponentially less likely. This is known as the gambler's fallacy [47].

Perceptions surrounding the mitigation behaviour itself include the efficacy and cost of the behaviour [41], [43]. Using cyclone shutters as an example, the perceived efficacy relates to the amount of damage the cyclone shutters are believed to be able to mitigate. The cost of cyclone shutters would include monetary costs, the time commitment to get the cyclone shutters installed, and knowledge requirements. Closely connected to cost, is the idea of self-efficacy. That is, whether the individual believes they can install the shutters or organise to have the cyclone shutters installed.

Beyond risk perceptions and mitigation perceptions, there are additional factors that are seen as significant contributors to protective motivation. One such factor is the individual's emotional response or sense of fear towards the threat. An individual who has a high sense of fear or dread towards cyclones may be more likely to react to the threat with maladaptive behaviours, such as avoidance, wishful thinking, or fatalism. An additional contributor is whether the mitigation behaviour has additional purposes. If a mitigation behaviour has other uses, an individual may be more likely to perform it, as performing the behaviour offers additional value. For example, any individual who believes cyclone shutters look good or improve the resell value of the house may be more likely to get cyclone shutters installed than someone who does not.

External factors can also play a role in determining motivation to perform mitigation behaviours. These factors can include social norms, perceived government support, and perception of other key stakeholders. One example is if many people in an individual's social circle have performed the behaviour the individual may be more likely to perform the behaviour. To be able to effectively segment audiences based on their likelihood to perform mitigation behaviours, each of the key factors needs to be measured and considered.

2.3 Summary

To effectively promote the performance of protective behaviours, social marketing methods need to be applied. An important concept of social marketing is segmenting the audience so that communication can be targeted based on the barriers an individual likely has to performing the behaviour. Effective audience segmentation requires is based on behavioural theory, so that the segments reflect an individual's values and goals and will give the best chance of producing behaviour change. There are multiple behavioural theories, however they are all based on EV theory, stating that the individual's expectations and values are what determines whether they will perform a given behaviour. Key factors shared by the behavioural theories surround perceptions of the risk and the protective behaviour.

By gathering data that reflects standing on these key factors in the behavioural models the audience will be able to be segmented based on these factors. Creating proficient segments is key to developing targeted communication and social marketing strategies to educate and align the mitigation behavioural with the values of the individuals. To allow these segments to be most effectively used, the audience segments need to be given in an understandable and digestible manner. That is, through the use of a technique such as a persona. A persona is a description of a fictitious person used to describe analytical data and audience segments in a manner that promotes empathy. The next chapter will look at the literature surrounding personas and persona development.

Chapter 3: PERSONA DEVELOPMENT

A persona is a description of a fictitious person used to describe analytical data and customer segments in a manner that emphasises human attributes to elicit empathy [23], [24]. To this end, personas usually include a name, photo, and a description of the persona's traits, attitudes, or behaviours [23]. Traditionally, designers use personas to represent a particular type of target user so that designs can be better tailored [23]. More recently, personas have been used across a variety of industries and applied as a method to facilitate communication between various stakeholders [23].

The strength of personas comes from their ability to humanise data and communicate information without the need for domain-specific knowledge or jargon [23]. While personas are often used to represent the primary segments within an audience, they are also useful to describe common outliers who may not be considered during design or development [48]. In both cases, personas humanise the audience and promote empathy, making the individuals the personas represent easier to consider [23], [48]. For example, the same information can be presented in two ways:

- **Without a persona:** Individuals within this segment have low self-efficacy and coping appraisal combined with a high level of risk perception regarding cyclone risk and cyclone shutters. As such, these individuals are more likely to tend towards maladaptive responses, such as fatalism, rather than protective behaviours.
- **Using a persona:** Jane is terrified to think about a cyclone occurring. During the last cyclone in her area, all her windows were smashed, and she believes if there is another cyclone, she will lose all her windows again, no matter what she does.

The research around personas can be split into three categories: 1) persona design and use, including the reception of personas within industry and how persona design affects perception; 2) persona development, covering the various methods to create personas; and 3) persona evaluation, how to validate and evaluate personas. This project focuses on persona development, specifically on automated and semi-automated methods.

3.1 Persona Design and Use

Persona design and use research focuses on determining and improving the usefulness of personas, primarily through the design and use of personas. Applied research around persona use commonly investigates how personas are used in industry and the nature of the opinions and complaints of people who regularly work with them [48]–[53]. As a part of persona use research, methods to increase uptake and usefulness are often proposed and tested [48]–[53]. Persona design research focuses on the layout of personas and the elements included, and how changing these aspects affects a persona's reception and use [54]–[58].

Research around the design of personas can be based on visual aspects, such as the number of images included [54] and whether images generated by artificial intelligence can be effectively used in personas [57], to features surrounding the nature of the persona, such as the effect of including or excluding gender [55] or personality traits [56]. The general method for investigating these factors is to develop two sets of personas that are identical except for the variable being investigated, and then to present them to individuals who regularly use personas for evaluation. Methods for evaluating the impact of the factors being investigated often include eye-tracking software [54], [55], [58] to determine what elements are focused on and interviews [54], [55], [58] or surveys [56], [57] to assess the opinions formed of each persona.

Anvari et al. [56] investigated the effect of including personality traits in personas used for software design and development. To evaluate the effect of doing so, 91 software engineers from Australia and Denmark were asked to design an app for a set of four personas, each with different personality traits, and then fill out a survey. The study found that the views of and perceived needs of the personas were influenced by the personality traits. For example, socialisation features were deemed more important for introverts than extroverts. The study also compared how these perceptions differed between countries by recruiting software engineers from Australia and Denmark and found that the different cultures tended to prefer different personality types.

Studies around the use of personas also investigate how personas are used in industry. Personas have been found to be treated within companies as actual people, commonly brought up during design meetings [48]. One new approach identified by Nielsen and Hansen [48] was role-playing. Designers would act out scenarios as a given persona during ideation and early development to explore how the persona would interact with the product. The practice of role-playing occurs as an extension of the persona method, rather than an activity explicitly inherent to the use of personas and those who partook in role-playing were often unaware of the technique's prevalence [48].

So and Joo [50] undertook two experiments, the first with 18 individuals and the second with 32, to identify whether using a persona increases the originality of ideas during ideation. Using personas was found to help develop more original ideas and overcome design fixation.

To identify the main benefits of personas, Miaskiewicz and Kozar [51] performed a Delphi study with a panel of 19 experts, the majority of whom had created and worked with personas during at least 8 design projects. The top 5 benefits found were [51]:

1. **Audience focus:** Personas help focus product development on users and their goals (rather than the specific limitations or opportunities presented by technology).
2. **Product requirements prioritisation:** Personas help prioritise product requirements and help to determine if the right problems are being solved.
3. **Audience prioritisation:** Personas help prioritise audiences and concentrate focus on the most important users.
4. **Challenge assumptions:** Personas assist in bringing to the surface and challenging long-standing organisational assumptions about the users.
5. **Prevention of self-referential design:** Personas help individuals realise how the users are different from themselves

Conversely, studies also focus on identifying the problems and pain points in persona use. A survey of 60 employees who work for one of a selection of software development companies by Billstrup et al. [49] found that many industry members either were not familiar with personas or did not find them useful. The problems with personas surrounded a lack of knowledge or understanding of them from industry members, a lack of resources in the support and creation of personas, poor development methods, and the use of personas not being integrated into the software development process beyond initial use [49]. These findings support the results of previous studies that found people did not understand how to effectively use personas, often due to how to use personas or the personas themselves being poorly communicated to those who are supposed to use them, and that the personas lacked high-level resources and support for their continued use [49].

Other complaints around personas were that they can be abstract or misleading, and some industry members disagree with the generalisations they represent [53]. One main complaint was that the interpretation and generalisations performed during the persona creation process can result in the complex information available in the raw data being lost [53]. The general personal details of personas were found to be least useful [53], which is supported through Hill et al.'s [55] findings that the inclusion of gender made no particular difference to persona perception. Personas were also found to fall out of date too quickly or fall victim to company policy [49], [52].

3.2 Persona Development

Persona development approaches vary along a scale from completely manual to completely automated. Historically, most persona development approaches have been manual, based on rich qualitative data, such as data from interviews or in-depth case studies, and utilise the deep interpretation and extrapolation able to be performed by individuals on this type of data [26], [27], [59]–[62]. In contrast, an automated approach requires large amounts of quantitative data and uses a clustering algorithm or similar method to automatically develop fully realised personas, resulting in a quicker and less resource intensive persona development process [26], [27], [63]–[66]. Semi-automated approaches sit between the manual and automated methods. Semi-automated persona development can range from an almost completely manual approach with the addition of statistical insights to almost fully automated methods only relying on manual intervention to create the final personas [26], [27], [67]–[70]. The approach chosen for a particular project is dependent on the used case and desired outcomes.

The trends within persona development literature are beginning to favour automated or semi-automated approaches [26]–[28]. Automated and semi-automated approaches have the benefit of being able to be more time and resource efficient and allowing for easy repeatability to keep personas up to date [26], [27]. However, automated approaches are often criticised as being unable to capture the complex concepts and opinions that makes manually developed personas so valuable [26]–[28]. As such, current approaches are rarely near fully automated, instead, semi-automated approaches rely on the manual creation of personas and/or prior data manipulation to mitigate the shallowness of automated results [28].

To develop deeper personas with automated methods Salminen et al. [28] suggests the use of more complex computational techniques is required. For example, using multiple techniques to identify the different elements of a persona. However, to treat the different elements of an individual's behaviours or perceptions as distinct would be a flawed approach, as psychological theory and findings report that elements of an individual do influence each other and cannot be easily separated. Furthermore, the majority of automated or semi-automated persona development approaches rely on one of a small set of clustering algorithms with limited prior analysis towards algorithm choice, rather than taking advantage of the differing nature of clustering approaches [28], [57].

3.2.1 Manual Persona Development

Manual persona development methods primarily employ an expert or group of experts to create personas. Although current trends are moving towards more automated methods, in 2012 more than 80% of papers on persona development used completely manual methods [27]. There are two primary steps in manual persona development:

1. **Data collection:** Manual persona development tends to prefer small amounts of rich qualitative data over larger quantities of quantitative data, such as data gathered through case studies or interviews [59]–[62]. Manual interpretation can identify the meaning and conceptual information captured in qualitative data [26], [27].
2. **Persona creation:** Manual persona creation can take many forms and is generally dependent on the type of data collected. Often, manual methods depend on an expert or small group of experts analysing the data and producing a set of personas they believe are most reflective of the data [59]–[62]. Tools such as sticky notes and mind maps are often used to keep track of

common themes and ideas. Basic statistics are also commonly used where available to comprehend larger data sets or more complex data.

A study performed by Terlouw et al. [60] that created a set of personas to reflect the needs and worldviews of children with autism exemplifies the process and importance of collecting rich qualitative data for persona development. Data was gathered through focus groups ran with a small selection of autistic children (n=8) and their parents (n=6), and interviews were performed with key stakeholders (n=7), such as teachers [60]. An alternate approach by Ward [59], who was creating personas to reflect library users, first performed a workshop with approximately 30 staff members who are associated with the library to gather the common perceptions and assumptions of library users. This data was then supplemented with relevant surveys, ethnographic research, and usability studies during the manual persona development process [59].

Quintana et al. [62] developed the "Persona Party," where small groups of stakeholders are asked to develop personas from the same set of data. Common elements across the personas created by each group are then identified and used to create a single set of personas. During the persona party case study 19 stakeholders were broken into four groups and asked to create persons based on survey data from 3700 individuals.

Of the completely manual persona development studies identified only one, a study by Hirskyj-Douglas et al. [61], performed any external evaluation of the personas developed. Hirskyj-Douglas et al. created a set of dog personas to be used during the development of technology for dogs. The personas created were sent to a panel of five animal computer interaction researchers alongside a short questionnaire to determine the strengths and weaknesses of the "dog-sonas" created.

The primary benefit of manual persona development is that the method allows for deeper interpretation and extrapolation during the data gathering process, which can lead to more complex ideas and concepts [26], [27]. Having a human interpret and analyse the data, also allows for nuance, that automated methods can struggle with, to be captured. Manual development methods are also able to effectively capture the struggles or unique requirements of edge cases. Interviews and case studies can give context and a thorough understanding of requirements that are difficult to capture through other mediums.

However, manual persona development is quite a resource intensive method as both steps, data collection and persona creation, require significant expertise and time dedication [23], [26], [27]. Salminen et al. [23] found that manual persona development often took months to complete and cost tens of thousands of dollars. The high cost of persona development can act as a barrier to persona use and makes maintaining and updating personas more difficult, potentially exacerbating problems identified in persona use research. There is also an increased risk of the individual's personal biases or prejudices affecting the personas that are created since they are often solely determined by an individual

[23], [27]. This effect was demonstrated in the study by Anvari et al. [56], which found that the attitudes towards a set of personas differed between cultures.

As the methods to collect qualitative data are more resource intensive, fewer samples tend to be collected compared to quantitative data. Data from millions of users can potentially be automatically scraped from websites, to reveal information such as how much time is spent on a page or the posts a user makes or interacts with. Capturing the same amount of data through interviews or case studies is generally not feasible, but the data collected can give insights not only into what behaviours are performed but the reasoning and motivations behind the behaviours. Additionally, larger or more quantitative data sets are used are more difficult for an individual to accurately conceptualise and analyse, especially when the data has large dimensionality [23], [27]. However, the smaller sample sizes are still critiqued as less representative of the entire population, meaning the personas developed from the data may not accurately represent the attitudes of the user base [23].

3.2.2 Automatic Persona Development

On the opposite end of the spectrum to manual persona development is automated persona development, where there is minimal, if any, input from an individual during the development process. There are very few fully automated persona development methods and the few that exist are still quite rudimentary with minimal evaluation of the quality and accuracy of the personas created. Fully automated methods include the generation or selection of names, images, and descriptions of the personas.

The general approach to automated persona development is to use a framework where the data is collected, processed, and then fed into a clustering algorithm that groups the data into clusters based on the similarity between data points [63]–[66]. The average of each feature in each cluster is then used as the basis for the personas [63]–[66]. Demographic features are used to generate a name and image to go with each persona and text descriptions are then generated from the key features based on the goals of the personas [63]–[66]. The persona generation methods range from simple approaches, such as using the values of key features to select one of a predefined set of statements, to complex approaches, such as using natural language processing generation [63]–[66]. The generated personas are often exported to a PDF but have also been integrated with a web app for regular use [66].

Automated persona development can address many of the problems found with manual approaches. Once developed, an automated persona development pipeline can be easily applied to a range of datasets and problems. The only requirements are having significant data and hardware capable of running the framework. As computers can easily analyse large amounts of data with high dimensionality, big data, such as data from web scraping, can be used [27]. Using a larger dataset also means the results are likely to be more representative of the audience. Additionally, algorithms do not have any inherent bias, so the only bias present in the personas would be from bias within the data.

However, automated processes have great difficulty capturing complex concepts and opinions [26]. The requirement for numerical data makes complex data difficult to represent and algorithms struggle to capture the nuance and context often present in data. Numerical data cannot capture the same information as textual data, and retaining semantic meaning during the conversion of textual data to numerical data is difficult and complex [27].

When developing personas from the cluster centroids, there is also the challenge of generating persona descriptions that sound believable and natural, rather than a description obviously generated by a computer. One of the key strengths of personas is that they can help elicit empathy and understanding towards the population the persona is designed to represent. If the descriptions are wooden or unbelievable, the persona will not be able to promote significant empathy.

Automated persona development methods have also relied on a small selection of clustering algorithms. Within the existing automated persona development methods, only two clustering algorithms have been applied: k-means [63], [64] and *Non-negative Matrix Factorisation* (NMF) [65], [66]. These algorithms are popular within persona development literature, and minimal or no analysis is performed prior to algorithm or parameter choice [28], [71].

3.2.3 Semi-Automated Persona Development

To combat the problems present in both manual and automated persona development, semi-automated methods are often used [26]–[28]. Semi-automated persona development approaches combine methods from the manual and automated approaches and attempt to take advantage of the benefits of each approach [26]. There are two main approaches to semi-automated persona development. The first approach is to have a primarily manual method that takes advantage of clustering or statistical algorithms to help with data analysis, such as reducing dimensionality, identifying correlations, or finding keywords. The second method is more like an automated approach, using a clustering algorithm as the primary method to identify the personas, but relying on manual intervention prior to and/or after applying the clustering algorithm in an attempt to retain depth [26]–[28]. Both methods sometimes include identifying individuals that closely represent each of the clusters and conducting interviews to get a more in-depth view of their behaviours and motivations, which are then incorporated into the personas developed [70].

Semi-automated methods can use any type of data, but qualitative data or data that contains natural language often requires more manual interpretation to convert the data into a format that can be used by the chosen algorithm [26], [69]. This includes applying algorithms such as *Latent Semantic Analysis* (LSA) to identify key topics within natural language or *Principal Component Analysis* (PCA) to reduce dimensionality and reveal relationships between variables [67], [68]. LSA and PCA are commonly used in approaches that rely on manual intervention to develop the personas, as they allow for large or dense data sets to be simplified while losing minimal detail [68]. The approaches that rely heavily on manual intervention for persona development often use manual persona development methods such as mind maps or keyword identification [68], [69].

When qualitative or natural language data is used with more complex algorithmic approaches, the data is first processed manually or by using an algorithm such as LSA. The study by Dupree et al. [69] is an example of manual data processing. Dupree et al. [69] analysed the results of interviews with 32 participants around attitudes towards online security to develop a set of key concepts and traits, which were then used with a clustering algorithm. One of the most well-known approaches to semi-automated persona development was the LSA methodology proposed by Miaskiewicz et al. [67]. The method has five steps: 1) collect data; 2) calculate cosines using LSA; 3) use cluster analysis to identify personas; 4) interpret results to create personas; and, 5) evaluate through additional interviews [67].

The LSA method also demonstrates the approach where manual intervention is performed after a clustering algorithm is applied to the data to allow believable, nuanced personas to be developed. The manual persona development step generally involves interpreting the cluster centroids to develop personas [67], [68], [69]. This process can include combining similar clusters, removing redundant clusters, and looking at the scope and size of each cluster [68]. Often the most important element of manual persona development is identifying the goals and motivations of the persona that reflects the centroids [70].

During a study to develop a set of personas for an online travel service business by Tu et al. [70], two participants who closely reflected the centroid of each of the clusters developed were contacted for an interview. The interview was used to collect in-depth data on their goals and motives, which was used to flesh out the personas [70]. A similar approach was taken by McGinn and Kotamraju [72], who performed interviews with 26 individuals that represented the 11 clusters generated for a training organisation. The interviews were found to assist in fleshing out the personas and to validate that the personas created reflected the user base [72].

The most prevalent clustering algorithm for semi-automated persona development is *Agglomerative Hierarchical Clustering* (AHC) [28], [71]. The LSA methodology was designed to work with any clustering approach, however, was demonstrated with AHC [67], and the studies by Dupree et al. [69] and Tu et al. [70] both used AHC. After AHC, statistical methods such as *Exploratory Factor Analysis* (EFA) and PCA tend to be more popular than other clustering algorithms. However, there is a move towards the incorporation of more clustering algorithms, as seen by the popularity of k-means and NMF in automated persona development approaches.

Semi-automated approaches can benefit from the strengths of both automated and manual persona development methods; personas can be created that effectively communicate the user's goals and motives while being based on large data sets that are more representative of the population. However, semi-automated approaches are also prone to fall into the pitfalls of both methods. Some semi-automated methods require considerable time and resources, while they are still unable to be used with qualitative data, potentially losing the nuance or complexity that qualitative data can capture [27]. Like the automated approaches, the semi-automated approaches tend to rely on a small, popular selection of clustering algorithms or statistical methods with little documented analysis before choosing an approach [26]–[28].

3.3 Persona Evaluation

Persona evaluation is one of the biggest challenges in persona development. The subjective nature of personas makes them difficult to quantitatively evaluate as there is no specified "correct" set of personas to compare the results of a given development method to. The use and understanding of personas differ across industry and use case, meaning methodologies to evaluate personas can have limited use and subjective opinion can be extremely varied. As such, objective, empirical evaluation of a set of personas or a persona development method is difficult. The evaluation and validation of personas and development approaches tend to be informal and limited [28].

Quantitative approaches to persona evaluation can be employed to evaluate how stable a persona development method is, or how well-formed a set of personas are according to various metrics. Personas created through automated or semi-automated persona development methods can be validated by re-running the development process. If a similar persona set is created, the method is deemed to be stable. Re-running the development process on an additional data set can also help verify the personas by determining whether personas with similar features are created from similar data.

Internal cluster evaluation metrics and statistical methods, such as conducting Chi-squared tests, can evaluate the clusters that a set of personas are based on to a certain extent [28]. However, quantitative evaluation methods generally provide shallow insights and comment more on the quality of the persona development method, rather than the personas created. As such, quantitative evaluation methods cannot comment on the usefulness or believability of a set of personas. Qualitative or mixed-methodology evaluation is required to get an insight into the quality of the personas developed for the given use case.

Most persona development research does not include any form of evaluation of the personas developed. Salminen et al. [28] identified that only 31% of research articles presenting an automated or semi-automated approach to persona development included qualitative or mixed-methodology evaluation. Most qualitative or mixed-methodology approaches to persona evaluation involve performing further interviews or case studies with a subset of the population originally surveyed [28]. These approaches will generally recruit users belonging to each persona for interviews to determine how well the user characteristics align with the characteristics within the personas [28]. The final, optional, step in the LSA driven method proposed by Miaskiewicz et al. [67] was to perform additional interviews to determine how well the personas aligned with the needs of the individuals.

Semi-automated persona development approaches that use additional interviews to flesh out the final personas are often able to use the additional interview step as a form of qualitative validation. McGinn and Kotamraju [72] reported that the additional interview step verified how well the personas developed align with the users they are based off.

Dupree et al. [69] performed an additional study on an additional 212 individuals, who did not participate in the original study. The original survey was re-run as part of this study and used to verify the stability of the study, finding that the same general personas persisted. The additional study also asked each individual which persona they identified most strongly with, how easy they found identifying with a persona, and how representative they felt the personas were. Dupree et al. [69] proposed that the easier it was for an individual to identify with a persona and the more representative they felt the persona was, the more realistic and accurate the personas generated were likely to be.

A small number of studies involve sending the personas for external evaluation by an expert in the field or someone intended to work with the personas. Salminen et al. [73] proposed the 'Persona Perception Scale' specifically for the external evaluation of personas. The scale is designed to focus on key criticisms of personas, such as credibility, completeness, clarity, usefulness, and willingness to use [73]. When the "dog-sonas" developed by Hirskyj-Douglas et al. [61] were sent to five experts in the field for evaluation, the questionnaire asked about the perceived usefulness, competence, and efficiency of the personas, as well as whether there was any room for improvement.

Wölckl et al. [63] performed a workshop with 6 future users of the personas developed, all of which were involved in developing ICT solutions for older adults, the population the personas were designed to reflect. During the workshop, individuals were first asked to rate the understandability, level of detail, level of reality, and intention to use on a 5-point Likert scale. This was followed by a discussion on missing information and ideas for improvements.

The primary weakness of qualitative persona evaluation is that it is often performed with a small selection of individuals, often less than 10. As such, the evaluation is subject to the biases and preferences of those individuals. Users evaluate personas based on how well they represent the users but cannot evaluate the personas usefulness or how likely the personas are to be used by experts cannot be determined. Whilst, when a set of personas is evaluated by experts in the field, their usefulness can be determined but how representative the personas are of the population is often based on anecdotal information.

3.4 Summary

Personas represent audience segments in a manner that facilitates communication and promotes empathy. The efficacy of a set of personas heavily depends on their use and integration. However, personas have been found to facilitate focussing on and prioritising the audience and their challenges. There is a range of approaches to developing effective personas that range from an individual who performs interviews and manually develops personas, to automated systems that scrape a large amount of data from websites and use clustering algorithms to automatically develop personas. Each approach to persona development has unique benefits and drawbacks. Personas are difficult to evaluate, as there are no correct answers, and the usefulness of a persona is often subjective, which exacerbates the difficulty in choosing a persona development method.

A key aspect to the automated or semi-automated development of a set of personas is the clustering algorithm used to develop the base personas. However, there are numerous approaches to clustering, each of which target's clusters of a specific nature. Similar to persona development, clustering algorithms are notoriously difficult to evaluate and face many challenges when applied to subjective fields.

Chapter 4: CLUSTERING ALGORITHMS

Clustering, or cluster analysis, attempts to partition or identify clusters within an unstructured data set so that elements within a cluster have a higher similarity than elements between clusters [74], [75], [29]. Clustering is seen as an area of unsupervised *Machine Learning* (ML) although many approaches originate in fields outside of ML. There are two overarching branches of ML; supervised and unsupervised [74]. Supervised ML occurs when the data is already labelled or structured so that a model can be trained to replicate the desired results or make predictions based on historical data [74]. Unsupervised ML occurs when there are no predefined answers and the job of the model is to find structure or patterns within the data given [74].

Unsupervised ML algorithms are notoriously difficult to evaluate as there is no ‘correct’ answer available [29]. This challenge is even more prominent in clustering, as the performance of a clustering algorithm is often highly dependent on the underlying data and the type of clusters desired [29]. Furthermore, there may be more than one valid set of so it is impossible to concretely say if one set of clusters is more accurate than another [29].

As a result, a new clustering approach or algorithm is often needed to fit a new application. Clustering is a diverse field that can have a wide range of applications across many industries; some of the most common applications of clustering algorithms include image segmentation, object recognition, information retrieval, data analysis, marketing, and bioinformatics [76], [77]. However, choosing a clustering algorithm to apply is difficult, as the lack of empirical evaluation makes applying methods such as hyperparameter tuning to automatically identify the best clustering algorithm impossible.

4.1 Clustering Approaches

The various approaches to clustering each attempt to identify clusters of a different nature [29]. The underlying clusters in a data set can vary in size, shape, and density. Not only the clustering approach but the specific algorithm and parameters used can all affect the clusters identified [29]. Due to the difference in underlying clusters, algorithm performance is highly dependent on the data set [29].

Figure 4-1 gives a comparison of clusters identified by Agglomerative Hierarchical Clustering (AHC) using four different similarity metrics, or linkages. Figure 4-1 exemplifies how even similar clustering algorithms can find vastly different sets of clusters of varying quality and the impact the dataset can have on an algorithm’s performance. In the leftmost column, AHC with single linkage identifies the desired clusters within the first two data sets but gives the worst performance on the third data set, clustering the entire data set together. Alternatively, AHC with Ward's linkage performs poorly on the first two data sets but develops the most accurate cluster set on data set three.

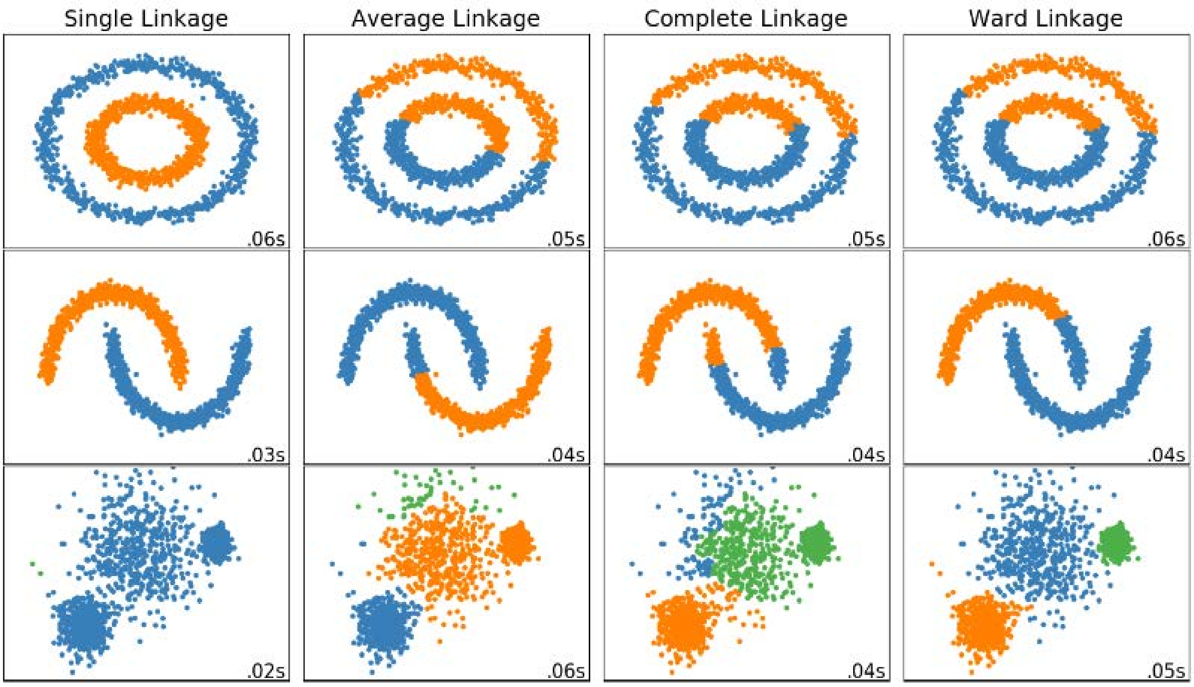


FIGURE 4-1: COMPARISON OF CLUSTERS DEVELOPED BY AHC BASED ON LINKAGE [78]

There are numerous approaches to clustering, these approaches fall under one of into two broad categories, hierarchical or partitioning [29], [75]–[77]. Hierarchical algorithms build up a tree of clusters based on similarity, which can be cut at any point to get the desired number of clusters [29], [75]–[77]. Hierarchical clustering approaches also include approaches based on graph theory. Partitioning algorithms attempt to find a clean division between elements to create the desired number of clusters [29], [77]. Any clustering approach that is not hierarchical is considered a partitioning algorithm. As the clustering algorithm has such a significant impact on the clusters identified, understanding the key approaches to clustering is imperative.

4.1.1 Hierarchical Approaches

Hierarchical clustering algorithms, usually depicted as a binary tree, can be split at any point to create the desired number of clusters [29], [75]–[77]. There are two approaches to hierarchical clustering; top-down, called divisive clustering, and bottom-up, called agglomerative clustering [29], [75]–[77]. Divisive clustering starts with all the data points in one cluster, which is recursively split until each value is in a cluster by itself [29], [75]–[77]. Whereas *Agglomerative Hierarchical Clustering* (AHC) starts with each datapoint as its own cluster and recursively combines the two most similar clusters until all of the data points are in a single cluster [29], [75]–[77]. AHC is one of the clustering algorithms most commonly applied to persona development [71].

AHC is often defined in terms of linkage, which is the metric used to determine the similarity of two clusters. There are four common methods of linkage:

1. **Single Linkage:** Measures the distance between the two closest data points [76], [77].
2. **Complete Linkage:** Measures the maximum distance between any two points within a cluster [76], [77].
3. **Average Linkage:** Measures the average distance between each value in both clusters [76], [77].
4. **Wards Linkage:** Measures the within-cluster variance [79].

The classical hierarchical clustering methods all have the benefits that the number of clusters does not need to be known before applying the algorithm and the hierarchical nature of the approach can provide insights into the relationships between and within clusters [75]–[77]. The algorithms based on hierarchical clustering also share in a series of criticisms. Namely, being sensitive to noise and outliers, and having high time and computational complexity, which limits the ability to apply these methods to large data sets [75]–[77]. They can also only be applied to numerical data since the similarity metrics are based on comparing the numeric values [75]–[77].

There are many extensions on AHC that attempt to solve the weaknesses of AHC, prime examples include; *Balanced Iterative Reducing and Clustering using Hierarchies* (BIRCH) [80], *Clustering Using Representatives* (CURE) [81], and ROCK [82]. BIRCH is an agglomerative hierarchical algorithm that was developed with the aim of improving efficiency for use on large data sets and reducing the impact of outliers [80]. The basis of BIRCH is the use of *Cluster Features* (CF) to represent a sub-cluster in a CF-tree, which greatly reduces the amount of data required, while still capturing vital information [80].

The primary aim of CURE was to find clusters of more sophisticated shapes, while still being able to deal with large data sets and outliers [81]. CURE's main point of difference from other hierarchical algorithms is that each cluster is represented with a fixed number of well-scattered points, which are moved towards the cluster centre by a specified fraction after generation [81]. CURE also uses a combination of random sampling and partitioning to reduce computational complexity [81]. Compared to BIRCH, CURE achieves a better cluster quality; however, BIRCH has a better time complexity [75], [76].

ROCK introduces the concept of links, which measure the similarity between data points in a manner that extends to non-metric similarity measures [82]. By introducing links, ROCK can be run on categorical data [82]. Like CURE, ROCK uses random sampling to improve the computational complexity, but, ROCK has a much worse time complexity than both BIRCH and CURE [75].

4.1.1.1 Graph Theory Approaches

The methods mentioned thus far have followed the traditional approach to hierarchical clustering. However, there are alternative approaches; primarily, approaches based on graph theory. Graph-theoretic clustering algorithms represent the data as a graph, where each node is a data point, and the edge between two nodes is the relationship between those data points [29], [75]–[77]. Clustering is then treated as a graph partitioning problem and clusters are created so that the edge density is higher within clusters than between clusters [29], [75]–[77].

One of the main motivations behind graph-theoretic clustering is that when dealing with spatial or geographic data the graphical representation of the data is closer to the real-life representation of data. Graph-theoretic algorithms also tend to have good scalability and accuracy, and can find clusters with arbitrary shapes [75]. However, the time complexity increases dramatically as the graph complexity increases, so graph-theoretic algorithms are not suitable for high dimensional data [75]. Graph-theoretic algorithms are also weak at identifying outliers and overlapping clusters [75], [76].

During the current study, two graph-theoretic clustering algorithms were used: *Minimum Spanning Tree* (MST) [83] based clustering and Spectral graph theory clustering [84], [85]. Single linkage and complete linkage AHC algorithms are technically examples of graph-theoretic clustering. MST based clustering first develops an MST, a subset of the whole graph that connects all the nodes without any cycles and the minimum possible total edge weight [83]. Once the MST has been extracted, the edges with the largest weights are cut to create the desired number of clusters [83].

Spectral graph theory is an extension of the graph theory approach to clustering [75], [76]. The defining difference of the spectral approach is that it performs clustering in a lower dimensional space by using eigenvectors based on a similarity matrix of the data [75], [76]. Once the similarity matrix has been developed there are three general steps to spectral clustering: 1) apply a graph Laplacian to the similarity matrix; 2) compute the eigenvectors of the Laplacian matrix; and, 3) create clusters based on the eigenvectors [86].

There are two main methods of normalized spectral clustering, one proposed by Shi and Malik [84] (SM), and the other by Ng, et al [85] (NJW). The methods differ in which graph Laplacian they use and how they create the clusters. The SM approach creates clusters by minimizing the normalized cut and uses the eigenvectors to generalize the problem [84]. While the NJW approach first applies a clustering algorithm, usually k-means, to the feature space corresponding to the first *k* eigenvectors [85]. In the current project, the SM method is used.

The main benefit of spectral graph algorithms is that they work well on clusters of an arbitrary shape and with high dimensionality, as no assumptions are made about cluster shape [75], [76]. Additionally, spectral algorithms always converge at the global optimum, are not sensitive to outliers, and can be implemented efficiently for large data sets [75], [76]. However, the main drawback of spectral clustering is that choosing what type of similarity graph to use is not a trivial issue and has a significant impact on algorithm performance [75], [76].

4.1.2 Partitioning Approaches

Partitioning algorithms attempt to identify all the clusters simultaneously by trying to find clean divisions between clusters, instead of developing a hierarchical structure [29], [77]. There are many different methods of identifying clean divisions, and most clustering algorithm approaches are a type of partitioning algorithm. The most famous and widely used clustering algorithm, both within persona development and other fields, is a simple partitioning algorithm; k-means [28], [29], [71], [76].

K-means was introduced over 50 years ago by multiple people in different industries [29] and is most often attributed to Steinhaus [87], Forgy [88], Ball and Hall [89], MacQueen [90], and Lloyd [91]. The approach behind k-means can be described as simple partitioning, and the premise is to find the optimal cluster centroids, the centre point of the cluster, to use to partition the data set. The optimal cluster centroids are identified through a four-step iterative process:

1. Initialize k random points as the initial cluster centroids, where k is the desired number of clusters.
2. Assign each data point to its closest centroid, creating the clusters.
3. Re-calculate the cluster centroids to be the mean of each cluster.
4. Repeat steps two and three until the centroids no longer change during step three.

Surveys performed by Xu and Tian [75], and Xu and Wunsch [77], detailed the primary strengths and weaknesses of the k-means algorithm. The strength of k-means lies in its simplicity, which makes it easy to understand and implement. K-means also has a good time complexity which lends itself well for use on large data sets. Due to these strengths, k-means is often the go-to clustering algorithm, being one of the most accessible and widely used clustering approaches. Although, k-means does have many drawbacks.

The main drawback of k-means is that it requires the value of k, the number of clusters, to be defined, which has a significant impact on the performance of the algorithm and is difficult to determine. This is a drawback of most partitioning clustering algorithms. Another drawback of k-means is that the performance is tied to the random initialization of the centroids, which tend to converge at the local optimum. K-means will also generally find spherical clusters, performing poorly on data sets with non-convex clusters. Since k-means is based on mean values, it shares the weaknesses that mean calculation has; sensitivity to noise and outliers and is only able to be performed on numerical values.

There have been numerous iterations of k-means, one of the most popular is k-means++ [92] which uses a different initialization method for identifying the initial centroids. K-means++ selected the initial centroid values based on their distance from other centroids, rather than through random initialisation as in the original algorithm [92]. As the centroids are ensured to be spread out K-means++ can lead to better performance than random initialisation but can still get stuck at a local optimum [92].

Another variant of k-means is k-medians which uses the medoid, the median point of the cluster, rather than the centroid. One key strength of the k-medians approach is that k-medians can be applied more accurately to discrete data [75], [77]. K-medians also benefits from the strengths the median value has over the mean, primarily being less sensitive to outliers and potentially being a more representative central measure of a cluster [75], [77]. However, K-medians has a worse time complexity compared to k-means. K-medians is also still confined to only finding convex clusters and tend to perform more poorly on high dimensional data.

There are many alternative approaches to partitioning based clustering that attempt to identify different types of clusters or are designed for different data sets. The following sections will address the most common clustering approaches and the algorithms that are used during the current study. The algorithms used have been selected to represent the range of approaches available, rather than comparing many clustering algorithms that are based on the same general approach.

4.1.2.1 Density Approaches

After AHC and K-means, the most widely known clustering algorithm is *Density Based Spatial Clustering of Applications with Noise* (DBSCAN) [93], a clustering algorithm that uses a density-based approach. The density approach to clustering asserts that clusters should be defined as regions of high density separated by regions of low density [75], [76]. By using this definition, clusters of any shape and size can be identified and the number of clusters does not need to be predefined [29], [75] Which addresses two of the primary weaknesses of k-means.

Density clustering algorithms are not affected by noise and outliers since they do not cluster values that are not in an area of high density [75]. The weakness of the density approach is that the algorithms perform quite poorly when the density of the space is not even [75]. Density methods have poor time and computational complexity [75].

DBSCAN defines three types of data points, core points, border points, and noise points. Core points are defined by having at least a specified minimum number of points within a specified radius [93]. A point that is reachable by a core point, but itself does not meet the criteria of a core point, is a border point [93]. Noise points are points that satisfy neither of these criteria [93]. The algorithm goes through every data point in the set, marking the data point as a core, border, or noise point, clusters are then defined by sets of connected core and border points [93].

Another clustering algorithm based on the density approach is *Ordering Points To Identify the Clustering Structure* (OPTICS) [94]. OPTICS is an extension of DBSCAN that aims to be able to identify clusters in data of varying density and overcome the parameter sensitivity of DBSCAN [94]. OPTICS achieves these goals by giving the clusters in an ordered manner so that the points that are closest to each other are neighbours and clusters can be found for different values of the radius [94].

4.1.2.2 Kernel Approaches

Similar to the density approach, the kernel-based approach also focusses on being able to find arbitrarily shaped clusters [75]. The data set can be converted into a higher dimensional space through the application of a non-linear kernel function, which allows approaches that can only identify convex clusters to identify non-convex clusters [75], [77]. There are a wide range of kernels available, each based on a mathematical function. Some of the most common functions used are based on the *Radial Basis Function* (RBF), polynomial function, sigmoid function, cosine function, or the Laplacian function. The kernel transformation makes the algorithm much more robust against noise and outliers [75], [77]. However, applying the kernel function comes at a considerable time cost, meaning that kernel-based methods are rarely applied to large or high dimensional data sets [75], [77].

There are two primary approaches to kernel-based clustering algorithms; applying a more traditional clustering algorithm to data that has been transformed through a non-linear kernel, and algorithms that have been specifically designed to use kernels. Kernel specific algorithms can deal with noise and outliers effectively and separate overlapping clusters [75], [77]. A weakness of both types of kernel methods is that they are highly sensitive to the type of kernel used. In the current project, both approaches to kernel clustering were used. K-means was applied to data transformed through one of a range of kernels to create kernel-k-means [95] and *Support Vector Clustering* (SVC) [96] represented the kernel specific approach.

SVC first transforms the original data space with the kernel function, usually the RBF which is the standard kernel function used with support vector machines [96]. SVC then finds the smallest sphere which can enclose all the data points within the transformed space [96]. When the sphere is translated back into the original data space, it creates a set of contours that act as cluster boundaries [96].

4.1.2.3 Distribution Approaches

The distribution approach, also referred to as the mixture model approach, is based on the idea that several data distributions exist within the data set, and values that belong to the same distribution are clustered together [29], [76], [77]. The different distributions within a data set can be from completely different functions or the same function with different parameters [29], [76], [77]. The most popular distribution-based clustering algorithm is *Expectation-Maximisation* (EM) clustering [97].

EM clustering operates very similarly to k-means. As described by Do and Batzoglou [98], EM Clustering consists of two steps; the Expectation step (E-step) and the Maximisation step (M-step). The distributions within the data set are represented by their mean and standard deviation, which are initially set to random values. During the E-step the probability of each data point belonging to each distribution is calculated, and each data point is assigned to the distribution the data point is most likely to belong to. Then, during the M-step the parameters of each distribution are re-calculated based on their current assignments. Each step is then repeated until the algorithm converges.

4.1.2.4 Model Approaches

Very similarly to the distribution approach, the model approach is based on the premise that there are separate models that make up the data set, each of which represents a distinct cluster [75], [76]. Most model-based approaches are based on a decision tree or a neural network as the base model [75], [76]. By some definitions, distribution approaches are categorized as a model approach, as the approach is based on a statistical model.

The general advantages of model approaches are that well developed models allow the data to be described effectively and particular models can have significant advantages in specific fields [75]. The drawbacks of model approaches are that they generally have a high time complexity, and in the same vein as particular models having advantages for specific fields, if the model's premise does not fit with the field it is being applied to, the method is at a significant disadvantage [75].

In the current project an approach based on a *Self-Organising Map* (SOM), a method using a neural network model defined by Kohonen [99], is used. A SOM, also known as a self-organising feature map or a Kohonen map, is a single layer, fully connected neural network, usually visualised as a lattice structure. In the map, a neuron has a series of weights attached that represent a potential data point. Each data point is iterated over, where the neuron that is closest to the data point 'wins' and all the neurons within a certain radius are updated to be slightly closer to the winning neuron. The SOM iterates over the data a set number of times, producing a lattice structure that can be more easily clustered. The most simplistic approach to SOM clustering is to set the number of nodes in the neural network to the number of desired clusters.

4.1.2.5 Metaheuristic Approaches

Metaheuristic algorithms is a category of optimisation algorithms that are inspired by nature or natural processes in some way [100], [101]. By treating centroid identification as an optimisation problem, metaheuristic algorithms can be applied to clustering. The general benefit of metaheuristic algorithms is that they tend to settle at the global optimum, instead of being stuck at a local optimum [75]. However, metaheuristic algorithms also have a very high time and computational complexity, which can make them unsuitable for large or high dimensional data sets [75]. There are many types of meta-heuristic algorithms and there is no agreed method of categorization. For the current study three main categories of approaches have been used:

1. **Evolutionary Theory Approaches:** Many metaheuristic algorithms are inspired by evolution or natural selection [100]. Evolutionary algorithms begin with a population, a set of potential solutions, which attempt to survive in an environment where survival is determined by a criterion of fitness [100]. The members of the population pass properties to their children through various mechanisms, such as genetic crossover and adaption [100].
2. **Swarm Intelligence Approaches:** Swarm intelligence algorithms are inspired by the collective intelligence of a population [100], [101]. Through the simple behaviours of many unsophisticated agents, complex behaviours can be performed [100], [101]. The behaviours of many swarming creatures, such as ant colonies, bee colonies, flocks of birds, or schooling fish, are often used as the basis for a swarm intelligence metaheuristic algorithm.
3. **Other Approaches:** Other approaches to metaheuristic algorithms can be based on a variety of natural processes, which can be biological, physical, or cultural in nature [100]. Some processes include annealing, immune systems, and the transfer of information in frogs.

In the current project, an algorithm from each category of metaheuristic algorithms was investigated. The *Genetic Algorithm* (GA) [102] from evolutionary theory, the *Artificial Bee Colony* (ABC) [103] algorithm from the swarm intelligence category, and the *Shuffled Frog-Leaping Algorithm* (SFLA) [31] were selected. All the algorithms are dependent on a fitness function, to determine if one solution is better than another. The selection of a fitness metric is affected by the same issues that plague cluster evaluation, in the current study Euclidean distance was used as it is industry standard for these algorithms.

The GA, applied to clustering by Maulik and Bandyopadhyay [102], mimics biological evolution, where the population is a set of chromosomes, strings of real numbers representing the cluster centres. Each chromosome is assigned a number of duplicates within the mating pool based on the chromosome's proportional fitness. Each chromosome in the mating pool then has a likelihood of experiencing crossover or mutation. During crossover, information is exchanged between two parent chromosomes to develop two child chromosomes. During mutation, one of the genes in the chromosome is multiplied by a random value. The process of selection, crossover, and mutation is repeated for a fixed number of generations or until a termination condition is satisfied.

ABC algorithms mimic the foraging behaviour of a honey bee swarm [103]–[105]. The foraging behaviours of honey bees rely on three types of bees, employed bees, onlooker bees, and scout bees, and their interactions with a food source [104]. The goal of the bees is to find the best food source, where the food source represents a potential solution [104]. Scout bees find new food sources, employed bees visit known food sources and adjust them based on fitness, and onlooker bees wait in the dance area [104]. When employed bees return to the dance area, they share their information with the onlooker bees, who become employed bees.

The SFLA is inspired by memetics, the evolution of information and culture, in frogs [31], [32]. SFLA was first applied to clustering by Amiri, Fathian, and Maroosi [32]. In the SFLA each frog represents a potential solution, and the population of frogs is divided into memeplexes. Within each memeplex, the frogs share their ideas through memetic evolution. A predefined number of phases of memetic evolution are performed before the memeplexes are shuffled together to create new memeplexes, effectively sharing the ideas between memeplexes. This process is repeated either until convergence or a set number of iterations have been performed.

4.1.2.6 Ensemble Approaches

Ensemble algorithms combine the results of multiple algorithms to develop more robust and stable clusters [29], [106]. There are two main parts to an ensemble algorithm; the set-up of the various clustering algorithms used, the ensemble, and how the results are combined to determine the final clusters, the consensus function [29], [75], [106]. Ensemble approaches have the benefit of being easy to scale and parallelize, and can take advantage of the strengths of the clustering algorithms used as part of the ensemble [75]. The primary drawback of ensembles is that there are a lot of parameters and setup to consider [75]. The consensus function used is also integral to the success of the algorithm and is often the weak link in ensemble algorithms [75].

There are numerous ways to configure a clustering ensemble and the two principal elements are how the cluster ensemble is set up and the consensus function used [106]. A cluster ensemble aims is to foster diversity within the ensemble, as that has been found to give the best results [106]. There are four general approaches to developing diverse cluster ensembles [29], [106]:

1. **Homogenous:** One clustering algorithm is used with different parameters, which can include the number of clusters, or can be run multiple times with the same parameters to account for random initialisation.
2. **Heterogeneous:** Multiple clustering algorithms are used.
3. **Random subspace/random sampling:** A subset of the features or rows is used with each algorithm. This can be used with multiple instances of the same algorithm, or alongside a homogeneous or heterogeneous approach.

4. **Mixed Heuristics:** Any combination of the above variations.

After the approach to creating the ensemble is selected, the consensus function must then be selected. The consensus function has a significant impact on the ensemble, as it determines how the final clusters are created [106]. All consensus functions use some form of information matrix to represent the clustering results [106]. There are four broad categories of consensus functions: direct, feature based, pairwise-similarity based, and graph based [106]. In the current project two consensus functions were used: 1) a simple voting direct consensus function, known as the basic consensus function, and 2) a graph-based consensus function based on *Non-negative Matrix Factorisation* (NMF).

4.1.2.7 One-Off Approaches

Some algorithms do not neatly fit into any of the general approaches to clustering. Two of these algorithms were used in the current project, *Affinity Propagation* (AP) [107] and *Non-negative Matrix Factorization* (NMF) [108]. NMF clustering is one of the most common techniques used for semi-automated and automated persona development [28], [71].

AP clustering regards all the data points as a potential cluster centre and the negative distance between two points as their affinity [107]. A process of message sending begins, where each data point sends a responsibility message to each nearby point, reflecting how suited the other point is to be their exemplar [107]. The point then receives a return availability message, reflecting how appropriate it would be for the data point to choose the second data point as its exemplar [107]. Through each point choosing an exemplar, the clusters are formed.

NMF finds two non-negative matrixes, W and H, whose product approximates the non-negative matrix of the data set [108]. Each row in W represents a data point in terms of its importance to component c, and each column in H gives the importance of a feature for component c [108]. By setting the number of components to the number of clusters desired, clusters can be created by determining the component each data point has the strongest affinity for [108].

4.2 Clustering of Categorical Data

Most clustering algorithms are designed for continuous numeric data, which can make applying clustering difficult in domains where categorical data is more common. However, multiple approaches can be taken to allow quality clustering to be performed on categorical data. The two general approaches are to design a clustering algorithm that is designed for categorical data or to convert the categorical data into numerical data.

An example of an algorithm designed for categorical data is ROCK [82], a hierarchical clustering algorithm that can be used on both categorical and numeric data. Most popular clustering algorithms have adaptions for use on categorical data and there are many algorithms and approaches to clustering categorical data that are unique. Another option is to use a similarity metric that supports categorical data. Most algorithms are based on Euclidean distance, which is restricted to continuous numeric data, but can be adapted to be based on any similarity metric. However, a drawback of using dedicated algorithms or similarity metrics is that they do not always support numerical data. Which makes them a good option for purely categorical data but causes the reverse problem for mixed data.

When the data contains both numeric and categorical data, the best option is often to convert the categorical data into numeric data. Ordinal data, such as $x = [low, medium, high]$, can often be directly mapped to numerical data, $x = [1, 2,3]$. Boolean data can be converted to numeric data using zero and one. For more complex categorical data one-hot encoding, or dummy variables, are often used. One-hot encoding creates a Boolean feature for each of the categorical values. Usually, one of the options is represented by a false in all columns, resulting in $n - 1$ columns being added. One hot encoding is the best option for categorical values that do not have a linear relationship. However, one hot encoding can result in many additional rows with little substance being added, which can create sparse data sets with high dimensionality.

4.3 Cluster Evaluation

Clustering algorithms are notoriously difficult to evaluate, as there is no ground truth available to compare results to and multiple sets of clusters created from the same data set could be equally valid [29]. Furthermore, due to the variability of algorithm performance based on the data set and types of clusters desired, some believe that clustering algorithms cannot be effectively evaluated independently of the context in which they will be used [30]. Domain-specific evaluation is often highly subjective, as it requires evaluating the usefulness of a cluster set in a use case and tends to require considerable time and resources to perform. Outside of domain-specific evaluation, there are three approaches to evaluating a set of clusters: 1) internal metrics; 2) external metrics; and, 3) meta-criteria [75], [76].

4.3.1 Internal Evaluation Metrics

Internal evaluation metrics measure the cluster quality with similarity metrics, usually measuring: the inter-cluster separability; the intra-cluster homogeneity; or a combination of both [76]. Internal metrics can be useful for comparing two sets of clusters, however, they reveal little about the general algorithm performance and cannot be used between data sets, as the values are often relative [30]. Furthermore, most internal metrics favour particular types of clustering algorithms, making them quite bias [30]. For example, a metric measuring cluster separability will prefer an algorithm with a similar basis, such as k-means, while a metric measuring the intra-cluster homogeneity will prefer an algorithm based on the density approach.

During the current study, three popular internal metrics were used to provide some general information and act as a basis for basic evaluation. The three internal metrics used were:

1. ***Silhouette Coefficient* (SC) [109]:** The SC is based on how well defined the clusters are, taking the intra-cluster distances and the distances between a given cluster and the next closest cluster into account. SC scores are bounded between -1 and 1, where -1 represents incorrect or overlapping clusters and 1 represents highly dense and well-separated clusters.
2. ***Calinski-Harabasz Index* (CHI) [110]:** The CHI also attempts to score a set of clusters based on cluster definition, using the ratio of the sum of between-cluster dispersion and the within-cluster dispersion. Higher CHI scores relate to better defined clusters, and the values are not bounded.
3. ***Davies-Bouldin Index* (DBI) [111]:** The DBI evaluates a set of clusters on how well separated they are, taking both the distance between clusters and cluster size into account. Lower DBI scores represent more distinct cluster partitions, with 0 being the best possible score.

4.3.2 External Evaluation Metrics

External evaluation metrics require having ground truth labels available and are most commonly used to verify a proposed clustering algorithm or approach [30], [112]. External evaluation metrics include metrics such as accuracy, completeness, and mutual information. However, external evaluation is often flawed as the set of clusters given as the 'correct' clustering may not reflect the 'best' or most 'natural' clusters for the data, or may be based on theoretical differences that are not sufficiently represented on the data [30], [112]. Within one data set there may be multiple correct sets of clusters, meaning just because an algorithm does not find the expected clustering, does not necessarily mean that the algorithm did not find a valid set of clusters [112].

Furthermore, external evaluation metrics often use artificially generated data sets that are not reflective of real-world data sets [30]. Even evaluations that use real-world data sets are only applicable to similar real-world applications or problem areas [30]. Labelled data often does not take noise into account, with all values assigned to a class, which makes the evaluation of clustering algorithms that account for noise more difficult [112].

4.3.3 Meta-Criteria Evaluation

Meta-criteria can be useful in determining the quality of a clustering algorithm rather than the quality of the cluster set developed. Stability is a popular method of evaluation. If a clustering algorithm re-run on the same data consistently develops the same clusters the algorithm is considered stable [30]. Unstable algorithms are considered unreliable, and generally unsuitable for further use [30].

Some meta-criteria border internal metrics, such as cluster size and cluster significance. Cluster size is an important meta-criteria as small clusters cannot be representative of a significant portion of the population and are likely to represent outliers rather than legitimate groups within the data [30]. Using statistical tests to determine if the clusters developed differ significantly from each other can also be useful [30]. If a test finds a pair of clusters do not deviate from each other, the algorithm has likely identified overlapping clusters rather than the desired, well-separated clusters [30].

4.4 Hyperparameter Tuning of Clustering Algorithms

There are numerous approaches to clustering and both the algorithm and parameters selected can have a drastic impact on the clusters developed. The selection of an algorithm and parameters, a process known as hyperparameter tuning, is a considerable challenge when applying clustering to a real-world problem. Multiple iterations and considerable domain knowledge is often required to find an optimal algorithm configuration, and the process is often long and tedious [113], [114]. In supervised problems, where the ground truth is available and evaluation is straightforward, hyperparameter tuning is often automated. However, the evaluation of clustering algorithms is complex and rarely objective.

Current automated hyperparameter tuning methods for clustering algorithms rely on internal evaluation metrics [113], [115], [116], or having some ground truth labels available so external evaluation metrics can be used [114], [117]. Hyperparameter tuning with external evaluation metrics becomes a semi-supervised problem, rather than an unsupervised problem. Both approaches to hyperparameter tuning fall victim to the same criticisms of the evaluation metrics used and cannot take the use case of the clusters into account [30].

The closest instance to a hyperparameter tuning approach for clustering algorithms found is Hypercluster, which is an framework designed to facilitate parameter selection for clustering algorithms developed by Blumenberg and Ruggles [115]. Hypercluster tests multiple clustering algorithms and parameter combinations and outputs a heatmap of each combination's performance in terms of a series of internal and external metrics. Hypercluster, however does not aid in the interpretation of the results beyond the development of a heatmap and, as such, parameter selection is still a manual process.

Domain-specific evaluation is impossible to integrate into a completely automated hyperparameter tuning solution. Furthermore, as the effect that hyperparameters have on clustering results cannot be described through a convex function, inferences about the effect of the hyperparameters cannot be drawn [115]. Thus, an exhaustive grid search is required to find the optimal hyperparameters, and shortcuts cannot be identified [115]. Manually performing an exhaustive grid search across an extensive range of algorithms and parameters would be a time-intensive and cumbersome process.

4.5 Summary

Clustering algorithms have significant potential to be applied to areas such as persona development. Clustering is a diverse problem with numerous approaches, and each approach is proficient at finding clusters of a particular nature. One of the primary challenges of clustering is that algorithm selection can have a drastic impact on performance and the performance of a particular algorithm is often dependent on the nature of the clusters in the data [29]. Even two similar algorithms may find completely different sets of clusters in the same data set [29]. However, cluster evaluation is notoriously difficult and useful evaluation is rarely empirical [30]. This is due to the unsupervised nature of clustering, and that there may be multiple valid cluster sets within one data set [30], [112]. As a result, quality automated hyperparameter tuning is almost impossible.

Due to the difficulty in evaluating and selecting a clustering algorithm, the first step of the current project is to develop a framework that will facilitate the identification of quality clustering algorithms. The framework will approach hyperparameter tuning from a semi-automated standpoint by finding the results of multiple clustering algorithm and parameter combinations, and automatically ruling out poor results. The semi-automated approach will allow for quality domain-specific evaluation. To further facilitate evaluation, the framework will perform the first steps of semi-automated persona development. The framework will also address SRQ1: How can a range of clustering algorithms and the clusters they develop be efficiently evaluated and compared?

Chapter 5: THE HYPERSONA FRAMEWORK

The clustering algorithm and parameters selected has a significant impact on the clusters developed from a data set, and thus the personas created. In the supervised machine learning fields, the process of selecting an algorithm and parameters, known as hyperparameter tuning, is often automated. Automated hyperparameter tuning methods rely on objective evaluation metrics to determine the best algorithm and parameters to select. However, there are no objective evaluation metrics for clustering algorithms as there are no ground-truth answers to compare results to. Current approaches towards automated hyperparameter tuning of clustering algorithms rely on internal metrics, which are often biased towards certain algorithms, or have some ground truth labels available, moving the problem into the semi-supervised space.

To facilitate the evaluation and selection of a clustering algorithm and parameter combination for persona development, and address the first SRQ, *How can a range of clustering algorithms and the clusters they develop be efficiently evaluated and compared?*, HyPersona was developed. HyPersona approaches hyperparameter tuning and persona development from a semi-automated standpoint, automating elements where possible, and creating outputs and graphs to minimise the effort required during manual domain-specific evaluation. The HyPersona framework applies an exhaustive grid search over a range of clustering algorithms and parameter combinations, followed by simple evaluation to rule out results that do not meet required thresholds. The use of simple evaluation to rule out invalid results minimises the amount of manual evaluation required. HyPersona develops graphs and primitive personas for each potential algorithm-parameter combination to simplify the manual evaluation and persona development processes.

During the development of HyPersona, a novel internal evaluation metric designed to reflect the quality of a cluster set for persona development was created. The internal evaluation metric, *Average Feature Significance* (AFS), is based on the premise that quality personas should have unique attributes that significantly differ between clusters.

5.1 The HyPersona Framework

The HyPersona framework went through multiple development iterations. An early version of HyPersona was presented at the 36th IEEE/ACM International Conference on *Automated Software Engineering* (ASE 2021) as a part of the Late Breaking Results track [118]. The early version did not yet involve AFS, the use of thresholds, or develop early-stage persona. The final iteration of HyPersona, as presented in this chapter, was published in Array [119] and the source code is available at [120]. The contents of this chapter are available in the related publications.

The following section is published in [119]

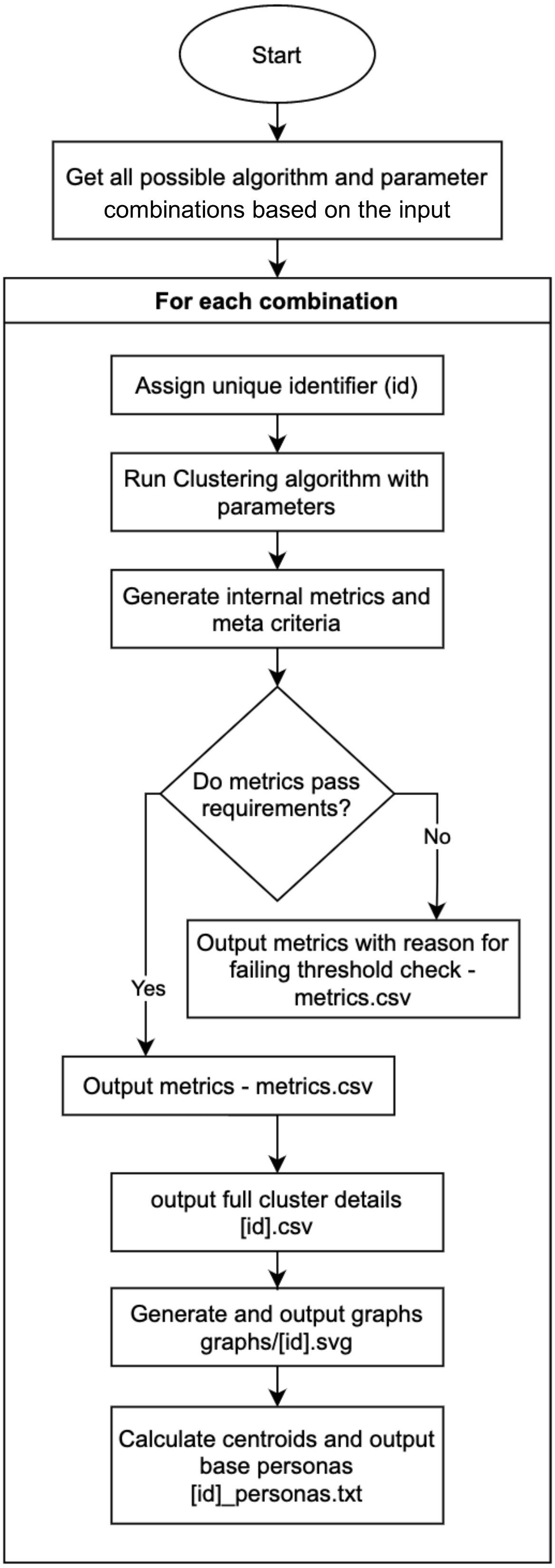


FIGURE 5-1: GRAPHICAL REPRESENTATION OF THE AUTOMATED PORTION OF THE HYPERSONA FRAMEWORK

The core of the HyPersona framework is to perform an exhaustive grid search over a range of algorithm and parameter combinations, calculate relevant metrics to be used for simple evaluation, and then output information on each combination that can be used to identify the most appropriate algorithm. HyPersona extends upon the semi-automated clustering algorithm hyperparameter tuning framework previously presented in [118] through the application of thresholds to rule out invalid cluster sets for persona development, the introduction of AFS, and the creation of early-stage personas. Figure 5-1 gives a graphical overview of the automated portion of HyPersona.

HyPersona takes a dictionary that details each of the algorithms and parameters to be tested, which is expanded into a list of all possible algorithm and parameter combinations based on that dictionary. The internal metrics, including the AFS, are calculated for each algorithm-parameter combination and then used to test the algorithm-parameter combination for validity based on whether their internal metrics meet certain thresholds, with any algorithm-parameter combinations that do not meet the threshold being dropped. The internal metrics and results of the simple evaluation are then outputted to a running CSV file which can be used to evaluate the performance of individual algorithm-parameter combinations and the overall algorithm performance. Graphs representing the key features of the clusters and early-stage personas are developed for each remaining algorithm-parameter combination to facilitate efficient domain-specific evaluation. The current iteration of HyPersona has been written in Python 3.8, utilizing the many computer-science and scientific libraries available.

5.1.1 Inputs and Data Prerequisites

Three main inputs can be passed to the HyPersona framework: 1) the data to be clustered; 2) a dictionary of algorithms and parameters to be tested; and, optionally, 3) a range of domain-specific information to be used in outputs. Before running the HyPersona some initial testing is required to configure the internal metric thresholds, as the values considered acceptable can vary based on the data set used. The data passed into HyPersona is expected to be clean, numeric data, free of nulls. This is largely due to the requirements of many clustering algorithms not handling non-numeric or null data. Domain-specific information can be passed into HyPersona to determine which features are included in the graphs and acronyms can be provided to simplify the graphs. Key features that should always be included in the early-stage personas developed, such as demographic factors, can also be defined. Finally, aggregate features can be set, where outputs should also include the average of a selection of features. The aggregate features do not affect the clustering process, they are only to give more concise outputs where there are multiple features relating to a single factor. For example, when there is multiple features all relating to different aspects of personal risk perception that have very high correlation to one another, a single value representing the overall personal risk perception can be used to provide more concise output.

The algorithm dictionary details every clustering algorithm and parameter to be considered and assigns an identifier to each set. HyPersona uses the identifier alongside a number to represent each parameter combination to act as a unique identifier of an algorithm-parameter combination. The schema for the algorithm map is given in Figure 5-2.

```
{
[identifier]: {
    "algorithm": [algorithm class/function reference],
    "type": [one of: class | function | ensemble],
    "params": {
        [parameter name]: [list of potential values],
        ...
     }
...
}
```

FIGURE 5-2: THE SCHEMA FOR THE ALGORITHM DICTIONARY

Each entry in the map applies to one clustering algorithm that can be provided as a class, which is the standard for the sklearn library, or a function, which is the standard for the pyclustering library. The '`type`' parameter defines what type of algorithm definition is given. When the type is ensemble value instead of including the algorithm and parameters directly, the map will include a set of dictionaries for each of the algorithms and parameters to be used as part of the ensemble. The params value contains another map of each parameter and the potential '`values`' to be used.

5.1.2 The Framework Core

The core of the HyPersona framework is an exhaustive grid search that runs each algorithm-parameter combination and outputs information about the results that can then be used to select the most appropriate algorithm. HyPersona first gets each possible algorithm-parameter combination from the algorithm dictionary and assigns the combination a unique identifier (id). The next step is to run each algorithm-parameter combination on the data and calculate the internal metrics of the cluster set results. The internal metrics are then compared to a set of predefined thresholds, with the cluster sets that do not meet the thresholds being dropped.

The internal metrics used are Silhouettes Coefficient (SC), Calinski-Harabasz Index (CHI), and Davies-Bouldin Index (DBI), as well as the proposed, purpose-specific internal metric, AFS. The internal metrics were selected as they all primarily measure the separability and definition of the clusters, with poor values usually indicating overlapping or indistinct clusters which are factors that are also important to persona development. As well as the internal metrics, the cluster size is considered. As the desired personas for the use cases HyPersona is targeted at would reflect the common, significant attitudes and beliefs within the population, rather than outlying opinions, the clusters should each contain a significant portion of the population.

The internal metrics of all algorithm-parameter combinations are outputted to a running CSV file (metrics.csv), for algorithm-parameter combinations that were dropped, the details of the threshold they did not meet are also included. For the algorithm-parameter combinations that were not dropped, graphs representing the cluster centroids and early-stage personas are developed. The graphs display the number of standard deviations each feature of the cluster centroid is from the population mean. When set, only key features are included in the graph and acronyms are used when available. A separate graph is given for each cluster centroid, and an SVG file containing the graphs is saved for the algorithm-parameter combination using the id ([id].csv). Similarly, the early-stage personas list all the values found to significantly differ from the population mean or between cluster centroids. For each feature the mean value for the cluster, the population mean, and the number of standard deviations the cluster mean differs from the population mean is listed. The early-stage personas are saved to a text file using the id ([id]_personas.txt).

5.1.3 Manual Evaluation and Persona Creation

The final aspect of the HyPersona framework is to use the outputs to facilitate the manual, domain-specific evaluation of the algorithm-parameter combinations so that the most appropriate algorithm-parameter combination can be selected. There can be multiple valid sets of clusters within one data set, and internal metrics can be biased towards particular clustering algorithms, rewarding algorithms based on similar premises. For example, the SC is generally higher for convex clusters, meaning algorithms, like k-means, which tend to develop convex clusters are more likely to perform well. Thus, the top performing algorithm-parameter combination according to the internal metrics should not be automatically chosen. Instead, the internal metrics are used as guides to direct which cluster sets should be considered first. As AFS was developed with the goals of persona development in mind, AFS is used as the primary indicator of the quality of a cluster set for persona development.

Some domain-specific expertise is required to evaluate the results, and if key features are being used some domain-specific expertise may also be required to identify which features qualify. The process of domain-specific evaluation will differ depending on the use case. However, HyPersona is designed to make evaluation more straightforward with graphs and simple metrics.

Identifying algorithm-parameter combinations that have developed significantly similar cluster sets is one of the first steps during domain-specific evaluation. A pair of cluster sets are significantly similar if they are identical, or the differences between the cluster sets would not affect the interpretation of the clusters during persona development. The graphs developed by HyPersona allow for efficient comparison of cluster sets to determine similarity. When two cluster sets are significantly similar, the internal metrics determine which cluster set would be used.

Once the domain-specific evaluation has been used to determine the best performing algorithm-parameter combination, the early-stage personas are then used as a base for the fully realised personas. The early-stage persona files are simple to allow for the results to be transferred into any desired persona format, with the focus on the features that significantly differ for each cluster and the features predetermined to be important for the persona creation. The early-stage personas minimise the amount of data interpretation required during the persona creation phase.

End of section is published in [119]

5.2 Average Feature Significance

AFS was developed as there were no existing internal evaluation metrics that could be applied to evaluate the statistical significance of a set of clusters. Whether or not the clusters within a set statistically differ from one another is an important indicator of cluster quality and may be required for many use cases, such as during persona development. As such, AFS is calculated based on the average statistical significance of the cluster features compared to each other and the population mean.

The definition of AFS provided in [119] is as follows. When the list of clusters is given as $c = \{c_1, \ldots, c_n\}$ and the distinct pairs of clusters, ${}_nC_2$, are given as $p = \{p_1, \ldots, p_m\}$. Let $t_1(c_i, \mu)$ return the number of features in the cluster c_i, that are significantly different compared to the mean μ using a one-sample t-test, and $t_2(p_j)$ return the number of features that are significantly different between a pair of clusters p_j, using a two-sample t-test. Then, AFS can be defined as:

$$AFS = \frac{\sum_{i=0}^{n} t_1(c_i, \mu) + \sum_{j-0}^{m} t_2(p_j)}{n + m} \tag{1}$$

A feature with a p-value less than 0.05 is considered statistically significant. The AFS is not bounded but will always be greater than 0, with higher values meaning that, on average, the features of the clusters are more significantly different.

5.3 Applying HyPersona with Behavioural Models

To apply the HyPersona framework to the current project the inputs had to be prepared. A key element of the inputs was the domain-specific information. As individuals have different attitudes and perceptions surrounding cyclones and preparatory behaviours, the domain specific evaluation should be based on how well the personas developed align with relevant behavioural models. The behavioural models used for domain-specific evaluations were PMT and PADM, each of which are discussed in Chapter 2.

The focus of the evaluation was on PADM. PADM was selected because the model has been successfully applied to explain motivation to perform cyclone damage mitigation behaviours, and the model's history of being used to design and target messaging around protective behaviours [121]–[123]. When preparing the data for use with HyPersona, key features from the data set were identified based on PMT and PADM, and where multiple elements were required to describe a single perception or belief, aggregate features were defined.

During the domain-specific evaluation the cluster sets were evaluated based on how well they align with PMT or PADM using the graphs and early-stage personas developed by HyPersona. That is, whether the features that indicate the individual's perceptions and attitudes toward cyclones and cyclone preparatory behaviours explain the individual's motivation to perform preparatory behaviours. The cluster sets that did not align with PADM were ruled out and the remaining cluster sets were ranked based on how well each cluster aligns with PADM and how distinct the clusters within each cluster set were in terms of the key features identified. After domain-specific evaluation, the algorithm-parameter combination that produced the cluster set that ranked highest was selected as the best performer.

5.4 Implementation of HyPersona

The HyPersona Framework was implemented in Python 3.8, utilizing the many computer-science and scientific libraries available. Python was selected due to the language's popularity, simplicity, and wide support within computer science and machine learning communities. Core mathematical and scientific libraries such as pandas [124], NumPy [125], and SciPy [126] were used. The plotting library, Matplotlib [127], was used to generate the graphs.

When running HyPersona with clustering algorithms, existing implementations of clustering algorithms were used where possible. Two primary python libraries were used: *scikit-learn* (sklearn) [128] and pyclustering [129]. Individual algorithm implementations or custom implementations were used for the clustering algorithms not available in sklearn [128] and pyclustering [129]. Further details on the implementation of each clustering algorithm used will be provided in Chapter 7.

5.5 Summary

The HyPersona framework begins to resolve the challenges involved in the tuning of clustering algorithms for automated persona development through the proposal of a semi-automated approach. HyPersona uses an exhaustive grid search to validate all possible algorithm-parameter combination against a set of naïve evaluation thresholds. Easy-to-use graphs and metrics are then outputted for each valid algorithm-parameter combination which can then be used for effective comparison and domain-specific evaluation. Furthermore, as part of HyPersona, a new internal metric, AFS, was proposed.

However, to apply HyPersona with confidence, and address SRQ1, the framework had to be validated first. Whether HyPersona and the application of thresholds to rule out invalid cluster sets was useful compared to existing approaches and techniques had to be evaluated. The quality of AFS as an internal evaluation metric also had to be determined. An additional key aspect of the evaluation of HyPersona was to verify whether clustering algorithms are potentially able to develop clusters that can be used to create deep and nuanced personas based on behavioural theory.

Chapter 6: DEMONSTRATION AND VALIDATION OF HYPERSONA

Validation of HyPersona was required before the framework could be applied to address the primary research question and the remaining SRQs. Additionally, validating HyPersona through a preliminary study allowed the premise that clustering algorithms can develop clusters that lead to effective personas to be verified. The preliminary study applied HyPersona to the data set collected by Scovell et al. [6], [25] to the three most commonly used clustering algorithms for persona development. The results of HyPersona were evaluated to determine whether AFS offers unique insights and indicates a cluster's quality for persona development. The algorithm-parameter combination selection process was compared to existing methods and frameworks, and the accuracy and usefulness of thresholds to rule out invalid cluster sets was determined. Lastly, the personas developed during the preliminary study were compared to those developed by Scovell et al. [6], [25].

The AFS was found to provide insights into the cluster quality that were not present with existing internal metrics and functioned as an indicator of cluster quality for persona development. K-means with random initialisation was found to have developed the best cluster set and a set of personas representing the prominent attitudes and risk mitigation behaviours was able to be developed. The HyPersona framework was validated against Hypercluster [115], an existing framework for the hyperparameter tuning of clustering algorithms, and the results of completely automatic methods using an individual internal evaluation metric. When compared to the personas developed by Scovell et al. [6], [25], the personas developed by HyPersona were found to have a similar level of depth and efficacy. As with the HyPersona framework, the results of the preliminary study have been published in Array [119].

The following sections are published in [119]

6.1 Application of HyPersona

To evaluate HyPersona, the framework was applied to a real-world use case for personas. The selected use case requires a set of personas to target communication around cyclone damage mitigation behaviours. The HyPersona evaluation was designed to answer a set of research questions:

RQ1. How effective is the use of thresholds based on internal metrics at ruling out algorithm-parameter combinations?

RQ2. Is AFS a useful internal metric that provides alternate insights to existing internal metrics?

RQ3. How does the selection of algorithm-parameter combination based on the HyPersona framework differ from that based on an automated framework using an internal metric?

6.1.1 Algorithms and Parameters

The algorithms selected to be compared were the three most prominent algorithms within the persona development field [71]: k-means, AHC, and NMF. The details of the algorithms and parameters used are given in Section 2.1. Table 6-1 gives the specifics of each of the algorithm-parameter combinations and the id assigned. Based on scope requirements and to facilitate comparison to the expert driven personas, the only number of clusters, k, used was 3.

TABLE 6-1: ALGORITHM-PARAMETER COMBINATIONS AND UNIQUE IDENTIFIERS

ID	Parameters
AHC based algorithm-parameter combinations	
agg_heir_v0	linkage: Ward's [79]
agg_heir_v1	linkage: complete
agg_heir_v2	linkage: average
agg_heir_v3	linkage: single
K-means based algorithm-parameter combinations	
kmeans_v0	initialization: k-means++ [92]
kmeans_v1	initialization: random [28], [29], [71], [76]
NMF based algorithm-parameter combinations	
nmf_v0	solver: cd [130], iterations: 100
nmf_v1	solver: cd [130], iterations: 500
nmf_v2	solver: cd [130], iterations: 1000
nmf_v3	solver: mu [131], iterations: 100
nmf_v4	solver: mu [131], iterations: 500
nmf_v5	solver: mu [131], iterations: 1000

6.1.2 Data

This study used survey responses from 519 NQ residents on cyclone preparatory behaviours, psychological characteristics, and demographics [6]. Informed consent was obtained before any data was collected and all possible steps were taken to protect the privacy of the individuals who participated. The survey covered key elements identified as part of the risk mitigation decision process, as well as the likelihood that they will perform some risk mitigation behaviours, as well as more general demographic details [6]. The data was prepared by first converting any non-numeric features either through directly mapping the values, i.e., $\{None,\ Low,\ Moderate,\ High\} = \{0, 1, 2, 3\}$, or one-hot encoding when the values were not ordinal. Then any null values were replaced using an iterative imputation [132].

Key features were identified based on the PADM and where multiple elements were required to describe a single perception or belief; aggregate features were defined. Each key or aggregate feature was assigned an acronym. The aggregate features and acronyms are available in Table 6-2, and the key individual behavioural features are given in Table 6-3. The values of each key feature reflect how strongly an individual agrees with the given statement, larger values always mean a stronger level of agreement.

TABLE 6-2: KEY AGGREGATE BEHAVIOURAL FEATURES AND ACRONYMS USED

Acronym	Feature Description
Eff	Encompasses the perceived effectiveness of cyclone shutters to reduce damage, keep family safe, to increase property value, and for other purposes.
C	Encompasses financial, time, effort, and knowledge cost of having cyclone shutters installed.
PR	Encompasses the perceived personal risk of a cyclone; how the individual's daily life, job, mental health, and physical health would be affected.
GR	Encompasses the perceived general risk of a cyclone, the likelihood of catastrophic destruction, widespread death, the financial threat, and the threat to future generations.

TABLE 6-3: KEY INDIVIDUAL BEHAVIOURAL FEATURES AND ACRONYMS USED

Acronym	Feature Description
S	How stressed thinking about the possibility of a cyclone makes the individual feel
F	How fearful thinking about the possibility of a cyclone makes the individual feel
H	How helpless thinking about the possibility of a cyclone makes the individual feel
D	How depressed thinking about the possibility of a cyclone makes the individual feel
1-2S	How much damage a category 1-2 cyclone would do
3-4S	How much damage a category 3-4 cyclone would do
5S	How much damage a category 5 cyclone would do
1-2C	Likelihood of a category 1-2 cyclone hitting
3-4C	Likelihood of a category 3-4 cyclone hitting
5C	Likelihood of a category 5 cyclone hitting
VA	How visually appealing cyclone shutters are

AO	Whether the individual feels they could organize to have cyclone shutters installed
GS	Whether the government would give financial support in the event of a cyclone
TF	How often the individual discusses or thinks about cyclones
IS	Whether the individual has actively looked for ways to minimize cyclone damage
How likely you are to perform the following next cyclone season/once a cyclone warning is declared	
TT	Trim treetops and branches
CR	Check property for rust, rotten timber, termite infestations and loose fittings
CW	Check that the walls, roof, and eaves of your home are secure
CF	Check fencing is not loose or damaged
CG	Clean gutters and downpipes
Ply	Put plywood up on glass windows/doors
SO	Secure outdoor furniture and garden items
CY	Clear yard of any loose items
Likelihood of the individual to install cyclone shutters	
XU	Extremely unlikely
MU	Moderately unlikely
SU	Slightly unlikely
N	Neither likely nor unlikely
SL	Slightly likely
ML	Moderately likely
XL	Extremely likely

6.1.3 Internal Metric Thresholds

The thresholds for each of the internal metrics and the cluster size had to be set before the HyPersona framework was run. The thresholds were designed not to be too strict, instead, to only rule out inadmissible results. Any cluster with less than 5% of the data points was considered too small, as such clusters were likely to be representing edge cases. The AFS threshold was 15, as there were more than 30 key features, and if there were, on average, less than 15 significantly different features between clusters, the personas created from them were unlikely to have significantly different behavioural features. For the other internal metrics, SC values less than 0, CHI values less than 10, and DBI values greater than 5 were all found to be indicative of poorly formed or overlapping clusters. Algorithm-parameter combinations that did not meet these thresholds were dropped by the HyPersona framework.

6.2 Results

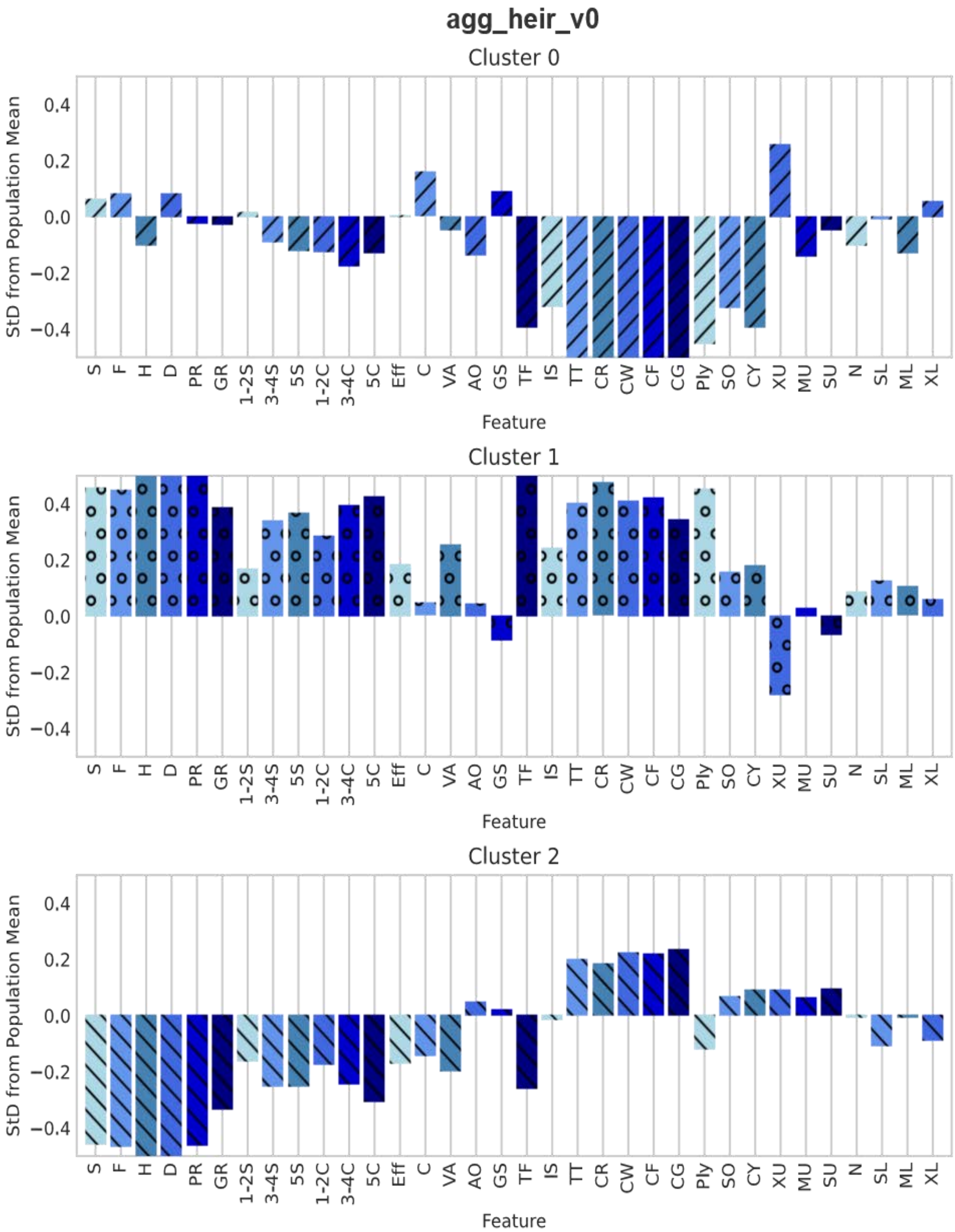
The internal metrics of the cluster sets developed by the algorithm-parameter combinations are given in Table 6-4. The top score of each metric is given in bold, and the second-best score is italicized. All five of the algorithm-parameter combinations that were dropped failed to meet multiple thresholds. The dropped algorithm-parameter combinations all failed to meet the CHI threshold of 10 and had created clusters that contained less than 5% of the total data points. Additionally, agg_heir_v3 also failed to meet the AFS threshold, with an average of 0 significant features.

TABLE 6-4: HYPERSONA FRAMEWORK RESULTS

ID	SC	CHI	DBI	AFS
agg_heir_v0	0.0663	38.141	3.3818	**67.33**
agg_heir_v1	0.0741	36.817	2.9288	53.67
agg_heir_v2*	**0.1742**	3.084	*1.2825*	16.50
agg_heir_v3*	*0.1684*	2.098	**0.6768**	0.00
kmeans_v0	0.0875	*47.084*	2.8595	58.00
kmeans_v1	0.0889	**47.095**	2.9145	*60.67*
nmf_v0	0.0429	26.947	3.3509	55.17
nmf_v1	0.0627	31.088	3.1030	56.33
nmf_v2	0.0655	30.114	3.0520	55.33
nmf_v3*	0.0207	6.873	3.5346	36.00
nmf_v4*	0.0207	6.873	3.5346	36.00
nmf_v5*	0.0207	6.873	3.5346	36.00

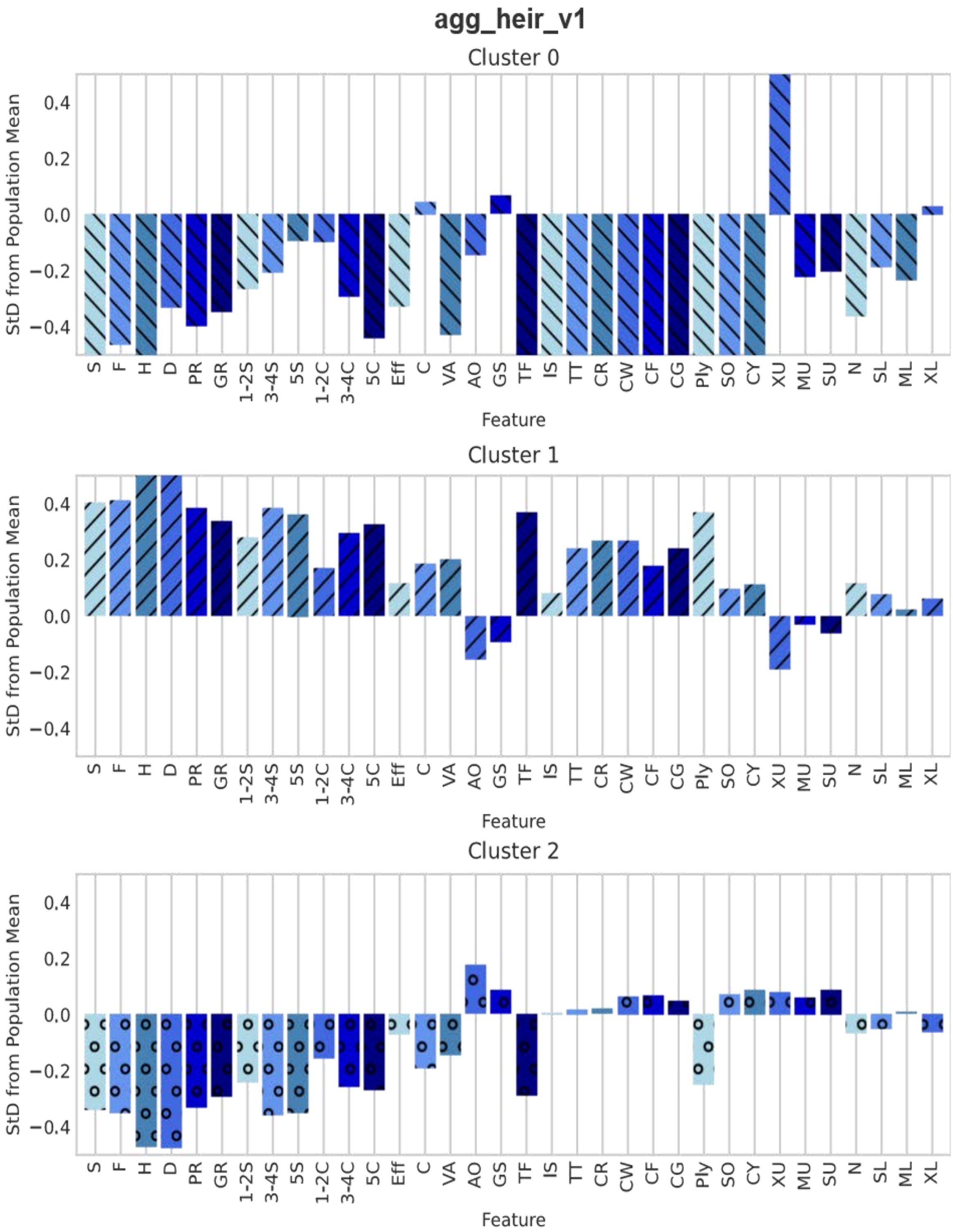
Using the graphs and early-stage personas developed by HyPersona, the clusters developed by kmeans_v1 and kmeans_v0 were found to be functionally identical, as the minor differences between the clusters developed would not have any impact on a set of personas developed. As such only kmeans_v1, which had the higher internal metric values, was considered. The graphs developed for agg_heir_v0, agg_heir_v1, kmeans_v1, and nmf_v1 are given in Figure 6-1, Figure 6-2, Figure 6-3, and Figure 6-4 respectively. The clusters have been re-ordered to allow for the most similar clusters to be compared to one another. The cluster sets developed by agg_heir_v0 and kmeans_v1 were quite similar, with each of the clusters following similar overall patterns, while the cluster set developed by nmf_v1 differed most greatly.

Based on the internal metrics, the domain-specific evaluation focused on agg_heir_v0 and kmeans_v1 first, followed by agg_heir_v1, and nmf_v1. Each cluster set was evaluated based upon how well each cluster aligned with behavioural theory, and how distinctive each prospective persona would be was also considered during the domain-specific evaluation. Through the domain-specific evaluation, kmeans_v1 was determined to be the best performer. Compared to agg_heir_v1, k-means_v1 was selected as the difference between likelihoods to install cyclone shutters was more significant, and the average risk perceptions of each cluster within the set were more distinct.



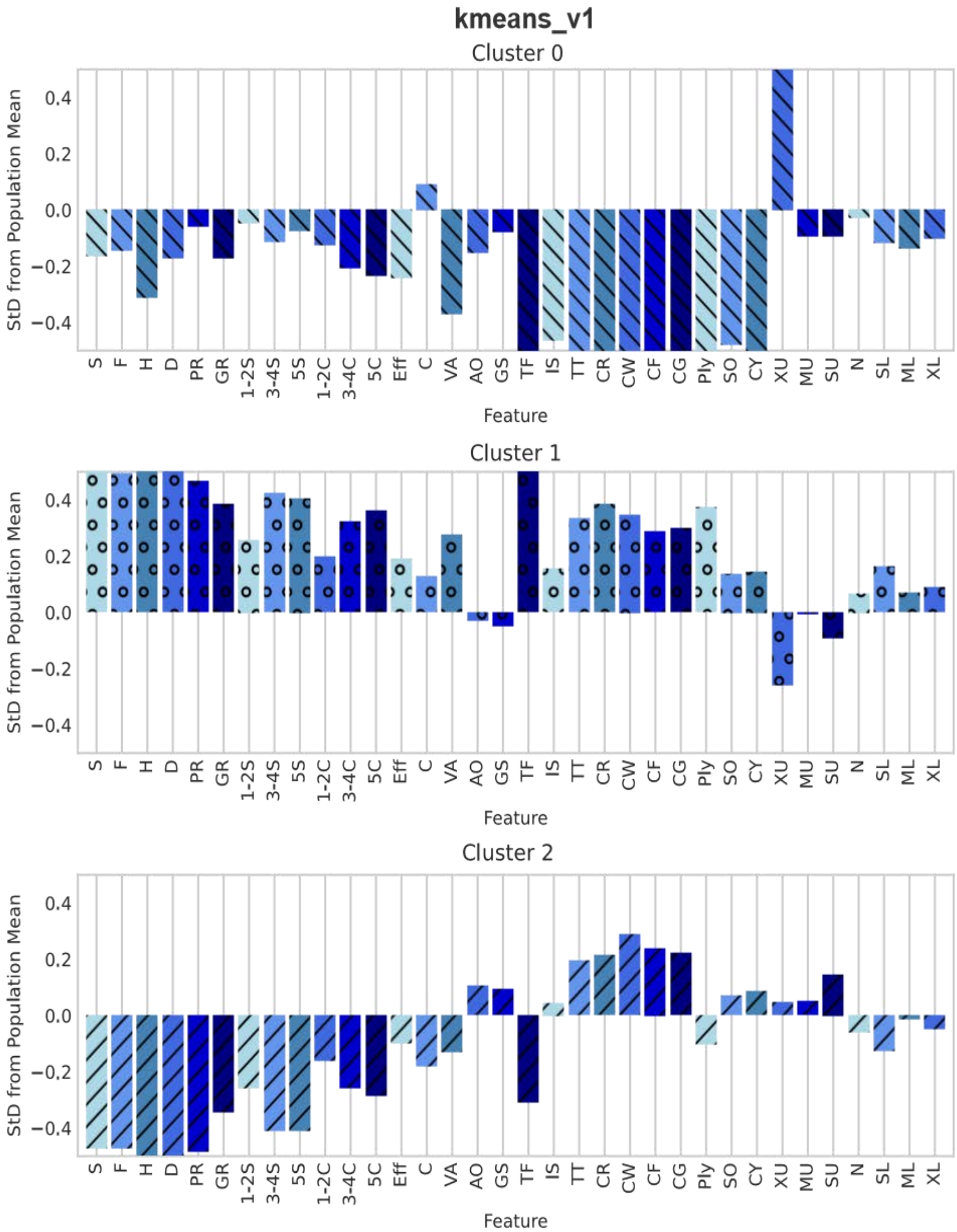
Each set of graphs give the number of standard deviations each of the key features were from the population mean for the centroid.

FIGURE 6-1: GRAPHS DEVELOPED BY HYPERSONA FOR AGG_HEIR_V0



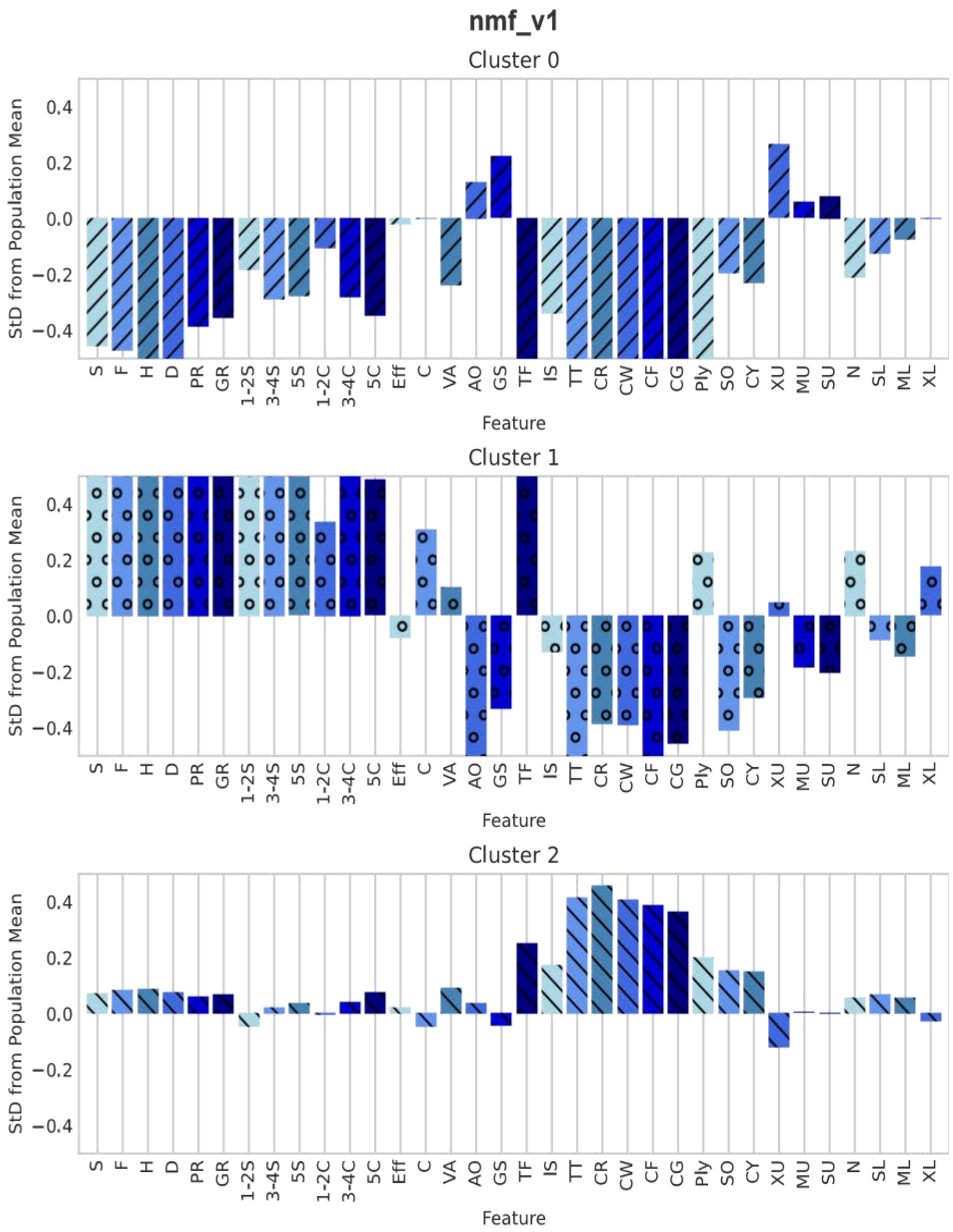
Each set of graphs give the number of standard deviations each of the key features were from the population mean for the centroid.

FIGURE 6-2: GRAPHS DEVELOPED BY HYPERSONA FOR AGG_HEIR_V1



Each set of graphs give the number of standard deviations each of the key features were from the population mean for the centroid.

FIGURE 6-3: GRAPHS DEVELOPED BY HYPERSONA FOR K-MEANS_V1



Each set of graphs give the number of standard deviations each of the key features were from the population mean for the centroid.

FIGURE 6-4: GRAPHS DEVELOPED BY HYPERSONA FOR NMF_V1

A set of three personas were developed based on the early-stage personas produced by HyPersona for kmeans_v1. As there were no significant differences in age, gender, marital status, or location between the clusters, those demographic factors were not included in the final personas. The most important demographic factor was found to be previous experience with cyclones and cyclone damage. The personas were developed manually based on the most significant features for each cluster. Each persona was assigned an epithet to describe their attitude towards performing damage mitigation behaviours leading up to a cyclone. The three personas created were:

The Unconcerned (Cluster 0)

Cyclones were not on the radar of the Unconcerned persona. They were the least likely to think about or discuss cyclones in their day-to-day life or to have looked for methods to help prevent cyclone damage. Of all the personas the Unconcerned persona rated themselves as having the least knowledge about cyclones. The Unconcerned persona was the least likely to have experienced a cyclone and had a below-average expectation of a cyclone occurring and a lower perception of the risks associated with cyclones than average. The Unconcerned persona had the lowest self-reported likelihood to perform any of the preparatory behaviours or install structural upgrades to their property.

The Concerned (Cluster 1)

The Concerned persona was the most worried about a future cyclone, with the highest expectation of a cyclone occurring and doing considerable damage. Significantly, they believed a cyclone would impact their mental and physical well-being and thinking about a cyclone occurring gave them feelings of helplessness and depression. They spent the most time thinking about and discussing cyclones and were most likely to have investigated ways to protect against cyclone damage. However, they had the lowest perception of their ability to get cyclone shutters installed. The Concerned persona was self-reported as most likely to perform all available preparatory behaviours leading up to the next cyclone and had the highest motivation to install structural upgrades such as cyclone shutters. They were more likely to have previously experienced a cyclone where they received moderate or severe damage.

The Confident (Cluster 2)

The Confident persona had the lowest perception of the risks associated with cyclones and expected far less damage from high category cyclones than average. Mirroring the Concerned persona, the Confident persona differed most in their feelings of helplessness and depression when thinking about a possible cyclone and the perceived impact of a cyclone on their mental and physical health. They self-reported being likely to perform simple preparatory behaviours, but less likely to perform more difficult behaviours, such as putting up plywood or installing structural upgrades. They were most likely to have experienced a cyclone without receiving any damage, and of those who experienced cyclone damage, they were least likely to have experienced considerable damage.

6.3 Framework evaluation

HyPersona was applied to a real-world use case to demonstrate the framework's ability to be effectively applied to persona development problems. Through the application of the HyPersona framework and the results found, HyPersona can be evaluated by how well it answers the research questions.

6.3.1 How effective is the use of thresholds based on internal metrics at ruling out algorithm- parameter combinations?

Five of the twelve algorithm-parameter combinations used, or just over 40%, were dropped as they did not meet the required thresholds. The dropped algorithm-parameter combinations had created heavily imbalanced clusters, with the clusters not meeting the minimum size threshold of 5% of the total data. Additionally, although not tested for, all the dropped algorithm-parameter combinations developed one cluster that contained more than 90% of the total data points. In comparison, the sizes of clusters that were developed by the algorithm-parameter combinations that were not dropped were more balanced. The size of each cluster created by each algorithm-parameter combination is given in Table 6-5.

TABLE 6-5: ALGORITHM-PARAMETER COMBINATION CLUSTER SIZES

ID	Cluster 0	Cluster 1	Cluster 2
agg_heir_v0	131	186	202
agg_heir_v1	245	242	32
agg_heir_v2*	515	3	1
agg_heir_v1*	517	1	1
kmeans_v0	80	223	216
kmeans_v1	93	223	203
nmf_v0	239	35	245
nmf_v1	143	52	324
nmf_v2	124	52	343
nmf_v3*	479	9	31
nmf_v4*	479	9	31
nmf_v5*	479	9	31

* The algorithm-parameter combination was dropped by HyPersona

As a result of the size imbalance, the clusters developed did not differ in a statistically significant way. The large clusters contained nearly all the data points, and thus sat extremely close to the population mean. While the values of the small clusters differed greatly, but due to their small sizes the differences were rarely statistically significant. This was reflected in the AFS of the dropped algorithm-parameter combinations being the lowest, with agg_heir_v3 not meeting the required threshold. The CHI also acted as an indicator of how well balanced the cluster sizes were, as none of the dropped algorithm-parameter combinations met the minimum CHI threshold.

Alternately, two of the dropped algorithm-parameter combinations performed quite well in terms of the SC and DBI, with agg_heir_v2 and agg_heir_v3 together achieving the best and second-best values for both DBI and SC. Without using the internal metrics or cluster sizes to rule out the incompatible algorithm-parameter combinations, agg_heir_v2 and agg_heir_v3 would have been considered based upon their SC and DBI scores. The dropped algorithm-parameter combinations would not have created quality personas. Thus, by automatically dropping 40% of the algorithm-parameter combinations, considerable manual time and effort was saved.

6.3.2 Is AFS a useful internal metric that provides alternate insights to existing internal metrics?

The algorithm-parameter combinations that scored best in AFS differ from the best performing algorithm-parameter combinations according to the other internal metrics. Table 6-6 gives the Pearson's Correlation Coefficient between each of the internal metrics based on the results of the HyPersona framework. A level of correlation was expected between the internal metrics as they were all rewarding similar traits in cluster sets. That is, all the internal metrics prefer well separated, convex clusters. The strongest correlation was between the SC and DBI, while the weakest correlation was between the SC and CHI. AFS had a moderate correlation to the SC and a strong correlation to both the CHI and DBI. However, the results of AFS were different enough to the existing internal metrics not to be redundant.

TABLE 6-6: PEARSON'S CORRELATION COEFFICIENT BETWEEN INTERNAL METRICS

	SC	**CHI**	**DBI**	**AFS**
SC	1.000			
CHI	-0.112	1.000		
DBI	-0.954	0.342	1.000	
AFS	-0.526	0.864	0.733	1.000

The algorithm-parameter combination that achieved the best AFS score, agg_heir_v0, achieved more mediocre scores in the other internal metrics, however, the domain-specific evaluation found agg_heir_v0 to be a serious contender. While, the algorithm-parameter combination that was determined to be the best performer, kmeans_v1, had the second best AFS score. The AFS score was also a primary reason nmf_v1 was considered, which demonstrated interesting differences from the other algorithm-parameter combinations.

AFS was found to provide insight and information on cluster quality not otherwise present in existing internal metrics that were important to selecting relevant cluster sets for persona development. As such, AFS proved to be a useful addition to the hyperparameter tuning framework. Other problem areas where having distinct cluster centroids is important, may also benefit from applying AFS.

6.3.3 How does the selection of algorithm-parameter combination based on the HyPersona framework differ from that based on an automated framework using an internal metric?

Without any ground-truth values available, a fully automated hyperparameter tuning framework relies on internal metrics to determine the best performing algorithm. Based purely on an individual internal metric, an automated hyperparameter tuning method using the SC would select agg_heir_v2, a method based on the DBI would select agg_heir_v3, and a method based on the CHI would select kmeans_v1. As such, a framework based on the SC or DBI would give poor results, as both algorithm-parameter combinations selected by these metrics were ruled out by the proposed framework. While acting as a very useful guide, the algorithm with the best AFS, agg_heir_v0, also was not chosen as the best performing algorithm although agg_heir_v0 did produce an acceptable set of clusters for persona development.

Alternately, the algorithm-parameter combination that performed the best according to the CHI was selected by the proposed framework as the best performer. Additionally, once minimum thresholds were applied the next best SC score was achieved by kmeans_v1, and the best DBI score was achieved by kmeans_v0 which was found to be almost identical to kmeans_v1. This suggests that by using a combination of internal metrics and ruling out algorithm-parameter combinations that did not meet minimum thresholds the internal metrics may be reliable predictors of persona quality.

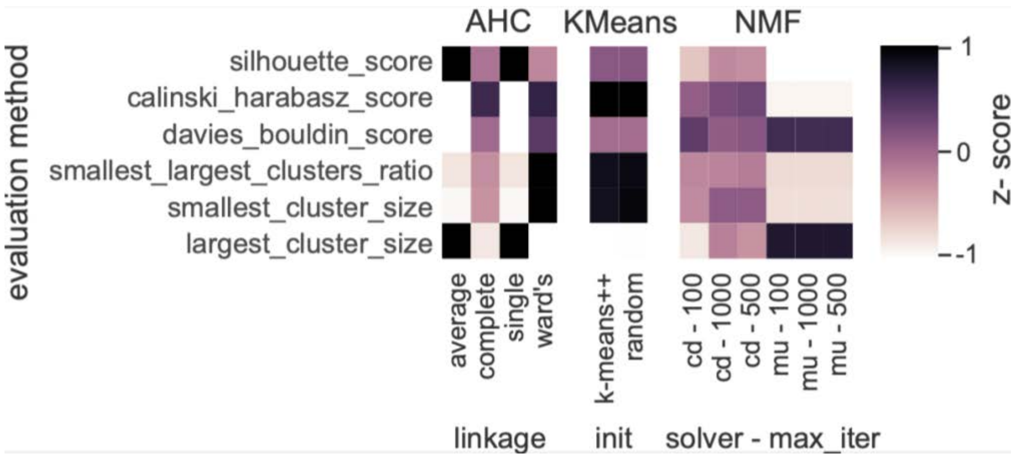


FIGURE 6-5: RESULTS OF THE HYPERCLUSTER [115] FRAMEWORK: HEAT MAP OF THE INTERNAL METRICS

Hyperparameter tuning was also performed on the data set using the Hypercluster [115] framework with the same algorithm and parameter combinations. Hypercluster develops a heat map to graphically display the quality of a range of internal metrics, which has been given in Figure 6-5. One important note in interpreting the heat map developed by Hypercluster was that the heat map does not adjust for the fact that the closer the DBI score was to 0, the better quality the clusters were, which was the opposite of the other internal metrics. As such, in only the DBI row, the lower z-score was the better result.

Based on the heat map, when considering all the evaluation methods either k-means initialization appears to be the best performer, followed by AHC with single or average linkage. This was quite similar to the results found by the HyPersona framework, which was expected as a similar set of internal metrics were applied to the data in both cases. Other than providing an in-built tool for visualisation of the internal metrics, Hypercluster does not provide any additional information to the proposed semi-automated framework and still relies on the manual identification and selection of the best performing algorithm-parameter combination. As such, the Hypercluster framework does not give any insights into the nature and content of the clusters.

As there were no minimum thresholds applied or other information given by Hypercluster, determining the best performer was more difficult. For example, AHC with Ward's linkage, agg_heir_v0, and AHC with single linkage, agg_heir_v3, performed well in different internal metrics but overall appear to have performed similarly. However, AHC with single linkage, agg_heir_v3, was dropped by HyPersona due to cluster size, the CHI score, and the AFS score. As such, when using Hypercluster manual evaluation of AHC with single linkage, agg_heir_v3, would be required. Additionally, the heat map developed by Hypercluster only shows minimal differences between AHC with wards linkage, agg_heir_v0, and AHC with complete linkage, agg_heir_v1. By comparing the graphs developed for agg_heir_v0, Figure 6-1, and agg_heir_v1, Figure 6-2, the significant impact of the difference in linkage used on the clusters developed and thus the personas that would be developed was apparent.

Internal metrics cannot be solely used to identify the quality of a set of clusters for a specific purpose, such as persona development. Applying the Hypercluster framework for hyperparameter tuning of clustering algorithms requires significantly more manual intervention than applying HyPersona. The graphs developed by HyPersona simplify the manual evaluation process, saving considerable time compared to methods that only provide insights into the internal metrics.

Existing methods such as Hypercluster [115] still require manual, domain-specific evaluation for their effective application however do not facilitate the required evaluation. HyPersona extends upon the current approach by outputting relevant information and visualisations to assist in the efficient domain-specific evaluation and streamline the evaluation process by eliminating cluster sets that were not appropriate for the use case. As such, the algorithm-parameter combination chosen by HyPersona was assured to be useful and relevant to the use case.

End of the sections published in [119]

6.4 Evaluation of Personas Create with HyPersona Compared to Expert-Driven Personas

The personas created by HyPersona during the preliminary study, given in Section 6.2, differ from the expert-driven persona set developed by Scovell et al. [6], [121]. The two persona sets were expected to differ as the expert-driven personas focused on motivation to install cyclone shutters and only used four features: personal risk perception, and the visual appeal, cost, and efficacy of cyclone shutters.

6.4.1 Expert-Driven Personas

This section details the process and findings of Scovell et al. [6], [121]. To develop the personas an analysis was performed on the survey data to identify which psychological variables acted as key factors in predicting an individual's intention to install cyclone shutters. During this particular analysis only a subset of the total data set, made up of the individuals who owned property in NQ, was used. The data set was narrowed down to the subset of 322 individuals, as only homeowners would have the ability to install cyclone shutters. Four key features for predicting motivation to install cyclone shutters were identified:

- **Personal Risk:** The expected personal consequences of a cyclone occurring.
- **Cost:** The perceived cost of cyclone shutters.
- **Visual Appeal:** The visual appeal of cyclone shutters.
- **Efficacy:** The perceived efficacy of cyclone shutters.

These features were made up of multiple questions asked during the survey. Table 6-7 gives each of the features and the factors that make up each feature. Each factor was measured on a 7-point Likert scale measuring how much the individual agrees or disagrees with a statement or how likely a factor is to occur, and the answers were summed and averaged to create each overarching feature.

TABLE 6-7: KEY FEATURES IDENTIFIED FOR PREDICTING INTENTION TO INSTALL CYCLONE SHUTTERS [6], [121]

Key Feature	Statements used for measurement
	Factors measured by expected likelihood: *If a cyclone were to occur in your area, how likely would it be that each of the following would occur?*
Personal Risk	Your property has been damaged

	Your, or a member of your household's, daily life is disturbed
	You, or a member of your household, are prevented from going to work or doing their job
	Your, or a member of your household's, mental health is negatively affected
	Your, or a member of your household's, physical health is negatively affected
	Factors measured by the extent of agreement: *Please indicate how strongly you agree with the following statements.*
Cost	Cyclone shutters are expensive to install considering my income and other expenses
	Cyclone shutters take a lot of time and effort to install considering my free time
	Cyclone shutters are difficult to get installed considering the knowledge and skill that is required
	It would require a lot of help/cooperation from others (family, friends, neighbours, or government) to install cyclone shutters
Visual Appeal	Shutters are visually appealing.
Efficacy	Cyclone shutters are effective for reducing damage and financial consequences of cyclones to my property and belongings.
	Cyclone shutters are effective in keeping my family safe during a cyclone.
	Cyclone shutters are useful for other purposes besides preventing cyclone damage
	Installing cyclone shutters increases property value

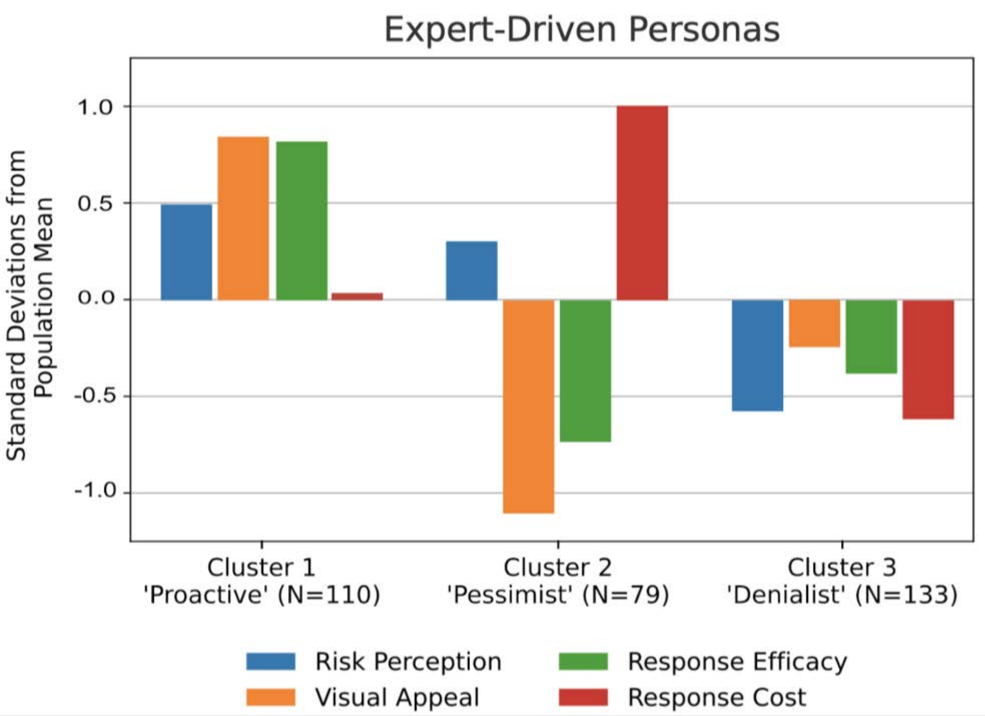


FIGURE 6-6: PREVIOUSLY DEVELOPED PERSONA CENTRES [6], [121]

Each of these features were used with the k-means clustering algorithm to create a set of 3 clusters. The number of standard deviations each key feature was from the population mean is given in Figure 6-6. The average demographic factors for each persona were then found for each cluster.

By combining the behavioural features of each cluster, demographic information, and the average intention to install cyclone shutters, the three personas were then developed. Each persona was given an alias based on the nature of the cluster's perceptions and behaviour. An overview of the personas follows:

The Proactive

The Proactive persona had the highest motivation to install cyclone shutters. The proactive persona had the highest risk perception, but also had a high positive opinion of the visual appeal and efficacy of cyclone shutters. Their perception of the cost of cyclone shutters was not significantly different to the population mean. The Proactive persona was more likely than average to have experienced a cyclone where they received moderate cyclone damage. Demographically, the Proactive persona was likely to have lived in NQ significantly longer than average and to have had less formal education.

The Pessimist

The Pessimist persona had the lowest motivation to install cyclone shutters. The Pessimist persona's risk perception was higher than average, although not as high as the Proactive persona. However, they had the lowest positive perception of the visual appeal and efficacy of cyclone shutters, and the perceived cyclone shutters as significantly more expensive than average. The Pessimist persona was less likely to have experienced a cyclone and was likely to have lived in NQ for the least amount of time.

The Denialist

The Denialist persona also had low motivation to install cyclone shutters. The Denialist persona had the lowest risk perception, they also had a less positive perception of the visual appeal and efficacy of cyclone shutters but found the cost of cyclone shutters more affordable. The Denialist persona was most likely to have experienced a cyclone where they received no damage, and demographically, the denialist persona had the highest level of education.

6.4.2 Comparison of Persona Sets

The two persona sets differed quite significantly despite being created from the same data set. However, a data set can contain multiple valid cluster sets [29]. The persona set developed by HyPersona was evaluated based on whether the personas developed could be explained through behavioural theory and used to effectively target communication. There were also some similarities between the persona sets developed by HyPersona and Scovell et al. [6], [25]. Primarily between the Proactive and Concerned personas and the Denialist and Confident personas.

Although none of the personas are identical, the Proactive and Concerned personas and the Denialist and Confident personas described similar states within the behavioural models. Both the Proactive and Concerned personas were most likely to perform the desired behaviours, and primarily required a motivator for them to act. The Denialist and Confident persona had more differences; however, both had a much lower-than-average perception of the risks associated with cyclones and had previous experience with cyclones where they received either no or minimal damage. Targeted communication for either the Denialist or Confident persona would need to focus on the potential risks surrounding cyclones.

In general, the expert created personas put more emphasis on the perceptions related to cyclone shutters. Such as the perceived efficacy and visual appeal of cyclone shutters, which was expected as that was the primary purpose the personas were developed for, and the variables selected reflected this purpose. While the personas developed by HyPersona had no defined target and the entire data set was used.

As the personas developed from HyPersona included the entire data set, the additional features available allowed for further insights to be drawn about the personas. For example, the Concerned persona had a below-average perception of their ability to install cyclone shutters, or to organise for cyclone shutters to be installed, which provided additional information about how communication towards them may need to be targeted. Additionally, the Concerned persona had significant feelings of helplessness and depression when thinking about a cyclone occurring which may need to be mitigated to avoid the Concerned persona turning to maladaptive behaviours.

The additional information within the Confident persona showed that they intended to perform simple damage mitigation behaviours, which may indicate that their intention to perform those behaviours, combined with their previous cyclone damage experience, may be significantly reducing their perception of the risks associated with cyclones. If that is the case, rather than more general information on cyclone damage risks, education on the types of damage that can only be mitigated by structural upgrades, or the more difficult preparatory behaviours may be more effective.

The biggest difference between the persona sets was in the Unconcerned persona and the Pessimist personas. Both personas were least likely to have experienced a cyclone previously and had very negative opinions of the efficacy, visual appeal, and cost of cyclone shutters. The key difference was between the perception of the risks associated with cyclones. The Pessimist persona had a higher perception of risk, which combined with the negative opinion of the efficacy of cyclone shutters was likely leading the Pessimist persona to maladaptive behaviours. While the Unconcerned persona had a below-average perception of risk, which combined with the lowest reported knowledge of cyclones and likelihood to think about or discuss cyclones, likely means the Unconcerned persona does not consider cyclones a significant threat.

The inclusion of the additional data allowed the HyPersona personas to contain meaningful insight into the motivation to perform protective behaviours. Resulting in a more robust persona set. As the HyPersona personas did not target any single behaviour, the personas, instead, represent the general attitudes and the related motivation to perform a range of protective behaviours. As such, the same persona set can be used when targeting multiple protective behaviours, rather than requiring the entire persona development process to be repeated saving considerable time and expertise.

6.5 Summary

The HyPersona framework was evaluated through comparison to existing methods and the personas developed by Scovell et al. [6], [25]. Through manual domain-specific evaluation, k-means with random initialisation was selected as the best performer and a set of personas was developed from the results. Comparing the results of the differing algorithm-parameter combinations once again reinforced the importance of hyperparameter tuning in persona development methodologies.

HyPersona was found to facilitate the development of relevant, deep personas while minimising the amount of manual intervention required. The HyPersona framework and personas developed were validated against an existing hyperparameter tuning framework for clustering, Hypercluster, and frameworks based on individual internal metrics. All the algorithm-parameter combinations that were ruled out by the framework were confirmed to have been invalid choices for the use-case, and the graphs developed were found to be more insightful than the heatmap developed by Hypercluster. As such, HyPersona facilitated a quicker evaluation and selection process. The internal metric developed for HyPersona, AFS, was found to be a useful indicator of the quality of a cluster set for persona development and gave alternate insights into cluster quality to existing internal metrics. Although targeted towards persona development, the HyPersona framework and the AFS metric could both be applied to a wide range of use-cases.

The personas developed by HyPersona were found to share attributes with the personas developed by Scovell et al. [6], [25] and to have a similar level of depth and nuance. The personas developed with HyPersona could be used to effectively target communication around preparatory behaviours. Based on the findings of the HyPersona validation, the framework can be confidently applied to a wide range of clustering algorithms to determine whether clustering algorithms can mimic expert decision making and automate persona development.

The success of HyPersona and AFS addresses the first SRQ as the use of HyPersona with AFS was found to enable the efficient evaluation of clustering algorithms to develop quality personas. The next chapter will detail the methodology with which the primary research question will be addressed.

Chapter 7: Methodology and Additional Data Collection

Previously the HyPersona framework for the semi-automated hyperparameter tuning and persona development has been introduced and demonstrated. By applying HyPersona to a wide variety of clustering algorithms, across two data sets, whether the machine learning driven approach can facilitate the development of deep, nuanced personas based on behavioural models could be determined. Which directly addressed the primary research question:

Can clustering algorithms facilitate the development of deep, nuanced personas based on behavioural models, replicating the decision making of experts, for the automation of persona development?

This chapter details the methodology undertaken to answer the primary research question and the sub-research questions identified. An additional data set was collected to act as a confirmatory sample to the data collected by Scovell et al. [6], [25] and to demonstrate the stability and consistency of the results found. This chapter also details the methodology undertaken to collect the additional data and present an overview the findings present in the additional data.

7.1 Methodology Overview

There are three remaining SRQs that need to be answered to adequately address the primary research question. Each of which required multiple steps to answer. The remaining SRQs are:

- SRQ2. Are the performances of clustering algorithms and approaches to clustering for persona development consistent?
- SRQ3. Does the selection of a clustering algorithm and parameters significantly impact the set of personas developed?
- SRQ4. How do personas created by clustering algorithms compare to behavioural theory and personas created through the application of behavioural theory?

The second *Secondary Research Question* (SRQ2) must be addressed first. As, if the personas developed are not consistent or stable, the results cannot be generalised beyond that individual run and the entire process would need to be re-performed each time with no insights into how the algorithms may perform. Clustering algorithm performance depends on multiple factors including the nature of the groups present in the data. However, in data sets that are expected to have groups of a similar nature present, a clustering algorithm should produce clusters of a similar quality.

There are multiple manners in which the consistency of a clustering algorithm or approach can be assessed. To determine the performance of a clustering algorithm, the nature of the clusters developed by the algorithm were evaluated alongside the internal metrics of the cluster set developed, and whether the clustering algorithm was dropped by the HyPersona framework. Three types of consistency were evaluated:

1. Whether the parameters had a consistent effect on algorithm performance.
2. Whether algorithms that belong to the same approach to clustering are consistent.
3. Whether an algorithm gives a consistent performance across data sets with the same underlying data.

If the parameters of an algorithm were found to have a consistent effect on performance, inferences could be made about how a change in parameter may affect performance. If random factors, such as random initialisation, had a greater impact on algorithm performance then running the algorithm with the same parameters multiple times may be beneficial to capture the different variations. Considering the consistency of algorithm performance based on approach allows for results of algorithms that were not tested to be inferred if the performances are found to be consistent. Lastly, an algorithm's performance between data sets that contain the same underlying information determines how well the findings can be applied to similar use cases and data sets.

SRQ3 was similarly important. If the results of the clustering algorithm did not have a significant impact on the personas developed, such as if all the clustering algorithms develop sets of clusters that are functionally the same, there was no point in selecting different algorithms. If the difference between the cluster sets developed by the algorithm-parameter combinations that perform well are only minor the algorithm with the best metrics can be selected with minimum evaluation. However, if there are major differences in the nature of the clusters developed between approaches, domain-specific evaluation is required to select the best performing algorithm. To determine the impact of the clustering algorithm on the clusters developed, cluster sets developed by different algorithms with internal metrics that indicate a similar cluster quality were compared.

The last SRQ, SQR4, addresses the bulk of the primary research question. Whether the personas developed by a clustering algorithm were able to mimic the nuance and depth of an expert-driven set of personas was determined by three key factors: 1) whether the persona set could be explained by behavioural theory; 2) the persona set's similarity to the personas developed by Scovell et al. [6], [25]; and 3) whether the persona set could be used to target communication. The persona set was not required to replicate the personas created by Scovell et al. [6], [25], rather the personas needed to provide a similar level of depth and insight into the data set. The third factor determined how useful the personas set developed were for the desired use case. To mimic expert driven personas in a manner that was meaningful and useful, the personas must be able to be used to target communication for the purpose of promoting the performance of cyclone damage mitigation behaviours.

Addressing the SRQs provided the findings and insights required to address the primary research question. To answer these research questions, the following steps were taken:

1. Collect additional data
2. Identify whether the additional data can be used as a confirmatory sample to the data collected by Scovell et al. [6], [25]
3. Select the algorithms and parameters to be tested
4. Prepare the data and other inputs to be used with HyPersona
5. Evaluate the consistency of each clustering algorithm and approach (SRQ 2)
6. Determine the impact of clustering algorithm selection on persona development (SRQ 3)
7. Select the best performing algorithm and develop the final set of personas
8. Compare the personas developed to behavioural theory and the personas developed by Scovell et al. [6], [25], and then determine whether the personas are sufficiently deep (SRQ 4)

7.2 Additional Data Collection – The 2021 Data Set

Before any further steps were undertaken, additional data needed to be collected. The additional data set, collected in 2021, acted as a confirmatory sample to the original survey by Scovell et al. [6], [25], collected in 2018, and assisted in determining the stability and consistency of the performances of the clustering algorithms. The survey was run using a modified version of the original survey by Scovell et al. [6], [25]. The primary addition to the original survey by Scovell et al. [6], [25] were questions that asked about the respondents insurance status. The information sheet and full survey are available in Appendix A: Survey Information Sheet and Appendix B: Survey. The original data set collected by Scovell et al. [6], [25] was referred to as the 2018 Data Set, and the additional data set collected was referred to as the 2021 Data Set. The ethics approval number was H8310.

The survey was run completely online through Qualtrics. The survey took approximately 15 minutes to complete and was open to people over the age of 18 years who lived in coastal North Queensland (between Rockhampton and Bamaga). The survey was disseminated primarily via a Facebook page that was created providing information about the study and a link to the survey. The Facebook page was shared by other weather-related Facebook pages (e.g., Oz Cyclone Chasers) to reach a broader audience. Information about the survey was also disseminated via local media outlets (newspaper, radio, and TV) in various locations throughout North Queensland.

7.2.1 Respondents

There was a total of 238 responses to the survey. The first page of the survey gave an overview of the survey content and the intended use of the results allowing for informed consent, and those who did not consent were thanked and exited the survey. Once the respondents who did not consent to the survey or completed less than 65% of the survey were removed, 211 respondents who consented to and completed the survey remained.

Of the 211 respondents, 156 (73.9%) stated they were female, 52 (24.6%) male, and 3 (1.4%) preferred not to say. The majority (N=128, 60.7%) had no dependent children. Participants' age ranged between 18 years and 77 years, with an average age of 44.1 years (SD=13.85). On average, participants had lived in NQ for 25 years (SD=16.19) and had lived in their current area for 19.9 years (SD=15.60) Of the final sample 147 (69.6%) reported having a partnered or being married. The following tables give other relevant demographic data. Table 7-1 and Table 7-2 gives a further breakdown of the location and home ownership status of the respondents.

TABLE 7-1: RESPONDENT LOCATION

Location	% (n)
Townsville	72.5 (153)
Cairns	20.9 (44)
Whitsundays	2.3 (5)
Mackay	1.4 (3)
Rockhampton	0.4 (1)
Other	1.4 (3)
No Answer	0.9 (2)

TABLE 7-2: RESPONDENT HOMEOWNERSHIP AND INSURANCE

Homeownership	% (n)	Insurance	% (n)
Owns and lives in their own home	61.6 (130)	Home and contents insurance	86.9 (113)
		No insurance	1.5 (2)
		Only contents insurance	0.8 (1)
		Only home insurance	4.6 (6)
		No Answer	6.2 (8)
Owns a home but not the home they live in	5.7 (12)	Home and contents insurance	66.7 (8)
		Only contents insurance	16.7 (2)
		Only home insurance	8.3 (1)

		No answer	8.3 (1)
Rent	32.7 (69)	Home and contents insurance	7.2 (5)
		No insurance	46.4 (32)
		Only contents insurance	26.1 (18)
		No Answer	20.3 (14)

Most participants, 89.1% (N=188), reported having experienced at least one cyclone, with almost half (N=101, 47.9%) having experienced 4 or more cyclones. Just over half (N=107, 56.9%) of respondents who had experienced a cyclone reported having received property damage. Of those who experienced a cyclone, 94.6% (N=178) reported performing cyclone preparation behaviours leading up to a cyclone.

7.2.2 Determining Whether the Additional Data was a Confirmatory Data Set

The findings of the additional data was compared to the data found by Scovell et al. [6], [25] and were found to be significantly similar, providing similar insights into the attitudes and behaviours of NQ residents around Cyclone preparedness. As such the data set was considered a confirmatory sample. Table 7-3 gives a comparison on the values of key features present in both the 2018 Data Set and the 2021 Data Set.

TABLE 7-3: COMPARISON OF FEATURES BETWEEN THE 2018 AND 2021 DATA SETS

Feature	2018 Data Set	2021 Data Set
Demographic Yes (1) or No (0) Questions		
Have you ever experienced a cyclone before?	0.91	0.88
Do you own a property in North Queensland?	0.74	0.67
Has your property (current or previous) received damage from a previous cyclone?	0.57	0.51
7-Point Likert Scale Questions: 1 = Very Low – 7 = Very High		
How strongly would you agree or disagree with the following statements?		
I think that cyclones may cause catastrophic destruction	6.22	5.91
I think that cyclones may cause widespread death	3.94	3.53
I think that cyclones pose great financial threat	6.13	5.66
I think that cyclones pose a threat to future generations	4.37	3.81
If a cyclone was to occur in your area, how likely would it be that each of the following would occur?		
Your property has been damaged	5.52	4.83
Your, or a member of your household's, daily life is disturbed	6.31	5.89

You, or a member of your household, are prevented from going to work or doing their job	6.10	5.83
Your, or a member of your household's, mental health is negatively affected	4.10	3.83
Your, or a member of your household's, physical health is negatively affected	3.60	3.28
How likely do you believe each of the following cyclone events are to occur in the next 5 years?		
A category 1-2 cyclone	6.32	6.45
A category 3-4 cyclone	5.71	5.74
A category 5 cyclone	4.74	4.72
If the following cyclone events were to occur next week, what level of property damage would you expect to receive?		
A category 1-2 cyclone	1.92	2.29
A category 3-4 cyclone	3.53	3.73
A category 5 cyclone	5.11	5.11
How likely you are to perform the following next cyclone season/once a cyclone warning is declared		
Trim treetops and branches	5.40	4.74
Check property for rust, rotten timber, termite infestations and loose fittings	5.50	4.75
Check that the walls, roof, and eaves of your home are secure	5.79	5.05
Check fencing is not loose or damaged	5.76	5.29
Clean gutters and downpipes	5.88	5.39
Put plywood up on glass windows/doors	3.73	3.40
Secure outdoor furniture and garden items	6.80	6.68
Clear yard of any loose items	6.86	6.69

As the 2021 Data Set was a confirmatory sample to the 2018 Data Set, the clustering algorithms that performed similarly on each data set can be considered to have consistent performance across similar data sets. While any algorithm that performed significantly different between the two data sets can be assumed to be more volatile. Further, the personas developed from each data set could be expected to have many similarities, which means the two sets of personas that were created could be compared to each other and the personas developed by Scovell et al. [6], [25]. As the 2021 Data Set contained features not included in the 2018 Data Set, the persona sets were not expected to be identical.

7.3 Algorithm and Parameter Selection

A variety of algorithms that are reflective of the seven most common approaches to clustering were selected, as well as a small selection of clustering algorithms with unique approaches. A total of 21 algorithms were used as well as a variety of ensemble approaches. For each algorithm, a set of reasonable parameters were selected. The only parameter left constant was the number of clusters, *k*, which was set to three. Three clusters were selected as the previous study identified three clusters, and we are attempting to mimic that study. Testing multiple values for *k* could provide additional insights but is beyond the scope of the current study.

The majority of algorithm implementations used were from the *scikit-learn* (sklearn) [128] or pyclustering [129] python libraries. The implementation used for each algorithm will be given, and specific implementation details will be given for algorithms that were not available in the sklearn or pyclustering libraries. Notably, there were no existing implementations in python of clustering based on the ABC optimisation algorithm [104], [105] or the SFLA optimisation algorithm [32]. The entire custom implementation used for ABC and SFLA clustering is available in [133].

7.3.1 Hierarchical Approach Clustering Algorithms

The hierarchies returned by each algorithm was cut at the point where there were 3 clusters to simulate k = 3. Four hierarchical clustering algorithms were selected:

1. **AHC** (sklearn implementation [128]): AHC is the most well-known hierarchical algorithm and one of the most common algorithms to be applied to persona development. The only parameter set for AHC is the linkage; single, complete, average or Ward's.
2. **BIRCH** (sklearn implementation [128]): Birch attempts to improve on AHC using a *Cluster Features* (CF) tree. The primary parameters for BIRCH are the branching factor and the entry diameter of the CF tree.
3. **CURE** (pyclustering implementation [129]): CURE extends on other hierarchical clustering algorithms by defining each cluster with a fixed number of well-scattered points which are incrementally moved towards the cluster centre. The two most important parameters are the number of representative points used and the rate at which the points are moved towards the cluster centre, the compression rate.
4. **ROCK** (pyclustering implementation [129]): ROCK was designed to perform better on categorical data through the introduction of links. The primary parameters for are the maximum distance two points can be to be considered neighbours and the degree of normalization that should be applied during the cluster merging process.

The values of the parameters used with BIRCH, CURE, and ROCK are given in Table 7-4.

TABLE 7-4: HEIRARCHICAL ALGORITHM PARAMETERS

Parameter	Values
BIRCH	
Branching Factor	20, 40, 50, 70
CF-entry diameter	0.3, 0.5, 0.7

CURE	
Representative points count	3, 5, 10, 15, 20
Compression rate	0.2, 0.5, 0.8
ROCK	
Connectivity radius (eps)	8, 10, 12, 15, 18, 25, 50, 100
Degree of normalization (threshold)	0.2, 0.4, 0.6, 0.8

7.3.2 Graph Theory Approach Clustering Algorithms

The two graph theoretic clustering algorithms used are MST based clustering and Spectral graph theory clustering.

1. **MST based clustering** (uses the mst_clustering library [134]): MST based clustering first creates an MST from the data and then cuts an amount of the connections to develop the clusters. The key parameters for MST are the number of edges to cut, which can be given as a fraction of the total edges or an integer, and the minimum cluster size. Clusters smaller than the minimum cluster size will be regarded as noise. The other parameters used by MST is the distance metric to use and whether the MST should be approximated using nearest neighbours. The parameter values are given in Table 7-5.

TABLE 7-5: MST BASED CLUSTERING PARAMETERS

Parameter	Values
Edges to cut	0.02, 0.1, 0.3, 0.5, 0.7, 0.9, 3, 5, 10, 25, 50
Minimum cluster size	5, 10, 15, 20
Distance metric (metric)	Cosine, Euclidean, L2, Minkowski, Manhattan, Chebyshev, Mahalanobis, Hamming, Canberra, Bray-Curtis, Standardised Euclidean, Correlation
Approximate	True, false

2. **Spectral graph theory clustering** (sklearn implementation [128]): Spectral graph theory clustering develops a similarity matrix of the data to convert the data to a lower dimensional space and then creates eigenvectors from the matrix that are used to develop the clusters. The primary parameters for spectral graph theory clustering are the eigenvalue solver used, which is either arpack or lobpcg, and the way the affinity matrix is developed, which can utilize kernel functions. The kernel functions used, and the parameters of the kernel functions are given in Table 7-6. Additional to the kernel functions listed in Table 7-6 the Nearest Neighbours kernel function was also used with default parameters.

TABLE 7-6: KERNEL FUNCTION PARAMETERS

Parameter	Values	Polynomial	Sigmoid	Chi-Squared	RBF	Laplacian	Cosine
Gamma	None, .5, 2, 5, 7.5, 10, 15, 25	X	X	X	X	X	
Coef0	0.5, 1, 2,4	X	X				
Degree	2, 3, 4, 6	X					

7.3.3 Simple Partitioning Approach Clustering Algorithms

The primary partitioning approach algorithm is K-means, the most common clustering algorithm. K-means uses an iterative process to identify the optimal cluster centroids. Multiple variants of k-means are used alongside the original the original version. Other than the number of clusters, k-means and its variants do not take any parameters. The four k-means based algorithms used are k-means, k-means++, k-medians, and fuzzy c-means. The sklearn implementation of k-means and k-means++ and the pyclustering implementation of k-medians and fuzzy c-means were used.

7.3.4 Density Approach Clustering Algorithms

Table 7-7 gives the parameters used with each of the density-based algorithms. Neither density approach takes the number of clusters as a parameter, instead the algorithm determines the number of clusters. Two clustering algorithms based on the density approach are used:

1. **DBSCAN** (sklearn implementation [128]): DBSCAN identifies clusters as groups of closely grouped points. As such, the three parameters are the distance metric to be used, how far two points can be from each other and still considered neighbours (eps), and the minimum number of points required for a point to be a core point.
2. **OPTICS** (sklearn implementation [128]): OPTICS extends upon DBSCAN, using a dynamic version of the eps. Instead, OPTICS takes xi, which gives the minimum steepness on the reachability plot that constitutes a cluster boundary.

TABLE 7-7: DENSITY ALGORITHM PARAMETERS

Parameter	Values
Min Points	2, 5, 10, 15
Distance Metrics	Cosine, Euclidean, L2, Minkowski, Manhattan, Chebyshev, Mahalanobis, Hamming, Canberra, Bray-Curtis, Standardised Euclidean
DBSCAN - eps	.5, .3, .7, 10, 15, 20
OPTICS - xi	0.005, 0.02, 0.05, 0.07, 0.1

7.3.5 Kernel Approach Clustering Algorithms

There are two primary approaches to kernel-based clustering algorithms: 1) adapting an existing clustering algorithm to use a kernel, and 2) developing a clustering algorithm specifically designed to leverage the benefits of kernel-based clustering. One clustering algorithm was selected to represent each approach:

1. **Kernel k-means** (based on the implementation by Blondel [135] which was based on the paper by [136]): Is an adaption of the classic k-means algorithm to use kernels. The kernels used with Kernel k-means are the same as those given in Table 7-6.
2. **Scalable Support Vector Clustering** (custom implementation based on [137]): *Scalable Support Vector Clustering* (SSCV) is a clustering algorithm specifically designed to leverage Gaussian kernels to develop non-convex clusters. SSVC takes a variety of parameters that determine the nature of the kernel used, the parameters and values used are given in Table 7-8.

TABLE 7-8: SSVC PARAMETERS

Parameter	Values
p	0.001, 0.0005
B	100, 300, 500
Q	7, 10, 15
Eps1	0.01, 0.05
Eps2	10 ** -4, 10 ** -6
Step Size	0.5, 0.2, 0.05
Epochs	50

7.3.6 Metaheuristic Approach Clustering Algorithms

Three clustering algorithms were selected, each representing a different category of metaheuristic algorithm:

1. **Genetic Algorithm** (pyclustering implementation [129]): The GA is based on the evolutionary theory approach which mimics biological evolution, inspired by the way chromosomes mutate and evolve over generations. The parameters for GA are the number of chromosomes in each population, the number of populations, number of genes mutated each step, and the exponential coefficient for the selection procedure (select coeff).
2. **ABC** (custom implementation based on the application of ABC for clustering as detailed in [104], [105]): The ABC algorithm is based on the swarm intelligence approach and mimics the

foraging behaviour of honeybee swarms. The three parameters ABC takes is the total number of bees, the maximum number of iterations, and the discard limit, which determines how many iterations should pass before a bee is reset for not improving performance.

3. **SFLA** (custom implementation based on the application of SFLA for clustering as detailed in [32]): The SFLA algorithm is neither an evolutionary or swarm intelligence approach, mimicking the memetics in frogs. The SFLA algorithm takes the number of frogs, the number of memeplexes, the number of iterations within each memeplex, and the total number of iterations shuffling memeplexes.

As ABC and SFLA should not get stuck at the local minima, and instead find the optimum solution, two sets of parameters were used. The first used reasonable parameters that should give reasonable results, and the second were 'long' versions with many more iterations and larger populations to increase the likelihood that the optimal solution was found. The parameters used, including during the 'long' versions of ABC and SFLA, are given in Table 7-9.

TABLE 7-9: METAHEURISTIC ALGORITHM PARAMETERS

Parameter	Values	'Long' Version Values
	Genetic algorithm	
Chromosome count	100, 150, 200	N/A
Population count	100, 250, 300	
Gene mutation count	1, 2, 4	
Select Coeff	0.008, 0.01, 0.012	
	ABC	
Number of bees	30, 60, 100, 150	75, 150, 250, 500
Max iterations	50, 100, 250, 500	250, 500, 1000, 5000, 15000
Discard limit	10, 30	10, 30, 50
	SFLA	
Number of frogs	30, 60, 100, 150	50, 100, 200, 350, 500
Number of memeplexes	3, 5, 10	5, 10, 20
Memeplex iterations	5, 10, 20, 50	10, 20, 50, 75, 100
Max iterations	20, 40, 75, 150	150, 300, 500, 1000

7.3.7 Individual Clustering Algorithms and One-Off Approaches

EMA and SOM were the only clustering algorithms selected to represent their respective clustering approaches: the distribution approach and the model approach. AP clustering and NMF were the two one-off clustering approaches that were selected. The parameter values for these algorithms are given in Table 7-10.

- **EMA** (pyclustering implementation [129]): EMA represents the distribution approach. The two parameters for EMA are the maximum iterations, and the tolerance, which determines the threshold at which the distance between current and previous log-likelihood estimation is considered negligible and the clustering is ended.
- **SOM** (pyclustering implementation [129]): SOM represents the model-based approach. The only parameter, other than the number of clusters, is the max number of iterations (epochs).
- **AP** (sklearn implementation [78]): AP clustering uses a unique message sending approach. AP does not take the number of clusters as a parameter. The parameters AP does take is the max iterations, and the extent to which the current value is maintained relative to incoming values (damping).
- **NMF** (custom implementation of clustering based on the sklearn implementation of the NMF decomposition algorithm [78]): NMF takes the max number of iterations, and the solver used.

TABLE 7-10: OTHER ALGORITHM PARAMETERS

Parameter	Values
EMA	
Tolerance	1e-05, 2e-05, 5e-06
Max iterations	100, 250, 500
SOM	
Number of iterations (epochs)	25, 50, 100, 200, 400
AP	
Damping	0.5, 0.6, 0.7, 0.8, 0.9
Max Iterations	200, 500
NMF	
Solver	MU, CD
Max Iterations	100, 500, 1000

7.3.8 Ensemble Approach Clustering Algorithms

Ensemble algorithms run multiple clustering algorithms and then use a consensus function to combine the results. Two consensus functions are used, the basic consensus function, which was custom implemented based on [106], [138] and the NMF consensus function from the implementation at [139] which was based on [140]. The nine algorithm combinations used are given in Table 7-11.

TABLE 7-11: ENSEMBLE COMBINATIONS

Name	Description
k-means	K-means run 5 times with random initialization – should develop different initial centroids.
k-means++	K-means run 5 times with k-means++ initialization – should develop different initial centroids.
NMF	NMF run twice with each solver (CD and MU)
Spectral	Spectral run with three different affinities (RBF, polynomial, and nearest neighbours)
Spectral and K-means	Spectral run with three different affinities (RBF, polynomial, and nearest neighbours) and k-means run once
K-means, AHC, and NMF	k-means run twice with k-means++ initialization, AHC run twice with Ward's linkage, and NMF run with each solver
ABC	ABC run 6 times, with the number of bees (30, 50, or 100) and max iterations (50 or 100) varied
SFLA	SLFA run multiple times with the number of frogs (30, or 60), number of memeplexes (3, 5, or 10), max iterations (10 or 25), and number of memeplex iterations (5, 10, or 20) varied
ABC and SFLA	ABC was run with the number of bees (30, 50, or 100) and max iterations (50 or 100) varied and SLFA was run with the number of frogs (30, or 60), number of memeplexes (5 or 10), max iterations (10 or 25), and number of memeplex iterations (5 or 10) varied

7.4 Data Preparation

The data from both data sets had to be prepared to be used with the framework, and the key features had to be identified. First, any non-numeric features were converted to numeric values, and one-hot encoding was used for non-ordinal values. The null values were then replaced using an iterative imputer. The data was then partially normalized using min-max normalization. Features that greatly differed in scale to the other features, such as age and years lived in NQ, were normalized to avoid having those values being unfairly weighted. Although only a subset of the data was used as the key features, the entire data set was used for clustering.

The key features, aggregate features, and acronyms were then selected. Table 7-12 gives the features and acronyms that are shared across both data sets. While Table 7-13 and Table 7-14 gives the key features and acronyms that are unique to the 2018 Data Set and the 2021 Data Set, respectively. Most of the differences can be attributed to additional data collected during the second survey. In the second survey the likelihood to install a complete roof replacement was asked about, however as most of the houses are newer and thus not requiring a complete roof replacement. The likelihood to install dead locks was decided to be used to replace the likelihood to install a complete roof replacement as it was relevant to a wider range of individuals. Likelihood to install deadlocks may also provide interesting insights as deadlocks have functionality beyond disaster mitigation, primarily for general security.

TABLE 7-12: KEY FEATURES IN BOTH DATA SETS

Acronym	Feature Description
F	How fearful thinking about the possibility of a cyclone makes the individual feel
H	How helpless thinking about the possibility of a cyclone makes the individual feel
D / De	How depressed thinking about the possibility of a cyclone makes the individual feel
1-2S	How much damage a category 1-2 cyclone would do
3-4S	How much damage a category 3-4 cyclone would do
5S	How much damage a category 5 cyclone would do
1-2C	Likelihood of a category 1-2 cyclone hitting
3-4C	Likelihood of a category 3-4 cyclone hitting
5C	Likelihood of a category 5 cyclone hitting
VA	How visually appealing cyclone shutters are
AO	Whether the individual feels they could organize to have cyclone shutters installed
GS	Whether the government would give financial support in the event of a cyclone
TF	How often the individual discusses or thinks about cyclones
IS	Whether the individual has actively looked for ways to minimize cyclone damage
ICS	*Likelihood to install:* Cyclone Shutters
How likely you are to perform the following next cyclone season/once a cyclone warning is declared	
TT	Trim treetops and branches
CR	Check property for rust, rotten timber, termite infestations and loose fittings
CW	Check that the walls, roof, and eaves of your home are secure
CF	Check fencing is not loose or damaged
CG / CD	Clean gutters and downpipes
Ply	Put plywood up on glass windows/doors
SO	Secure outdoor furniture and garden items
CY	Clear yard of any loose items
Aggregate Features	
Eff	Encompasses the perceived effectiveness of cyclone shutters to reduce damage, keep family safe, to increase property value, and for other purposes.
C	Encompasses financial, time, effort, and knowledge cost of installing cyclone shutters.
PR	Encompasses the perceived personal risk of a cyclone; how the individual's daily life, job, mental health, and physical health would be affected.
GR	Encompasses the perceived general risk of a cyclone, the likelihood of catastrophic destruction, widespread death, the financial threat, and the threat to future generations.

TABLE 7-13: ADDITIONAL KEY FEATURES IN THE 2018 DATA SET

Acronym	Feature Description
S	How stressed thinking about the possibility of a cyclone makes the individual feel
IRR	*Likelihood to install:* Roof Replacement

TABLE 7-14: ADDITIONAL KEY FEATURES IN THE 2021 DATA SET

Acronym	Feature Description
Dr	How much dread thinking about the possibility of a cyclone makes the individual feel
W	How worried thinking about the possibility of a cyclone makes the individual feel
IDL	*Likelihood to install:* Dead Locks
RSS	*How likely you are to* Remove shade sails *next cyclone season/once a cyclone warning is declared*

7.5 Internal Metric Thresholds

The final step before running the framework was to set the internal metrics. For both data sets the minimum data size threshold was set to 5% of the total data set, as smaller clusters are likely to be too imbalanced, representing edge cases. Additionally, for both data sets the minimum SC value was set to 0, and the maximum DBI value was set to 5, as poorer performances indicate overlapping clusters. For the 2018 Data Set, the minimum CHI value was set to 10 and the minimum AFS value was set to 15. While in the 2021 Data Set the minimum CHI value was set to 7 and the minimum AFS value was set to 20. The AFS threshold was set to be equal to approximately half the significant features, as if less than half of the features significantly differed the personas created from them were unlikely to have significantly different behavioural features. The thresholds of SC, DBI, and CHI were identified through a small amount of preliminary testing.

7.6 Summary

To facilitate the answering of the remaining SRQs an additional, confirmatory, data set was collected. This chapter also laid out each of the steps to be taken to address each of the SRQs, including the key features identified within each data set, the parameter values that were used with each algorithm, and the internal metric thresholds that were set. The next chapters detail the results of running HyPersona with these settings and begin to address each of the remaining SRQs.

Chapter 8: COMPARISON OF ALGORITHM PERFORMANCE

The results of the HyPersona framework on each data set across a wide variety of algorithms and parameters were compared and analysed to address the primary research question. To address the primary research questions, the second and third SRQs were first addressed:

SRQ2. Are the performances of clustering algorithms and approaches to clustering for persona development consistent?

SRQ3. Does the selection of a clustering algorithm and parameters significantly impact the set of personas developed?

These SRQ focussed on the performance of the clustering algorithms and the clusters developed and were required to facilitate the selection of the best performing algorithm-parameter combination, the creation of the persona sets, and to determine the applicability of the results.

The second SRQ focussed on the consistency of algorithm performance. The performances of two algorithm-parameter combinations were considered consistent if they were dropped for the same reason, produced similar clusters, or produced clusters with similar internal metrics. A series of algorithm-parameter combinations were also considered consistent if the clusters developed by them were identical or functionally identical. The outputs of two algorithm-parameter combinations were considered identical if they produce the exact same set of clusters and functionally identical when the differences between the cluster sets were minor enough that a set of personas developed from each cluster set would be identical. That is, the differences were not great enough to affect the interpretation of the clusters. As whether a pair of cluster sets are functionally identical is based on the values of the features, not the quality of the clusters, two cluster sets could be functionally identical with significantly different metrics.

The third SRQ focussed on the impact of the algorithm-parameter combination on the cluster set developed, as if there is no significant difference between the cluster sets developed, there is no point in comparing algorithm-parameter combinations. Beyond not being identical or functionally identical, two cluster sets are considered significantly different if their interpretation would lead to significantly different persona sets. Two personas were most considerably different when they represent completely different states within the behavioural theory. Prior to addressing the two SRQs, the results of HyPersona first had to be detailed and analysed.

8.1 Overview of HyPersona Results

HyPersona was applied to both the 2018 and 2021 Data Sets, using the identical algorithm-parameter combinations. Over 20 algorithms each with a variety of parameter options were used, resulting in a total of 3,404 algorithm-parameter combinations run across each data set. The implementation details of each clustering algorithm and the list of parameters used are given in Chapter 7.

Not all algorithm-parameter combinations developed valid sets. On each data set a small number of algorithm-parameter combinations were unable to converge or resulted in another error, which stopped the clustering algorithm from developing a set of clusters.

TABLE 8-1: OVERVIEW OF THE RESULTS OF THE HYPERSONA FRAMEWORK

	Valid		Dropped		Errored		Total	
2018 Data Set	1,085	(31.9%)	2,155	(63.3%)	164	(4.8%)	3,404	(100%)
2021 Data Set	1,050	(30.8%)	2,230	(65.5%)	124	(3.6%)	3,404	(100%)

From those that developed a cluster set, most algorithm-parameter combinations were dropped as the cluster set did not meet the necessary thresholds as defined in Section 7.5. The cluster count and cluster size thresholds were the most common reasons for an algorithm-parameter combination to be dropped. Many clustering algorithms had the tendency to cluster all the data points together. Which resulted in only one cluster being developed by the clustering algorithms which determined the ideal number of clusters, such as DBSCAN and OPTICS, or extremely imbalanced clusters being developed by the clustering algorithms which developed a set number of clusters, such as CURE. Table 8-1 gives the distribution of algorithm-parameter results.

Table 8-1 shows that between the 2018 Data Set and the 2021 Data Set a similar number of algorithm-parameter combinations were dropped and errored, resulting in less than one-third of the algorithm-parameter combinations remaining as producing valid clusters from both data sets. As such, the HyPersona framework eliminated a large amount of potential manual analysis. However, over 1,000 potentially valid algorithm-parameter combinations were remaining that still required analysis.

The internal metrics were used as an indicator of algorithm performance to further narrow down the potential algorithm-parameter combinations. An overall performance metric was calculated to allow the overall performance of each algorithm-parameter combinations in terms of the four internal metrics used to be compared. The overall performance metric was calculated based sum each internal metric's 'rank' for the algorithm-parameter combination. Where the 'rank' of a metric is the min-max normalised value of the metric. As such, the overall performance metric value is bounded between zero and four, where four is the best possible value. The overall performance metric could then be used with the individual internal metrics to identify potential best performers.

8.2 Result Analysis Approach

The results of each clustering algorithm and parameter combination were analysed to address SRQ two and three. SRQ2 focuses on the consistency of the performance of each clustering algorithm while SRQ3 focuses on whether significantly different cluster sets of a similar quality can be produced by different algorithm-parameter combinations. Before either SRQ could be addressed, whether the internal metrics could be used to identify functionally identical cluster sets was determined. The first phase on analysis involved analysing random subsets of the algorithm-parameter combinations.

To determine whether the internal metrics could be reasonably used to identify whether two algorithm-parameter combinations were functionally identical, sets of algorithm-parameter combinations with similar internal metrics were analysed. Each set of algorithm-parameter combinations contained combinations where each individual internal metrics was significantly similar and combinations where only the overall performance metrics were similar. Through identifying that the individual internal metrics functioned as a strong indicator of functional identicality the amount of manual analysis required was significantly reduced.

The process of determining if the internal metrics could indicate cluster set similarity, that is, how similar the clusters in two cluster sets were, was also a key step in addressing SRQ3. As not all the results were functionally identical, the clustering algorithm and parameters had an impact on the cluster sets developed. However, a key element to addressing SRQ3 is whether there is a significant difference between cluster sets of similar quality, especially between the cluster sets with the best overall metric performances. As, if there are no significant differences, the internal metrics alone would be able to identify the best performer without need for domain-specific evaluation.

SRQ3 was addressed in two phases: first during the evaluation of the algorithm-parameter subsets, and then during the selection of the best overall performing algorithm-parameter combination on each data set. If the algorithm-parameter combinations within the subsets did not significantly differ, then the algorithm-parameter combination selected did not have a significant impact on the cluster sets developed when considering overall metric performance developed.

The second phase of analysis evaluated the overall performance and consistency of each algorithm to address SRQ2. The subset evaluation was performed prior to addressing SRQ2 to determine whether the internal metrics could be used to estimate cluster quality, which greatly reduced the amount of manual analysis required. There were multiple manners in which a clustering algorithm could be considered consistent. Three types of consistencies that were focussed on were:

1. **Algorithm Consistency:** Whether the performances of each algorithm-parameter set for a given algorithm is consistent. Where a particular parameter has considerable impact on the performance of an algorithm, the algorithm consistency may be considered regarding that parameter.
2. **Approach Consistency:** Whether the performances of each algorithm within an approach is consistent. For algorithms that are the sole representative for their approach, the approach consistency was not considered.
3. **Across Data Set Consistency:** Whether an algorithm's performance is consistent over both the 2018 Data Set and the 2021 Data Set.

Each type of consistency gave different insight into the overall performance of the clustering algorithm and how the results could be generalised and applied to differing algorithms, use cases, and data sets. The performance of an algorithm-parameter combination was determined by whether the algorithm-parameter combination was dropped or errored, the internal metrics of the cluster set, and the quality of the cluster set for persona development. Two algorithm-parameter combinations were also considered consistent if they produced identical or functionally identical results based on the internal metrics.

8.3 Comparison of Algorithm-Parameter Combination Subsets

To begin to address SRQ3 and determine whether the internal metrics could be used to indicate whether two algorithm-parameter combinations were functionally identical, subsets of algorithm-parameter combinations were analysed. To quantify the metric similarity, the Euclidian distance between the ranked metrics was used and a distance less than 0.1 was considered almost identical. Each subset of algorithm-parameter combinations was selected to represent a particular range of metric performances. The subsets also focused on algorithm-parameter combinations that gave a similar performance across both data sets. Three algorithm-parameter combination subsets were defined:

1. **The Good Overall Performance Subset:** Four algorithm-parameter combinations were selected that had a good overall performance metric. Three of which had similar performances across all the internal metrics, and one which had a similar overall performance metric value but each of the internal metric ranks differed.
2. **The Top AFS Performance Subset:** Three algorithm-parameter combinations with the best AFS values were selected, two of which were among the top performers for AFS across both data sets and were both algorithm-parameter combinations using the genetic clustering algorithm. The third algorithm-parameter combination was the best performer for AFS that did not use the genetic clustering algorithm.
3. **The Poor Overall Performance Subset:** Three algorithm-parameter combinations that consistently gave the poor overall performances without being dropped were selected.

The Good Overall Performance Subset was focused on for determining whether the internal metrics could suggest whether two algorithm-parameter combinations were functionally identical. The Top AFS Performance Subset was also important in determining the strength of AFS as an indicator of cluster quality. The Poor Overall Performance Subset was important to determine whether the internal metrics and overall performance metric could accurately suggest cluster quality and whether the cluster sets developed differed significantly as the metric performance differed.

8.3.1 Comparison of the Good Overall Performance Algorithm-Parameter Combination Subset

The Good Overall Performance Subset contained four algorithm-parameter combinations, each of which used a different base clustering algorithm but achieved a similar overall performance score. The overall performance score for the algorithm-parameter combinations within the Good Overall Performance Subset were amongst the best achieved. The four algorithm-parameter combinations selected were:

1. **k-means v0:** k-means with k-means++ initialisation
2. **k-means++ ensemble:** An ensemble of 5 k-means++ algorithms using the basic consensus function to combine results.
3. **SOM v0:** SOM clustering algorithm with the epoch set to 25. All the variations of SOM, with different epoch values, developed identical cluster sets.
4. **SFLA long v212:** SFLA based clustering with one of the 'long' parameter sets; 350 frogs, 10 memeplexes, 75 memeplex iterations, and 150 top level iterations.

SFLA long v212 was the algorithm-parameter combination with a similar overall performance metric but differing internal metric values, while the other three algorithm-parameter combinations had almost identical internal metrics. In comparison to the other three algorithm-parameter combinations, SFLA long v212 achieved a higher AFS score but a lower CHI and DBI values. Table 8-2 gives the internal metrics of each algorithm-parameter combination on each data set, including both the literal metric values and the normalized metric rank. Between k-means v0, the k-means++ ensemble, and SOM v0 the differences in internal metrics were extremely minor.

TABLE 8-2: GOOD OVERALL PERFORMANCE ALGORITHM-PARAMETER COMBINATION SUBSET INTERNAL METRICS

Data Set	Algorithm	SC		CHI		DBI		AFS		Overall
		Value	Rank	Value	Rank	Value	Rank	Value	Rank	
2018 Data Set	k-means v0	0.0860	0.92	47.08	1.00	2.94	0.91	48.33	0.44	3.26
	k-means++ ensemble	0.0864	0.93	47.07	1.00	2.94	0.91	48.50	0.44	3.27

	SOM v0	0.0864	0.93	46.43	0.98	2.88	0.94	46.17	0.37	3.21
	SFLA long v212	0.0845	0.90	42.77	0.87	3.35	0.73	57.67	0.74	3.24
2021 Data Set	k-means v0	0.0607	0.87	15.43	1.00	3.25	0.89	78.33	0.64	3.40
	k-means++ ensemble	0.0600	0.85	15.36	0.99	3.25	0.89	78.00	0.64	3.37
	SOM v0	0.0587	0.83	15.27	0.98	3.27	0.88	76.50	0.61	3.30
	SFLA long v212	0.0406	0.51	14.51	0.89	3.15	0.94	93.00	0.87	3.22

The primary focus of the Good Overall Performance Subset was to determine whether algorithm-parameter combinations with significantly similar internal metrics were more likely to be functionally identical. Specifically, to address the hypotheses:

> **$H1_0$**: There is no relationship between the internal metrics of a pair of cluster sets and how similar the cluster sets are.
>
> **$H1_A$**: The internal metrics of a pair of cluster sets indicate how similar the cluster sets are.

If the alternate hypothesis, $H1_A$, is true then a pair of cluster sets with identical or almost identical internal metrics can be expected to be functionally identical. Based on $H1_A$ and the Euclidian distances, given in Table 8-3, the k-means v0, k-means++ ensemble, and SOM v0 cluster sets were expected to be functionally identical. Additionally, as the internal metrics of the SFLA long v212 cluster set differs more significantly to the other cluster sets, if $H1_A$ is true then the cluster set developed by SFLA long v212 could be expected to significantly differ to the cluster sets developed by k-means v0, the k-means++ ensemble, and SOM v0. If the cluster set developed by SFLA long v212 is significantly different to those developed by the other algorithm-parameter combinations although having a similar overall performance metric, that would begin to allow SRQ2 to be answered positively.

TABLE 8-3: GOOD OVERALL PERFORMANCE ALGORITHM-PARAMETER COMBINATION SUBSET EUCLIDIAN DISTANCE BETWEEN INTERNAL METRICS

	k-means v0	**k-means++ ensemble**	**SOM v0**	**SFLA long v212**
k-means v0	0.000			
k-means++ ensemble	0.010	0.000		
SOM v0	0.079	0.079	0.000	
SFLA long v212	0.374	0.374	0.440	0.000

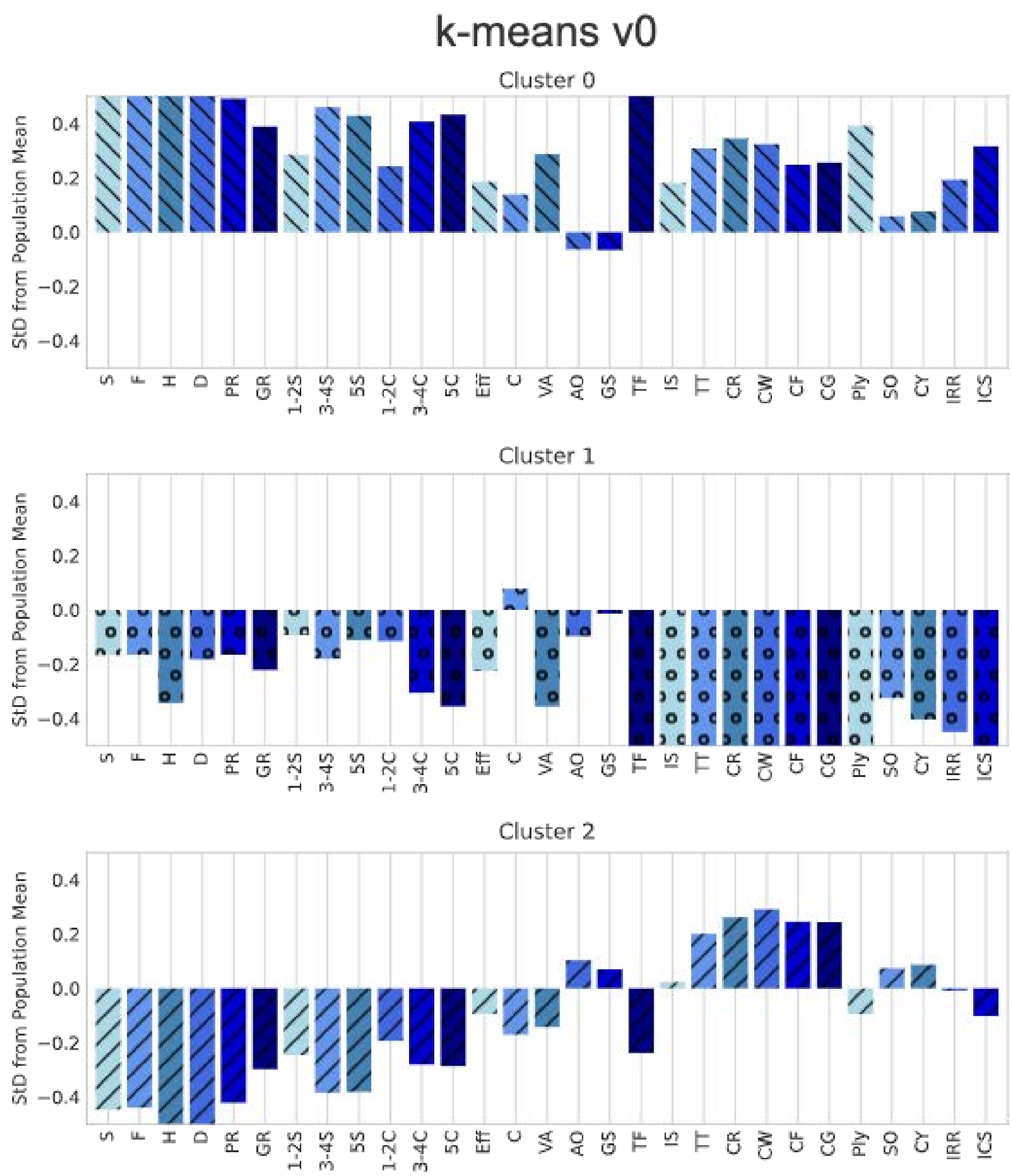


FIGURE 8-1: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2018 DATA SET – K-MEANS V0

To evaluate the results in terms of the hypotheses, the clusters developed by each algorithm-parameter combination were compared using the graphs developed by the HyPersona framework. To allow for the most straight forward comparison, the clusters were re-ordered. Figure 8-1, Figure 8-2, Figure 8-3, and Figure 8-4 gives the graphs for the clusters developed by each of the algorithm-parameter combinations on the 2018 Data Set and Figure 8-5, Figure 8-6, Figure 8-7, and Figure 8-8 gives the same graphs as developed from the 2021 Data Set. The cluster sets were then further evaluated using the CSV files and early-stage personas produced by HyPersona to compare the values and significance of each key feature.

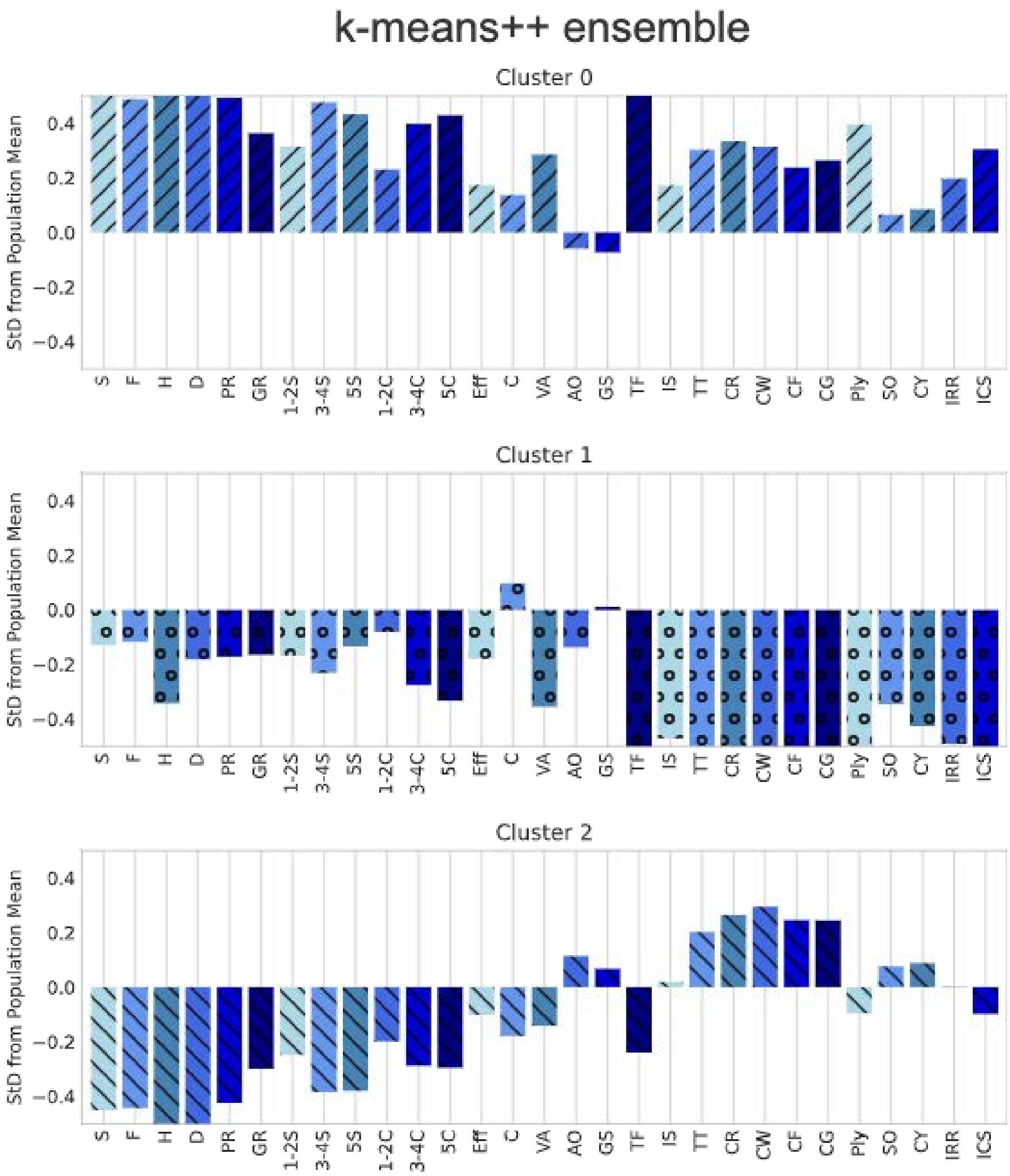


FIGURE 8-2: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2018 DATA SET – K-MEANS++ ENSEMBLE

Table 8-4 gives the sizes of the clusters developed by each of the algorithm-parameter combinations on the 2018 Data Set, as detailed in the CSV files developed by HyPersona. The cluster sizes indicated that there was only one data point clustered differently between the k-means v0 and the k-means++ ensemble cluster sets. Further, through comparing the early-stage personas only very few values of the clusters were identified that differed, most of which did not differ by more than 0.01 standard deviations.

TABLE 8-4: SIZES OF CLUSTERS DEVELOPED BY THE GOOD OVERALL PERFORMANCE ALGORITHM-PARAMETER COMBINATION SUBSET ON THE 2018 DATA SET

Cluster	k-means v0	k-means++ ensemble	SOM v0	SFLA long v212
Cluster 0	216	215	221	226
Cluster 1	88	88	73	109
Cluster 2	215	216	225	184

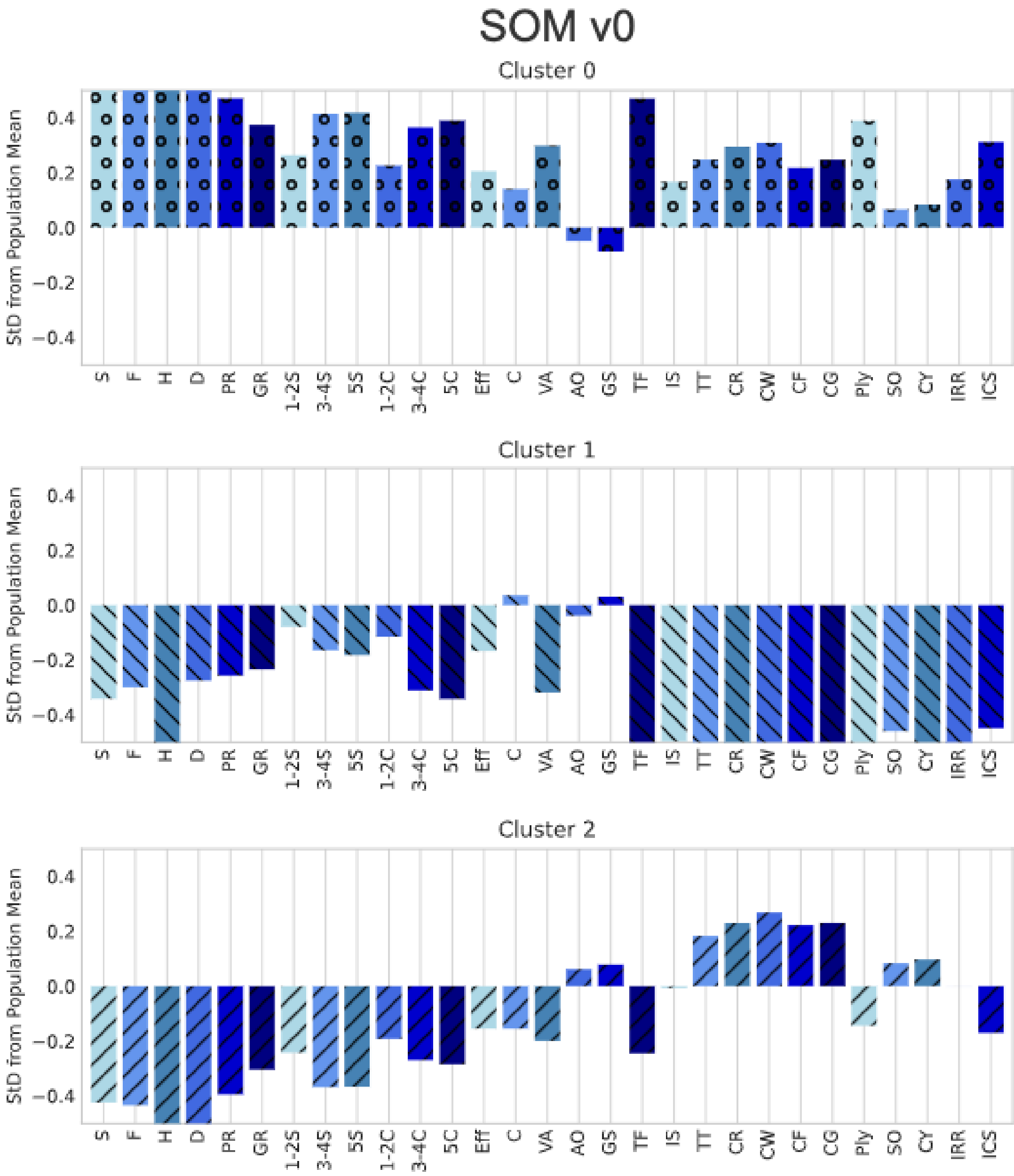


FIGURE 8-3: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2018 DATA SET – SOM V0

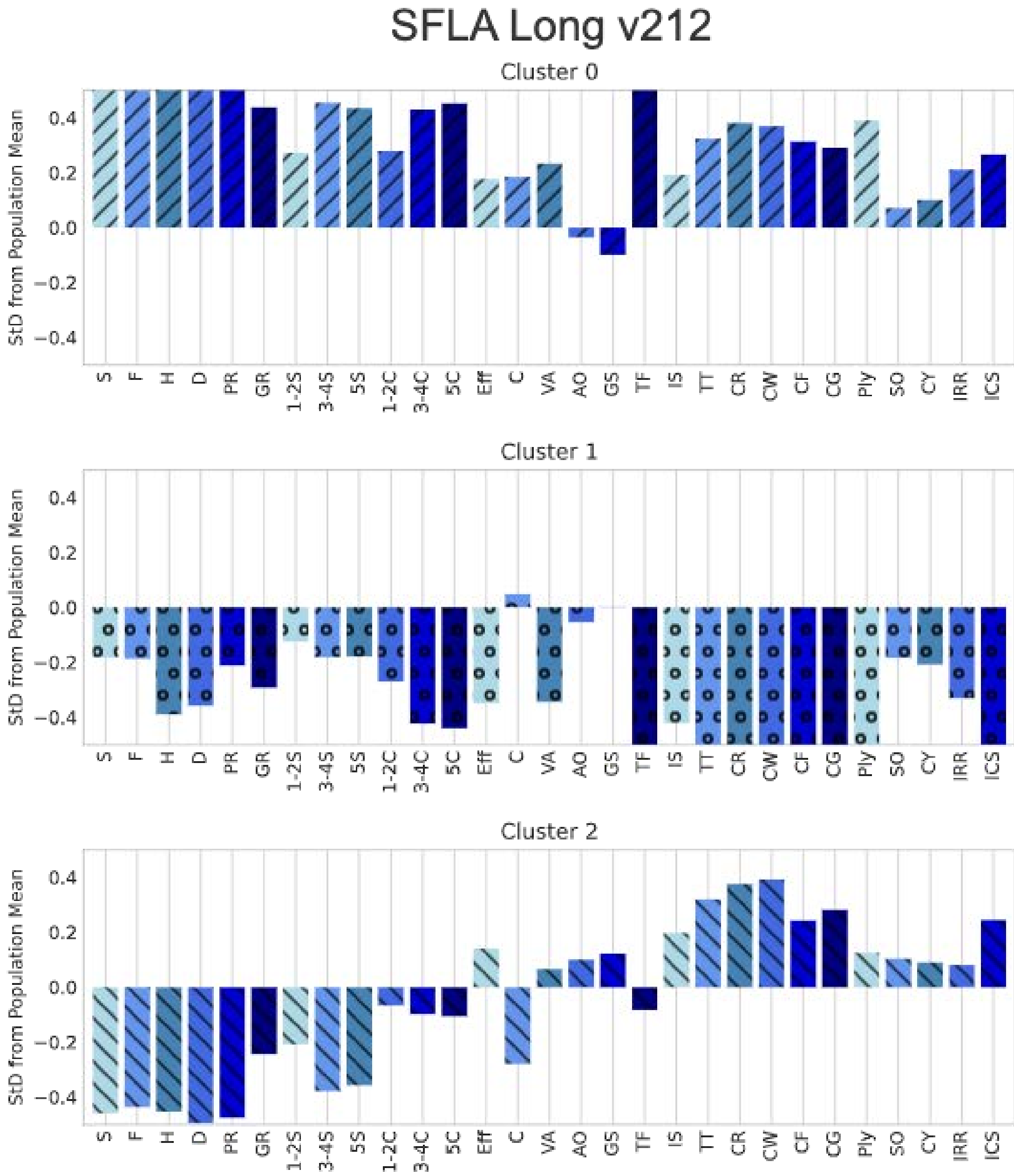


FIGURE 8-4: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2018 DATA SET – SFLA LONG V212

There were also only minor differences seen in the graphs of the k-means v0 and the k-means++ ensemble cluster sets on the 2021 Data Set. Based on these factors, k-means v0 and the k-means++ ensemble on both the 2018 Data Set and the 2021 Data Set were determined to be functionally identical. This aligned with $H1_A$, as the internal metrics of k-means++ and the k-means++ ensemble were the most similar.

There were more prominent differences between the cluster sets developed by SOM v0 compared to those developed by k-means v0 and the k-means++ ensemble. The graphs given in the graphs only show very minor differences between the cluster sets and Table 8-4 shows that the cluster sizes were quite

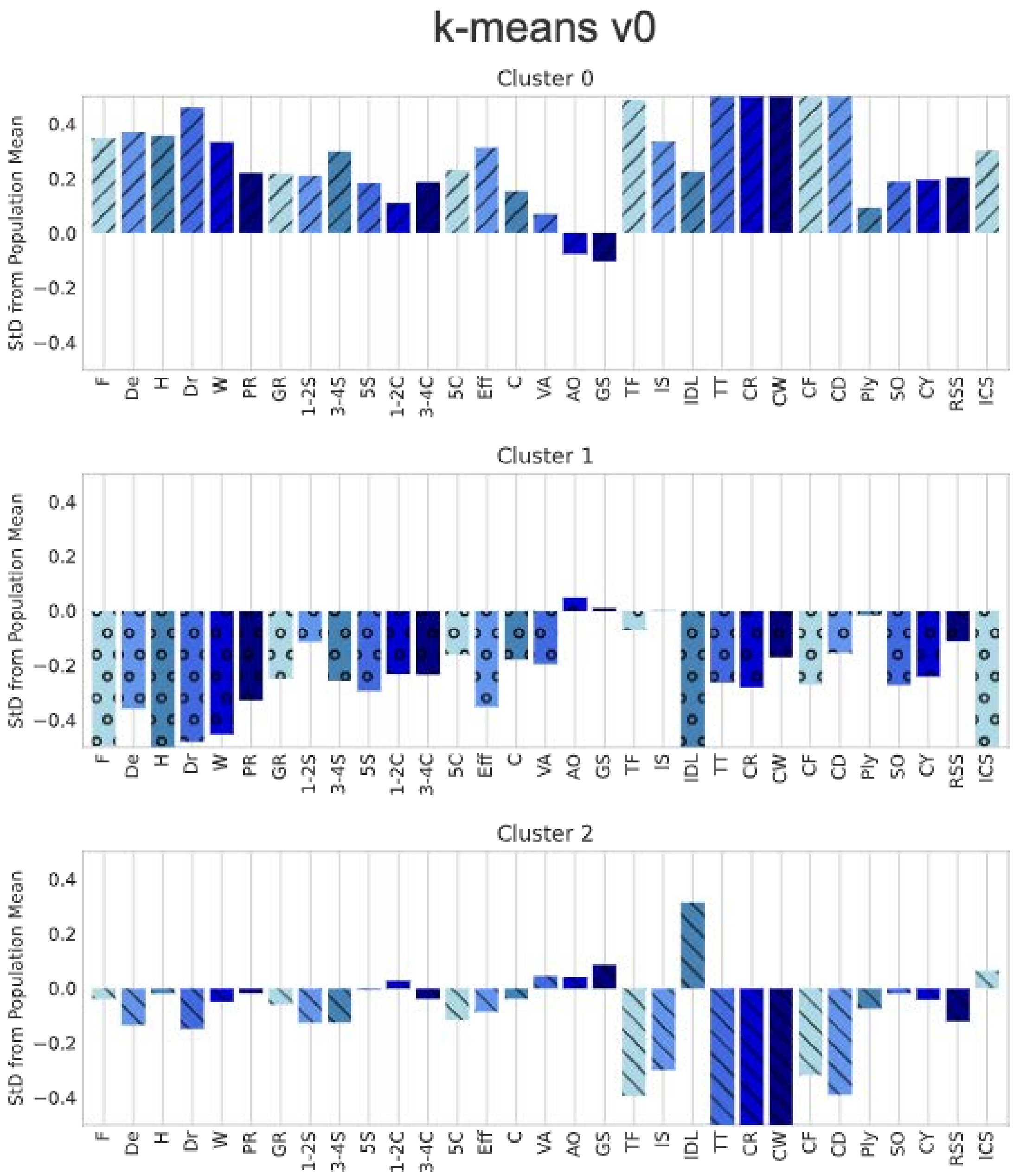


FIGURE 8-5: CLUSTER SET GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2021 DATA SET – K-MEANS V0

similar. To provide further confidence to whether the cluster sets were functionally identical, the value and significance of each feature of each cluster were compared using the CSV files developed by HyPersona. When the features of the clusters developed by SOM v0 were directly compared to those developed by k-means v0, there were a total of 23 features across the three clusters that significantly differed. None of which differed by more than 0.5 standard deviations. Despite the differences between the cluster sets, a persona set developed from the SOM v0, k-means v0, or the k-means++ ensemble clusters would not meaningfully differ to one another. As such, the cluster sets developed by SOM v0 were deemed to be functionally identical to the cluster sets developed by k-means v0 and the k-means++ ensemble.

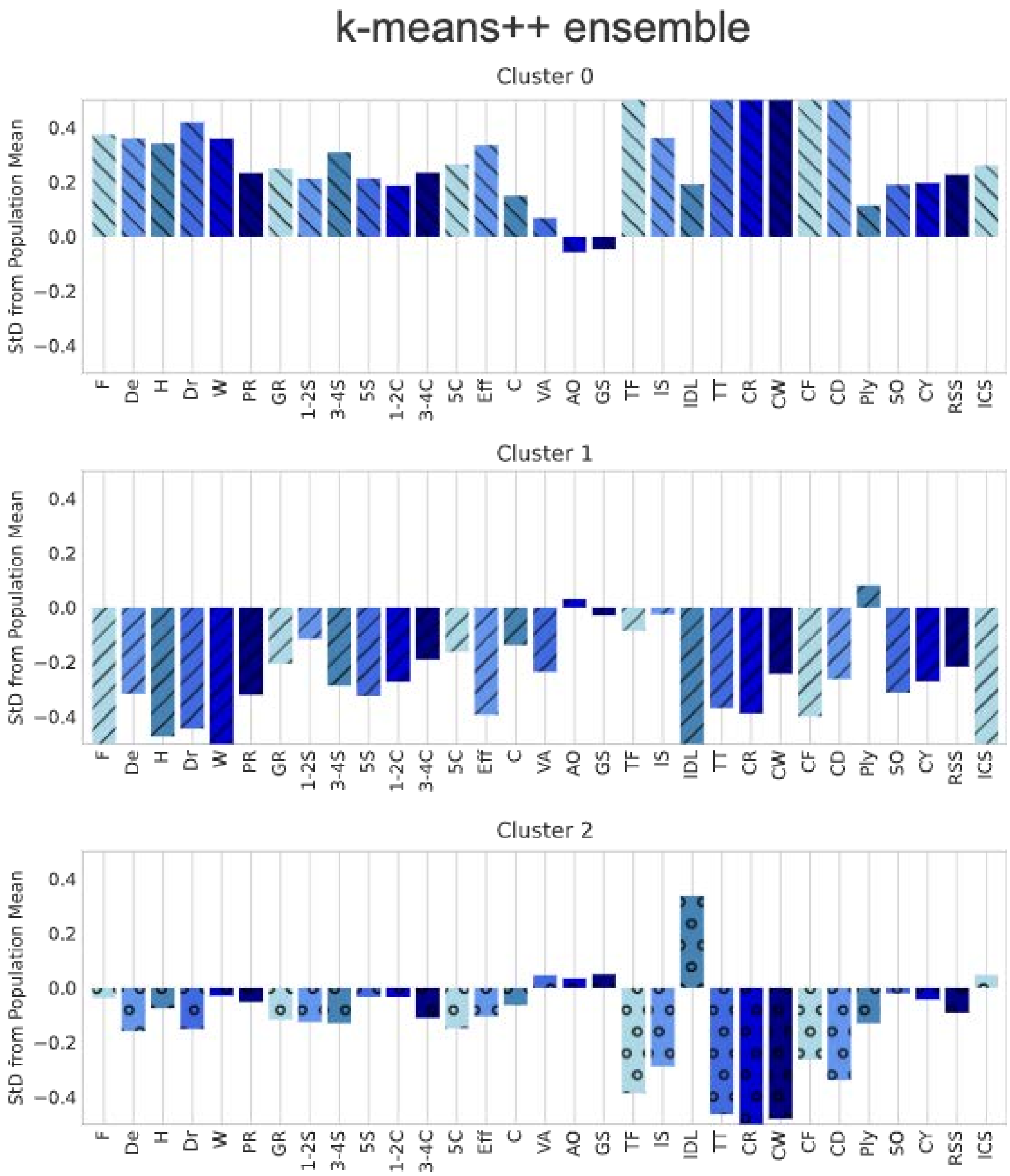


FIGURE 8-6: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2021 DATA SET – K-MEANS++ ENSEMBLE

Alternately, significant differences were seen between the cluster sets developed by SFLA long v212 and the cluster sets developed by k-means v0, the k-means++ ensemble, and SOM v0. The differences were apparent from the graphs developed for each cluster set, with the most significant differences between the 2021 Data Set clusters. Most prominently, when Cluster 2 was compared nearly all the features of the clusters were completely different. Cluster 2 developed by k-means v0, the k-means++ ensemble, and SOM v0 had risk perceptions slightly below the mean and had a much lower likelihood to perform most the preparatory behaviours. While the Cluster 2 developed by SFLA long v212 had a much higher than average level of risk perception, and was likely to perform most preparatory behaviours, especially the installation of cyclone shutters. The difference in Cluster 2 alone was enough

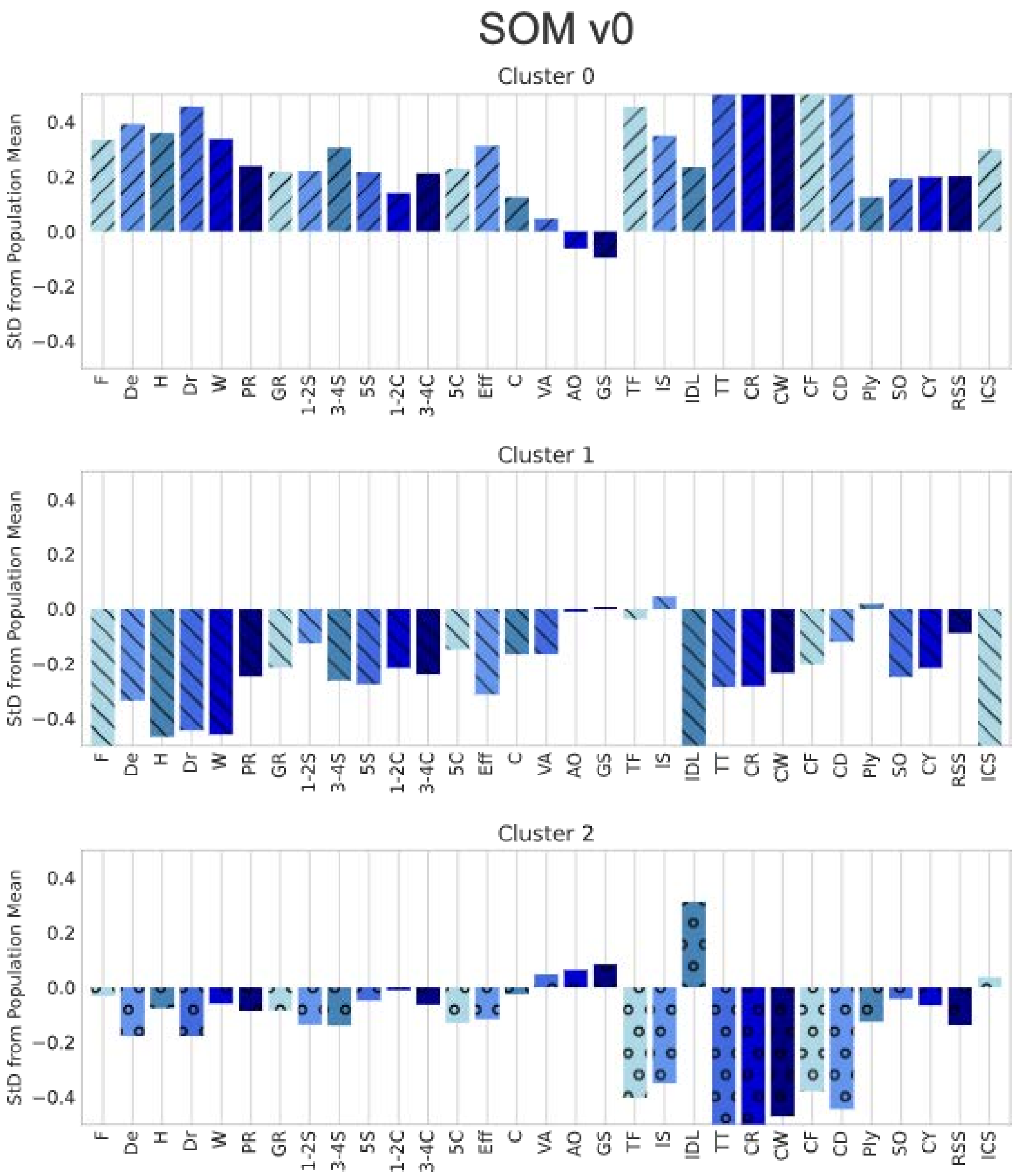


FIGURE 8-7: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2021 DATA SET – SOM V0

to determine that SFLA long v212 cannot be functionally identical to k-means v0, the k-means++ ensemble, and SOM v0.

There were also significant differences between the cluster set developed on the 2018 Data Set by SFLA long v212 compared to those developed by k-means v0, the k-means++ ensemble, and SOM v0. Again, the differences were primarily found in Cluster 2, however there were features across all clusters differed significantly enough that the differences could affect the interpretation of the clusters. As such, the cluster set developed by SFLA long v212 on the 2018 Data Set was also considered functionally independent.

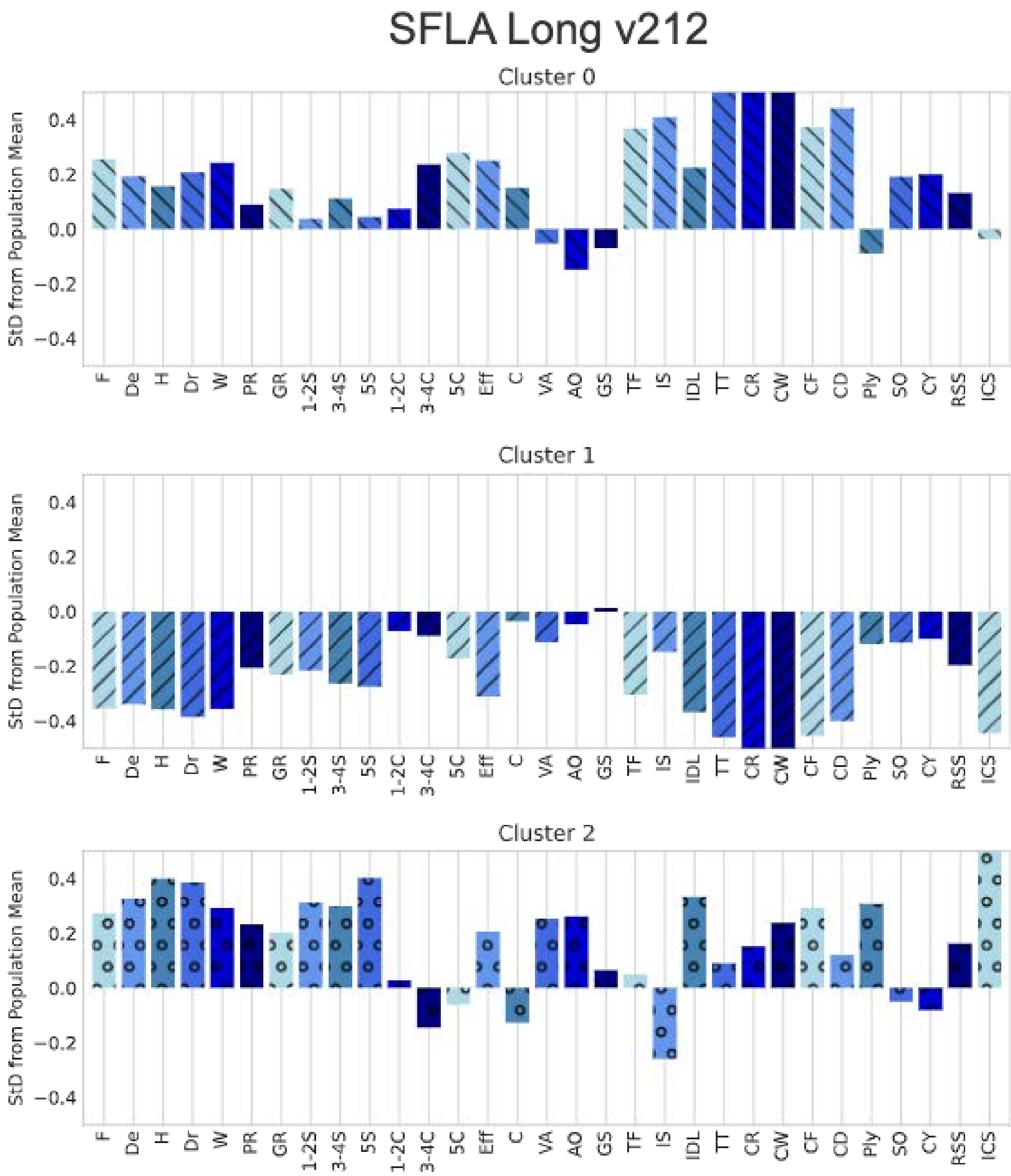


FIGURE 8-8: GRAPHS FOR THE GOOD OVERALL PERFORMANCE SUBSET ON THE 2021 DATA SET – SFLA LONG V212

K-means v0, the k-means++ ensemble, and SOM v0 were all found produce functionally identical cluster sets, supporting $H1_A$, that there was a relationship between the values of the internal metrics and the similarity of clusters. While SFLA long v212 was significantly different, further supporting $H1_A$, and suggesting that overall performance metric alone cannot indicate whether two algorithm-parameter combinations are functionally identical. Based on the analysis of the Good Overall Performance Subset the null hypothesis, $H1_0$, was rejected and the alternate hypotheses, $H1_A$, was accepted.

8.3.2 Comparison of the Top AFS Performance Algorithm-Parameter Combination Subset

The Top AFS Performance Subset contained the algorithm-parameter combinations that created the cluster sets with the highest AFS scores. The algorithm-parameter combinations based on the genetic clustering algorithm consistently developed the clusters with the highest AFS scores. Two of the algorithm-parameter combinations that consistently had the highest AFS score were selected to be a part of the Top AFS Performance Subset, both of which were based on the genetic algorithm. The third algorithm-parameter combination was the algorithm-parameter combination with the highest AFS score that was not based on the genetic algorithm. The three algorithm-parameter combinations in the Top AFS Performance Subset were:

1. **genetic v50:** Genetic clustering with 150 chromosomes, 300 populations, 2 genes mutated at each step, and a selection coefficient of 0.012.
2. **genetic v74:** Genetic clustering with 200 chromosomes, 300 populations, 1 gene mutated at each step, and a selection coefficient of 0.012.
3. **spectral v18:** Spectral graph theory clustering with the ARPACK eigen solver and the Laplacian kernel function with the gamma set to 'None' which defaults to 1.0/number of features.

All the algorithm-parameter combinations in the Top AFS Performance Subset had similar AFS scores and overall performance metric, but the other internal metrics differed more. The values of the internal metrics for each algorithm-parameter combination on each data set is given in Table 8-5. Based on the internal metrics and the findings of the Good Overall Performance Subset, the cluster sets developed by genetic v50 and genetic v74 can be assumed to be functionally identical. While the cluster sets developed by all three algorithm-parameter combinations on the 2018 Data Sets can be assumed to be similar but not functionally identical.

TABLE 8-5: TOP AFS PERFORMANCE ALGORITHM-PARAMETER COMBINATION SUBSET INTERNAL METRICS

Data Set	Algorithm	SC		CHI		DBI		AFS		Overall
		Value	Rank	Value	Rank	Value	Rank	Value	Rank	
2018 Data Set	genetic v50	0.0551	0.57	33.86	0.62	4.41	0.26	65.83	1.00	2.44
	genetic v74	0.0460	0.46	30.42	0.52	3.85	0.51	65.83	1.00	2.48
	spectral v18	0.0496	0.50	32.05	0.56	3.80	0.53	61.00	0.84	2.43
2021 Data Set	genetic v50	0.0425	0.55	13.09	0.72	3.46	0.78	99.17	0.97	3.02
	genetic v74	0.0492	0.66	12.93	0.70	3.42	0.80	100.33	0.99	3.15
	spectral v18	0.0137	0.04	13.58	0.78	3.33	0.85	97.00	0.94	2.60

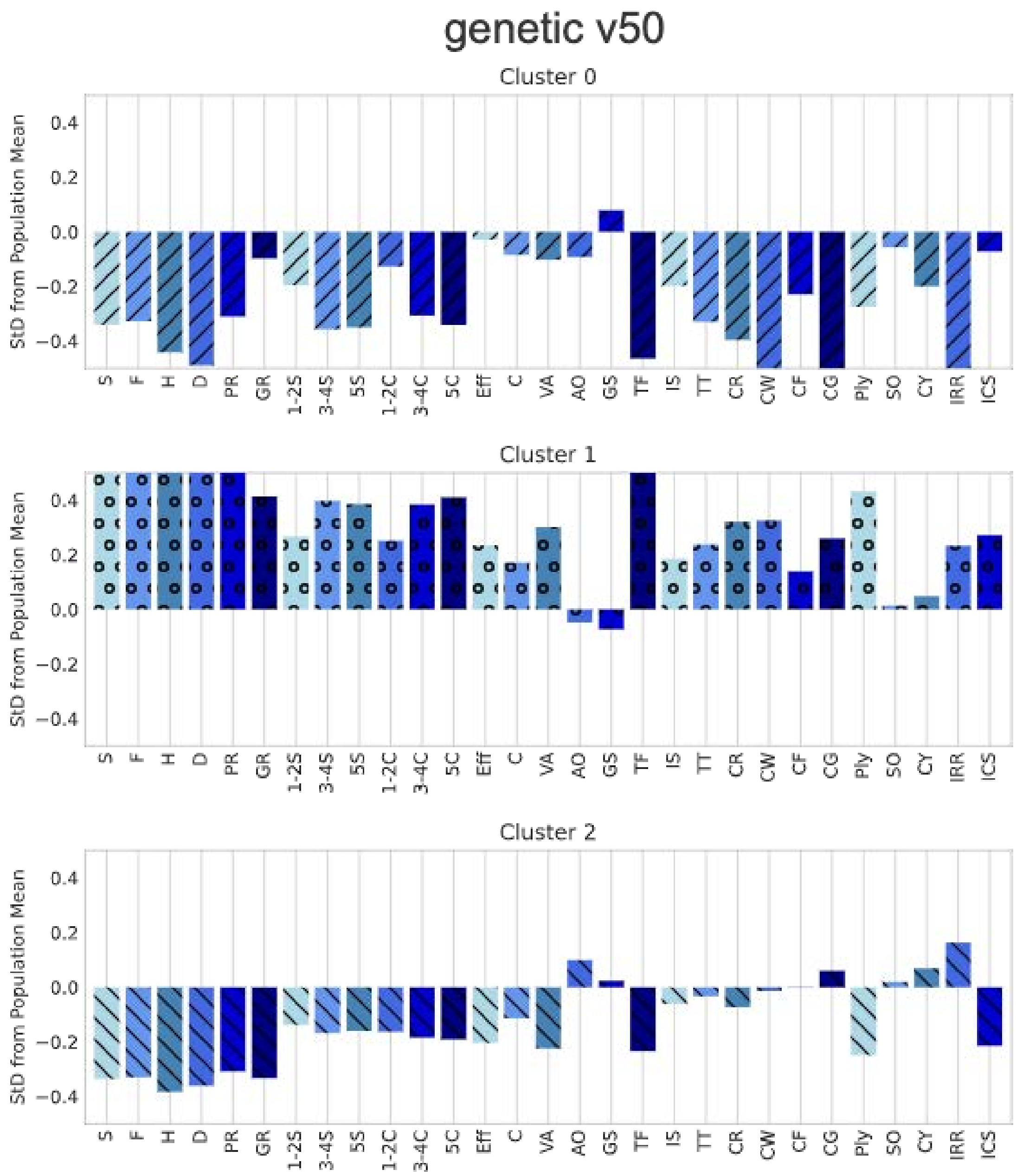


FIGURE 8-9: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2018 DATA SET – GENETIC V50

The primary purpose of examining the Top AFS Performer Subset was to identify the strength of AFS as an indicator of cluster quality, beyond the findings of the preliminary study in Chapter 6. The AFS was designed to indicate how distinct the clusters developed were in comparison to each other and the general population. Based on AFS, the cluster sets within the Top AFS Performance Subset were expected to be distinct and able to be used to develop a set of three significantly different personas.

To determine the quality and similarity of the cluster sets developed, the graphs created by HyPersona for the cluster sets developed by genetic v50, genetic v74, and spectral v18 on the 2018 Data Set, given

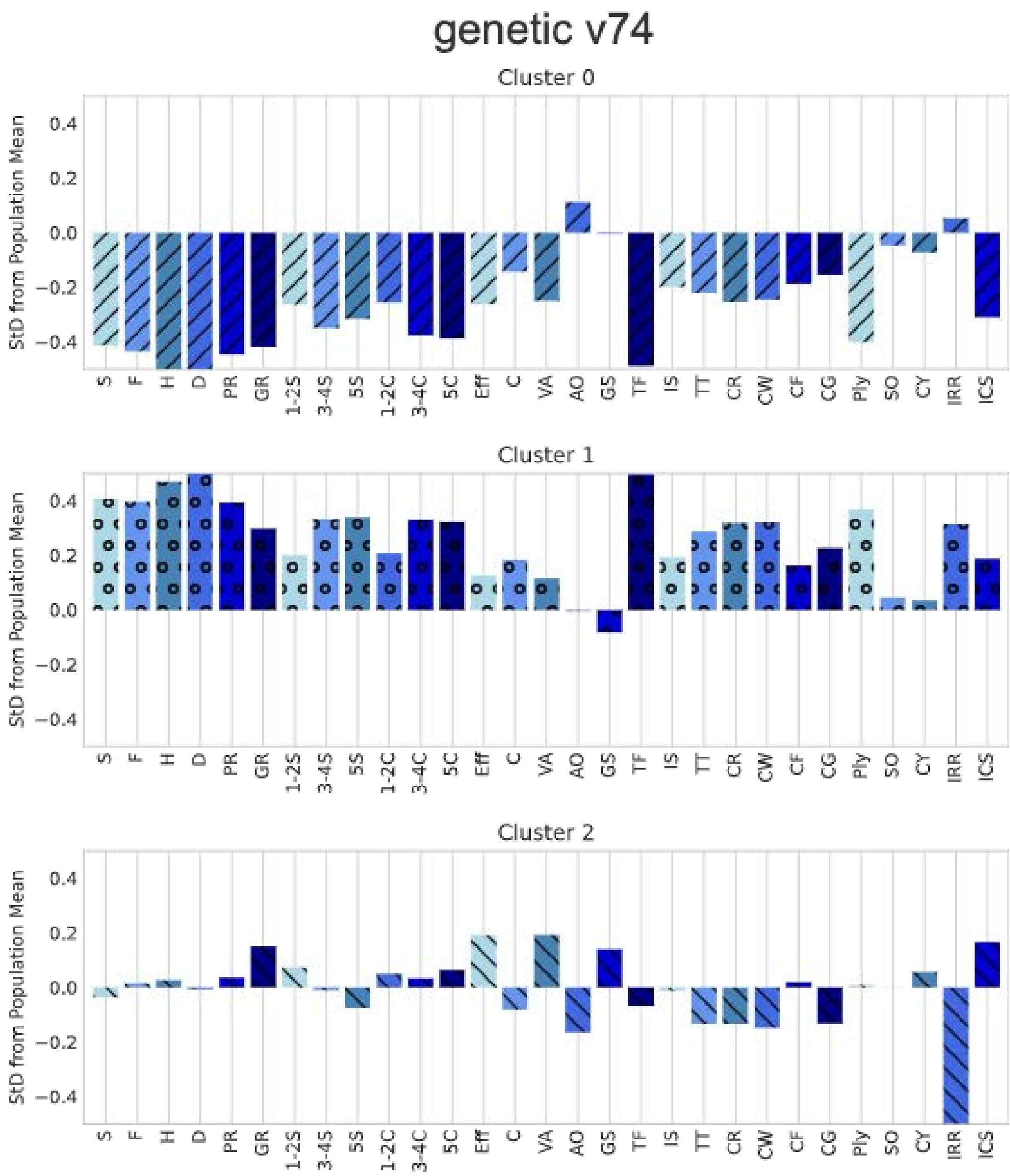


FIGURE 8-10: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2018 DATA SET – GENETIC V74

in Figure 8-9, Figure 8-10, and Figure 8-11, and the 2021 Data Set, given in Figure 8-12, Figure 8-13, and Figure 8-14, were first analysed. Despite having the same AFS, the cluster sets developed by genetic v50 and genetic v74 on the 2018 Data Set were significantly different. Which further supports the premise that all the internal metrics must be significantly similar to indicate that two cluster sets are functionally identical.

The differences between the cluster sets developed by genetic v74 and spectral v18 on the 2018 Data Set were more minor, but the cluster sets still were not considered functionally identical. The primary

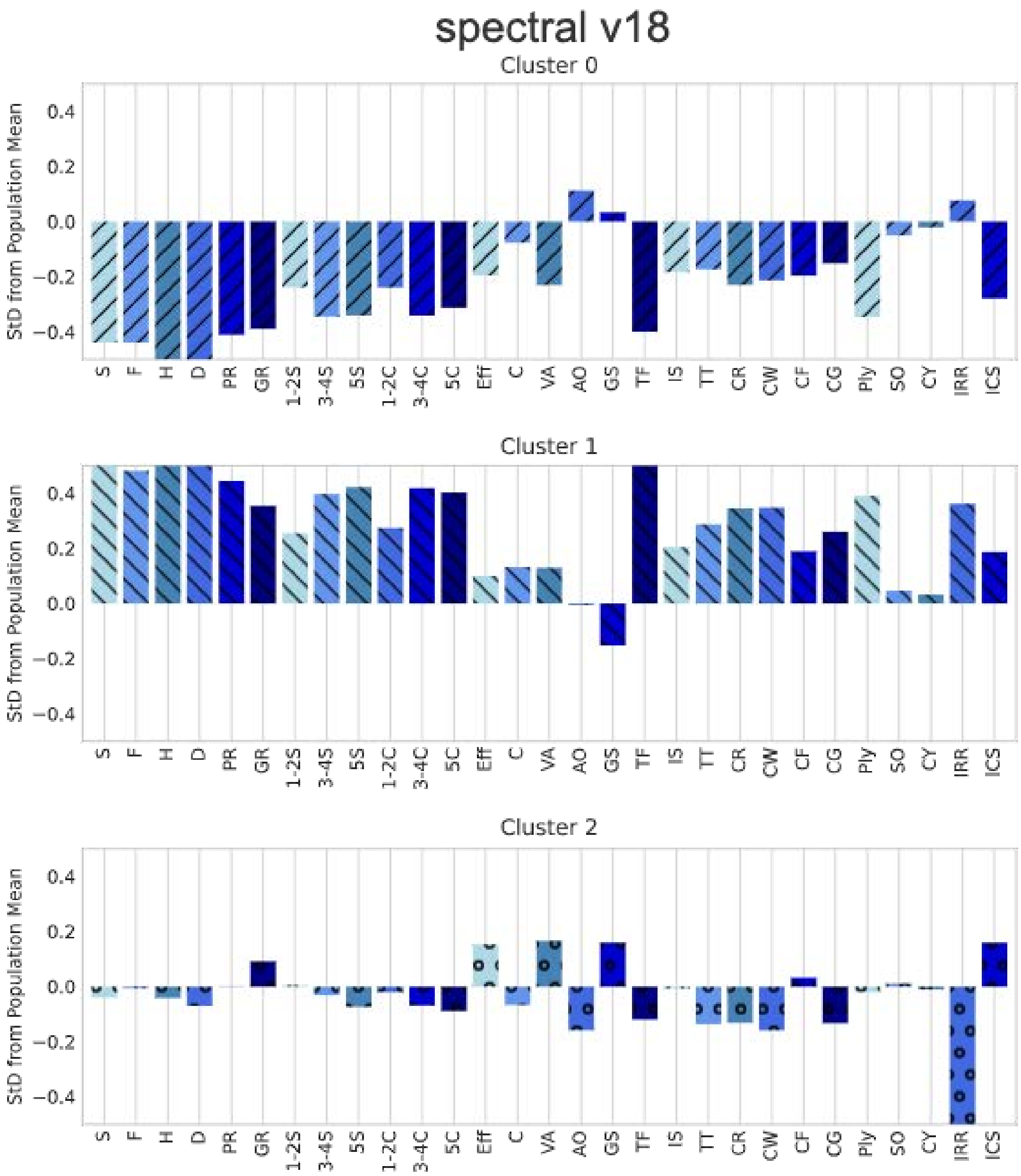


FIGURE 8-11: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2018 DATA SET – SPECTRAL V18

differences between the cluster sets were in Cluster 2, where Cluster 2 from genetic v74 had a higher-than-average perception of the likelihood of a cyclone occurring and a slightly stronger than average emotional response to the prospect of a cyclone occurring, the Cluster 2 from spectral v18 had a lower-than-average perception of the probability of a cyclone occurring and a slightly weaker than average emotional response to the prospect of a cyclone occurring. Although these differences were not huge, as both sets of perceptions were close to the population mean, the differences were significant enough that they altered the interpretation of a persona created based on them. For example, a persona based on Cluster 2 from spectral v18 may be described as not perceiving cyclones as a high risk, but as they think that cyclone shutters are effective and visually appealing, they would consider installing cyclone

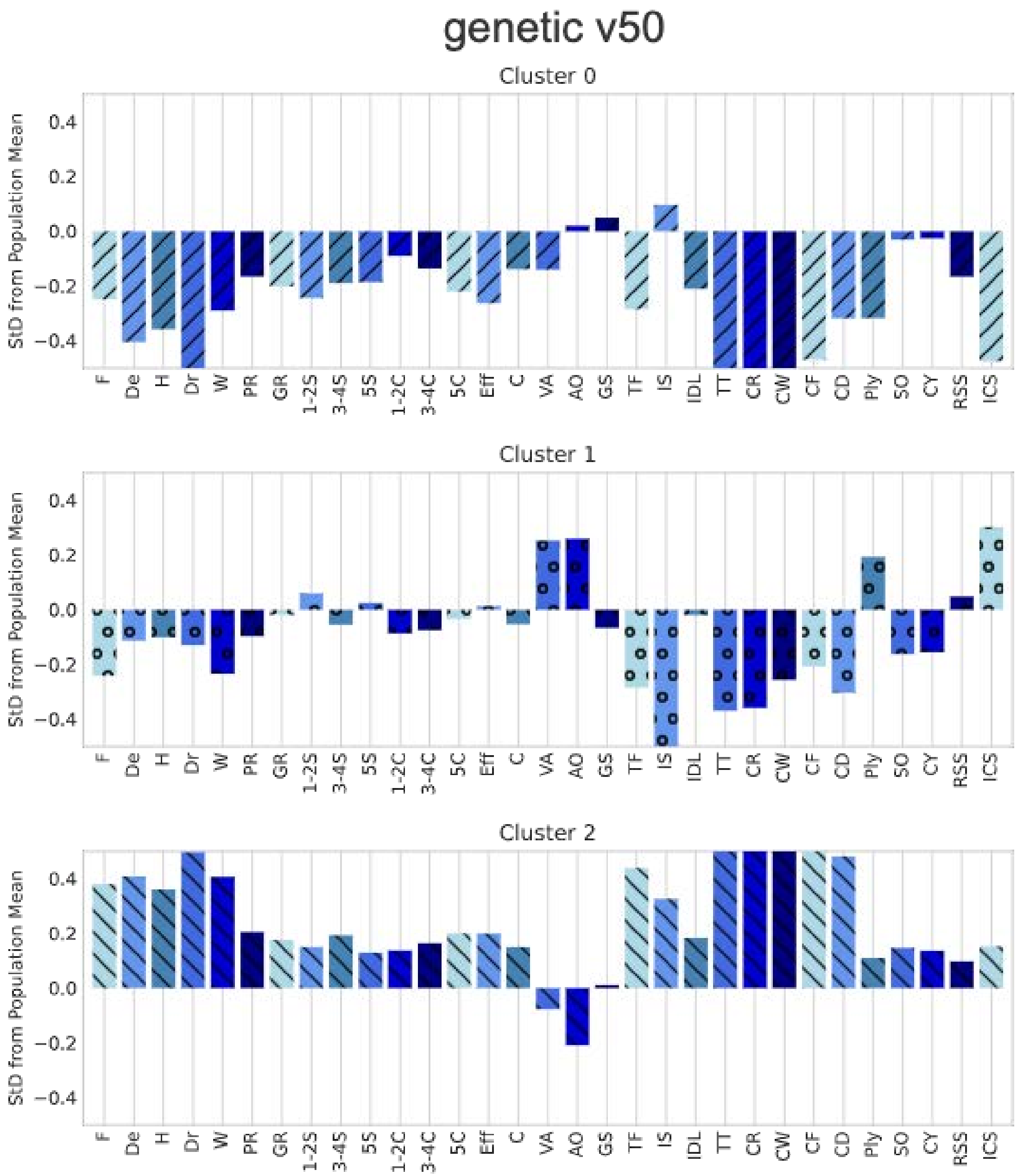


FIGURE 8-12: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2021 DATA SET – GENETIC V50

shutters. While a persona based on Cluster 2 from genetic v74 would mention that they think that a cyclone occurring is likely and could impact them personally.

Based on the graphs developed by HyPersona the cluster sets developed by genetic v50, Figure 8-12, and genetic v74, Figure 8-13, on the 2021 Data Set were functionally identical. The minor differences between the cluster sets were similar in severity to the differences seen between the cluster sets developed by SOM v0 and k-means v0. The most prominent differences between the cluster sets were the likelihood to have sought out information, IS, in Cluster 0 and the likelihood to remove shade sails,

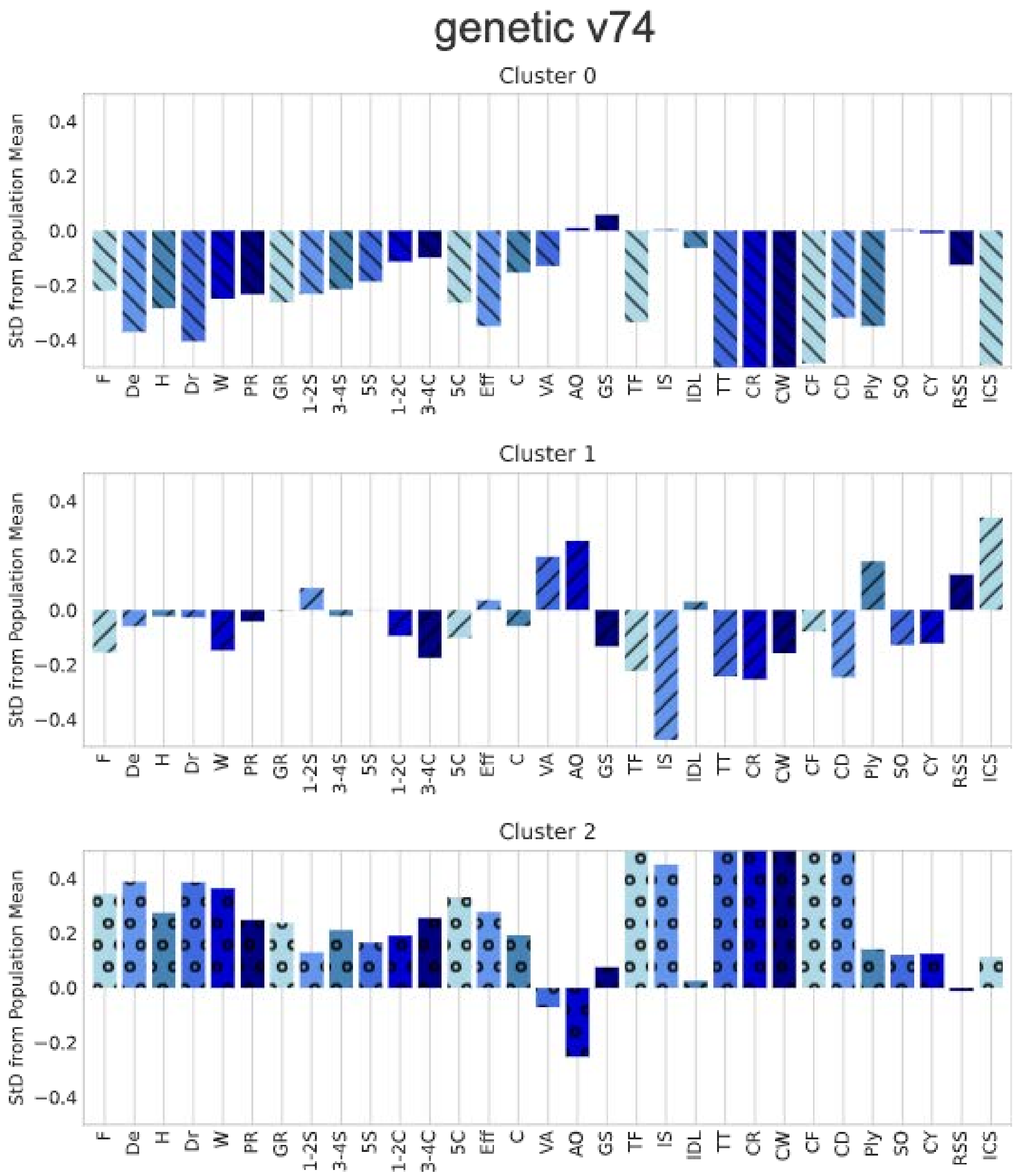


FIGURE 8-13: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2021 DATA SET – GENETIC V74

RSS, in Cluster 2. However, neither of these differences nor any of the other differences are statistically significant or would significantly impact how the clusters could be interpreted.

The differences between the cluster sets developed by genetic v50 and genetic v74 compared to the cluster set developed by spectral v18 were far more significant. Primarily, Cluster 1 was very different. The likelihood to perform protective behaviours and perceptions of cyclone shutters were very similar across the second clusters. However, the perceptions and attitudes around cyclones were quite different. The Cluster 1 developed by spectral v18 had a much stronger emotional response to the prospect to a cyclone occurring, particularly with feelings of helplessness. The Cluster 1 developed by spectral v18

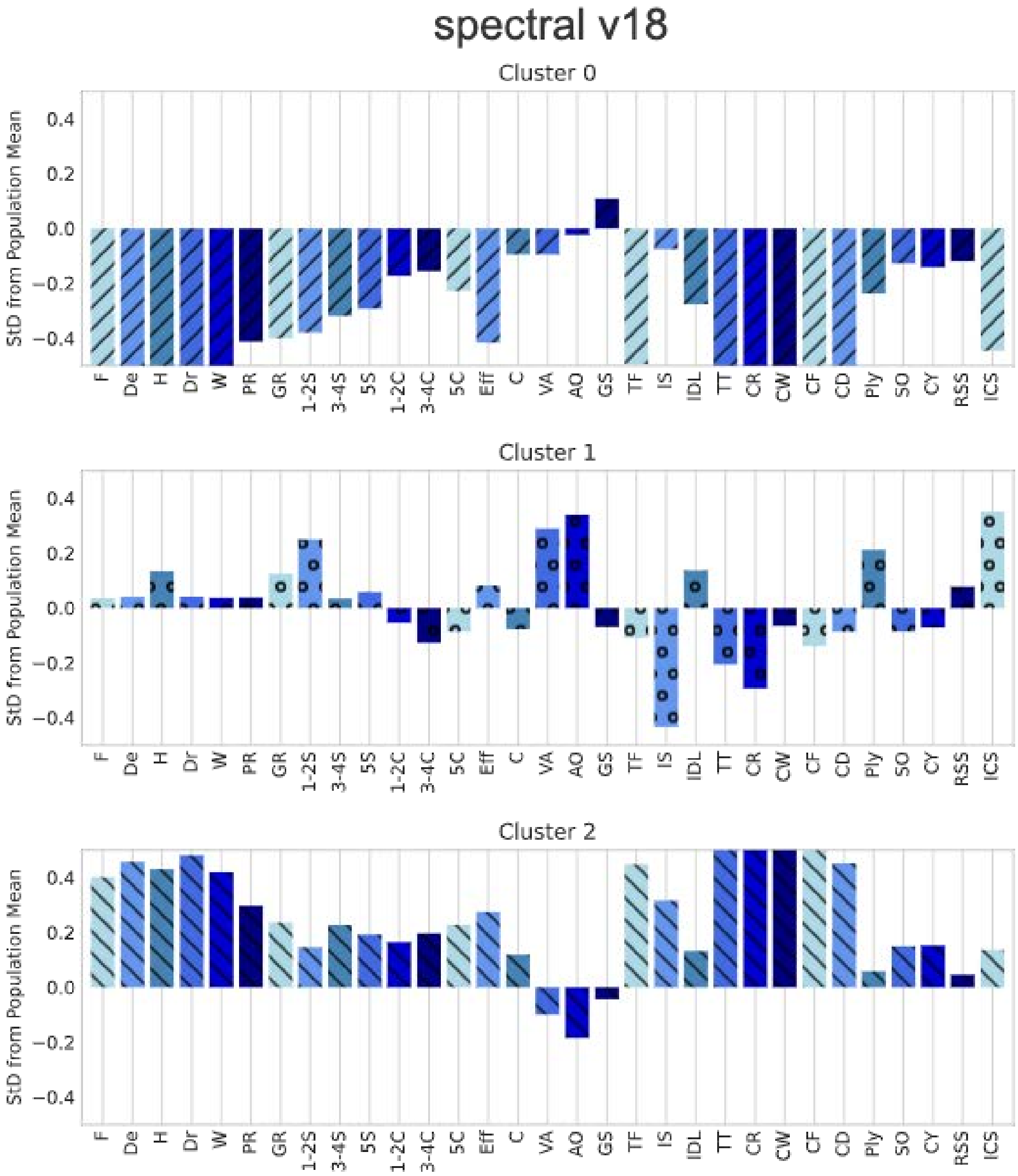


FIGURE 8-14: GRAPHS FOR THE TOP AFS PERFORMANCE SUBSET ON THE 2021 DATA SET – SPECTRAL V18

also had a much higher perception of the general risk proposed by a cyclone, and particularly in the severity of a category one or two cyclone. These differences greatly impacted the interpretation of the cluster set and, thus, impact any persona set developed based on the cluster set. The similarity of the cluster sets developed by the Top AFS Performance Subset further supported the acceptance of $H1_A$, as defined during the analysis of the Good Overall Performance Subset.

The primary purpose of the Top AFS Performance Subset was to determine the strength of AFS for indicating cluster set quality. The cluster sets developed by the Top AFS Performance Subset all contained distinct clusters that would develop significantly different personas. As defined by AFS, the cluster sets developed by the Top AFS Performance Subset had the most features that were significantly different to each other, or the population mean. However, to confidently say a set of personas developed by the Top AFS Performance Subset were objectively more distinct than a set of personas developed by the Good Overall Performance Subset would be difficult. Especially when the cluster sets developed on the 2021 Data Set were considered, as there were significant similarities between the cluster sets.

One bias of AFS was that the AFS value rewarded algorithms for splitting clusters based on an element that effected multiple features. For example, home ownership was a feature of the 2021 Data Set, and a subset of the questions within the survey relied on whether the individual was a homeowner.

That is, questions such as “Did you build your home?” or “Have you installed cyclone shutters” were all answered in the same manner by non-homeowners, which meant a set of clusters that were based on a significant difference in home ownership had a higher AFS score than a cluster set based on a different feature, despite the impact on the cluster set’s interpretation being equally affected.

When the cluster sets developed by genetic v50 on the 2021 Data set were compared to the cluster set developed by k-means v0, all the internal metrics differed quite significantly with the cluster set developed by genetic v50 performing the worst in all internal metrics other than AFS. One key difference between the genetic v50 and k-means v0 cluster sets was that the homeownership differed more greatly between the genetic v50 cluster set. On average the homeownership of the clusters developed by k-means v0 differed from the population mean by 0.1 standard deviations, while the clusters developed by genetic v50 differed on average by 0.6 standard deviations. As a result, many of the features that were found to be differ significantly between the genetic v50 clusters that did not significantly differ between the k-means v0 clusters were features where home ownership was important.

Developing clusters based on features such as home ownership is not an invalid approach, as such features have a significant impact on the data and provides important context to an individual’s behaviour. Instead, the fact that AFS will reward creating the clusters based on particular types of features should be considered when using AFS, in the same manner that most internal metric’s preference for convex clusters must be considered. Despite this consideration, AFS is useful for indicating how statistically significant a set of clusters are, and thus the quality of the cluster set for persona development. AFS should not be used as a standalone metric, however, still provides useful, additional information to the other internal metrics.

8.3.3 Comparison of the Poor Overall Performance Algorithm-Parameter Combination Subset

The final subset is the Poor Overall Performance Subset, which contained three algorithm-parameter combinations that achieved a poor overall performance metric across both data sets. The three algorithm-parameter combinations selected were:

1. **kernel_kmeans_v103:** kernel k-means using the polynomial kernel function with gamma set to 4, coef0 set to 4, and degree set to 4.
2. **genetic_v48:** Genetic clustering with 150 chromosomes, 300 populations, 2 genes mutated at each step, and a selection coefficient of 0.008.
3. **abc_long_v4:** ABC based clustering with 75 bees, a discard limit of 30, and 500 maximum iterations.

The primary purpose of analysing the Poor Overall Performance Subset was to confirm whether the internal metrics, and more specifically the overall performance metric, indicate the quality of a cluster set for persona development. The hypothesis and null hypothesis being addressed are:

> **$H2_0$:** The overall performance metric has no relationship with the quality of a cluster set for persona development.
> **$H2_A$:** As the overall performance metric increases, the quality of the cluster set for persona development will also increase.

TABLE 8-6: POOR OVERALL PERFORMANCE ALGORITHM-PARAMETER COMBINATION SUBSET INTERNAL METRICS

Data Set	Algorithm	SC		CHI		DBI		AFS		Overall
		Value	Rank	Value	Rank	Value	Rank	Value	Rank	
2018 Data Set	Kernel k-means v103	0.0304	0.28	38.27	0.74	3.78	0.54	50.83	0.52	2.08
	genetic v48	0.0376	0.36	32.51	0.58	4.63	0.16	60.00	0.81	1.91
	ABC long v4	0.0239	0.21	22.85	0.30	4.30	0.31	42.67	0.25	1.06
2021 Data Set	Kernel k-means v103	0.0283	0.30	12.89	0.70	4.05	0.48	78.00	0.64	2.11
	genetic v48	0.0393	0.49	10.35	0.39	4.51	0.25	89.00	0.81	1.94
	ABC long v4	0.0126	0.02	7.90	0.10	4.17	0.42	68.17	0.48	1.02

If $H2_A$ were accepted, all the algorithm-parameter combinations from the Good Overall Performance Subset would be expected to be of a better quality for persona development than the algorithm-parameter combinations belonging to the Poor Overall Performance Subset. As such, the internal

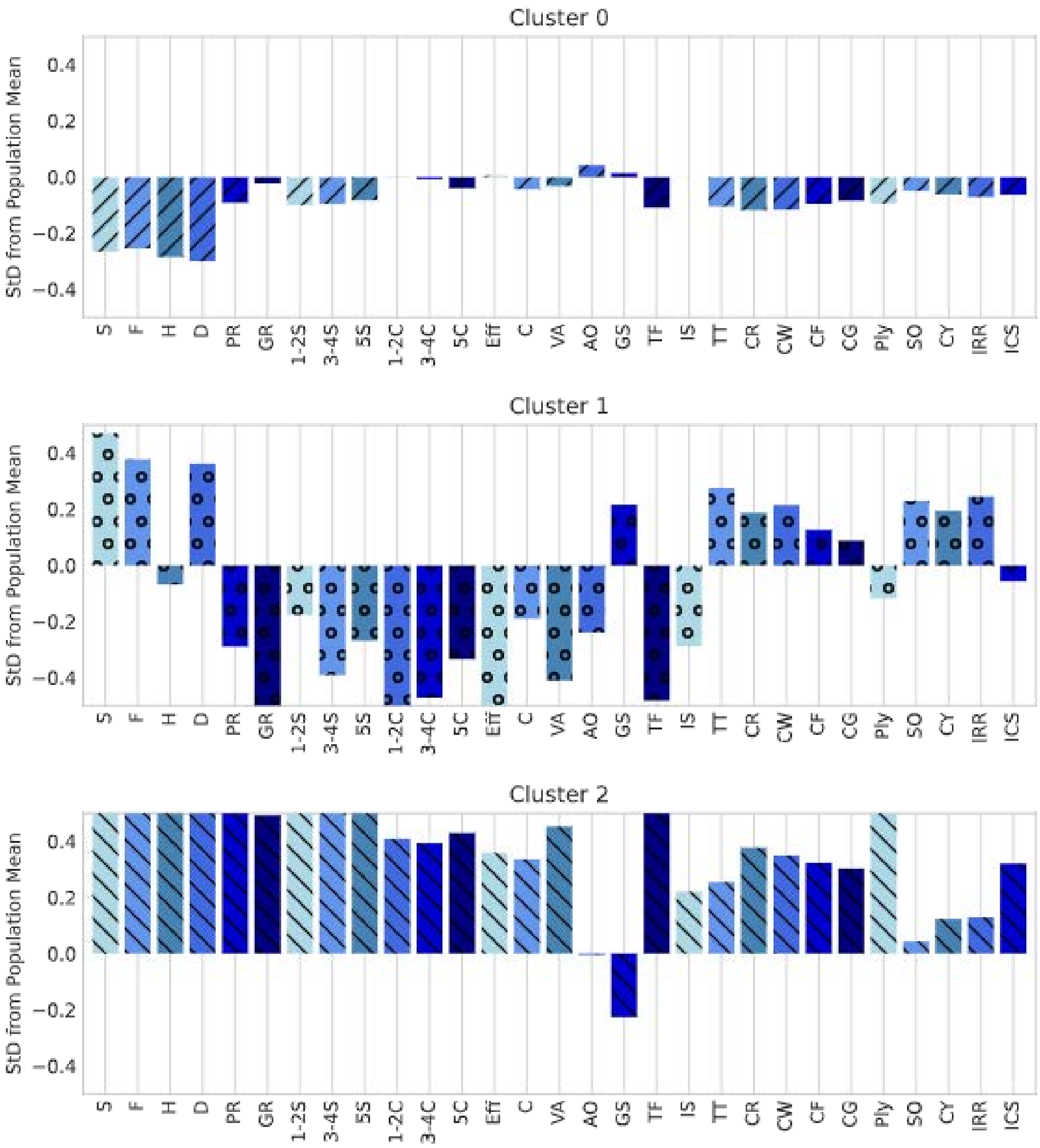


FIGURE 8-15: CLUSTERS DEVELOPED BY ABC LONG V4 ON THE 2018 DATA SET

metrics of the Poor Overall Performance Subset, given in Table 8-6, were not significantly similar to one another and did not have similar overall performance metrics. Instead, the selected algorithm-parameter combinations represented a range of overall performance metrics. The highest scoring algorithm-parameter combination, in terms of overall performance, was kernel k-means v103 which had an overall performance metric close to 2.10 and the lowest scoring algorithm-parameter combination was ABC long v4, with an overall performance metric around 1.05.

The graphs for the poorest performer according to the internal metrics, ABC long v4, for the 2018 Data Sets is given in Figure 8-15, and for the 2021 Data set in Figure 8-16. There was little to no resemblance between the two cluster sets, with the primary similarity being that they each contained a cluster that sat extremely close to the mean, Cluster 0.

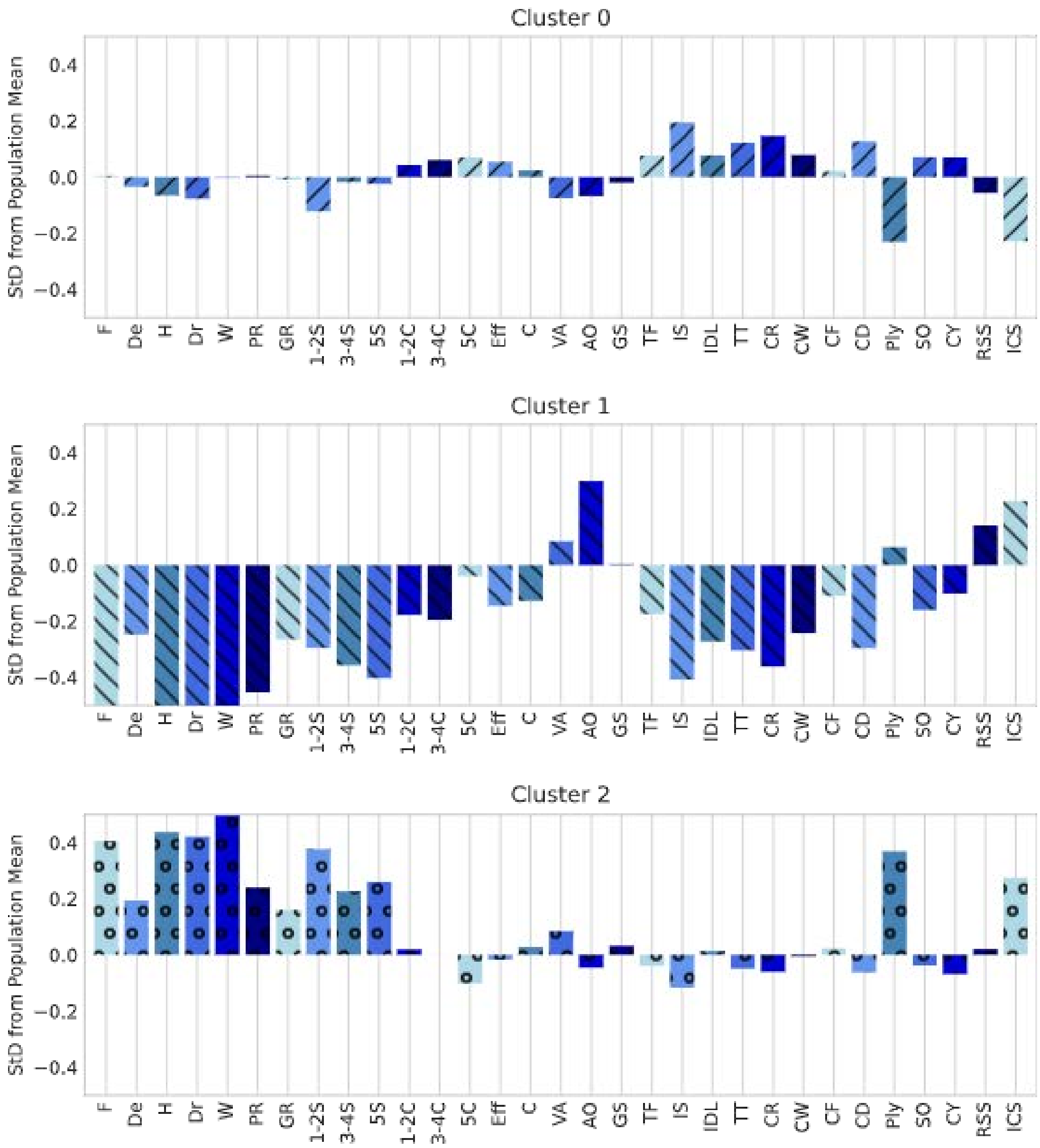


FIGURE 8-16: CLUSTERS DEVELOPED BY ABC LONG V4 ON THE 2021 DATA SET

Cluster 0 was a result of ABC long v4 developing relatively imbalanced clusters. Cluster 0 from the 2018 Data Set contained 371 data points (71.5%), and Cluster 0 from the 2021 Data Set contained 112 data points (53%). The imbalance was captured in the poor AFS score, as there were few features in Cluster 0 that significantly differed to the population mean, and the small clusters were less likely to be statistically significant.

As a result of the imbalance, Cluster 0 did not offer any particular insight into the audience segment the cluster represented, other than being close to the population average. Cluster 1 and Cluster 2 were distinct from Cluster 0, but the differences were not statistically significant and, especially in the case of Cluster 2 on the 2021 Data Set, were unlikely to be result in meaningful or nuanced personas. The

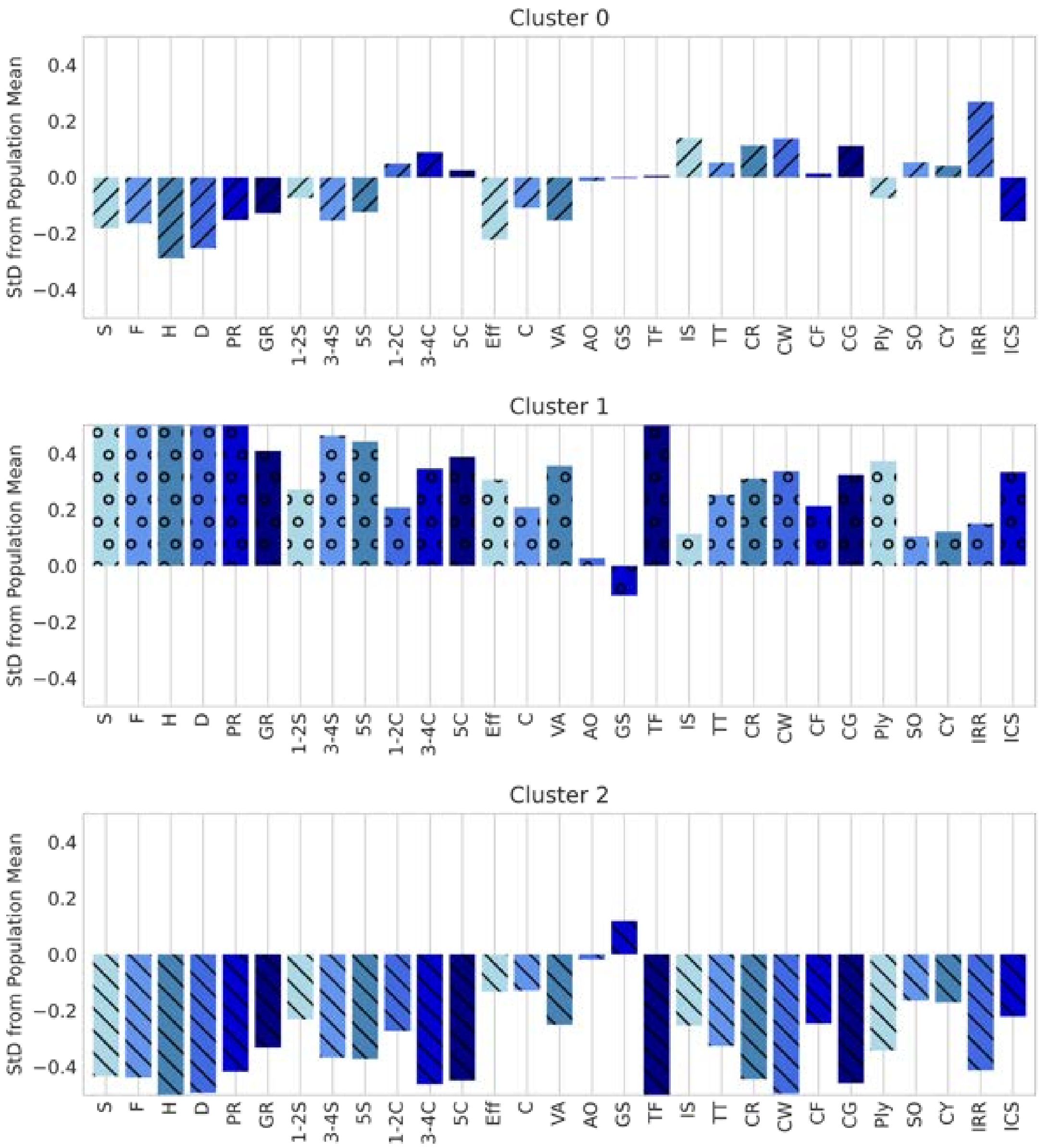


FIGURE 8-17: CLUSTERS DEVELOPED BY GENETIC V48 ON THE 2018 DATA

fact that the cluster set developed by ABC long v4 was a much weaker choice for persona development supported the alternate hypothesis, $H2_A$.

In comparison, the cluster sets developed by genetic v48 and kernel k-means v103 had much higher overall performance metrics, and based on the alternate hypothesis, $H2_A$, were expected to develop cluster sets that would result in better quality personas. Figure 8-17 and Figure 8-18 give the graphs created for the cluster sets developed by genetic v48 and kernel k-means v103 on the 2018 Data Set. Figure 8-19 and Figure 8-20 give the graphs created for the cluster sets developed by genetic v48 and kernel k-means v103 on the 2021 Data Set.

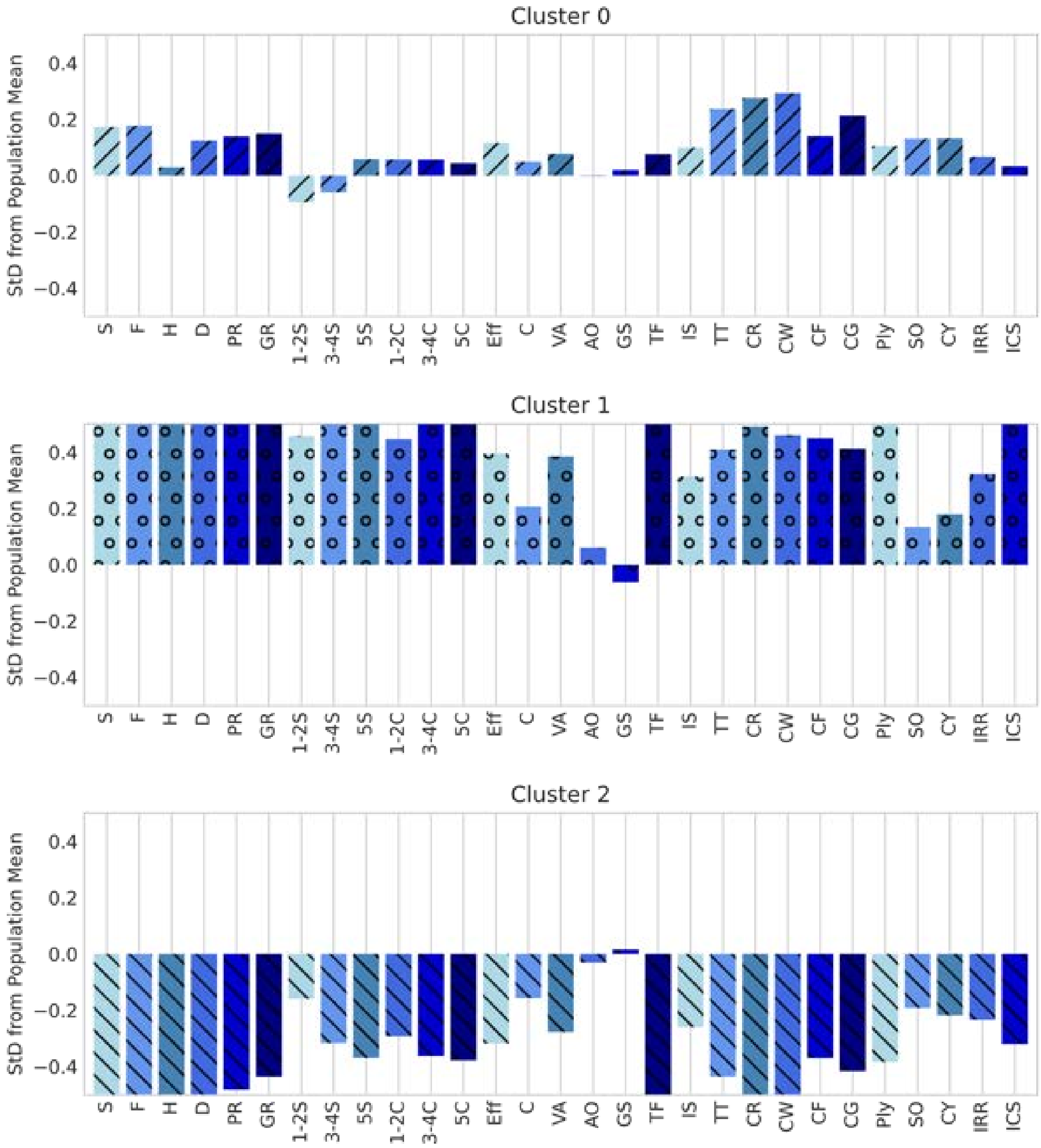


FIGURE 8-18: CLUSTERS DEVELOPED BY KERNEL K-MEANS V103 ON THE 2018 DATA

The clusters developed by genetic v48 and kernel k-means v103 were more balanced in size compared to ABC long v4, though each cluster set still tended to have one cluster that sat closer to the mean. Due to the difference in balance and having more significant features, the cluster sets developed by genetic v48 and kernel k-means v103 would be easier to use to develop a set of personas than ABC long v4.

Interestingly, the both cluster sets developed by genetic v48 and kernel k-means v103 had higher AFS scores than k-means v0, the k-means++ ensemble, and SOM v0 from the Good Overall Performance Subset. Despite this, the cluster sets presented in the Good Overall Performance Subset, were better for developing unique personas. The primary reason the Good Overall Performance Subset algorithm-parameter combinations were better for persona development was that there was a more overlap

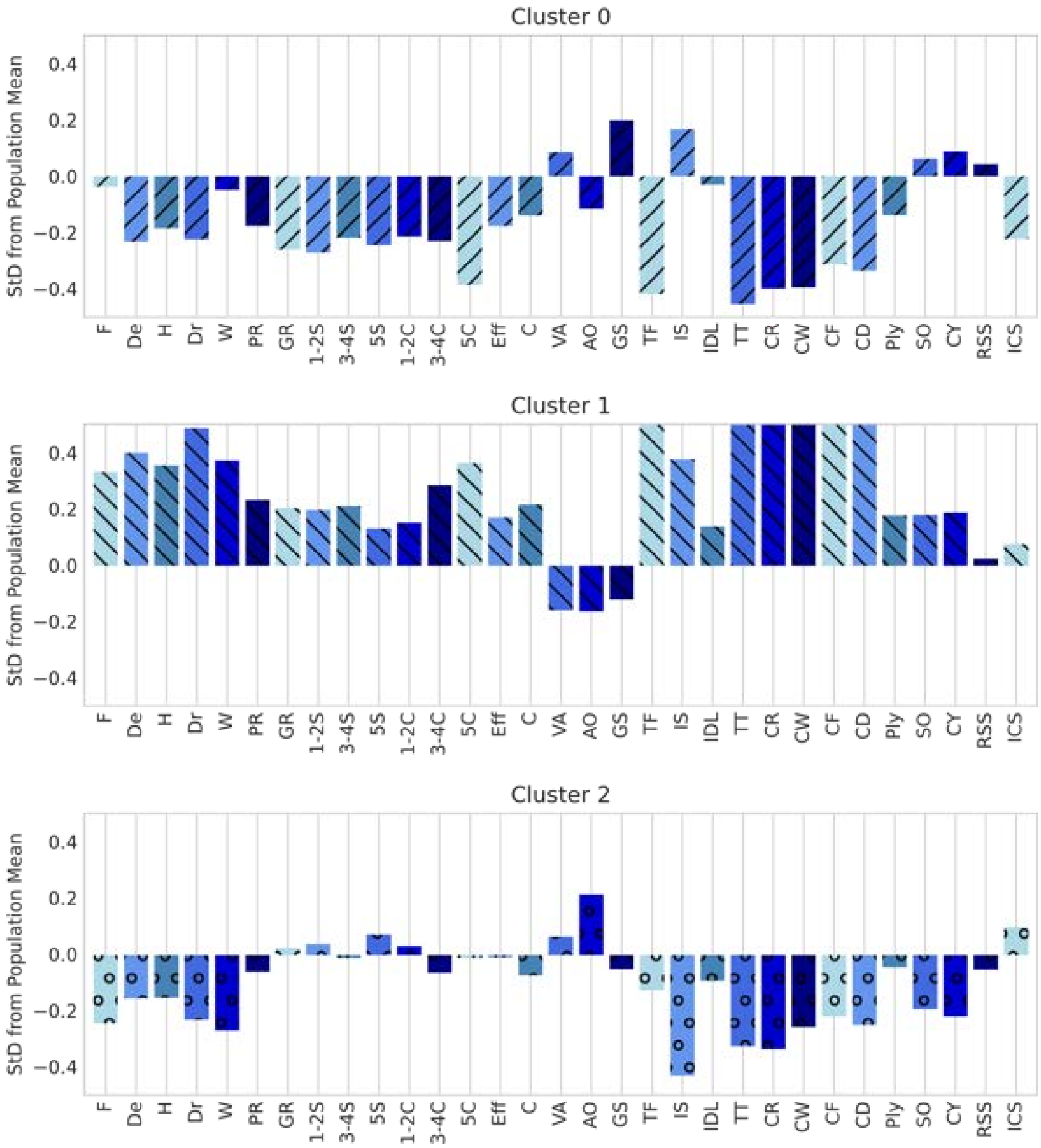


FIGURE 8-19: CLUSTERS DEVELOPED BY GENETIC V48 ON THE 2021 DATA SET

between the clusters presented by genetic v48 and kernel k-means v103, which would make developing distinct personas more difficult.

The overlap between clusters was particularly apparent between Cluster 0 and Cluster 1 developed by kernel k-means v103. Both clusters had positive risk perceptions and perceptions of cyclone shutters, and a higher-than average likelihood to perform preparatory behaviours. The primary difference between the clusters, based on the graphs, was in the scale to which features differed from the mean. This indicated that the features that significantly differed between the clusters, which were causing the higher AFS, were features that were not considered "key features". Whereas the few key features of the cluster sets developed by k-means v0, the k-means++ ensemble, and SOM v0 were more likely to be key features.

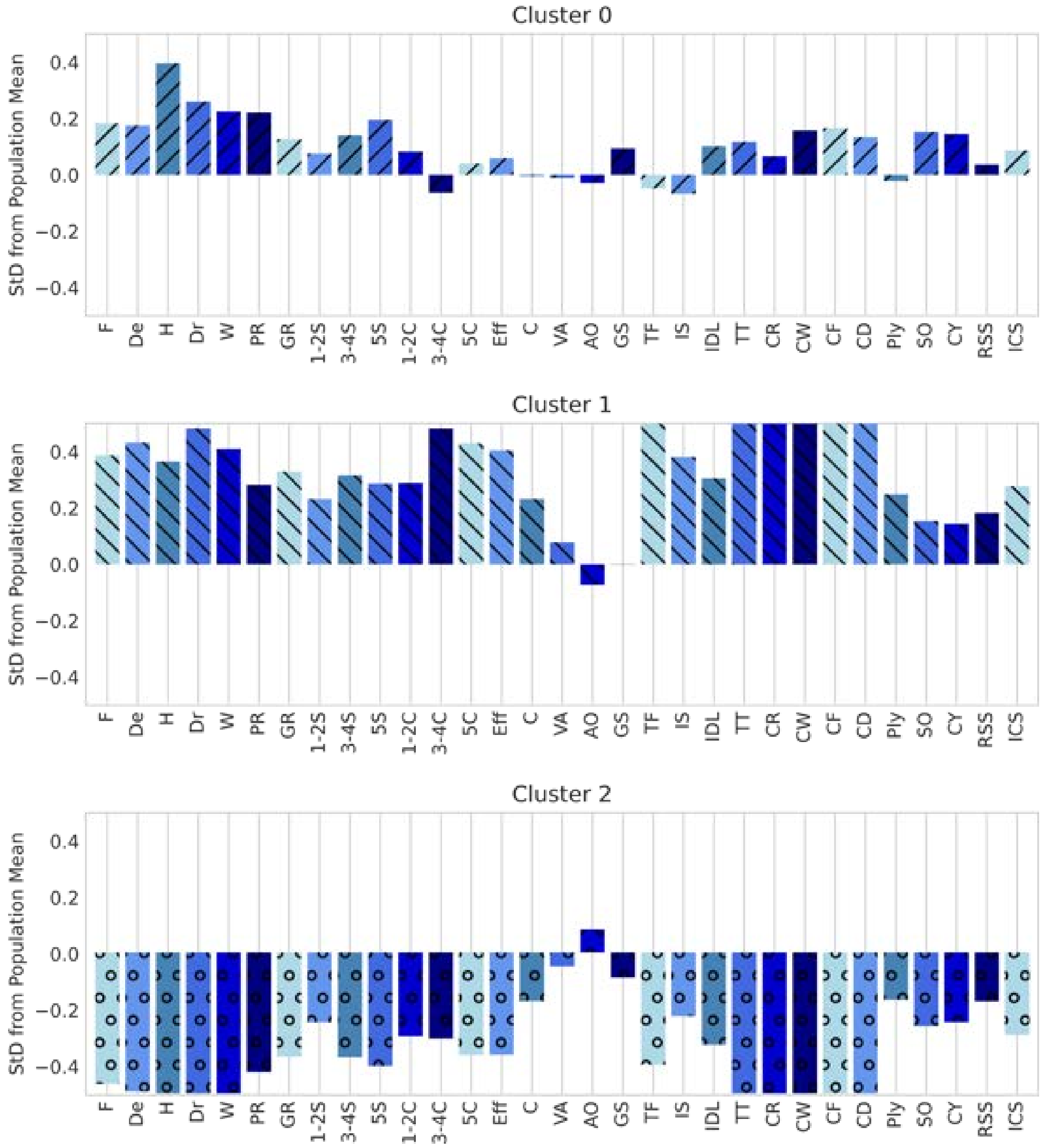


FIGURE 8-20: CLUSTERS DEVELOPED BY KERNEL K-MEANS V103 ON THE 2021 DATA SET

The cluster sets developed by the Poor Overall Performance Subset were all poorer choices for persona development than those developed by the Good Overall Performance Subset. Furthermore, the cluster sets that had higher overall performance metrics were found to be of better quality of those with lower overall performance metrics. Although, between those with similar overall performance metrics there was no significant difference in suitability. Based on these findings the null hypothesis, $H2_0$, can be rejected and the alternate hypothesis, $H2_A$, accepted.

8.3.4 Summary of Algorithm-Parameter Combination Subset Findings

Three representative subsets of the more than 1,000 valid algorithm-parameter combinations were analysed in depth to address a selection of premises. Based on these analyses two hypotheses were accepted, and their null hypotheses rejected:

> **$H1_A$**: The internal metrics of a pair of cluster sets indicate how similar the cluster sets are.
>
> **$H2_A$:** As the overall performance metric increases, the quality of the cluster set for persona development will increase.

The acceptance or rejection of these hypotheses was integral to determining how the overall best performing algorithm-parameter combination can be determined. As the overall performance metric was found to indicate cluster quality, the potential algorithm-parameter combinations were able to be narrowed down to those with the top overall performance metrics. The potential algorithm-parameter combinations could then be further narrowed down to rule out cluster sets that are functionally identical to another cluster set; first through the Euclidian distance between the internal metric ranks and second through using the graphs developed. Euclidian distances less than 0.1 were determined to indicate functionally identicality. The subset findings also began to address SRQ2, as in every subset there were cluster sets that had similar overall performance metrics but were significantly different and would develop significantly different personas.

8.4 Evaluation of Clustering Approach Consistency

To address SRQ2 the performances of each clustering algorithm and each clustering approach must be evaluated to determine consistency. The consistency of each algorithm was evaluated in terms of the algorithm consistency, approach consistency, and across data set consistency. The consistency was then classified into one of four categories:

1. **High:** The consistency high if the cluster sets were all dropped for the same reason, developed identical or functionally identical clusters, or developed cluster sets within a similar range of internal metrics.
2. **Moderate:** Moderate consistency is indicative of cluster sets that are dropped for a variety of reasons, an algorithm performance that is consistent when a particular parameter is controlled for, cluster sets that are consistent within a wider, but still similar, range of internal metrics, or cluster sets that are consistent across a particular internal metric.
3. **Low:** Low consistency indicates there is very little consistency, such as none of the cluster sets being dropped.
4. **None:** A consistency of None indicates the performances were not at all consistent.

While determining the consistency of each algorithm, the overall performance of the algorithm is determined and was also classified. The performance was classified into four categories:

1. **Dropped:** All the algorithm-parameter combinations were dropped for one reason or another.
2. **Poor:** The algorithm-parameter combinations achieved poor overall performance, often accompanied by a considerable percentage of the algorithm-parameter combinations being dropped.
3. **Mixed:** The performance of the algorithm-parameter combinations was mixed, with some performing quite well, and other performing more mediocrely. May also have a small number of algorithm-parameter combinations that were dropped
4. **Good:** The algorithm-parameter combinations performed amongst the best consistently.

8.4.1 Consistency of Hierarchical Approach Clustering Algorithms

Overall, the hierarchical approach gave a consistent performance. A summary of the consistencies and performance is given in Table 8-7. The performance of AHC was moderately consistent, AHC tended to develop imbalanced cluster sets that were all dropped due to cluster size, except for AHC with Ward's linkage on the 2021 Data Set, which performed relatively well. The algorithm-parameter combinations based on BIRCH all developed identical cluster sets, all of which were identical to the results of AHC using Ward's linkage. All the cluster sets developed by CURE were extremely imbalanced, consistently creating two clusters containing a single data point and a third cluster containing the entirety of the remaining data set. All the algorithm-parameter combinations based on ROCK were also dropped, however some were dropped due to cluster size and others were dropped for not meeting the internal metric thresholds.

TABLE 8-7: CONSISTENCY OF HIERARCHICAL APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Hierarchical	AHC	Moderate	High	High	Poor
	BIRCH	High		High	Poor
	CURE	High		High	Dropped
	ROCK	Moderate		High	Dropped

As the general performance of the algorithm-parameter combinations and the reason behind the dropped algorithm-parameter combinations were consistent between algorithms, the overall approach consistency was determined to be high. The only algorithm-parameter combination to perform differently on each data set being AHC with Ward's linkage, with all the other algorithm-parameter combinations performing the same on each data set. Thus, overall, the consistency of the hierarchical approach across data sets was high.

8.4.2 Consistency of Graph Theory Approach Clustering Algorithms

The two graph-theory approach clustering algorithms performed very differently. Table 8-8 gives an overview of the consistency and performance of the graph theoretic based clustering algorithms. MST tended to make very imbalanced clusters that got dropped due to cluster size, while the performance of the algorithm-parameter combinations based on Spectral Graph Theory depended heavily on the parameters, with some algorithm-parameter combinations being dropped and others being amongst the best performers. Between MST and the Spectral graph theory, there was no relationship or consistency between the performances.

TABLE 8-8: CONSISTENCY OF GRAPH THEORETIC APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Graph Theory	MST	Moderate	None	Moderate	Dropped
	Spectral	Moderate		High	Mixed

There was a total of 1056 algorithm-parameter combinations for MST, of which only 16 were not dropped from the 2021 Data Set for cluster size or the SC score not reaching the threshold. With more of the algorithm-parameter combinations being dropped for cluster size on the 2018 Data Set than the 2021 Data Set. Of the 16 algorithm-parameter combinations, most were identical to one another, with only two unique cluster sets developed. Both cluster sets had very poor overall internal metrics, with respective overall performance metrics of 0.78 and 0.71, and were found to be functionally identical.

The performance of the algorithm-parameter combinations using the spectral graph theory algorithm varied primarily based on the kernel used. Across both data sets the algorithm-parameter combinations that used the Sigmoid and Chi-Squared kernels were dropped. On the 2018 Data Set the algorithm-parameter combinations with the Nearest-Neighbour kernel was also dropped. For the algorithm-parameters that used a Laplacian or RBF kernel, only those with the gamma set to ‘None’ were not dropped. None of the algorithm-parameter combinations that used the cosine kernel were dropped. Only five of the algorithm-parameter combinations using the polynomial kernel were dropped, all of which were dropped from the new data set. The reasons why each of the algorithm-parameter combinations were dropped varied significantly.

Of the algorithm-parameter combinations not dropped, the internal metrics varied greatly. On the 2018 Data Set the overall performance metric ranged between 1.09 and 3.19, and on the 2021 Data Set the overall performance metric ranged between 0.61 and 3.1. Many of the cluster sets developed by the algorithm-parameter combinations that were not dropped developed identical clusters. On the 2021 Data Set, 287 algorithm-parameter combinations using the spectral clustering algorithm were not dropped, however there were only 59 unique cluster sets. When the parameters of the spectral graph theory algorithm, particularly the kernel used, were controlled for the algorithm performance was moderately consistent.

8.4.3 Consistency of Simple Partitioning Approach Clustering Algorithms

The consistency of the simple partitioning approach, given in, was mixed. Both k-means and k-means++, which were grouped together under the same algorithm (k-means) performed very well on both data sets, ranking in the top 10 overall on the 2021 Data Set and in the top 20 overall in the 2018 Data Set. There were no parameter variations of the k-medians algorithm, so algorithm consistency cannot be determined, however k-medians was dropped from the 2021 Data Set while k-medians performed moderately well on the 2018 Data Set. The algorithm-parameter combinations based on fuzzy c-means all developed imbalanced cluster sets that were dropped due to cluster size. An overview of the consistency and performance is given in Table 8-9.

TABLE 8-9: CONSISTENCY OF SIMPLE PARTITIONING APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Simple Partitioning	k-means	High	None	High	Good
	k-medians	N/A		None	Mixed
	Fuzzy c-means	High		High	Dropped

8.4.4 Consistency of Density Approach Clustering Algorithms

The performance of the density approach clustering algorithms was very consistent; however, the performance was very poor. An overview is given in Table 8-10. The performance of the density approach algorithms was expected to be quite poor due to the difference between the basis of the density approach and the requirements of persona development. Although, the reason for poor performance was not related to that difference.

TABLE 8-10: CONSISTENCY OF DENSITY APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Density	DBSCAN	High	High	High	Dropped
	OPTICS	Moderate		High	Dropped

Most algorithm-parameter combinations based on DBSCAN did not create enough cluster either clustering all the data points together or classifying all the data points as noise. Of the 240 algorithm-parameter combinations that used DBSCAN, 210 were dropped for only creating one cluster when applied to the 2018 Data Set and 179 were dropped for only creating one cluster when applied to the 2021 Data Set. The remainder of the algorithm-parameter combinations were dropped for creating too many or not enough clusters or creating imbalanced clusters. The majority of the algorithm-parameter combinations that used OPTICS were also dropped due to not creating enough clusters. The remaining algorithm-parameter combinations were dropped due to developing too many clusters or imbalanced clusters. The performances across both algorithms were consistent across data sets, with a similar percentage of algorithm-parameter combinations being dropped for each reason.

8.4.5 Consistency of Kernel Approach Clustering Algorithms

The two kernel approach clustering algorithms performed quite differently, which was expected as the two algorithms selected represent opposing approaches to clustering with kernels. Table 8-11 gives an overview of the consistency and performance of Kernel k-means and SVC.

TABLE 8-11: CONSISTENCY OF KERNEL APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Kernel	Kernel k-means	Moderate	None	Moderate	Mixed
	SVC	Low		None	Poor

The performance of kernel k-means was quite like the performance of spectral clustering, as the performance was dependent on the kernel used. Across both data sets the algorithm-parameter combinations based on kernel k-means that used the Polynomial or Cosine kernels were not dropped. Of the algorithm-parameter combinations that used the RBF or Laplacian kernels, only those with the gamma set to ‘None’ were not dropped. The algorithm-parameter combinations that used the Sigmoid kernel or Chi-Squared kernel were all dropped. On the 2018 Data Set all the algorithm-parameter combinations were dropped due to not reaching the AFS threshold, while on the 2021 Data Set the AFS threshold, SC threshold, and cluster size were all reasons that algorithm-parameter combinations were dropped. There was also considerable variance in the overall performance metrics of the kernel k-means based algorithm-parameter combinations that were not dropped.

Although so many of the algorithm-parameter combinations performed poorly, there was also a subset of algorithm-parameter combinations that were amongst the best performers. On the 2021 Data Set, of the algorithm-parameter combinations that were not dropped, the overall performance metrics ranged between 1.60 and 3.38. Due to the variance the performance was classified as Mixed.

More than half of the algorithm-parameter combinations based on SVC errored and did not produce a result, as the algorithm either did not converge or an error occurred. Only 11 of the algorithm-parameter combinations based on SVC were not dropped from each data set. There was minimal consistency in the reason the algorithm-parameter combinations were dropped, with most being dropped due to cluster count or size, and there was little overlap between which algorithm-parameter combinations were not dropped. The performance of the algorithm-parameter combinations that were not dropped was consistently poor.

8.4.6 Consistency of Metaheuristic Approach Clustering Algorithms

When considering the consistency of the metaheuristic approach clustering algorithms the 'long' parameter combinations were not considered separately from the other parameter combinations. The classified consistency and performance of each metaheuristic algorithm is given in in Table 8-12. Due to the metaheuristic algorithms relying on random initialisation to lead to global optima, the results of the metaheuristic algorithms were expected to be varied, unless the algorithm converged. The importance of random initialisation on the performance of the metaheuristic algorithms also had the effect of obfuscating the impact of the parameters on the performance. However, the 'long' algorithm-parameter combinations did tend to perform better which was likely due to the 'shorter' or 'smaller' algorithm-parameter combinations using fewer random initialisations and having fewer iterations, which minimises the opportunity to get close to the global optima.

TABLE 8-12: CONSISTENCY OF METAHEURISTIC APPROACH CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Metaheuristic	ABC	Moderate	None	Low	Mixed
	Genetic	High		High	Mixed
	SFLA	Low		High	Mixed

Of the metaheuristic approach algorithm-parameter combinations, those that were based on ABC were most likely to be dropped. All the dropped algorithm-parameter combinations based on ABC were dropped due to imbalanced cluster sizes. The algorithm-parameter combinations based on ABC performed better on the 2018 Data Set, as less of the algorithm-parameter combinations were dropped and the overall performance metrics were higher. On the 2018 Data Set the overall performance metrics of the algorithm-parameter combinations based on ABC ranged between 0.78 and 3.09, while on the

2021 Data Set the overall performance metrics of the algorithm-parameter combinations based on ABC ranged between 1.02 and 2.41.

The algorithm-parameter combinations that used the genetic algorithm consistently gave the best performances regarding the AFS metric. However, the performance of the other internal metrics was far more mixed, and generally poor. A small number of the algorithm-parameter combinations using the genetic algorithm were dropped on each data set, all due to not meeting the DBI threshold. Of the 81 total algorithm-parameter combinations based on the genetic algorithm, 30 were dropped when applied to the 2018 Data Set and 6 were dropped when applied to the 2021 Data Set. The performance of the algorithm-parameter combinations based on the genetic algorithm were very consistent across data sets.

There were consistently algorithm-parameter combinations based on SFLA were amongst the top overall performers. Of the top 10 algorithm-parameter combinations based on the overall performance metrics, seven were based on SFLA for the 2018 Data Set and six were based on the SFLA for the 2021 Data Set. However, as with all the metaheuristic algorithms, the performances of the algorithm-parameter combinations based on SFLA varied significantly. On the 2018 Data Set the overall performance metrics ranged from 1.45 to 3.40 and on the 2021 Data Set the overall performance metrics ranged from 0.90 to 3.46.

8.4.7 Consistency of Ensemble Approach Clustering Algorithms

The various clustering algorithm ensembles were not expected to perform consistently, as each ensemble is made up of different clustering algorithms, often including algorithms based on different approaches within one ensemble. However, the consensus function can be treated as a primary differentiator between algorithm-parameter combinations, with the algorithms and parameters chosen to make up the ensemble treated as the parameters. Based on this definition, the consistency and performance of the cluster ensembles are given in Table 8-13.

TABLE 8-13: CONSISTECY OF THE CLUSTER ENSEMBLES

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Ensemble	Basic Consensus	Low	Low	Moderate	Mixed
	NMF Consensus	Low		Moderate	Mixed

The algorithm-parameter combinations that used the basic consensus function were more likely to develop imbalanced clusters and get dropped, as five of the nine algorithm-parameter combinations that used the basic consensus function were dropped from each data set. While only two algorithm-parameter combinations using the NMF consensus function were dropped from the 2018 Data Set and none of the algorithm-parameter combinations using the NMF consensus function were dropped from the 2021 Data Set. The performance of the ensembles that were not dropped were mixed, as expected, some of which performed very well. However, the best performers were most commonly those based solely on k-means that produced cluster sets functionally identical to those developed by k-means without the ensemble.

8.4.8 Consistency of Individual Clustering Algorithms and One-Off Approaches

The remaining clustering algorithms either did not belong to an approach or were selected as the sole representative of their approach. As such, approach consistency is not applicable. Table 8-14 gives the summary of consistency and performance for the individual approaches and one-off algorithms.

TABLE 8-14: CONSISTECY OF INDIVIDUAL CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
N/A	AP	High	N/A	High	Dropped
	EMA	Moderate		None	Poor
	NMF	High		High	Poor
	SOM	High		High	Mixed

Across both data sets, the algorithm-parameter combinations based on AP were dropped as they developed far too many clusters. Alternately, the algorithm parameter combinations based on EMA developed only one or two clusters on the 2021 Data Set. While, on the 2018 Data Set, all the cluster sets except one performed very well for all the internal metrics except AFS. The one exception performed averagely across all the internal metrics. Due to the variance between giving a 'Dropped' performance on the 2018 Data Set and a 'Mixed' performance on the 2021 Data Set, EMA was assigned a performance classification of 'Poor'. All algorithm-parameter combinations based on SOM developed identical cluster sets on each data set. Like EMA, the cluster sets developed by SOM gave good performances for all the internal metrics except AFS.

The performance of the algorithm-parameter combinations based on NMF depended heavily on the solver used. All the algorithm-parameter combinations that used the MU solver were dropped due to cluster size, and those developed with the CD solver were also generally imbalanced. Table 8-15 gives the sizes of each of the clusters developed by algorithm-parameter combinations that used NMF with the CD Solver. The cluster set developed by NMF v2 on the 2021 Data Set was the only cluster set that had balanced cluster sizes. As such, the performance of NMF was very consistent but poor.

TABLE 8-15: NMF WITH THE CD SOLVER CLUSTER SIZES

	2018 Data Set			2021 Data Set		
	NMF v0	NMF v1	NMF v2	NMF v0	NMF v1	NMF v2
Cluster 1	392	207	227	164	113	82
Cluster 2	38	51	47	18	34	54
Cluster 3	89	261	245	29	64	75

8.4.9 Summary of Clustering Algorithm Consistency and Performance

Overall, there was considerable consistency within the performances of individual algorithms both between parameter combinations and across data sets. However, there was only minimal consistency between the performances of the clustering algorithms within an approach and the quality of the cluster sets developed varied greatly between the algorithms and approaches. Table 8-16 gives the collated tables of clustering algorithm consistency and a summary of each consistency and the performance.

Of the 22 algorithms, including the ensemble approach as two algorithms, 14 (63.64%) had a high consistency across data sets and a further four (18.18%) had a moderate consistency across data sets. Combined, more than 80% of the algorithms performed consistently across data sets. Of the 21 clustering algorithms that algorithm consistency was applicable to, nine (42.86%) had high consistency and eight (38.10%) had moderate consistency, resulting in just over 80% of the algorithms performing consistently. The algorithm consistency would not be as high if the impact of kernel choice for kernel k-means and spectral graph theory was not being controlled for.

Alternatively, there was very little consistency between the algorithms that belonged to the same approach to clustering. Four of the seven approaches (57.14%) had no internal consistency. Of those with a high consistency, the performance of the algorithms was extremely poor, with the reason for high consistency largely due to all or nearly all the algorithm-parameter combinations were dropped.

The performance of the clustering algorithms, according to the internal metrics, were generally mixed. The only algorithm that consistently performed well was k-means, however that does not mean k-means gave the best performance. SFLA, the ensembles based on k-means, and kernel k-means were all able to give performances as good as, if not better, than k-means, often developing functionally identical cluster sets. Nine of the clustering algorithms (40.91%) gave mixed performances, demonstrating the impact of parameter selection on the performance of the algorithm. While seven of the algorithms (31.82%) had all their algorithm-parameter combinations dropped. Combined with the five algorithms (22.73%) which gave a poor performance, more than half of the algorithms performed consistently poorly.

TABLE 8-16: OVERVIEW OF THE CONSISTENCY OF ALL CLUSTERING ALGORITHMS

Approach	Algorithm	Consistency			Performance
		Algorithm	Approach	Across Data Set	
Hierarchical	AHC	Moderate	High	High	Poor
	BIRCH	High		High	Poor
	CURE	High		High	Dropped
	ROCK	Moderate		High	Dropped
Graph Theory	MST	Moderate	None	Moderate	Dropped
	Spectral	Moderate		High	Mixed
Simple Partitioning	k-means	High	None	High	Good
	k-medians	N/A		None	Mixed
	Fuzzy c-means	High		High	Dropped
Density	DBSCAN	High	High	High	Dropped
	OPTICS	Moderate		High	Dropped
Kernel	Kernel k-means	Moderate	None	Moderate	Mixed
	SVC	Low		None	Poor
Metaheuristic	ABC	Moderate	None	Low	Mixed
	Genetic	High		High	Mixed
	SFLA	Low		High	Mixed
Ensemble	Basic Consensus	Low	Low	Moderate	Mixed
	NMF Consensus	Low		Moderate	Mixed
N/A	AP	High	N/A	High	Dropped
	EMA	Moderate		None	Poor
	NMF	High		High	Poor
	SOM	High		High	Mixed
Overall		**Moderate**	**None**	**High**	**Mixed**

8.5 Are the Performances of Clustering Algorithms and Approaches to Clustering for Persona Development Consistent?

The primary purpose of evaluating the consistency of the algorithm performance was to address SRQ2 and determine whether the results of this study could be inferred to have broader implications. Three types of consistency were evaluated: 1) algorithm consistency; 2) approach consistency; and 3) across data set consistency. The clustering algorithms were found to be consistent both in terms of algorithm consistency and across data set consistency, and not to have approach consistency. The approaches within an algorithm not being consistent means the results of the cluster algorithms cannot be inferred to be indicative of the performance of other algorithms belonging to the same approach.

The algorithm consistency was important as the consistency indicates the impact of the parameters on the algorithm performance, as the more consistent the algorithm performance is, despite the parameters, the smaller the impact of the parameters are. In algorithms such as BIRCH and SOM, the parameters were found to have no effect. While the kernel selected for kernel k-means, or the spectral graph theory algorithm had a significant impact. In algorithms where the consistency was low and the performance is mixed, evaluating a higher number of parameter combinations would be beneficial to determine the potential range of performance. While algorithms that had a high algorithm consistency and less variable performance likely do not need to have many parameter combinations assessed to determine performance.

The across data set consistency was potentially the most important consistency to determine, as the across data set consistency indicates whether the performance of the algorithms can be expected to be applied to similar data sets and use cases. As such, when approaching clustering for a similar use case or across a similar data set the clustering algorithms that performed poorly on both the 2018 Data Set and the 2021 Data Set can be ruled out with relative confidence. The algorithms with high across data set consistency and mixed or good performance would be recommended to focus on. Additionally, algorithms such as SVC and EMA should also be considered, despite their poor performance, as they had very low consistency across the data sets.

Although SRQ3 could not be completely addressed, the results of the subset and consistency evaluation had the potential to negatively answer SRQ3 and thus partially address SRQ3. SRQ3 focusses on the impact of the algorithm and parameters selected on the cluster set and personas developed. The differences in performance between algorithms demonstrate the impact algorithm selection has on the cluster sets developed. Further, even between cluster sets with similar metrics, such as in the Good Overall Performance Subset, the cluster sets, and thus the personas that would be developed from them significantly differed. During the selection of the best cluster set for the development of the personas, the importance of the domain-specific evaluation was determined, addressing the impact of the algorithm and parameter choice between the top performers.

8.6 Summary

The results of the HyPersona framework were evaluated to determine the consistency of the performance of the algorithms, answering SRQ2. Three subsets of algorithm-parameter combinations were evaluated to address two key points: 1) whether the internal metrics, including the overall performance metric, were indicative of cluster quality; and 2) whether the internal metric values can suggest that two cluster sets are functionally identical. By identifying that the internal metrics were indicative of both cluster quality and functional identicality, the internal metrics could be used confidently to determine algorithm performance consistency. The internal metrics were also used as a guide to identify the top performing algorithm-parameter combinations, so that the best cluster sets could be identified and used to develop persona sets. The next chapter will develop the personas and answer the remaining SRQs.

Chapter 9: PERSONA DEVELOPMENT AND EVALUATION OF RESULTS

Before the persona sets could be developed, the best cluster set developed from each data set had to be identified. The best algorithm-parameter combination for each data set was found through use of internal metrics and domain-specific evaluation. Identifying the best cluster set required all the top performing algorithm-parameter combinations to be compared, which allowed for the impact of algorithm and parameter choice on the clusters developed to be further investigated, thus addressing SRQ3. Through addressing SRQ3 the importance of domain specific evaluation, and therefore whether the hyperparameter tuning process could be further automated, was determined. As domain specific evaluation was found to be essential, the persona development and hyperparameter tuning process could not be further automated.

Based on the two cluster sets selected, a pair of persona sets were created by using the early-stage personas developed by HyPersona. The persona sets were found to be significantly similar, so the 2021 Data Set personas were selected to represent both persona sets. The 2021 persona set was then evaluated to address SRQ4. The evaluation was based on whether the persona set could be validated by the behavioural models that attempt to describe protective action motivation, how similar the personas were to those created by Scovell et al [6], [25] on the 2018 Data Set, and whether the personas could effectively facilitate the development of targeted communication. Finally, the primary research question was able to be addressed based on the results of all four SRQs.

9.1 Identifying the Importance of Hyperparameter Tuning and Domain Specific Evaluation

As described in the previous chapter, SRQ3 focussed on whether the clustering algorithm and parameter selected significantly changed the cluster set developed. As, if there was not a significant difference between the cluster sets developed, hyperparameter tuning was unnecessary. During the analysis of the algorithm consistency, the impact of algorithm and parameter choice was apparent, as some algorithms did not develop any valid cluster sets and algorithm performance often varied significantly based on the parameters.

The secondary purpose of SRQ3 was to determine whether algorithm and parameter selection still had a considerable impact on the cluster set developed once the overall performance was controlled for. That is, whether there was a considerable difference between the cluster sets developed by the overall best algorithm-parameter combinations. To determine the impact of clustering algorithm choice, the differences between the functionally independent top cluster sets were analysed and the impact the differences would have on persona sets developed compared. Based on the requirements of SRQ3 the following hypothesis and null hypothesis were developed:

> **$H3_0$**: There is no significant difference between cluster sets with top overall performance metrics.
>
> **$H3_A$**: The cluster sets with the top overall performance metrics differ significantly.

SRQ3 was then able to be addressed by accepting or rejecting the null hypothesis.

Extending upon SRQ3, whether domain-specific evaluation was required to select the best cluster set was also assessed. The importance of domain-specific evaluation was first determined during the validation of the HyPersona framework, and then re-assessed using the larger variety of algorithm-parameter combinations and additional data set available during the selection of the best cluster sets. Whether or not domain-specific evaluation was necessary was important as the requirement for domain-specific evaluation was the primary barrier to developing a completely automated framework.

The importance of domain specific evaluation relied on the rejection of $H3_0$, as if there was no significant difference between cluster sets with similar overall performance metrics, there was no point performing domain-specific evaluation. To address the importance of domain specific evaluation, the cluster sets selected were compared to the cluster sets with the best overall performance metric and then evaluated in terms of the following hypothesis and null hypothesis:

> **$H4_0$**: Cluster sets with better internal metrics would develop as good or better personas for the given use case as the selected cluster set.
>
> **$H4_A$**: Domain-specific evaluation is required to determine which algorithm-parameter combination gives the best performance for the given use case.

In addition to being compared to the cluster sets with the top overall performance metric, the selected cluster set was also compared to the cluster sets that gave the best performance regarding each individual performance metric. Before any of the hypotheses could be addressed, the best algorithm-parameter combination on each data set had to be identified.

9.2 Selecting the Best Algorithm-Parameter Combination

To select the best algorithm-parameter combination for persona development, the cluster sets with the top 20 overall performance metrics were compared. Each cluster set was evaluated based on the cluster set's quality. Where the quality of a cluster set was defined by whether the cluster set:

- Had good scores across all the internal metrics,
- Was distinct and significant, that is, had a high AFS metric,

- Could be explained through behavioural theory, and
- Was useful for the purpose of targeted communication around the performance of disaster mitigation behaviours.

Prior to evaluating the quality of the cluster sets, the functionally identical cluster sets were identified using the Euclidean distance between internal metric ranks. From a set of functionally identical cluster sets, the cluster set with the best overall performance metric was selected as the representative. As the selected cluster sets all had an overall performance metric amongst the top 20 for their dataset, the first two points of the definition of a quality cluster set were automatically fulfilled. As such, the evaluation of the functionally independent cluster sets focussed on the third and fourth points.

9.2.1 Selection of the Best Algorithm-Parameter Combination Applied to the 2018 Data Set

The algorithm-parameter combinations that achieved the top 20 overall performance metrics on the 2018 Data Set were identified, and the internal metrics of each are given in Table 9-1. Of the 20 algorithm-parameter combinations, nine were identified as functionally identical to SFLA Long v122 and six were identified as functionally identical to the basic k-means ensemble. Which left four cluster sets that were functionally independent from one another based on the algorithm-parameter combinations:

1. **SFLA Long v122:** SFLA based clustering with one of the ‘long’ parameter sets; 200 frogs, 5 memeplexes, 10 memeplex iterations, and 500 top level iterations.
2. **SFLA Long v274:** SFLA based clustering with one of the ‘long’ parameter sets; 500 frogs, 10 memeplexes, 75 memeplex iterations, and 500 top level iterations.
3. **Basic k-means ensemble:** An ensemble of 5 k-means algorithms with random initialisation using the basic consensus function to combine results.
4. **SFLA Long v99:** SFLA based clustering with one of the ‘long’ parameter sets; 100 frogs, 10 memeplexes, 100 memeplex iterations, and 3000 top level iterations.

Of the four algorithm-parameter combinations, SFLA Long v122 had the best overall metric and SC values, SFLA Long v274 had the best AFS value, and the basic k-means ensemble had the best CHI and DBI values. As such, these algorithm-parameter combinations were favoured going into the further stages of analysis. Although not functionally identical, the algorithm-parameter combinations shared many similarities to each other.

TABLE 9-1: TOP 20 ALGORITHM-PARAMETER COMBINATIONS ON THE 2018 DATA SET

Algorithm	SC		CHI		DBI		AFS		Overall
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Metric
SFLA Long v122	0.0874	0.926	45.46	0.952	3.09	0.843	55.33	0.682	3.403
SFLA Long v170 *	0.0873	0.924	45.62	0.957	3.09	0.839	55.17	0.677	3.397
SFLA Long v274	0.0831	0.877	45.07	0.941	3.19	0.795	58.17	0.768	3.380
SFLA Long v11 *	0.0834	0.880	45.82	0.963	3.12	0.829	56.17	0.707	3.379
Basic k-means ensemble	0.0872	0.924	47.11	1.000	2.89	0.929	48.67	0.480	3.333
NMF k-means ensemble ^	0.0872	0.924	47.11	1.000	2.89	0.929	48.67	0.480	3.333
SFLA Long v54 *	0.0853	0.902	46.34	0.978	3.07	0.851	52.50	0.596	3.327
SFLA Long v15 *	0.0839	0.886	46.55	0.984	3.08	0.845	53.00	0.611	3.326
sfla_v47 *	0.0821	0.866	45.85	0.964	3.15	0.815	55.33	0.682	3.326
Basic k-means++ ensemble ^	0.0862	0.912	47.11	1.000	2.93	0.911	49.33	0.500	3.322
SFLA Long v55 ^	0.0852	0.901	46.96	0.996	3.01	0.878	50.33	0.530	3.305
SFLA Long v147 *	0.0848	0.896	44.82	0.934	3.23	0.779	55.67	0.692	3.301
SFLA Long v271 *	0.0825	0.870	46.06	0.970	3.13	0.825	53.67	0.631	3.295
EMA v4 ^	0.0867	0.918	46.74	0.989	2.93	0.912	48.50	0.475	3.295
k-means ^	0.0874	0.926	47.08	0.999	2.90	0.925	47.50	0.444	3.294
SFLA Long v163 *	0.0834	0.880	45.27	0.947	3.24	0.775	55.67	0.692	3.293
SFLA Long v19 ^	0.0853	0.902	46.63	0.986	3.03	0.866	50.50	0.535	3.289
SFLA Long v49 *	0.0800	0.842	46.12	0.971	3.13	0.822	54.33	0.652	3.287
k-means++ ^	0.0860	0.911	47.08	0.999	2.94	0.906	48.33	0.470	3.286
SFLA Long v99	0.0831	0.878	46.30	0.976	3.09	0.840	52.33	0.591	3.285

* Functionally Identical to SFLA Long v122 ^ Functionally Identical to Basic k-means ensemble

Each of the four functionally unique algorithm-parameter combinations were analysed in terms of their internal metrics, the graphs of the clusters, and the additional information available in the full CSV files. An overview of the graphs that represent each of the cluster sets developed are given in Figure 9-1. Larger, more detailed versions of the graphs are available in Appendix C: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2018 Data Set. The clusters in the graphs were re-ordered to facilitate straightforward comparisons. Figure 9-1 shows the strong similarities between cluster sets, with the primary differences present between the third clusters.

Inspection of the demographic factors was a key element in determining which algorithm-parameter combination would be selected. As some cluster sets had distinct demographics, that accounted for a large portion of the AFS value while the behavioural features accounted for very little of the AFS. Demographics, such as previous cyclone experience, were also used to determine how well the clusters could be explained by behavioural theory.

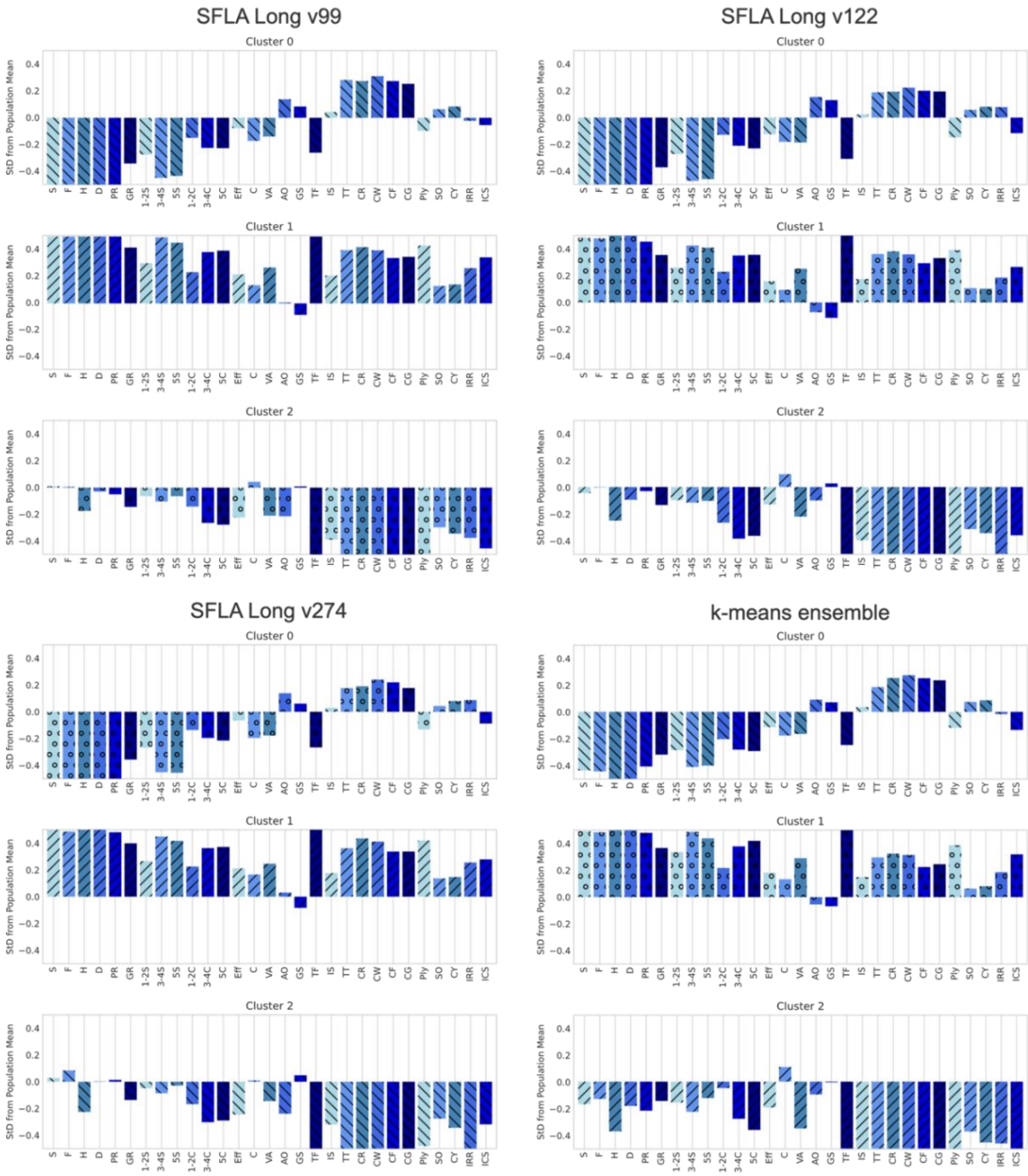


FIGURE 9-1: GRAPHS OF THE TOP FUNCTIONALLY UNIQUE CLUSTER SETS ON THE 2018 DATA SET

As the cluster sets developed were so similar, minor differences determined the best cluster set developed on the 2018 Data Set. Based on the definition of a quality cluster set for persona development, the cluster set developed by SFLA Long v274 was selected. The fact that the SFLA Long v274 cluster set had the highest AFS score, with a large portion of the significant features being behavioural features, was a key aspect in selecting SFLA Long v274 as the best performer. Furthermore, the clusters developed by SFLA Long v274 were slightly more distinct than the other functionally independent cluster sets.

Based on the differences between the four functionally independent cluster sets from the 2018 Data Set, the importance of domain specific evaluation was unclear. Although there were some differences between the cluster sets, the differences were not considerably significant which would support $H3_0$. However, the cluster set developed by SFLA Long v274 was able to be selected as the best cluster set demonstrating that there was some difference between the cluster sets developed, even if not significant.

9.2.2 Selection of the Best Algorithm-Parameter Combination Applied to the 2021 Data Set

The internal metrics for each of the cluster sets with the top overall performance metrics on the 2021 Data Set are given in Table 9-2. The top 20 cluster sets on the 2021 Data Set were more distinct than those developed on the 2018 Data Set. However, once the Euclidean distance between algorithm-parameter combinations internal metric ranks was used to rule out the functionally identical algorithm-parameter combinations, like the 2018 Data Set, only four functionally independent cluster sets remained. The algorithm-parameter combinations that developed the four functionally independent cluster sets were:

1. **k-means:** k-means with random initialisation.
2. **SFLA Long v31:** SFLA based clustering with one of the 'long' parameter sets; 50 frogs, 10 memeplexes, 50 memeplex iterations, and 1000 top level iterations.
3. **SLFA Long v152:** SFLA based clustering with one of the 'long' parameter sets; 200 frogs, 10 memeplexes, 75 memeplex iterations, and 150 top level iterations.
4. **Kernel k-means v57:** Kernel k-means using the polynomial kernel, with a gamma of 'None', coef0 of 0.5, and a degree of 2.

Of the four functionally unique algorithm-parameter combinations k-means had the best DBI, AFS, and overall internal metric values, while SFLA Long v31 had the best CHI value, and Kernel k-means v57 had the best SC value. As such, these three algorithm-parameter combinations were favoured going into the further stages of analysis. As with the 2018 Data Set algorithm-parameter combinations, each of the four functionally unique algorithm-parameter combinations were analysed in terms of their internal metrics, the cluster graphs, the cluster demographics, and the additional information available in the full CSV files. Figure 9-2 gives an overview of the graphs developed for each of the six functionally unique algorithm-parameter combinations and shows the variation between the cluster sets developed. Large, detailed versions of these graphs are available in Appendix D: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2021 Data Set. As with the top performing algorithm-parameter combinations on the 2018 Data Set, the clusters sets were re-ordered within the graphs to facilitate straightforward comparisons.

TABLE 9-2: TOP 20 ALGORITHM-PARAMETER COMBINATIONS ON THE 2021 DATA SET

Algorithm	SC		CHI		DBI		AFS		Overall
	Value	Rank	Value	Rank	Value	Rank	Value	Rank	Metric
k-means	0.0469	0.623	15.12	0.963	3.08	0.976	96.83	0.934	3.497
SFLA Long v139 *	0.0505	0.687	14.93	0.940	3.16	0.932	95.33	0.910	3.469
NMF k-means ensemble *	0.0461	0.609	15.05	0.954	3.09	0.969	95.17	0.908	3.440
SFLA Long v271 *	0.0489	0.658	14.84	0.930	3.15	0.941	94.17	0.892	3.420
SFLA Long v31	0.0601	0.856	15.33	0.987	3.27	0.880	81.83	0.697	3.420
SFLA Long v152	0.0514	0.702	14.91	0.938	3.19	0.918	91.00	0.842	3.400
NMF k-means++ ensemble ^	0.0621	0.891	15.40	0.996	3.23	0.896	76.67	0.615	3.399
SFLA Long v15 ~	0.0518	0.710	15.02	0.950	3.14	0.942	88.00	0.794	3.396
k-means++ ^	0.0607	0.866	15.43	1.000	3.25	0.888	78.33	0.641	3.395
SFLA Long v35 ^	0.0603	0.859	15.37	0.993	3.26	0.883	79.33	0.657	3.392
Kernel k-means v57	0.0620	0.889	14.30	0.865	3.30	0.861	86.17	0.765	3.379
Kernel k-means v173 *	0.0479	0.642	14.54	0.894	3.14	0.943	94.67	0.900	3.378
Basic k-means++ ensemble ^	0.0600	0.854	15.36	0.992	3.25	0.888	78.00	0.636	3.369
SFLA Long v226 †	0.0625	0.898	14.48	0.887	3.34	0.839	84.83	0.744	3.368
Kernel k-means v141 †	0.0626	0.900	14.36	0.873	3.35	0.838	85.50	0.755	3.366
SFLA Long v59 ^	0.0590	0.836	14.78	0.922	3.34	0.844	85.83	0.760	3.361
SFLA Long v47 ^	0.0579	0.817	15.26	0.979	3.31	0.857	82.50	0.707	3.360
SFLA Long v58 ~	0.0468	0.622	14.97	0.945	3.12	0.954	90.00	0.826	3.347
SFLA Long v279 ~	0.0522	0.716	14.79	0.924	3.23	0.899	88.83	0.807	3.346
SFLA Long v19 ^	0.0576	0.811	15.24	0.977	3.33	0.849	82.17	0.702	3.339

* Functionally identical to k-means
^ Functionally Identical to SFLA Long v31
~ Functionally Identical to SFLA Long v152
† Functionally Identical to Kernel k-means v57

Some similarities were found between the first and second clusters; primarily all the first clusters had a much lower than average risk perception and likelihood to perform most damage mitigation behaviours, while all the second clusters had a higher-than-average risk perception and general likelihood to perform damage mitigation behaviours. Although not identical, the cluster sets developed by k-means and SFLA Long v152 were very similar. The primary differences between the two cluster sets were not in the behavioural features, but in the demographic features.

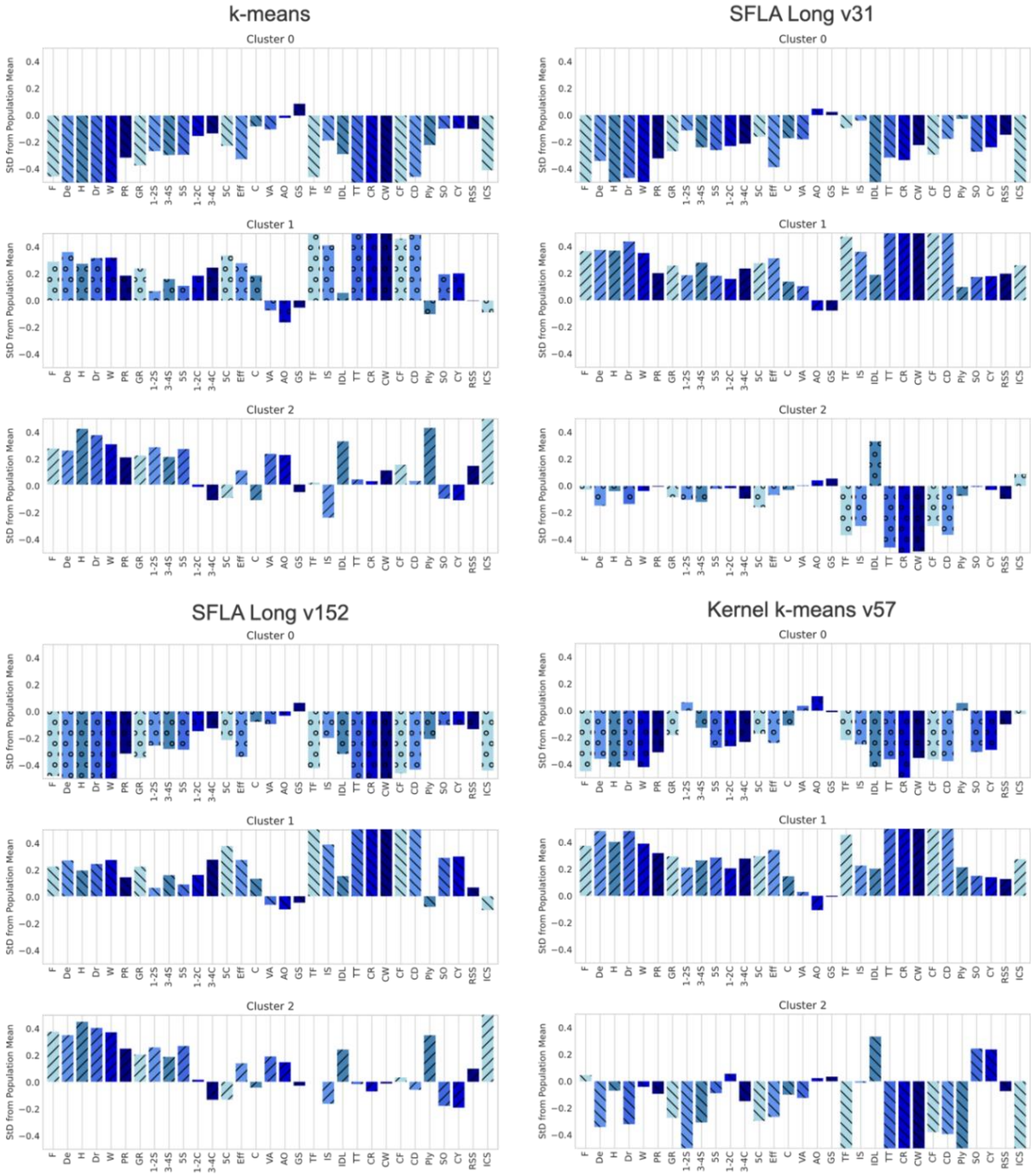


FIGURE 9-2: GRAPHS OF THE TOP FUNCTIONALLY UNIQUE CLUSTER SETS ON THE 2021 DATA SET

Based on the definition of best performer, the cluster set developed by kernel k-means v57 was selected. The kernel k-means v57 cluster set was selected due to the greater differences in the nature of the clusters and how the potential personas would be used to target communication. The potential personas based off the kernel k-means v57 cluster set were identified to be more meaningful in terms of the behavioural theory and ability to be used to target communication compared to the personas potentially developed from the other cluster sets.

The differences between the four functionally independent cluster sets on the 2021 Data Set were far more distinct than those from the 2018 Data Set. There was a blatant difference between all the functionally independent cluster sets, except between those developed by k-means and SFLA Long v152. The differences demonstrated the need for domain-specific evaluation, as the cluster set with the top overall performance metric, k-means, was found to be less suitable for the use case than the cluster set developed by kernel k-means v57. As such, the top cluster sets from on the 2021 Data Set would support the rejection of both null hypotheses.

9.3 The Importance of Algorithm and Parameter Selection

To finish addressing SRQ3 a hypothesis, $H3_A$, and null hypothesis, $H3_0$, were developed. Across both data sets the cluster sets with overall performance metrics amongst the top 20 differed, resulting in 4 functionally unique cluster sets for each data set. The cluster sets on the 2018 Data Set still shared many similarities, while the cluster sets from the 2021 Data Set differed far more significantly.

Considering the cluster sets developed on the 2018 Data Set alone, whether to accept the null hypothesis would be a difficult decision. The differences between the cluster sets were present enough that SLFA Long v274 could be selected as the best performer over the algorithm-parameter combination with the top overall performance metrics, SLFA Long v122. However, Cluster 0 and Cluster 1 developed by each algorithm-parameter combination were almost identical, and the differences between the variations of Cluster 2 were very minor.

The cluster sets developed on the 2021 Data Set differed far more significantly. There were some similarities that could be drawn between the cluster sets developed by each functionally independent algorithm-parameter combination, however there were no clusters that had significant similarities across all the functionally independent cluster sets. Based on the cluster sets developed by the 2021 Data Set the null hypothesis, $H3_0$, was rejected and the alternate hypothesis, $H3_A$, accepted. SRQ3 asked:

SRQ3. Does changing the clustering algorithm and parameters used to develop a cluster set significantly alter the cluster set developed from a data set?

Based on the findings in the previous chapter, the impact that algorithm choice can have on the cluster set developed was evident. The cluster sets developed on each data set varied greatly in quality and content between algorithm parameter combinations. The purpose of continuing to investigate SRQ3 was to determine whether the algorithm and parameters selected meaningfully alter the cluster set when the cluster quality was controlled for. The cluster sets developed on the 2021 Data Set supported the idea that multiple quality cluster sets could exist in a single data set, and that different clustering algorithms and parameters were able to identify the different cluster sets. As such, even when cluster quality was controlled for, the clustering algorithm and parameters used to develop a cluster set could significantly alter the results.

9.4 The Importance of Domain-Specific Evaluation

The existence of multiple quality cluster sets within a single data set indicates the need for domain specific evaluation to identify the most appropriate cluster set for a given use case. The hypotheses, $H4_0$ and $H4_A$ were developed to address whether domain-specific evaluation was required, or whether the cluster sets with better internal evaluation metrics would be the better cluster set for a use case. As the cluster sets developed on the 2021 Data Set differed the most greatly, those cluster sets were focussed on. To determine the importance of domain-specific evaluation, the selected cluster set on the 2021 Data Set, developed by kernel k-means v57, was compared to the cluster set with the top overall performance metric, and the cluster sets with the best score for each of the internal metrics.

9.4.1 Domain-Specific Evaluation vs. the Overall Performance Metric

On the 2021 Data Set, the selected algorithm-parameter combination, kernel k-means v57, and the overall best performer, k-means, differ significantly. As such, the persona sets that would be developed from each set of results would significantly differ. The primary reason kernel k-means v57 was selected over k-means was due to kernel k-means v57 aligning better with behavioural theory and being more useful for the current use case. The clusters developed by k-means, provided interesting insights to the data set and may be more useful than the clusters developed by kernel k-means v57 in some use cases. However, during the current use case the insights and information present in the kernel k-means v57 cluster set were found to be more useful.

The most interesting cluster developed by k-means was the third cluster, Cluster 2. The cluster was, by a significant margin, the most likely to install cyclone shutters. On average, Cluster 2 also perceived a higher-than-average level of risk associated with cyclones, found cyclone shutters to be visually appealing, and rated themselves as able to organise to have cyclone shutters installed. All these elements aligned with what would be expected based on the behavioural theory. However, most of the cluster (89.65%) were renters, and of the renters less than half (47.62%) reported having any form of insurance. This statistic combined with the cluster having the lowest likelihood to have sought out information on how to prevent cyclone damage, and a lower-than-average likelihood to secure outdoor furniture or clear the yard leading up to a cyclone occurring led to an alternate interpretation.

The individuals of the third cluster had rated themselves as less likely than average to perform the behaviours they were able to perform as renters, such as securing outdoor furniture, looking into methods to prevent cyclone damage, and getting insured. However, they had rated themselves as the most likely to perform the mitigation behaviours they were unable to perform as renters, such as installing cyclone shutters, roller door bracing, or deadlocks. As such, the perceived likelihood of the individuals of the third cluster to install structural upgrades was determined as more likely to be a result of idealisation rather than realistic intention.

While such cluster sets represented an interesting and valid group within the data, the cluster was not particularly useful for the purpose of communication targeting based on likelihood to perform protective behaviours. As such, the selected algorithm-parameter combination, kernel k-means v57, gave a better performance for the use case despite not being the top performer in terms of the overall performance metrics. Based on these findings, domain-specific evaluation was required to identify the best cluster set for the use case compared to using the overall performance metric for selecting a cluster set.

9.4.2 Domain-Specific Evaluation vs. Individual Internal Evaluation Metrics

To determine whether domain-specific evaluation led to the selection of a better cluster set for the given use case, compared to basing cluster set selection on and individual internal metric, the selected cluster set from the 2021 Data Set, kernel k-means v57, was compared to the cluster sets with the best individual score for each internal metric. The cluster sets with the best score for each of the individual internal evaluation metrics on the 2021 Data Set were:

- **Top SC value – SFLA v327:** SFLA based clustering with 150 frogs, 10 memeplexes, 10 memeplex iterations, and 150 top level iterations.
- **Top CHI value – k-means++:** k-means with k-means++ initialisation.
- **Top DBI value – ABC v49:** ABC based clustering with 100 bees, a discard limit of 30, and 500 maximum iterations.
- **Top AFS value – Genetic v37:** Genetic clustering with 150 chromosomes, 250 populations, 1 gene mutated at each step, and a selection coefficient of 0.01.

The cluster set with the best SC value was SFLA v327, the graphs created for SFLA v327 are given in Figure 9-3. The graphs created for Kernel k-means v57 are given in Figure 9-4 to allow for easier comparison. Although not functionally identical, there were strong similarities between Cluster 0 and Cluster 1 developed by Kernel k-means v57 and SFLA v327. However, SFLA v327 was slightly imbalanced, with Cluster 2 only containing 27 (12.8%) of the data points. As the information present in the cluster set developed by SFLA v327 did not appear to be any more meaningful and was more imbalanced than the Kernel k-means v57 cluster set, the Kernel k-means v57 cluster set was determined to provide more insight for the given use case.

The cluster set with the top CHI value, k-means++, was amongst the cluster sets with the top 20 overall performance metrics. The k-means++ cluster set was found to be functionally identical to the cluster

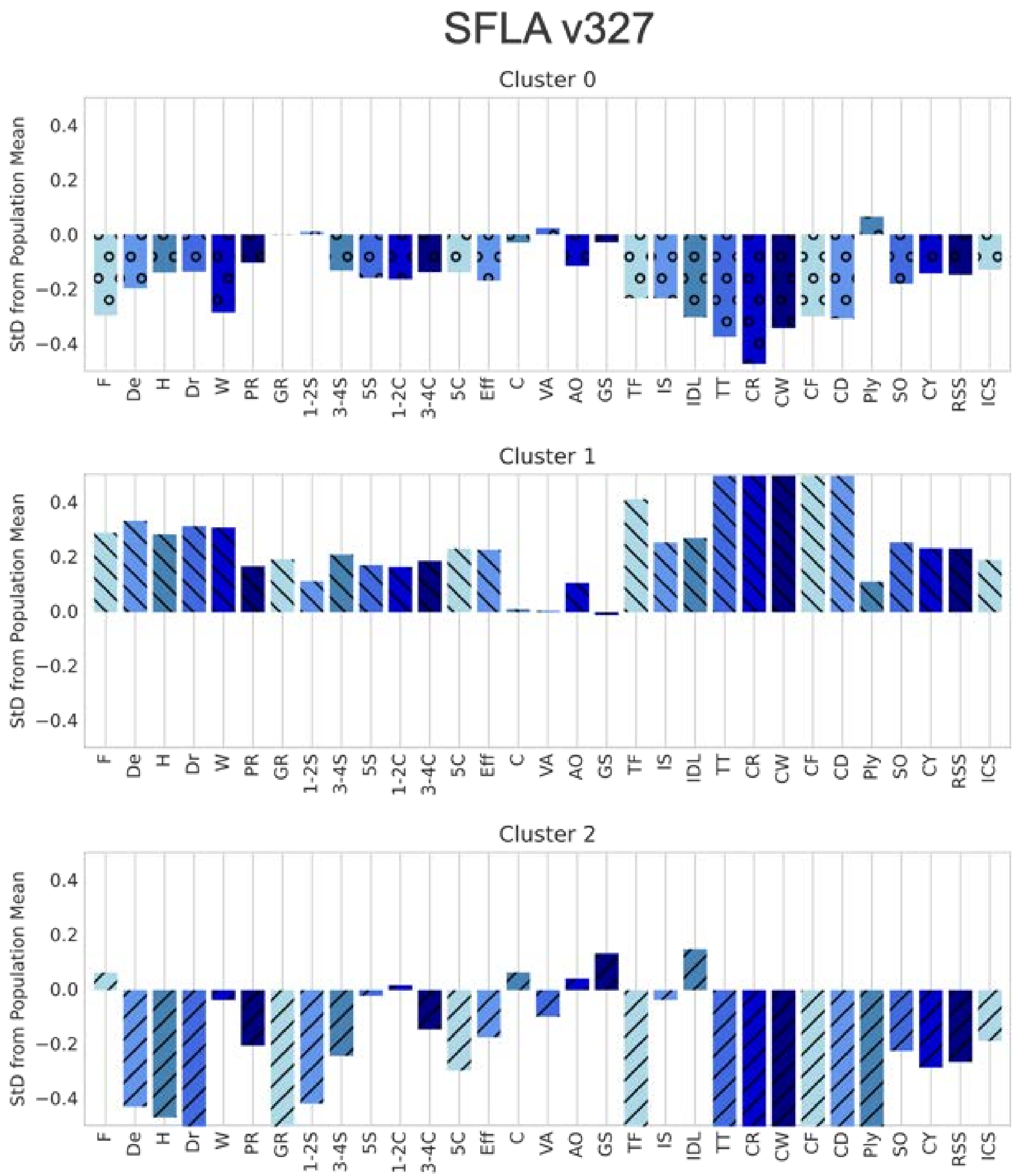


FIGURE 9-3: THE CLUSTER SET WITH THE TOP SC VALUE – SFLA V327

set developed by SFLA Long v31 based on the Euclidean distance between metric ranks. There were multiple reasons that SFLA Long v31, and thus k-means++, was not selected as the best performer.

Primarily, Cluster 1 did not align with behavioural theory. Cluster 1 had a generally lower than average perception of cyclone risks, severity, and likelihood, less cyclone and cyclone damage experience, and a much lower than average likelihood to perform protective behaviours, except for installing cyclone shutters. Behavioural theory would expect, based on the perceptions and experiences, that such an individual would have a lower-than-average likelihood to install cyclone shutters. Furthermore, the only cluster with a below average likelihood to perform all protective behaviours, Cluster 0, only contained

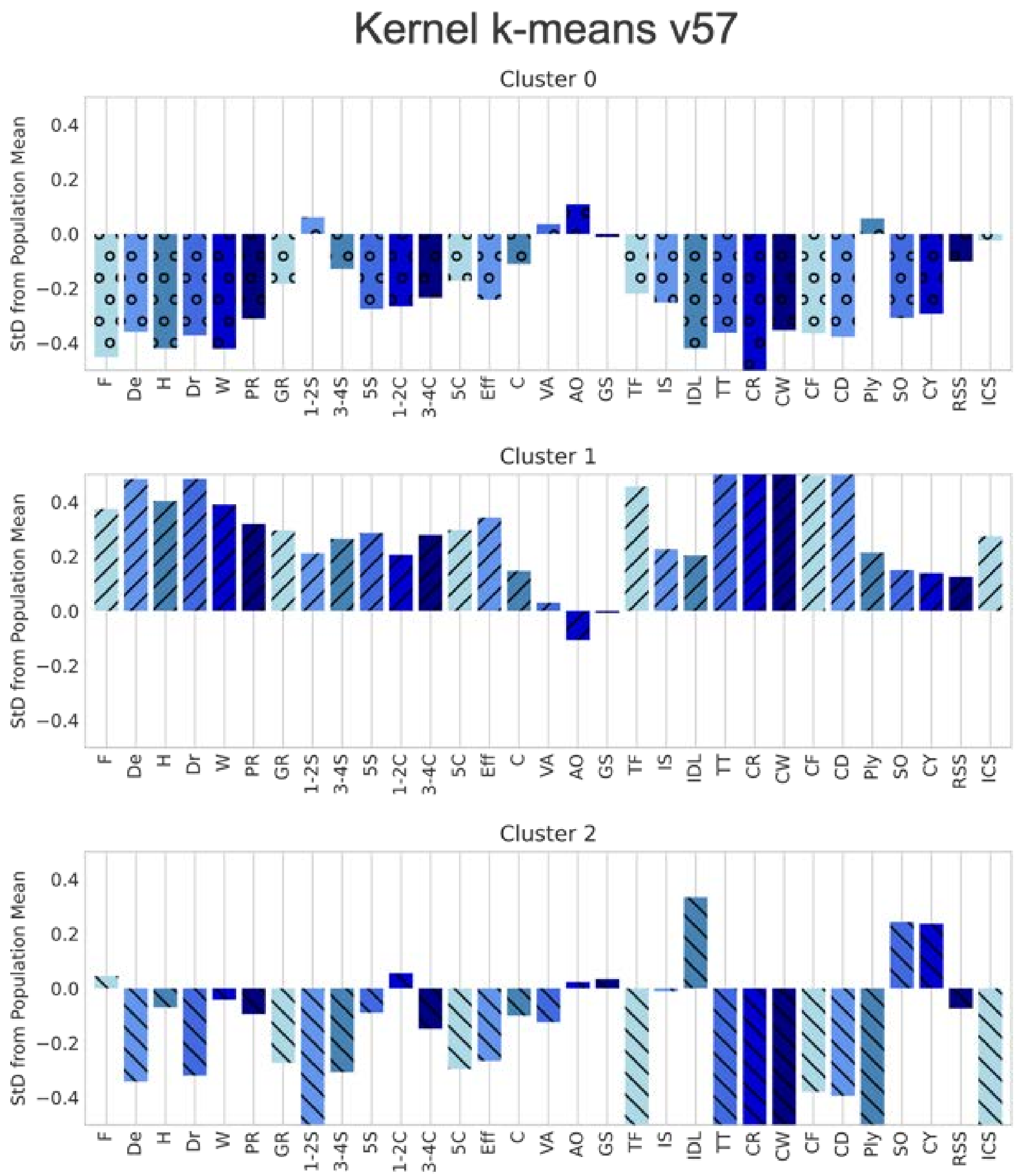


FIGURE 9-4: THE CHOSEN CLUSTER SET – KERNEL K-MEANS V57

45 statistically significant features. Which was far less than the 78 statistically significant features of the kernel k-means v57 cluster with the fewest statistically significant features, Cluster 2. The cluster set developed by k-means++ was interesting and may be worth further investigation. However, for the current use case, targeting customised communication to promote uptake of protective behaviours, the cluster set developed by kernel k-means v57 was more useful.

The cluster set with the top DBI value was developed by ABC v39, which is given in Figure 9-5. Across the board, there was a pattern of cluster sets that achieved the best DBI values performing poorly over the other internal metrics. ABC v39 was no exception to the trend, with an overall performance metric of 2.1. Furthermore, the cluster set developed by ABC v39 were more imbalanced, containing one

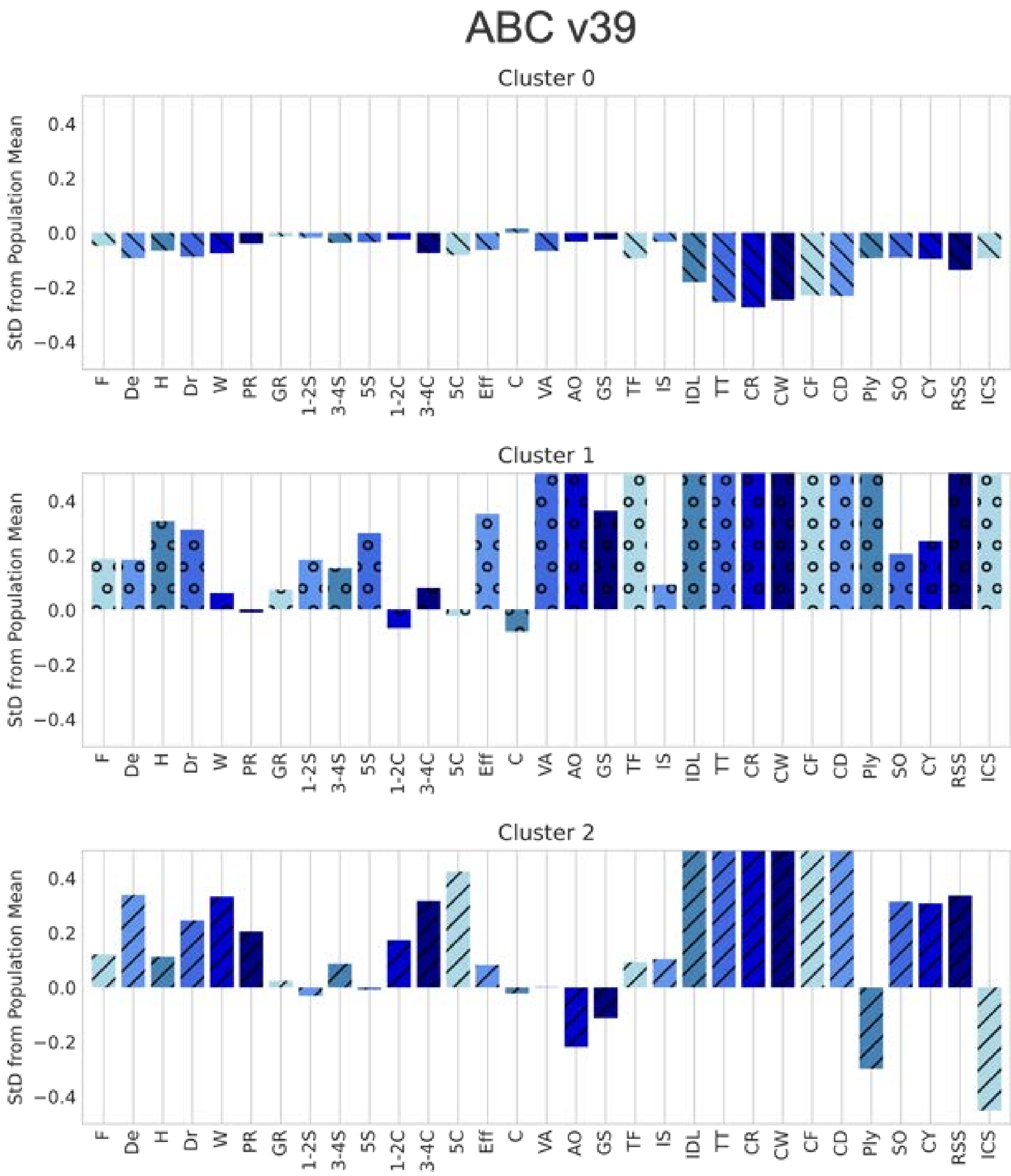


FIGURE 9-5: THE CLUSTER SET WITH THE TOP DBI VALUE – ABC V39

cluster, Cluster 0, with 158 data points (74.88%). Interestingly, the two small clusters, Cluster 1 and Cluster 2, were very similar except for a few key behavioural features and the likelihood to put up plywood and install cyclone shutters.

A primary difference between Cluster 2, which had a much lower than average likelihood to install cyclone shutters, and Cluster 1, which had a much higher than average likelihood to install cyclone shutters, was that Cluster 1's perception of the efficacy and visual appeal of cyclone shutters and their ability to have cyclone shutters installed were much higher than average. Conversely to the behavioural models, Cluster 2 had a higher perception of the likelihood of a cyclone occurring and a lower perception of the likelihood of government support, both factors that would be expected to lead to a

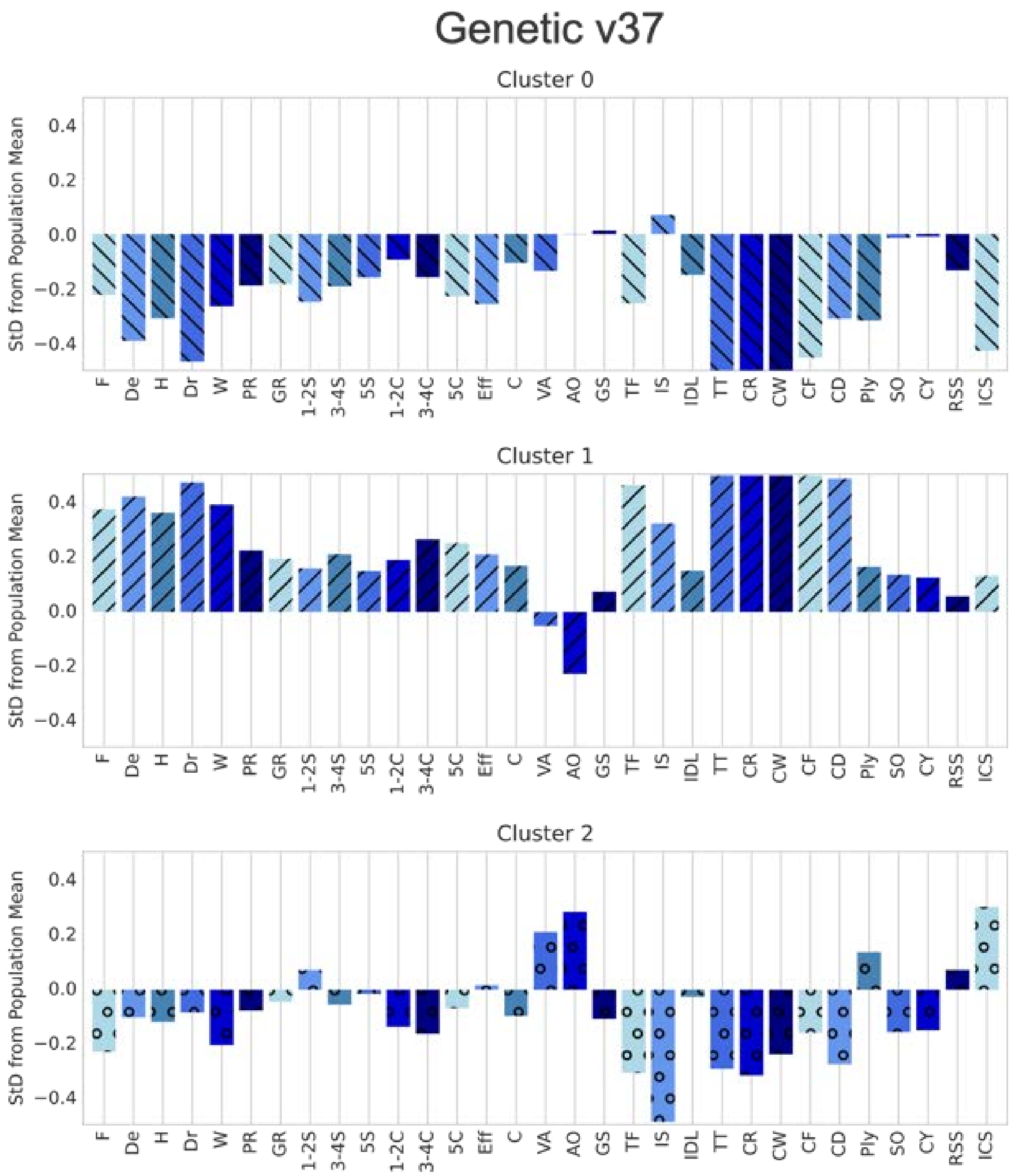


FIGURE 9-6: CLUSTER SET WITH THE TOP AFS VALUE – GENETIC V37

higher motivation to install cyclone shutters. However, they also had stronger feelings of worry and depression when thinking about a cyclone occurring, and lower than average perception of their ability to have cyclone shutters installed, which were all factors that could lead to a maladaptive response. As with the previous cluster sets, there were some interesting insights to be gained from the cluster set developed by ABC v39, however, for the current use case the cluster set developed by kernel k-means v57 would be more applicable.

The final internal metric to consider was AFS, the internal evaluation metric designed for persona development. The algorithm-parameter combination that developed the cluster set with the best AFS was genetic v37, which is given in Figure 9-6. Like the cluster set developed by k-means++, the genetic v37 cluster set contained two clusters with a higher-than-average likelihood to install cyclone shutters. However, compared to k-means++ and SFLA v31, the cluster sets developed by genetic v37 were more explainable through behavioural theory. Primarily as Cluster 2 had a much higher than average perception of the visual appeal of cyclone shutters and their ability to have cyclone shutters installed. As with the cluster set developed by k-means++, whilst interesting, the cluster set developed by kernel k-means v57 was more applicable to the current use case than the cluster set developed by genetic v37.

9.4.3 Summary of the Importance of Domain-Specific Evaluation

To determine the importance of domain-specific evaluation and address H4, the cluster set selected as the best performer, kernel k-means v57, was compared to the cluster sets with the best performance according to the internal metrics. All the cluster sets considered were valid cluster sets that were found to offer some insight into the data set. As such, using the overall performance metric or any of the individual evaluation metrics would produce an interesting and valid cluster set.

Each of the cluster sets developed were quite different, and other than relying on the internal metrics, domain-specific evaluation was the only way to determine which cluster set was the most useful for the given problem. When evaluated, kernel k-means v57 was consistently found to be the more applicable cluster set, despite not giving the top performance according to any of the metrics. As such, the null hypothesis, $H4_0$, must be rejected as cluster sets with better internal metrics would not develop better personas for the given use case. The alternate hypothesis, $H4_A$, that domain-specific evaluation is required to determine which algorithm-parameter combination gives the best performance for the given use case was then accepted.

9.5 Persona Creation

A set of personas was developed from each of the selected cluster sets. To create the personas the early-stage personas developed by HyPersona were used as the base, and the features that differed most greatly were identified. The most important features, defined as differing the most significantly and giving the most insight into the persona, were selected to be prominent details of the final personas. Each persona was assigned an epithet, rather than a name, and most demographic factors were left out, as factors such as age, gender, location, and marital status were generally not significant.

As part of determining the key features, Chi Squared tests were performed to determine the independence of the categorical elements. The categorical elements found to be statistically independent were cyclone damage experience, likelihood to install cyclone shutters, and, with the 2021 Data Set, insurance status. The results of the Chi Squared tests are given in Table 9-3. From the personas developed on the 2018 Data Set both the likelihood to install cyclone shutters and cyclone damage experience were similarly significant. While, on the 2021 Data Set personas the insurance status was the most significant feature.

TABLE 9-3: CHI SQUARED TEST RESULTS

	2018 Data Set Personas	2021 Data Set Personas
Likelihood to Install Cyclone Shutters	$\chi^2(6) = 28.01, p<0.001$	$\chi^2(6) = 21.49, p<0.002$
Cyclone Damage Experience	$\chi^2(10) = 71.71, p<0.001$	$\chi^2(12) = 23.16, p<0.05$
Insurance status		$\chi^2(10) = 48.59, p<0.001$

Other key elements for each set of personas included home ownership, risk perceptions, emotional response, perceived likelihood and severity of cyclones, and likelihood to perform preparatory behaviours. These elements were each included in the persona in an easy to interpret manner, such as pie charts to show cyclone damage experience. Based on the statistically significant features of each persona, a short blurb was written up for each persona which was then synthesised further into an overview sentence.

The blurbs and overview sentences were used to develop the epithets. Although the two persona sets were not identical, the same three epithets and the associated overview sentences were able to be used for each persona set. The three epithets and overview sentences were:

1. **Proactive:** They are actively worried about the possibility of a cyclone occurring and are willing to perform any behaviours to protect themselves, their family, and their property.
2. **Confident:** They are aware of the risks associated with cyclones but are confident they know what to do to mitigate any damage – as they have done before.
3. **Unconcerned:** They do not really consider cyclones as a pressing issue, whether due to lack of experience, the age of their house, or being a renter.

The persona sets did, however, differ enough that the blurbs created differed. Both the differences in the clusters developed and the data available impacted the blurbs created for the personas. One primary difference in the data available between the cluster sets was the presence of the insurance information and whether the individual previously played a role in cyclone preparation.

The personas created from the 2021 Data Set tended to have less distinct demographic factors, such as education and income, compared to the personas created from the 2018 Data Set. However, home ownership and the likelihood that the home was built before 2012, were more distinct in the 2021 Data Set personas. Across all three personas, the personas from the 2021 Data Set had a lower intention to install cyclone shutters and were less likely to have cyclone shutters already installed compared to the 2018 Data Set personas. The other differences between the persona sets tended to be more minor or reflect a feature of one persona being present in a different persona on the alternate data set. For example, in the 2018 Data Set personas, the proactive persona was most likely to be a homeowner, while in the 2021 Data Set the confident persona was most likely to be a homeowner.

Despite the differences, the persona sets were similar enough that the conclusions that could be drawn about one set of personas, could be drawn about the other, and communication for each set of personas would be targeted in the same manner. The similarity between the two data sets may have been, in part, due to the domain-specific evaluation favouring certain aspects within the personas. However, the 2021 Data Set was also a confirmatory study of the 2018 Data Set, so the information captured in both data sets was expected to be very similar. The 2021 Data Set personas were selected to be the main persona set discussed, predominantly as they contained additional data, such as insurance status.

Graphical representations of both persona sets were manually developed to facilitate their interpretation, particularly with audiences that were non-technical or unfamiliar with the behavioural theory behind preparation motivation. The graphical representations focused on the key traits of the personas and facilitating interpretation informed by behavioural theory. The complete personas developed from the 2018 Data Set, based on the cluster set created by SFLA long v274, are available in Appendix E: 2018 Data Set Personas. The complete personas developed from the 2021 Data Set using the kernel k-means v57 cluster set are given in:

Figure 9-7, the Proactive persona; Figure 9-8, the Confident persona; and Figure 9-9, the Unconcerned persona. Coloured versions of the personas are available in Appendix F: 2021 Data Set Coloured Personas.

9.6 Persona Evaluation

Persona evaluation was important to determine whether a useful persona set was able to be developed through the application of HyPersona, and to address SRQ4:

SRQ4. How do personas created by clustering algorithms compare to behavioural theory and personas created through the application of behavioural theory?

Through answering SRQ4, whether the HyPersona framework was able to develop a set of personas that successfully mimicked expert decision making could be determined. The personas were not expected to be an exact copy of the persona set developed by Scovell et al. [6], [25], especially given the difference in data used. Instead, to be most successful, the persona set was desired to have a similar level of connection to behavioural theory and provide a similar level of insight into the data set. Similarities in the nature of the personas developed was also seen as a positive.

Proactive Persona

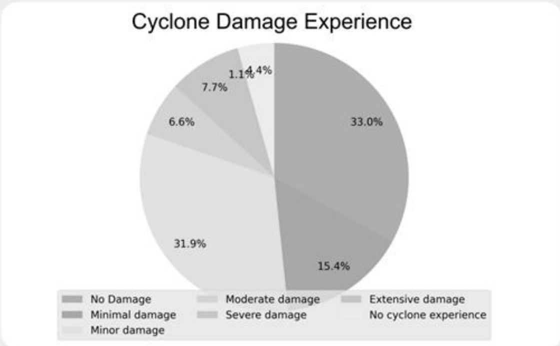
They are actively worried about the possibility of a cyclone occuring and are willing to perform any behaviours to protect themselves, their family, and their property.

They are the most concerned about the possibility of a cyclone, with the highest expectation of both a cyclone occurring and doing considerable damage. They have the strongest emotional response to the possibility of a cyclone and are most likely to spend time thinking about or discussing the possibility of a cyclone. They are the most likely to perform preparatory behaviours leading up to the next cyclone, including more difficult behaviours such as installing cyclone shutters, and to have looked up ways to prevent cyclone damage. They are the most likely to have previously played a role in cyclone preparation and have the strongest belief that a cyclone occurring could damage their property.

About Them

They are probably a homeowner (73.03%) of a home built before 2012 (62.92%), and intend to live in their house for 3-4 more years.

They are most likely to have played a role in the preparation leading up to previous cyclones (94.38%)



Attitudes towards cyclones

They are likely to discuss and think about cyclones somewhat regularly and to have sought out more information on ways to reduce cyclone damage.

They consider themselves as fairly knowledgeable about cyclones and the related risks.

They think cyclone shutters are likely to be somewhat effective

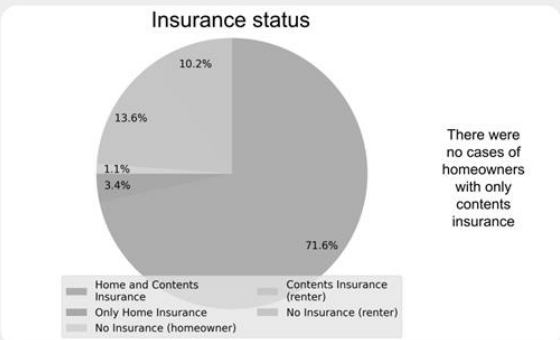
Beliefs about Cyclones:

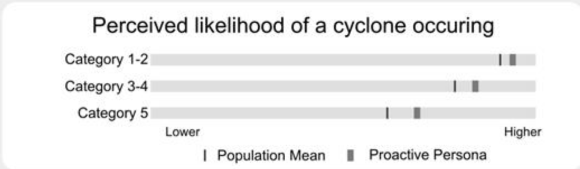
They believe a cyclone occuring would:

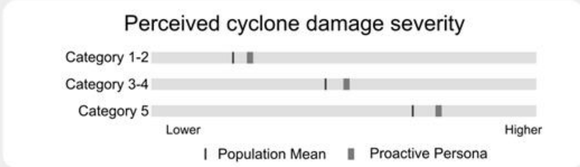
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

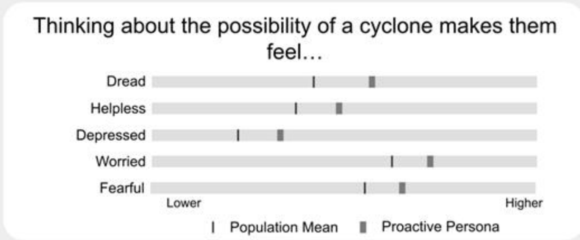
And might also:

- Damage their home









Before the next cyclone they are:

Very likely to:

- Secure outdoor furniture and items
- Clear yard of any loose items
- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes

Are unsure whether they would:

- Put plywood up on glass windows/doors

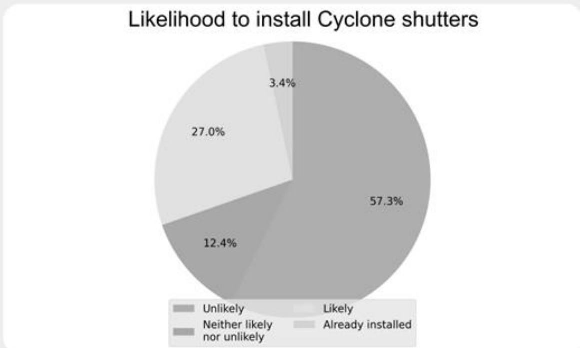


FIGURE 9-7: PROACTIVE PERSONA DEVELOPED FROM THE 2021 DATA SET

Confident Persona

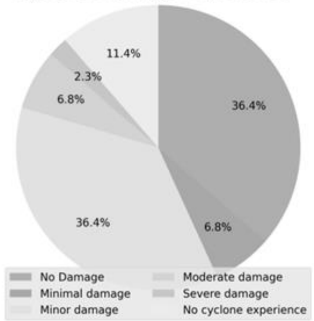
They are aware of the risks associated with cyclones but are confident they know what to do to mitigate any damage – as they have done before.

They perceive the least amount of risk associated with a cyclone occurring, only having a strong belief that a cyclone occuring would disrupt their daily life. They are most likely to have been through a cyclone before where they received minor damage and the most likely to be fully insured. They rate themselves as the most knowledgeable about cyclones, although they are the least likely to think about or discuss cyclones. They are likely to do simple mitigation behaviours leading up to the next cyclone but are unlikely to perform any of the more difficult tasks or install structural upgrades, although they are the most likely to be homeowners and living in a property built before 2012.

About Them

They are most likely to be a homeowner (97.73%) of a home built before 2012 (72.73%), and intend to live in their house for 5+ more years.
They likely played a role in the preparation leading up to previous cyclones (86.36%)

Cyclone Damage Experience



There were no cases of extreme cyclone damage

Attitudes towards cyclones

They are unlikely to discuss or think about cyclones regularly but might have sought out information on ways to reduce cyclone damage. They consider themselves very knowledgeable about cyclones and the related risks and are unsure the effectiveness of cyclone shutters.

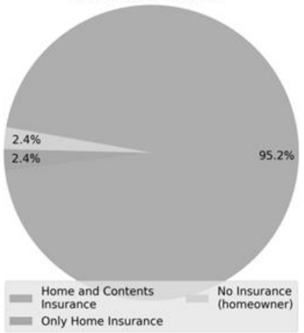
Beliefs about Cyclones:

They believe a cyclone occuring will:
- Disturb daily life

Might also:
- Cause catastrophic destruction
- Pose a great financial threat
- Disrupt their ability to work

But probably wouldn't:
- Cause widespread death
- Pose a significant threat to future generations
- Negatively affect physical health

Insurance status

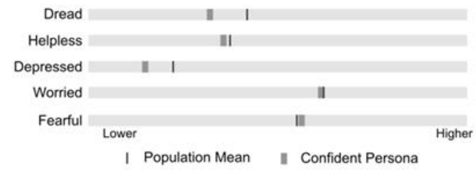


There were no cases of homeowners with only contents insurance and no insurance information from renters

Perceived likelihood of a cyclone occuring

Perceived cyclone damage severity

Thinking about the possibility of a cyclone makes them feel…



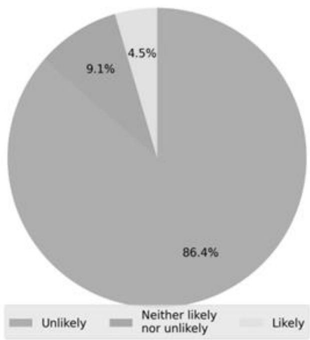
Before the next cyclone they are:

Very likely to:
- Secure outdoor furniture and items
- Clear yard of any loose items

Are unsure whether they would:
- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes

Are unlikely to:
- Put plywood up on glass windows/doors

Likelihood to install Cyclone shutters



There was no one who already had cyclone shutters installed

FIGURE 9-8: UNCONCERNED PERSONA DEVELOEPD FROM THE 2021 DATA SET

Unconcerned Persona

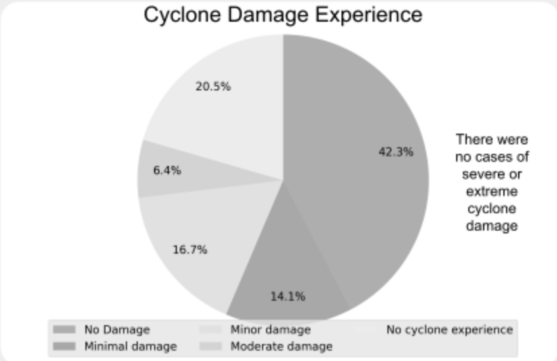
They don't really consider cyclones as a pressing issue, whether due to lack of experience, the age of their house, or being a renter.

They are unlikely to spend time discussing or thinking about cyclones, rate themselves the least knowledgeable about cyclones, and are least likely to have looked up ways to reduce cyclone damage. While they do believe a cyclone occurring would have significant consequences, they have the weakest emotional response to the prospect of a cyclone occuring. They are likely to perform the very simple mitigation behaviours but are unsure about the more difficult behaviours. They are the least likely to have experienced a cyclone, or to have been previously involved in cyclone preparation, and of those with cyclone experience are the most likely to have received no damage. They are most likely to be renting, living in a more recently built property, and to be uninsured.

About Them

They are probably a renter (56.41%) of a home built after 2012 (69.23%), and only intend to live in their house for 2-3 more years.

They are least likely to have played a role in the preparation leading up to previous cyclones (71.79%)



Attitudes towards cyclones

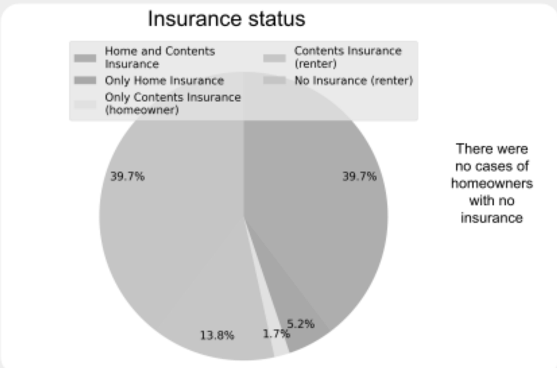
They are less likely to discuss and think about cyclones regularly and to have not sought out information on ways to reduce cyclone damage. They don't consider themselves as knowledgeable about cyclones and the related risks. Are unsure whether cyclone shutters would be effective

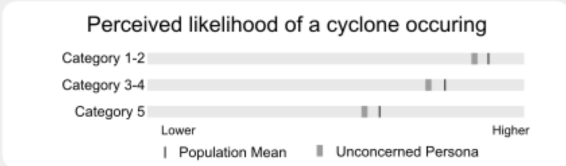
Beliefs about Cyclones:

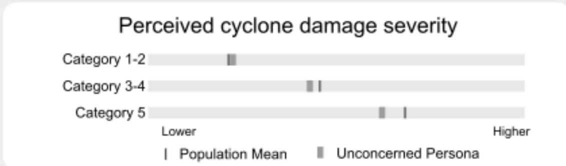
They believe a cyclone occuring might:

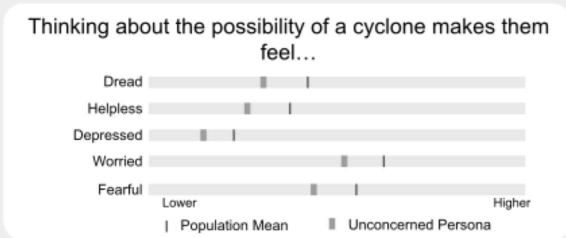
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

But they are unsure on any other possible effects.









Before the next cyclone they are:

Very likely to:

- Secure outdoor furniture and items
- Clear yard of any loose items

Are unsure whether they would:

- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes
- Put plywood up on glass windows/doors

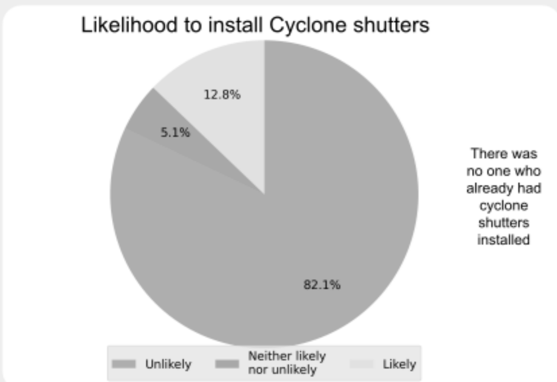


FIGURE 9-9: UNCONCERNED PERSONA DEVELOPED FROM THE 2021 DATA SET

Although not explicit in the SRQ, to have successfully mimicked expert decision making the personas had to be useful. If the personas could not be used to effectively target communication, they were not useful for the given use case despite any internal metrics or links to behavioural theory. The more proficient the persona set was for the given purpose the closer the persona set mimicked expert decision making. Based on these requirements, three key elements were identified to evaluate the personas:

1. Can the personas be explained through behavioural theory?
2. Is the persona set comparable to the expert created persona set developed by Scovell et al. [6], [25] on the 2018 Data Set?
3. Can the persona set be used to effectively target communication for the purpose of promoting the uptake of protective behaviours, such as cyclone shutters?

9.6.1 Validating the Personas Through Behavioural Theory

Quality personas for targeted communication needed to be explainable through existing behavioural models, so that behavioural theory could be applied to customise the communication. Two of the primary behavioural models around predicting motivation to perform risk mitigation behaviours, PADM and PMT, were applied to the personas to determine how well the personas could be explained through behavioural theory. The most straightforward explanation of the 2021 personas was through PMT.

PMT proposes that an individual first appraises their perception of the threat, known as threat appraisal. Then, if the threat is considerable, they will go on to assess the protective behaviour and their ability to perform the behaviour, known as the coping appraisal. When the coping appraisal is sufficient, the individual is the most likely to be motivated to perform the behaviour. However, if the coping appraisal is too low the individual is more likely to turn to maladaptive behaviours, and if the coping appraisal is too high threat appraisal can be reduced even though the protective behaviour was not performed.

Figure 9-10 gives a simple overview of how PMT was used to explain the 2021 personas. Based on PMT, the Unconcerned persona was a result of low threat appraisal, and the Proactive persona was a result of sufficient threat appraisal and sufficient coping appraisal. The most difficult persona to explain through PMT was the Confident persona. There were potentially multiple ways in which the Confident persona could be interpreted, the chosen interpretation of the Confident persona was that they had sufficient threat appraisal, however their coping appraisal was so high that their perception of the threat was mitigated. As the Confident persona had the most experience with cyclones with no to minor damage, they were likely to be confident that they know how to deal with any prospective cyclone and that performing the behaviours they had performed previously would mitigate damage, which reduced their perception of the threat associated with a cyclone occurring.

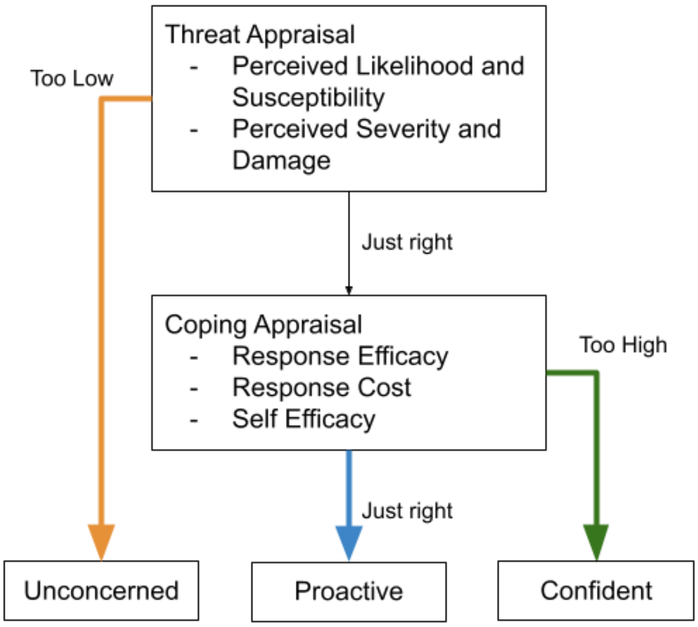


FIGURE 9-10: EXPLAINING THE PERSONAS THROUGH PMT

Factors from the PADM could also be used to describe the personas and provide extra insight. The elements in the pre-decisional phase of the PADM were found to be of particular importance as both exposure and attention functioned as key differentiators between the personas. In the 2021 Data Set exposure was operationalised as the individual's previous cyclone experience, including whether they received damage, and whether they have played a role in the preparation for a previous cyclone. While attention was operationalised in the 2021 Data Set as the topic frequency, that is, how often the individual spends time thinking about or discussing cyclones and cyclone related risks. Both the exposure and attention of each persona aligned with what would be expected by the PADM, and thus further validated the personas through behavioural theory. The way in which each persona was explained with the pre-decisional phase of the PADM was:

- **Protective Persona:** The Proactive persona had the highest level of exposure to and attention towards cyclones and was the most likely to perform protective behaviours. Which was also demonstrated in the Proactive persona reporting the highest chance of having sought out information on ways to mitigate cyclone damage as the PADM lists searching for additional information as one of the possible positive behavioural responses that can be undertaken prior to performing the protective behaviour.
- **Unconcerned Persona:** The Unconcerned persona, on the other hand, had the lowest average exposure to cyclones and cyclone damage, and paid lower than average attention to cyclones and cyclone risks. The Unconcerned persona also rated themselves as having the least knowledge about cyclones and cyclone damage and were least likely to have previously performed information seeking behaviour.

- **Confident Persona:** The Confident persona had lower than average exposure to cyclones and cyclone damage and, most importantly, paid the lowest attention to cyclones. They rated themselves as the most knowledgeable about cyclones and cyclone damage and had a moderate likelihood of having performed an information seeking behaviour, which suggests the low attention to cyclones may be from a confidence in their existing knowledge of cyclones and the related risks.

As such, the personas developed on the 2021 Data Set were found able to be meaningfully explained through both PMT and the PADM. Which indicated that the personas would allow for the meaningful targeting of communication. However, the personas first had to be evaluated compared to the expert created persona set by Scovell et al. [6], [25].

9.6.2 Evaluating the Personas against the Expert Created Persona Set

There were strong similarities between the persona sets developed by HyPersona and the expert created persona set developed by Scovell et al. [6], [25]. Although the two persona sets were not identical, there were similarities in the way that the data set was divided, and between the personas themselves. In particular, the Proactive persona developed by Scovell et al. [6], [25] and the Proactive persona developed with HyPersona shared many similarities. To a lesser degree there were similarities between the Denialist persona developed by Scovell et al. [6], [25] and the Confident and Unconcerned personas developed by HyPersona. The Pessimist persona developed by Scovell et al. [6], [25] was the most different to any of the personas developed with HyPersona, however the persona does share some similarities with the Unconcerned persona.

Both the Proactive personas were most likely to perform the desired behaviours, with the highest risk perception and the highest likelihood to have previously experienced a cyclone where they received property damage. The Proactive personas were similar enough that both would be targeted with the same type of communication and treated similarly. The Denialist and Confident personas were the next most similar, although the Denialist had many of the properties of both the Confident and Unconcerned personas. The primary similarity between the Denialist and Confident personas was that both were likely to have experienced a cyclone where they received no, or minimal, damage and their lack of risk perception was proposed to come from this experience.

However, unlike the two Proactive personas, the Denialist persona and the Confident persona would need to have communication customised for them differently. In terms of the targeted communication required, the Denialist persona was most like the Unconcerned persona, as both had low perception of the risks surrounding cyclones and primarily require their threat appraisal to first be adjusted. The Unconcerned and Pessimist persona were alike in that they were both the least likely persona to have previously experienced a cyclone, were relatively unlikely to install cyclone shutters, and had a low perception of the efficacy of cyclone shutters. The Pessimist persona represents an individual who had responded to the cyclone risk with maladaptive behaviours, due to a high threat appraisal and low coping appraisal, which none of the HyPersona personas represented.

In general, the expert created personas put more emphasis on the perceptions related to cyclone shutters, such as the perceived efficacy and visual appeal of cyclone shutters, which was expected as that was the primary purpose the personas were developed for. While the HyPersona personas focused more on previous experiences, attitudes towards cyclones, and the attention paid to cyclones. However, both persona sets equally represented states which the behavioural models can be in, and both could be used to a similar extent to develop targeted communication. There were not any major differences in the quality, depth, or nuance, between the persona sets or the way in which the personas could be explained through behavioural models. As such, the persona set developed through the application of HyPersona was found to be comparable to the expert created persona set.

9.6.3 The Personas' Use for Targeted Communication

The persona set developed by HyPersona were found to be able to be used to effectively target communication. Figure 9-11 outlines how the explanation of the personas through behavioural theory could be directly used to effectively target communication. Through extending the model used to explain the personas through PMT the primary needs of each persona could be identified, where the need was the primary barrier between the individual and performing the desired protective behaviour according to the behavioural model. Communication that targets those needs has, in theory, the greatest potential to impact the individual's motivation to perform a given behaviour.

In each case, the example messaging gave a 'fun fact' about cyclones or a preparatory behaviour to capture the individual's attention and then prompted the individual to follow a URL to find out more information. The fun fact and call to action were both designed to address the persona's need. As the Proactive persona was the most likely to perform the protective behaviours, the primary purpose of the messaging was to act as a call to action. Additionally, as the Proactive persona had a significant emotional response to the prospect of a cyclone occurring, the fun fact was targeted at increasing coping appraisal and avoided reinforcing the threat appraisal to circumvent any possibility of the individual turning towards maladaptive behaviours. The Unconcerned persona was least worried about a cyclone occurring, so the messaging focussed on reminding the individual of the risk and likelihood of a cyclone occurring. The Confident persona also had a low perception of the risk surrounding cyclones, but likely already intended to perform the simple damage mitigation behaviours. As such, the messaging focussed on the specific types of damage that could only be mitigated through the behaviours they were less likely to perform.

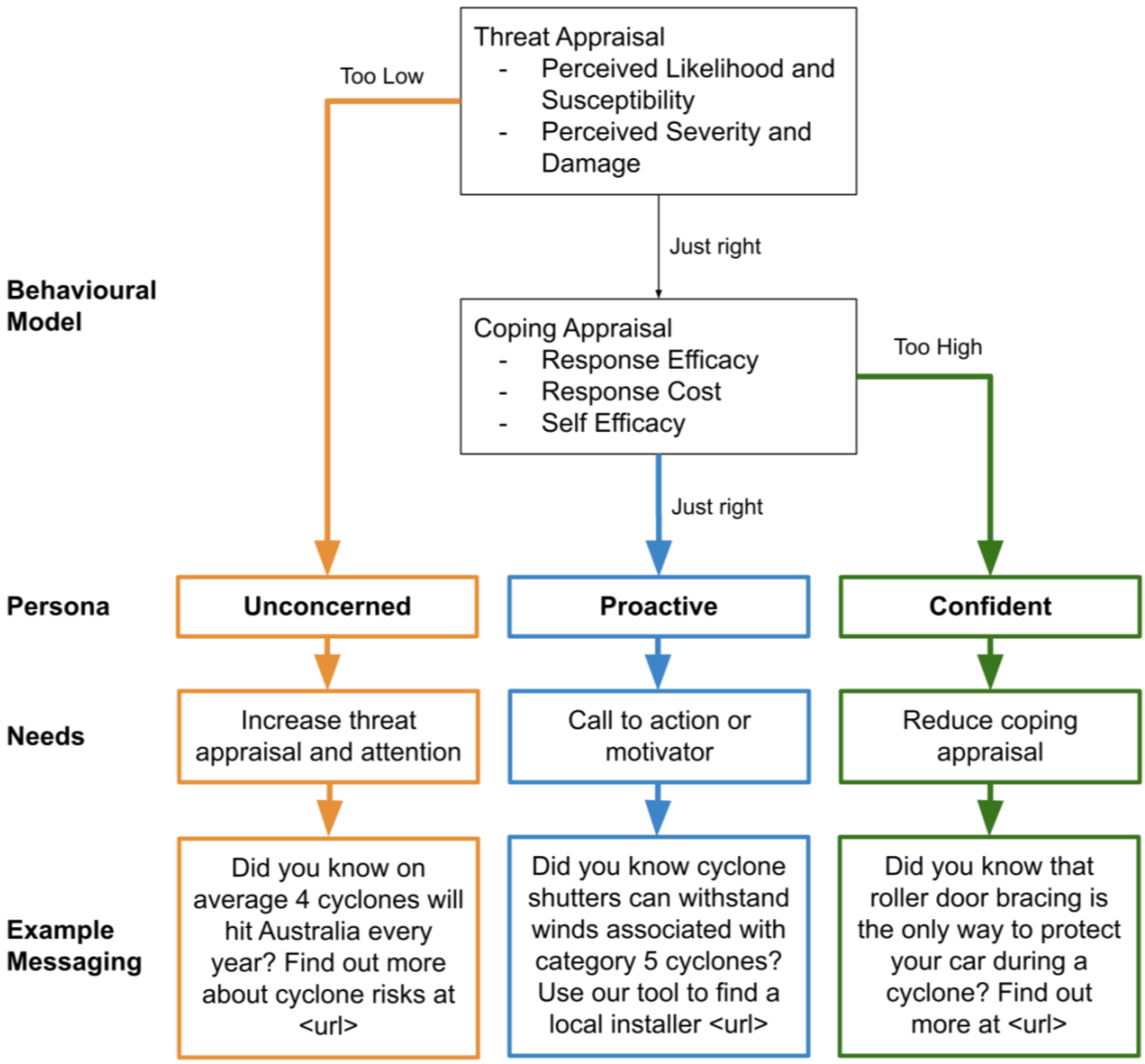


FIGURE 9-11: USING THE PERSONAS AND PMT TO TARGET COMMUNICATION

9.6.4 Summary of Persona Evaluation

The personas developed based on the HyPersona results were evaluated in terms of their ability to be explained through behavioural theory, how they compared to the persona set developed by Scovell et al. [6], [25], and how effectively they could be used to target communication. The persona set was validated by behavioural theory and could effectively and easily be used to target communication in a manner that, in principle, would result in a higher uptake of the desired protective behaviours. The persona set developed also shared significant similarities to the expert created persona set developed by Scovell et al. [6], [25] and had a similar connection to the behavioural models.

As such, SRQ4 was addressed positively as the persona set developed through application of a clustering algorithm was significantly supported by behavioural theory and effectively mimicked the depth and nuance of the expert created persona set. SRQ4 was integral to determining how applicable HyPersona and the findings of this research would be to future applications. As HyPersona and the personas developed would only be useful if the persona development methodology mimicked expert decision making in a manner that resulted in meaningful personas.

9.7 The Primary Research Question

Only once the three secondary research questions were addressed could the primary research question be adequately addressed. The primary research question was:

Can clustering algorithms facilitate the development of deep, nuanced personas based on behavioural models, replicating the decision making of experts, for the automation of persona development?

The secondary research questions identified that the clustering algorithm and parameters selected had a significant impact on the clusters identified and that, while some clustering algorithms did not develop useful cluster sets for persona development, many did. Furthermore, the clustering algorithms were found to give a generally consistent performance across similar data sets. Which indicates that algorithms found to perform well for the current use case, such as SFLA or Kernel k-means, could be inferred to also perform well on similar datasets for a similar purpose. Finally, the personas that were developed from the best performing clustering algorithm-parameter combinations were found to be of the same quality as those created by a behavioural theory and allowed for communication to be effectively targeted.

Based on the answers to the three secondary research questions, the primary research question was able to be positively answered, as yes, clustering algorithms were found to effectively replicate the decision making of experts for persona development. The only caveat was that to identify the best clustering algorithm for the given use case, manual, domain-specific evaluation was required which stops the persona development process from being completely automated. The requirement of domain-specific evaluation was based on the analysis of the top performing algorithm-parameter combinations which confirmed that there can be multiple valid and useful cluster sets present within a data set.

9.8 Summary

The algorithm-parameter combinations with the top overall internal metrics on each data set were evaluated so that the best performer could be identified and used for persona development. During the analysis of the algorithm-parameter combination SRQ3 was positively answered and the importance of domain-specific evaluation was reinforced. The persona sets developed on the 2018 Data Set and the 2021 Data Set were found to be similar enough that they could be treated as a single persona set. The personas were then evaluated based on how well they could be explained through behavioural theory, whether they were of a similar quality to the persona set developed on the same data set by a behavioural expert, and whether the personas could be used to effectively customise communication. Based on the evaluation of the personas SRQ4 could also be addressed, which allowed for the primary research question to be addressed positively. The findings of this research have many implications across the persona development, clustering algorithm, and cyclone damage mitigation fields.

Chapter 10: DISCUSSION AND CONCLUSION

The primary aim of this study was to determine whether machine learning techniques, specifically clustering algorithms, could be employed to facilitate the development of deep, nuanced personas based on behavioural models. Existing approaches developed solutions that lacked the necessary complexity or required a high level of time, resources, and expertise that reduced the feasibility of persona development and maintenance. The motivation of this project was the need for targeted communication to encourage the uptake of protective behaviours, specifically cyclone damage mitigation behaviours.

Previous studies indicate that identifying audience segments based on the behavioural models that attempt to describe the decision processes behind the performances of protective behaviours would allow for the most effective targeting of communication. As such, for a clustering algorithm to facilitate the development of a relevant persona set, the clustering algorithm must mimic the expert decision making that is employed to identify segments based on behavioural theory. There are several benefits to applying clustering algorithms to persona development. Primarily, the benefits are a reduction of resources required to develop a persona set and the fact that larger, more complex data sets can be automatically analysed without the need to set a particular target for segmentation.

Alongside identifying that clustering algorithms are able to mimic expert decision making, there were two main outputs of this research. The first output is the HyPersona framework, a semi-automated framework for the hyperparameter tuning of clustering algorithms and development of personas. The other output was *Average Feature Significance* (AFS), a novel internal evaluation metric specific to the requirements of persona development. A key finding of applying HyPersona with AFS, alongside common internal evaluation metrics, was that hyperparameter tuning, the practice of evaluating the performance of multiple algorithms and parameters for a use case, is required for effective persona development with clustering algorithms.

10.1 Overview of Findings

There were a number of findings as a result of this study, some of which reinforce existing knowledge within the clustering field, primarily:

- Internal metrics alone cannot accurately comment on the quality of a cluster set for a given use case.
- Algorithm performance is dependent on the data set and the use case.

As part of addressing and testing these statements, a novel internal metric, AFS, and a hyperparameter tuning framework, HyPersona were developed. Through analysis of the results of HyPersona several conclusions could be drawn:

- Cluster performance is generally consistent across related data sets.
- The most widely used algorithms do not perform any better than other algorithms.
- Internal metrics can indicate similarities between cluster sets, but cluster sets of similar quality, according to the internal metrics, can significantly differ.

Most importantly, the culmination of the study found that, with domain-specific evaluation, clustering algorithms can mimic expert decision making for persona development.

10.1.1 Findings Surrounding Internal Metrics

Von Luxburg et al. [30] criticised internal evaluation metrics as revealing little about general algorithm performance and claimed that a clustering algorithm cannot be evaluated independently of the desired use case. To determine whether domain-specific evaluation was required, or internal evaluation metrics could be used to automatically identify the best cluster set, the top cluster sets according to the internal metrics were evaluated. The cluster set identified as the best for the current use case was compared to the cluster sets with the best individual internal metrics and the cluster set with the best overall performance metric.

The five cluster sets compared to the selected cluster set were all valid cluster sets that provided some insight into the population. Each of the cluster sets were significantly different from one another and the selected cluster set, demonstrating how each internal metric has different preferences. When considering the use case, however, none of the cluster sets were found to be more useful. For example, the cluster set with the best overall performance metric on the 2021 Data Set identified a subset of the population who rated themselves as likely to install cyclone shutters not out of realistic intention but out of idealisation. The subset offers interesting insight into the population, and with a different use case could be the most applicable. Nonetheless, for the current use case, the idealistic subset does not provide as much use.

Conversely, when evaluating the results as a whole, the cluster sets with poor internal metrics were generally found to be of poorer quality than those with better internal metrics. As, although internal metrics cannot comment on the usefulness of a cluster set, they do comment on quality of the cluster set. Where the quality of a cluster set refers to how distinct and well separated the clusters are. Overlapping or poorly formed clusters are unlikely to be useful.

Based on these findings, the recommendation would be that internal metrics can be used to rule out the overall worst performers, as is done in HyPersona, and potentially narrow down the cluster sets considered. However, internal metrics, whether alone or combined, should not be used to identify the best performing clustering algorithm or the best cluster set.

When the individual internal metrics of two cluster sets were found to be identical, or almost identical, the cluster sets were found to be functionally identical. Where two cluster sets are considered functionally identical if the personas developed from each cluster set would not differ. That is, the differences between the cluster sets would not impact the interpretation of the clusters. However, the similarity of one internal metric or the overall metric performance is not enough to distinguish any similarity between the cluster sets. Specifically, the Euclidian distance between the "ranks" of each of the internal metrics was found to be able to be used to identify whether two cluster sets were functionally identical. However, further research investigating the pattern's consistency across alternate data sets would be recommended before using internal metrics to identify cluster set similarity in a broader or more automated fashion.

Once the internal metrics differ further the cluster sets differ more. The overall performance metric, which gives a single value to represent how the cluster set performed over all the individual evaluation metrics, cannot be used to determine similarity. As discussed, the overall metric performance can indicate cluster quality. However, the overall metric performance alone cannot indicate similarity between cluster sets. As such, multiple cluster sets can significantly differ while having a similar overall cluster quality, reinforcing the importance of domain-specific evaluation.

10.1.2 Findings Surrounding the Biases of Internal Metrics

The analysis of a range of clustering algorithm results, and their performance in terms of various internal metrics supported the idea that internal metrics are bias towards certain types of clustering algorithms and clusters with specific traits. The pre-existing internal metrics used, SC, DBI, and CHI, all had known biases towards convex clusters. The new internal metric proposed in this study, AFS, also has a slight bias towards convex clusters due to being based on the cluster centroids.

The cluster centroid represents the average values of the cluster, without any further details pertaining to the shape and nature of the cluster. Thus, non-convex clusters can be misrepresented by the centroid. As the current use case going to use the centroids, quality persona development is also bias towards convex clusters. As such, the bias of using centroids was acceptable for evaluation with the use case of persona development. However, this is dependent on the nature of the use case.

AFS also had a bias towards algorithms that determined clusters based on an element that effected multiple features. That is features of a cluster that impact or inform multiple other features, for example home ownership, have a bigger impact on the AFS which meant that algorithms which split the clusters based on those features are rewarded for that.

One-hot encoding should also be taken into account when interpreting AFS. As if a feature that is one-hot encoded into five features is statistically different from the mean or other clusters, that may still count as two features being statistically different while the remaining three are not. This means that the choice of whether to one-hot encode an ordinal or interval feature may have a significant impact of the AFS value. However, the choice to one-hot encode a variable will also have an impact on how a clustering algorithm will perform and the values of other internal metrics.

10.1.3 Findings Surrounding Hyperparameter Tuning and Algorithm Selection

As indicated by the importance of domain-specific evaluation, the cluster sets developed from the data set varied significantly based on the algorithm and parameters used. Cluster sets of a similar overall quality developed by the same clustering algorithm could significantly differ due to the parameters used or the influence of random initialisation. The difference between clustering algorithm performances reinforces the importance of hyperparameter tuning.

Existing research in persona development usually lacks any documented process for selecting a clustering algorithm and parameters. As a result, the most accessible or common algorithms are used. However, these algorithms are not necessarily going to give quality results. One key finding of the current research is that to develop quality personas, the clustering algorithm used must be carefully selected. That is, hyperparameter tuning should always be performed prior to applying a clustering algorithm.

To facilitate the hyperparameter tuning process HyPersona was developed. HyPersona attempts to simplify the hyperparameter tuning process using naïve thresholds and graphs and allows the hyperparameter tuning process to lead directly into the persona development process. Multiple parameter combinations should be investigated, and where randomisation plays a large role in the cluster identification process, such as with the meta-heuristic algorithms, running the algorithm with the same parameters multiple times may also be beneficial.

10.1.4 Findings Surrounding Algorithm Performance

The consistency of each algorithm's performance was important, as the algorithm consistency determined the extent to which the results could be generalised. The algorithms were found to perform consistently across the data sets. So, the results and algorithm performances can be generalised across similar datasets and potentially applied to similar domains. However, there was very little consistency between algorithms belonging to the same approach and mixed consistency between parameter sets. As such, the recommendation would be that future problems in a similar domain or using a similar type of data should focus on algorithms that had good or mixed performance across both data sets. While the algorithms that consistently performed poorly would not be recommended, as they would be likely to continue to perform poorly. A range of algorithms and parameters should still be tested.

The algorithms that tended to consistently perform well were k-means, including k-means++ and k-means ensembles, SFLA, and Kernel k-means. SOM and the ensemble with both spectral and k-means also did moderately well, and EMA did moderately well on the 2018 Data Set. The Genetic algorithm also consistently performed the best in terms of AFS, and the Spectral Graph Theory algorithm tended to do well regarding the SC and DBI, however the algorithm's performance on the other metrics, especially AFS, affected the overall performance.

Alternately, DBSCAN and OPTICS both performed very poorly across both data sets. However, this was expected based on the nature of the clusters found by the density approach not aligning well with the requirements of clustering for persona development. Other algorithms that performed consistently poorly included: Affinity Propagation, AHC (especially with any linkage other than Ward's), cure, MST, NMF, ROCK, and SVC. The remaining algorithms, such as ABC, BIRCH, and AHC with Ward's Linkage, and the remaining ensembles, performed more averagely or gave a more inconsistent performance.

As seen by the results, the performance of the most commonly applied algorithms to persona development, AHC, NMF, and k-means, was quite mixed. Both variations of k-means consistently performed well, however the performance did vary during testing due to the differences caused by random initialisation. Furthermore, despite performing well, the cluster sets developed by k-means were not selected as the best. Only AHC with Ward's linkage performed okay, with the other variations being dropped for developing imbalanced clusters. Similarly, only some parameter combinations of NMF performed okay. Most cluster sets developed by NMF were dropped, and the remaining cluster sets had average metrics. One interpretation of the poor performance of NMF and AHC was that there were no clusters present in the data set of the nature targeted by NMF and AHC.

10.1.5 Findings Surrounding the Personas Developed

Most importantly, the personas developed through the application of clustering algorithms using HyPersona were found to be of a similar depth and quality as those developed by a behavioural expert. The personas were found to be directly explainable with behavioural theory. With close similarities to PMT and PADM. The persona set also offered a similar level of depth as those created by Scovell et al. [25], [121].

Primarily, the persona set developed could be used to effectively target communication. Through the use of the relationship between each persona and the behavioural theory, the key barriers stopping each persona from performing the desired mitigation behavioural could be identified. Through targeting these elements communication and incentives can be effectively targeted. As such, the personas were found to be useful and relevant. The key to developing deep and nuanced clusters was identified to be to perform hyperparameter tuning and use domain specific evaluation to identify the best cluster set for the use case.

10.2 Novel Contributions

Beyond the research findings, there were two novel contributions of this research: the AFS internal evaluation metric and the HyPersona framework. Average Feature Significance, AFS, is a novel internal evaluation metric which is based on how statistically significant a set of clusters are. AFS is targeted towards the requirements of persona development as AFS gives a measure of how distinct the clusters developed are from each other and the population, an important factor for developing a set of quality audience segments. However, AFS can be applied to any clustering application. AFS was found to provide useful and unique insights into cluster quality, especially for the purpose of persona development.

The HyPersona framework automatically runs a series of clustering algorithms and parameter combinations on a data set and evaluates each cluster set against a series of metrics. These metrics are used to rule out poor cluster sets and rank the remaining cluster sets. HyPersona also develops a set of graphs and early-stage personas to facilitate meaningful domain-specific evaluation and simplify persona development.

The HyPersona Framework was developed with several targets. First, HyPersona begins to address the problem of hyperparameter tuning for clustering algorithms through the application of a semi-automated methodology. Secondly, the HyPersona framework develops early-stage personas, acting as a semi-automated persona development framework where, once an algorithm is selected, the only manual intervention required is to interpret the early-stage personas and graphs to create the fleshed-out personas. Lastly, the HyPersona framework allowed for all the clustering algorithms and parameters selected to be compared to one another and evaluated.

HyPersona was found to be a useful method for both hyperparameter tuning and persona development. Through use of internal metric threshold, a key feature of HyPersona, was found to be effective at ruling out inadmissible cluster sets. During the final stage of research HyPersona was run on 3,404 algorithm-parameter combinations, and the internal metrics were used to rule out over 2,150 on both data sets, ruling out 63%-65% of the algorithm-parameter combinations. Due to the consistency of the algorithm performances, once a set of algorithms have been identified to perform well on a data set, they can be re-applied to similar data or within a similar domain with HyPersona. Additionally, the thresholds could be made stricter once the metrics of the previous best performers are known to further reduce the evaluation required. As a result, subsequent processes to maintain, update, or expand a persona set are less resource intensive than the initial persona set development process.

10.3 Implications of Findings

The findings of this research have several implications which can impact several fields. These implications primarily affect methodologies for persona development, the application of personas, and fields where personas with a behavioural basis are required. However, there are also implications for the clustering and disaster preparation messaging fields.

The primary finding of the current research was that clustering algorithms can be used to develop deep and nuanced personas, however, hyperparameter tuning and domain-specific evaluation is required. The key implication of this finding is that hyperparameter tuning is required to achieve quality results. Clustering algorithms should not be selected based purely on popularity or accessibility. None of the clustering algorithms that are commonly applied to persona development were found to produce the best results for the current use case. Further, two of the most popular algorithms for persona development, AHC and NMF, consistently performed poorly, with many of the parameter combinations being dropped. The current research suggests that evaluating the performance of a range of clustering algorithms and parameters is essential to produce quality results.

The findings around internal metrics reinforce the idea that internal evaluation metrics should only be used as an approximate guide to cluster quality and cannot be used to identify the best cluster set for a given use case. Domain-specific evaluation was found to be the only method of reliably identifying the best cluster set for the use case. Further, meta-criteria such as cluster count and size were found to be more important factors for ruling out inadmissible cluster sets. Poor metric performance, especially poor performance according to the CHI metric, was strongly correlated to poor performance according to the meta-criteria. The implication of these findings is that internal metrics should not be used as the sole determiners of cluster quality, and relevant meta-criteria for the use case are likely to be more useful.

There are many practical implications for the findings of this research. Primarily, that semi-automatic methodologies for persona development can be applied to domains where complex personas based on behavioural theory are required. The ability to automate, even partially, the persona development process reduces the resources, time, and expertise required to create a persona set. Further, the removal of any requirement for data analysis prior to applying a clustering algorithm means there is no expertise required until the hyperparameter tuning and cluster set evaluation process. As such, larger and more complex data sets can be used, as the data size or complexity is not limited by the abilities of an individual to analyse and comprehend the data set.

Large data sets allow for additional features to be taken into account during persona creation. The additional features can add unique insights. Such features may not have been the most significant for predicting the given behaviour, however, they could be key in identifying one cluster or add important insights into the behaviour which can lead to a better or more nuanced understanding of the persona. By using the entire data set, the semi-automated approach was able to reveal unexpected findings in the data. Notably there were additional insights into the perspectives and behaviours of renters found as a result of including data from renters, home ownership information, and insurance status alongside the features used by Scovell et al. [25], [121].

The inclusion of additional features compared to the persona set developed by Scovell et al. [25], [121] allowed for the HyPersona persona set to contain more insight into the reasoning behind the motivation to perform protective behaviours. The additional information, particularly information around insurance status and topic frequency, provided insight into the personas that allow for messaging to be more effectively targeted.

Further, as the entire data set was used, the personas developed did not particularly target a single behaviour. Instead, the personas describe audience segments which can be used to describe the general attitudes and the related motivation to perform a range of protective behaviours. The difference in purpose allows for the personas to be reused when targeting alternate protective behaviours, rather than requiring the entire persona development process to be repeated. As such, considerable time and expertise is saved.

Beyond the implications for persona development and clustering algorithms, the persona set developed has implications for how communication and incentives encouraging protective behaviours should be customised. One example was that topic frequency, which is how often the individual thinks about or discusses cyclones, differed more greatly between personas than the perception of the costs associated with installing cyclone shutters. Which indicates that to increase the uptake of cyclone shutters, focusing on bringing awareness to the efficacy of cyclone shutters and the importance of mitigating cyclone damage may be more effective than incentives that reduce the cost of cyclone shutters.

10.4 Limitations

There are limitations to this study. Primarily, when comparing the clustering algorithms and determining which algorithms and parameters performed the best, not all algorithms and parameters could be tested. The algorithms were selected to be representative of a wide range of clustering approaches and algorithms, and the parameters were selected to give a comprehensive overview of the parameters available for each algorithm. However, neither the algorithms nor parameters selected were exhaustive. There are numerous clustering algorithms and parameters can be minutely adjusted, as such, an exhaustive comparison is impossible. Further insight may have been gotten if multiple numbers of clusters were tested, however such testing was out of the scope of this study.

Further limitations are that, although the consistency of algorithm performance was thoroughly evaluated, conclusions on how a given algorithm may perform in future for similar use cases or on similar data sets is only inference and performance still depend heavily on the nature of the clusters present within the data. Additionally, the personas developed, and the algorithm-parameter combinations selected are a result of manual evaluation. Although every effort was made to be systematic and objective, there is always the possibility for human error or unconscious bias to affect the results.

The requirement for manual, domain-specific evaluation could not be worked around at this point. Further automation would require a way to codify a cluster set's usefulness for a given task or a metric to determine how well a cluster set aligns with behavioural theory, both of which were out of scope for the current project. Future work would also benefit from testing more varied data sets and potentially larger data sets. Future work may also apply HyPersona to alternate persona development problems and evaluate how the clustering algorithms found to perform well in for the current use case perform across a wider variety of domains and data sets.

10.5 Conclusion

Persona development is a complicated task that currently requires a large amount of manual intervention and is difficult to automate. However, automation is possible and the first requirement to assert that is to determine that the current clustering algorithms can produce cluster sets that lead to personas that are just as deep and nuanced as those created by an expert in the field. To that end, the current project determined whether clustering algorithms can mimic expert decision making.

Personas are required to facilitate the promotion of damage mitigation behaviours in NQ. The NQ region is at particular risk for cyclones, each of which has the potential to cause from millions to billions of dollars in damage. HyPersona, a framework for testing and evaluating a series of clustering algorithms for persona development, and AFS, an internal evaluation metric for clusters, were developed to facilitate the development of a set of clusters for the targeted communication around cyclone damage mitigation strategies in NQ and compare the performance of over 3,000 algorithm-parameter combinations for this purpose.

The clustering algorithms were found to perform consistently and were able to develop deep and nuanced personas that mimicked expert decision making. Thus, reducing the required resources and expertise to create a persona set and producing a repeatable method for developing and updating personas. The findings of this study have implications across a broad range of fields but have direct implications for the customisation of communication around a range of threats and mitigation behaviours, thus facilitating the adoption of risk mitigation strategies that would lead to safer populations that have a lower risk of receiving damage during the next natural disaster.

REFERENCES

[1] D. Smith, C. McShane, A. Swinbourne, and D. Henderson, "Towards effective mitigation strategies for severe wind events," *Australian Journal of Emergency Management*, vol. 31, no. 3, pp. 33–39, Jul. 2016.

[2] Bureau of Meteorology, "What is a Tropical Cyclone?," 2022. http://www.bom.gov.au/cyclone/tropical-cyclone-knowledge-centre/understanding/tc-info/ (accessed Oct. 08, 2021).

[3] Bureau of Meteorology, "Australian tropical cyclone season outlook." http://www.bom.gov.au/climate/cyclones/australia/ (accessed Feb. 02, 2022).

[4] Australian Bureau of Statistics, "National, state and territory population, March 2021," Sep. 16, 2021. https://www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release (accessed Oct. 08, 2021).

[5] Australian Bureau of Statistics, "Story Map Series." https://absstats.maps.arcgis.com/apps/MapSeries/index.html?appid=b2fa123c0032456a8d47fbd0203a3dec (accessed Oct. 08, 2021).

[6] M. Scovell, C. McShane, A. Swinbourne, and D. Smith, "North Queenslanders' perceptions of cyclone risk and structural mitigation intentions. Part I: psychological and demographic factors." Jul. 04, 2018.

[7] Bureau of Meteorology, "Severe Tropical Cyclone Niran." http://www.bom.gov.au/cyclone/history/niran21.shtml (accessed Feb. 02, 2022).

[8] K. J. E. Walsh *et al.*, "Tropical cyclones and climate change," *Wiley Interdisciplinary Reviews: Climate Change*, vol. 7, no. 1, pp. 65–89, 2016, doi: 10.1002/wcc.371.

[9] C. L. Parker, C. L. Bruyère, P. A. Mooney, and A. H. Lynch, "The response of land-falling tropical cyclone characteristics to projected climate change in northeast Australia," *Clim Dyn*, vol. 51, no. 9, pp. 3467–3485, Nov. 2018, doi: 10.1007/s00382-018-4091-9.

[10] Bureau of Meteorology, "Tropical Cyclone Debbie." http://www.bom.gov.au/cyclone/history/debbie17.shtml (accessed Oct. 08, 2021).

[11] Inspector-General Emergency Management, "The Cyclone Debbie Review: Lessons for delivering value and confidence through trust and empowerment." Inspector-General Emergency Management, Aug. 10, 2017. Accessed: Oct. 08, 2021. [Online]. Available: https://documents.parliament.qld.gov.au/tableoffice/tabledpapers/2017/5517t2058.pdf

[12] N. Klein, "Banana drama over: prices crash," *Daily Telegraph*, Oct. 19, 2011. Accessed: Oct. 08, 2021. [Online]. Available: https://www.dailytelegraph.com.au/prices-of-bananas-fall-after-crops-recover-from-cyclone-yasi/news-story/a426ad5906a2d0e86f72578c0629f2d8

[13] Bureau of Meteorology, "Snow, hail, rain and icy winds overnight in Victoria - Bureau of Meteorology Newsroom." https://media.bom.gov.au/releases/829/severe-tropical-cyclone-seroja-final-update/ (accessed Feb. 03, 2022).

[14] T. Logan, "Cyclone Seroja's destruction prompts move to have area north of Perth classified as cyclonic," *ABC News*, May 28, 2021. Accessed: Feb. 03, 2022. [Online]. Available: https://www.abc.net.au/news/2021-05-28/cyclone-seroja-damage-report-suggests-region-classified-cyclonic/100170602

[15] H. McNeill, "Cyclone Seroja damage bill estimated at $200 million with 115 homes destroyed," *WAtoday*, Apr. 15, 2021. https://www.watoday.com.au/national/western-australia/cyclone-seroja-damage-bill-estimated-at-200-million-with-115-homes-destroyed-20210415-p57jly.html (accessed Feb. 03, 2022).

[16] G. R. Q. Queensland State Government, "House Maintenance and Preparation." https://www.getready.qld.gov.au/get-prepared/house-maintenance-and-preparation#cyclones-and-storms-and-flooding (accessed Apr. 12, 2022).

[17] Queensland State Government, "Prepare your home | Preparing for disasters." https://www.qld.gov.au/emergency/dealing-disasters/prepare-for-disasters/prepare-home (accessed Apr. 12, 2022).

[18] CEO, Queensland Reconstruction Authority, "Cyclone Resilient Building Guidance for Qld Homes." Queensland Reconstruction Authority, Dec. 2019. [Online]. Available: https://www.qra.qld.gov.au/sites/default/files/2019-12/0321%20Cyclone%20Resilient%20Building%20Guidance%20for%20Qld%20Homes%20FINAL.pdf

[19] D. Smith and D. Henderson, "Insurance Claims Data Analysis for Cyclones Yasi and Larry," p. 53.

[20] M. Scovell, C. McShane, A. Swinbourne, and D. Smith, "Investigating factors that influence cyclone mitigation behaviour: a pilot study," presented at the 6th Australian and New Zealand Disaster & Emergency Management Conference, Gold Coast, QLD, Australia, May 2017. Accessed: Apr. 12, 2019. [Online]. Available: https://researchonline.jcu.edu.au/49758/

[21] S. Milne, P. Sheeran, and S. Orbell, "Prediction and Intervention in Health-Related Behavior: A Meta-Analytic Review of Protection Motivation Theory," *Journal of Applied Social Psychology*, vol. 30, no. 1, pp. 106–143, 2000, doi: 10.1111/j.1559-1816.2000.tb02308.x.

[22] U. Trautwein, H. W. Marsh, B. Nagengast, O. Lüdtke, G. Nagy, and K. Jonkmann, "Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study.," *Journal of Educational Psychology*, vol. 104, no. 3, pp. 763–777, 2012, doi: 10.1037/a0027470.

[23] J. Salminen, B. J. Jansen, J. An, H. Kwak, and S. Jung, "Are personas done? Evaluating their usefulness in the age of digital analytics," *Persona Studies*, vol. 4, no. 2, pp. 47–65, Nov. 2018, doi: 10.21153/psj2018vol4no2art737.

[24] V. Thoma and B. Williams, “Developing and validating personas in e-commerce: A heuristic approach,” in *Human-Computer Interaction – INTERACT 2009*, vol. 5727, T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, and M. Winckler, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 524–527. doi: 10.1007/978-3-642-03658-3_56.

[25] M. Scovell, C. McShane, A. Swinbourne, and D. Smith, “North Queenslanders’ perceptions of cyclone risk and structural mitigation intentions. Part II: cluster analysis and personas.” Oct. 10, 2018.

[26] M. Mesgari, C. Okoli, and A. O. de Guinea, “Affordance-based user personas : A mixed-method approach to persona development,” in *AMCIS 2015 Proceedings*, Puerto Rico, Jun. 2015. [Online]. Available: https://aisel.aisnet.org/amcis2015/HCI/GeneralPresentations/1

[27] J. Brickey, S. Walczak, and T. Burgess, “Comparing semi-automated clustering methods for persona development,” *IEEE Transactions on Software Engineering*, vol. 38, no. 3, pp. 537–546, May 2012, doi: 10.1109/TSE.2011.60.

[28] J. Salminen, K. Guan, S.-G. Jung, and B. J. Jansen, “A Survey of 15 Years of Data-Driven Persona Development,” *International Journal of Human–Computer Interaction*, vol. 0, no. 0, pp. 1–24, Apr. 2021, doi: 10.1080/10447318.2021.1908670.

[29] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognition Letters*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.

[30] U. Von Luxburg, R. C. Williamson, and I. Guyon, “Clustering: Science or art?,” presented at the Proceedings of ICML Workshop on Unsupervised and Transfer Learning, 2012, pp. 65–79.

[31] M. Eusuff, K. Lansey, and F. Pasha, “Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization,” *Engineering optimization*, vol. 38, no. 2, pp. 129–154, 2006.

[32] B. Amiri, M. Fathian, and A. Maroosi, “Application of shuffled frog-leaping algorithm on clustering,” *The International Journal of Advanced Manufacturing Technology*, vol. 45, no. 1–2, pp. 199–209, 2009.

[33] S. Grier and C. A. Bryant, “Social Marketing in Public Health,” *Annual Review of Public Health*, vol. 26, no. 1, pp. 319–339, 2005, doi: 10.1146/annurev.publhealth.26.021304.144610.

[34] M. L. Rothschild, “The Road Crew Final Report,” 2003.

[35] S. C. Moser, “Communicating climate change: history, challenges, process and future directions,” *WIREs Climate Change*, vol. 1, no. 1, pp. 31–53, 2010, doi: https://doi.org/10.1002/wcc.11.

[36] P. Bubeck, W. J. W. Botzen, and J. C. J. H. Aerts, “A Review of Risk Perceptions and Other Factors that Influence Flood Mitigation Behavior,” *Risk Analysis*, vol. 32, no. 9, pp. 1481–1495, 2012, doi: 10.1111/j.1539-6924.2011.01783.x.

[37] Department of Transport and Main Roads, Queensland Government, “Speeding fines and demerit points,” Nov. 02, 2020. https://www.tmr.qld.gov.au/Safety/Driver-guide/Speeding/Speeding-fines-and-demerit-points (accessed Feb. 18, 2021).

[38] Department of Transport and Main Roads, Queensland Governement, “Anti-speeding public education 2010/2011: Slow down stupid.” Accessed: Feb. 18, 2021. [Online]. Available: http://www.tmr.qld.gov.au/~/media/Safety/safetycampaigns/Anti%20speeding/Anti_speeding_campaign_summary.pdf

[39] N. D. Weinstein, “Testing four competing theories of health-protective behavior.,” *Health psychology*, vol. 12, no. 4, p. 324, 1993.

[40] R. W. Rogers, “A protection motivation theory of fear appeals and attitude change,” *The Journal of Psychology*, vol. 91, no. 1, pp. 93–114, Sep. 1975, doi: 10.1080/00223980.1975.9915803.

[41] T. Grothmann and F. Reusswig, “People at risk of flooding: why some residents take precautionary action while others do not,” *Natural hazards*, vol. 38, no. 1–2, pp. 101–120, 2006.

[42] M. K. Lindell and R. W. Perry, *Behavioral foundations of community emergency planning*. Washington, DC, US: Hemisphere Publishing Corp, 1992.

[43] M. K. Lindell and R. W. Perry, “The protective action decision model: theoretical modifications and additional evidence,” *Risk Analysis: An International Journal*, vol. 32, no. 4, pp. 616–632, 2012.

[44] T. Terpstra and M. K. Lindell, “Citizens’ perceptions of flood hazard adjustments: an application of the protective action decision model,” *Environment and Behavior*, vol. 45, no. 8, pp. 993–1018, 2013.

[45] M. Scovell, C. McShane, A. Swinbourne, and D. Smith, “How fringe cyclone experience affects predictions of damage severity,” *Disaster Prevention and Management: An International Journal*, vol. ahead-of-print, no. ahead-of-print, Jan. 2020, doi: 10.1108/DPM-07-2019-0228.

[46] G. Wachinger, O. Renn, C. Begg, and C. Kuhlicke, “The Risk Perception Paradox—Implications for Governance and Communication of Natural Hazards,” *Risk Analysis*, vol. 33, no. 6, pp. 1049–1065, 2013, doi: https://doi.org/10.1111/j.1539-6924.2012.01942.x.

[47] H. Kunreuther *et al.*, “High Stakes Decision Making: Normative, Descriptive and Prescriptive Considerations,” *Marketing Letters*, vol. 13, no. 3, pp. 259–268, Aug. 2002, doi: 10.1023/A:1020287225409.

[48] Lene Nielsen and Kira Storgaard Hansen, “Personas is applicable: a study on the use of personas in Denmark,” presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada, 2014, pp. 1665–1674. doi: 10.1145/2556288.2557080.

[49] J. Billestrup, J. Stage, L. Nielsen, and K. S. Hansen, “Persona usage in software development: advantages and obstacles,” *Proc. of ACHI*, pp. 359–364, 2014.

[50] C. So and J. Joo, “Does a Persona Improve Creativity?,” *The Design Journal*, vol. 20, no. 4, pp. 459–475, Jul. 2017, doi: 10.1080/14606925.2017.1319672.

[51] T. Miaskiewicz and K. A. Kozar, “Personas and user-centered design: How can personas benefit product design processes?,” *Design Studies*, vol. 32, no. 5, pp. 417–430, Sep. 2011, doi: https://doi.org/10.1016/j.destud.2011.03.003.

[52] John Pruitt and Jonathan Grudin, “Personas: practice and theory,” presented at the Proceedings of the 2003 conference on Designing for user experiences, San Francisco, California, 2003, pp. 1–15. doi: 10.1145/997078.997089.

[53] T. Matthews, T. Judge, and S. Whittaker, “How do designers and user experience professionals actually perceive and use personas?,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: Association for Computing Machinery, 2012, pp. 1219–1228. Accessed: Aug. 19, 2021. [Online]. Available: https://doi.org/10.1145/2207676.2208573

[54] Joni Salminen, Lene Nielsen, Soon-Gyo Jung, Jisun An, Haewoon Kwak, and Bernard J. Jansen, “‘Is More Better?’: Impact of Multiple Photos on Perception of Persona Profiles,” presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada, 2018, pp. 1–13. doi: 10.1145/3173574.3173891.

[55] Charles G. Hill *et al.*, “Gender-inclusiveness personas vs. stereotyping: Can we have it both ways?,” presented at the Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, Colorado, USA, 2017. doi: 10.1145/3025453.3025609.

[56] F. Anvari, D. Richards, M. Hitchens, M. A. Babar, H. M. T. Tran, and P. Busch, “An empirical investigation of the influence of persona with personality traits on conceptual design,” *Journal of Systems and Software*, vol. 134, pp. 324–339, Dec. 2017, doi: https://doi.org/10.1016/j.jss.2017.09.020.

[57] J. Salminen, S. Jung, A. M. S. Kamel, J. M. Santos, and B. J. Jansen, “Using artificially generated pictures in customer-facing systems: an evaluation study with data-driven personas,” *Behaviour & Information Technology*, vol. 0, no. 0, pp. 1–17, Nov. 2020, doi: 10.1080/0144929X.2020.1838610.

[58] J. Salminen, S. Şengün, S.-G. Jung, and J. Jansen, *Design Issues in Automatically Generated Persona Profiles: A Qualitative Analysis from 38 Think-Aloud Transcripts*. 2019, p. 229. doi: 10.1145/3295750.3298942.

[59] J. L. Ward, “Persona development and use, or, how to make imaginary people work for you,” 2010.

[60] G. Terlouw, J. T. B. van’t Veer, D. A. Kuipers, and J. Metselaar, “Context analysis, needs assessment and persona development: towards a digital game-like intervention for high functioning children with ASD to train social skills,” *Early Child Development and Care*, pp. 1–16, 2018, doi: 10.1080/03004430.2018.1555826.

[61] I. Hirskyj-Douglas, J. Read, and M. Horton, “Animal personas: Representing dog stakeholders in interaction design,” presented at the HCI 2017 - Digital make-believe, 2017, pp. 1–13. doi: 10.14236/ewic/HCI2017.37.

[62] Rebecca M. Quintana, Stephanie R. Haley, Adam Levick, Caitlin Holman, Ben Hayward, and Mike Wojan, “The persona party: Using personas to design for learning at scale,” presented at the Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, Colorado, USA, 2017, pp. 933–941. doi: 10.1145/3027063.3053355.

[63] Bernhard Wöckl, Ulcay Yildizoglu, Isabella Buber, Belinda Aparicio Diaz, Ernst Kruijff, and Manfred Tscheligi, “Basic senior personas: a representative design tool covering the spectrum of European older adults,” presented at the Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility, Boulder, Colorado, USA, 2012, pp. 25–32. doi: 10.1145/2384916.2384922.

[64] J. An, H. Cho, H. Kwak, M. Z. Hassen, and B. J. Jansen, “Towards automatic persona generation using social media,” in *2016 IEEE 4th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)*, Aug. 2016, pp. 206–211. doi: 10.1109/W-FiCloud.2016.51.

[65] J. An, H. Kwak, S. Jung, J. Salminen, and B. J. Jansen, “Customer segmentation using online platforms: isolating behavioral and demographic segments for persona creation via aggregated user data,” *Social Network Analysis and Mining*, vol. 8, no. 1, p. 54, 2018.

[66] Soon-gyo Jung, Joni Salminen, Haewoon Kwak, Jisun An, and Bernard J. Jansen, “Automatic Persona Generation (APG): A rationale and demonstration,” presented at the Proceedings of the 2018 Conference on Human Information Interaction & Retrieval, New Brunswick, NJ, USA, 2018. doi: 10.1145/3176349.3176893.

[67] Tomasz Miaskiewicz, Tamara Sumner, and Kenneth A. Kozar, “A latent semantic analysis methodology for the identification and creation of personas,” presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Florence, Italy, 2008, pp. 1501–1510. doi: 10.1145/1357054.1357290.

[68] Rashmi Sinha, “Persona development for information-rich domains,” presented at the CHI ’03 Extended Abstracts on Human Factors in Computing Systems, Ft. Lauderdale, Florida, USA, 2003. doi: 10.1145/765891.766017.

[69] Janna Lynn Dupree, Richard Devries, Daniel M. Berry, and Edward Lank, “Privacy personas: clustering users via attitudes and behaviors toward security practices,” presented at the Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, San Jose, California, USA, 2016, pp. 5228–5239. doi: 10.1145/2858036.2858214.

[70] N. Tu, X. Dong, P. P. Rau, and T. Zhang, “Using cluster analysis in persona development,” in *2010 8th International Conference on Supply Chain Management and Information*, Oct. 2010, pp. 1–5.

[71] J. Salminen, K. Guan, S.-G. Jung, S. A. Chowdhury, and B. J. Jansen, “A Literature Review of Quantitative Persona Creation,” in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Honolulu HI USA, Apr. 2020, pp. 1–14. doi: 10.1145/3313831.3376502.

[72] J. (Jen) McGinn and N. Kotamraju, “Data-driven persona development,” in *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI ’08*, Florence, Italy, 2008, p. 1521. doi: 10.1145/1357054.1357292.

[73] Joni Salminen *et al.*, “Persona perception scale: Developing and validating an instrument for human-like representations of data,” presented at the Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal QC, Canada, 2018, pp. 1–6. doi: 10.1145/3170427.3188461.

[74] Z. Zhao and H. Liu, *Spectral Feature Selection for Supervised and Unsupervised Learning*. 2007.

[75] D. Xu and Y. Tian, “A comprehensive survey of clustering algorithms,” *Ann. Data. Sci.*, vol. 2, no. 2, pp. 165–193, Jun. 2015, doi: 10.1007/s40745-015-0040-1.

[76] A. Saxena *et al.*, “A review of clustering techniques and developments,” *Neurocomputing*, vol. 267, pp. 664–681, Dec. 2017, doi: 10.1016/j.neucom.2017.06.053.

[77] Rui Xu and D. Wunsch, “Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, May 2005, doi: 10.1109/TNN.2005.845141.

[78] scikit-learn, *Comparison of Agglomerative Clustering Linkage Strategies*.

[79] G. J. Szekely and M. L. Rizzo, “Hierarchical Clustering via Joint Between-Within Distances: Extending Ward’s Minimum Variance Method,” *Journal of Classification*, vol. 22, no. 2, pp. 151–183, Sep. 2005, doi: 10.1007/s00357-005-0012-9.

[80] T. Zhang, R. Ramakrishnan, and M. Livny, “BIRCH: An Efficient Data Clustering Method for Very Large Databases,” in *Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 1996, pp. 103–114. doi: 10.1145/233269.233324.

[81] S. Guha, R. Rastogi, and K. Shim, “CURE: An Efficient Clustering Algorithm for Large Databases,” in *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 1998, pp. 73–84. doi: 10.1145/276304.276312.

[82] S. Guha, R. Rastogi, and K. Shim, “Rock: A robust clustering algorithm for categorical attributes,” *Information Systems*, vol. 25, no. 5, pp. 345–366, Jul. 2000, doi: 10.1016/S0306-4379(00)00022-3.

[83] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering: A Review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999, doi: 10.1145/331499.331504.

[84] Jianbo Shi and J. Malik, “Normalized cuts and image segmentation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000, doi: 10.1109/34.868688.

[85] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On Spectral Clustering: Analysis and an algorithm,” in *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, S. Becker, and Z.

Ghahramani, Eds. MIT Press, 2002, pp. 849–856. Accessed: May 31, 2019. [Online]. Available: http://papers.nips.cc/paper/2092-on-spectral-clustering-analysis-and-an-algorithm.pdf

[86] U. von Luxburg, "A tutorial on spectral clustering," *Stat Comput*, vol. 17, no. 4, pp. 395–416, Dec. 2007, doi: 10.1007/s11222-007-9033-z.

[87] H. Steinhaus, "Sur la division des corp materiels en parties," *Bull. Acad. Polon. Sci*, vol. 1, no. 804, p. 801, 1956.

[88] E. W. Forgy, "Cluster analysis of multivariate data: efficiency versus interpretability of classifications," *biometrics*, vol. 21, pp. 768–769, 1965.

[89] G. H. Ball and D. J. Hall, "ISODATA, a novel method of data analysis and pattern classification," Stanford research inst Menlo Park CA, 1965.

[90] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967, vol. 1, pp. 281–297.

[91] S. Lloyd, "Least squares quantization in PCM," *IEEE transactions on information theory*, vol. 28, no. 2, pp. 129–137, 1982.

[92] D. Arthur and S. Vassilvitskii, "k-means++: The advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, 2007, pp. 1027–1035.

[93] M. Ester, H.-P. Kriegel, and X. Xu, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," vol. 96, p. 6, 1996.

[94] M. Ankerst, M. M. Breunig, H.-P. Kriegel, and J. Sander, "OPTICS: Ordering Points to Identify the Clustering Structure," in *Proceedings of the 1999 ACM SIGMOD International Conference on Management of Data*, New York, NY, USA, 1999, pp. 49–60. doi: 10.1145/304182.304187.

[95] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[96] A. Ben-Hur, D. Horn, H. T. Siegelmann, and V. Vapnik, "Support vector clustering," *Journal of machine learning research*, vol. 2, no. Dec, pp. 125–137, 2001.

[97] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, vol. 382. John Wiley & Sons, 2007.

[98] C. B. Do and S. Batzoglou, "What is the expectation maximization algorithm?," *Nature Biotechnology*, vol. 26, no. 8, pp. 897–899, 2008, doi: 10.1038/nbt1406.

[99] T. Kohonen, "The self-organizing map," *Proceedings of the IEEE*, vol. 78, no. 9, pp. 1464–1480, Sep. 1990, doi: 10.1109/5.58325.

[100] S. J. Nanda and G. Panda, "A survey on nature inspired metaheuristic algorithms for partitional clustering," *Swarm and Evolutionary Computation*, vol. 16, pp. 1–18, Jun. 2014, doi: 10.1016/j.swevo.2013.11.003.

[101] A. José-García and W. Gómez-Flores, "Automatic clustering using nature-inspired metaheuristics: A survey," *Applied Soft Computing*, vol. 41, pp. 192–213, Apr. 2016, doi: 10.1016/j.asoc.2015.12.001.

[102] U. Maulik and S. Bandyopadhyay, "Genetic algorithm-based clustering technique," *Pattern Recognition*, vol. 33, no. 9, pp. 1455–1465, Sep. 2000, doi: 10.1016/S0031-3203(99)00137-5.

[103] B. Basturk, "An artificial bee colony (ABC) algorithm for numeric function optimization," in *IEEE Swarm Intelligence Symposium, Indianapolis, IN, USA, 2006*, 2006.

[104] D. Karaboga and C. Ozturk, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," *Applied soft computing*, vol. 11, no. 1, pp. 652–657, 2011.

[105] M. Fathian, B. Amiri, and A. Maroosi, "Application of honey-bee mating optimization algorithm on clustering," *Applied Mathematics and Computation*, vol. 190, no. 2, pp. 1502–1513, 2007.

[106] T. Boongoen and N. Iam-On, "Cluster ensembles: A survey of approaches with recent extensions and applications," *Computer Science Review*, vol. 28, pp. 1–25, May 2018, doi: 10.1016/j.cosrev.2018.01.003.

[107] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, 2007.

[108] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[109] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, Nov. 1987, doi: 10.1016/0377-0427(87)90125-7.

[110] T. Caliński and J. Harabasz, "A Dendrite Method for Cluster Analysis," *Communications in Statistics - Theory and Methods*, vol. 3, pp. 1–27, Jan. 1974, doi: 10.1080/03610927408827101.

[111] D. L. Davies and D. W. Bouldin, "A Cluster Separation Measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 2, pp. 224–227, Apr. 1979, doi: 10.1109/TPAMI.1979.4766909.

[112] I. Färber *et al.*, "On using class-labels in evaluation of clusterings," presented at the MultiClust: 1st international workshop on discovering, summarizing and using multiple clusterings held in conjunction with KDD, 2010, p. 1.

[113] X. Fan, Y. Yue, P. Sarkar, and Y. X. R. Wang, "On hyperparameter tuning in general clustering problems," in *Proceedings of the 37th International Conference on Machine Learning*, Jul. 2020, vol. 119, pp. 2996–3007. [Online]. Available: http://proceedings.mlr.press/v119/fan20b.html

[114] T. Van Craenendonck and H. Blockeel, "Constraint-based clustering selection," *Mach Learn*, vol. 106, no. 9, pp. 1497–1521, Oct. 2017, doi: 10.1007/s10994-017-5643-7.

[115] L. Blumenberg and K. V. Ruggles, "Hypercluster: a flexible tool for parallelized unsupervised clustering optimization," *BMC Bioinformatics*, vol. 21, no. 1, p. 428, Sep. 2020, doi: 10.1186/s12859-020-03774-1.

[116] V. Shalamov, V. Efimova, S. Muravyov, and A. Filchenkov, “Reinforcement-based Method for Simultaneous Clustering Algorithm Selection and its Hyperparameters Optimization,” *Procedia Computer Science*, vol. 136, pp. 144–153, Jan. 2018, doi: 10.1016/j.procs.2018.08.247.

[117] L. L. Minku, “A novel online supervised hyperparameter tuning procedure applied to cross-company software effort estimation,” *Empir Software Eng*, vol. 24, no. 5, pp. 3153–3204, Oct. 2019, doi: 10.1007/s10664-019-09686-w.

[118] E. Ditton, A. Swinbourne, T. Myers, and M. Scovell, “Applying Semi-Automated Hyperparameter Tuning for Clustering Algorithms,” *arXiv:2108.11053 [cs]*, Aug. 2021, Accessed: Apr. 03, 2022. [Online]. Available: http://arxiv.org/abs/2108.11053

[119] E. Ditton, A. Swinbourne, and T. Myers, “Selecting a clustering algorithm: A semi-automated hyperparameter tuning framework for effective persona development,” *Array*, vol. 14, p. 100186, Jul. 2022, doi: 10.1016/j.array.2022.100186.

[120] E. Ditton (Forest), “HyPersona.” Jun. 21, 2022. Accessed: Jul. 20, 2022. [Online]. Available: https://github.com/ElizabethForest/HyPersona

[121] M. Scovell, C. McShane, D. Smith, and A. Swinbourne, “Personalising the message: promoting cyclone protection in North Queensland,” *Australian Journal of Emergency Management*, vol. 34, no. 4, pp. 48–53, Oct. 2019.

[122] J. L. Doermann, E. D. Kuligowski, and J. Milke, “From Social Science Research to Engineering Practice: Development of a Short Message Creation Tool for Wildfire Emergencies,” *Fire Technol*, vol. 57, no. 2, pp. 815–837, Mar. 2021, doi: 10.1007/s10694-020-01008-7.

[123] M. Scovell, C. McShane, A. Swinbourne, and D. Smith, “Applying the Protective Action Decision Model to Explain Cyclone Shutter Installation Behavior,” *Natural Hazards Review*, vol. 22, no. 1, p. 04020043, Feb. 2021, doi: 10.1061/(ASCE)NH.1527-6996.0000417.

[124] The pandas development team, “pandas-dev/pandas: Pandas.” pandas, Feb. 2020. Accessed: May 23, 2022. [Online]. Available: https://doi.org/10.5281/zenodo.3509134

[125] C. R. Harris *et al.*, “Array programming with NumPy,” *Nature*, vol. 585, no. 7825, Art. no. 7825, Sep. 2020, doi: 10.1038/s41586-020-2649-2.

[126] P. Virtanen *et al.*, “SciPy 1.0: fundamental algorithms for scientific computing in Python,” *Nat Methods*, vol. 17, no. 3, pp. 261–272, Mar. 2020, doi: 10.1038/s41592-019-0686-2.

[127] J. D. Hunter, “Matplotlib: A 2D Graphics Environment,” *Computing in Science Engineering*, vol. 9, no. 3, pp. 90–95, May 2007, doi: 10.1109/MCSE.2007.55.

[128] F. Pedregosa *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011.

[129] A. Novikov, “PyClustering: Data Mining Library,” *Journal of Open Source Software*, vol. 4, no. 36, p. 1230, Apr. 2019, doi: 10.21105/joss.01230.

[130] A. Cichocki and A.-H. Phan, “Fast local algorithms for large scale nonnegative matrix and tensor factorizations,” *IEICE transactions on fundamentals of electronics, communications and computer sciences*, vol. 92, no. 3, pp. 708–721, 2009.

[131] C. Févotte and J. Idier, “Algorithms for nonnegative matrix factorization with the β-divergence,” *Neural computation*, vol. 23, no. 9, pp. 2421–2456, 2011, doi: https://doi.org/10.1162/NECO_a_00168.

[132] S. van Buuren and K. Groothuis-Oudshoorn, “mice: Multivariate Imputation by Chained Equations in R,” *Journal of Statistical Software*, vol. 45, no. 1, Art. no. 1, Dec. 2011, doi: 10.18637/jss.v045.i03.

[133] E. Ditton (Forest), “Metaheuristic Clustering.” Sep. 08, 2021. Accessed: Apr. 11, 2022. [Online]. Available: https://github.com/ElizabethForest/metaheuristic_clustering

[134] J. VanderPlas, “mst_clustering: Clustering via Euclidean Minimum Spanning Trees,” *Journal of Open Source Software*, vol. 1, no. 1, p. 12, May 2016, doi: 10.21105/joss.00012.

[135] M. Blondel (mblondel), “Kernel K-means.” Nov. 14, 2013. Accessed: Mar. 10, 2022. [Online]. Available: https://gist.github.com/mblondel/6230787

[136] I. S. Dhillon, Y. Guan, and B. Kulis, “Kernel k-means: spectral clustering and normalized cuts,” in *Proceedings of the 2004 ACM SIGKDD international conference on Knowledge discovery and data mining - KDD ’04*, Seattle, WA, USA, 2004, p. 551. doi: 10.1145/1014052.1014118.

[137] T. Pham, T. Le, and H. Dang, “Scalable Support Vector Clustering Using Budget,” p. 23.

[138] A. P. Topchy, M. H. C. Law, A. K. Jain, and A. L. Fred, “Analysis of consensus partition in cluster ensemble,” in *Fourth IEEE International Conference on Data Mining (ICDM’04)*, Nov. 2004, pp. 225–232. doi: 10.1109/ICDM.2004.10100.

[139] T. Sano, “ClusterEnsembles.” Aug. 2021. Accessed: Mar. 10, 2022. [Online]. Available: https://github.com/tsano430/ClusterEnsembles

[140] T. Li, C. Ding, and M. I. Jordan, “Solving Consensus and Semi-supervised Clustering Problems Using Nonnegative Matrix Factorization,” in *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, Oct. 2007, pp. 577–582. doi: 10.1109/ICDM.2007.98.

APPENDIX A: SURVEY INFORMATION SHEET

JAMES COOK UNIVERSITY AUSTRALIA

INFORMATION SHEET

PROJECT TITLE: **Investigation of Cyclone Preparatory Behaviour**

Hi, my name is Elizabeth Forest and I am a postgraduate research student at James Cook University. I am currently looking for participants to take part in a study investigating factors that influence the performance behaviours aimed at reducing cyclone damage and would like to invite you to participate. This study is being conducted with Dr Anne Swinbourne and Prof Trina Myers and will contribute to my PhD thesis at James Cook University.

Participation is open to people over the age of 18 years who currently live in coastal North Queensland (between Rockhampton and Bamaga). If you choose to participate, you will be asked to fill out an online questionnaire. The questionnaire will ask you about your perception of cyclones, your perceptions of specific cyclone damage preparation behaviours, and your previous experience with cyclones and performing preparation behaviours. You will also be asked for information about yourself. This does not include your name. We are interested in what sort of people do what sort of behaviours. We don't need to know names. The questionnaire should take approximately 15 minutes.

Participation in the questionnaire is completely voluntary and you can stop taking part in the study at any time without explanation or prejudice. Once you start the questionnaire you can choose to stop taking part at any point by closing your internet browser window. The responses are entirely anonymous, and you will be unidentifiable in any reports or publications based on this data.

The survey is available at https://jcuchs.qualtrics.com/jfe/form/SV_8CY0Ypi50sXqV4V

You can also get access to the survey and further information through our public Facebook page, Cyclone Preparation in 2021 @CyclonePreparation2021. You do not need a Facebook account to access the page. Preliminary results of the study will be made available through the Facebook page during May 2021.

If you know anybody who would be interested in participating, please let them know and direct them to this letter or the Facebook page, Cyclone Preparation in 2021 @CyclonePreparation2021, for more information.

Thank you for your interest in completing this study and contributing to my research! Your time is greatly appreciated. If you have any questions or concerns, please do not hesitate to contact me.

Principal Investigator:
Elizabeth Forest
College of Science and Engineering
James Cook University
Elizabeth.Forest@jcu.edu.au

Supervisor:
Professor Trina Myers
Science and Engineering Faculty
Queensland University of Technology
Trina.Myers@qut.edu.au

Supervisor:
Dr Anne Swinbourne
College of Healthcare Sciences
James Cook University
Phone: 07 4781 4809
Anne.Swinbourne@jcu.edu.au

If you have any concerns regarding the ethical conduct of the study, please contact:
Human Ethics, Research Office
James Cook University, Townsville, Qld, 4811
Phone: (07) 4781 5011 (ethics@jcu.edu.au)

APPENDIX B: SURVEY

By continuing to the questionnaire, you declare that you understand the research aim, consent to participate in the questionnaire, and allow the data collected to be used in the manner described above.

- O I do not consent. Exit the survey
- O I consent. Proceed to the survey.

This section will ask about some general information about yourself.

What is your age in years?

____

Sex

- O Male
- O Female
- O Other
- O Prefer Not to Say

Where in the North Queensland Region do you live? If you do not live in any of the listed zones, please enter your post code.

- O Cairns and Surrounds (North of Ingham)
- O Townsville and Surrounds (Ingham to Ayr)
- O Whitsunday Region (south of Ayr and north of Mackay)
- O Mackay and Surrounds (to St. Lawrence)
- O Rockhampton and Surrounds
- O Other ____

How many years have you lived in your current city/area?

____

How many years have you lived in the NQ area?

____

How many dependent children do you have?

____

What is your current marital status?

- O Single
- O Partnered
- O Married
- O Divorced
- O Separated
- O Widowed

This section will ask you about your previous cyclone experience

Have you ever experienced a cyclone before?

- O No
- O Yes

How many cyclones have you experienced?

- O 1
- O 2
- O 3
- O 4
- O 5+

Did you experience Cyclone Yasi or Cyclone Debbie?

- O I experienced both Cyclone Yasi and Cyclone Debbie
- O Just Cyclone Yasi
- O Just Cyclone Debbie
- O Neither

Did you play a role in your household's preparation for any of the cyclones you have experienced?

- O Yes
- O No

Has your property received damage from previous cyclone?

- O Yes
- O No

How would you rate the amount of damage your property has received from previous cyclones?

	Minimal Damage	Minor Damage	Moderate Damage	Severe Damage	Extensive Damage
From all previous cyclones	O	O	O	O	O
From Cyclone Yasi (if applicable)	O	O	O	O	O
From Cyclone Debbie (if applicable)	O	O	O	O	O

When thinking about a cyclone you have experienced, do you remember feeling any of the following feelings? Rate your level of these feelings.

	None	Low	Moderate	High
Stressed	O	O	O	O
Fearful	O	O	O	O
Helpless	O	O	O	O
Depressed	O	O	O	O
Dread	O	O	O	O

Based on what you can remember about Cyclone Yasi, what was the highest category winds that your city experienced due to Cyclone Yasi based on the below descriptions of each category?

O Category 5

O Category 4

O Category 3

O Category 2

O Category 1

People have different kinds of *emotional responses* to the threat of a cyclone. In thinking about the possibility of your location being hit by a major cyclone with the potential for widespread damage, how strongly would you agree or disagree with the following statements?

Thinking about a cyclone makes me feel...

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
Fearful	O	O	O	O	O	O	O
Worried	O	O	O	O	O	O	O
Depressed	O	O	O	O	O	O	O
Helpless	O	O	O	O	O	O	O
Dread	O	O	O	O	O	O	O

People *understand* cyclones in different ways. In thinking about the nature of cyclones generally, how strongly would you agree or disagree with the following?

I think that cyclones...

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
may cause catastrophic destruction	O	O	O	O	O	O	O
may cause widespread death	O	O	O	O	O	O	O
pose great financial threat	O	O	O	O	O	O	O
pose a threat to future generations	O	O	O	O	O	O	O

If a cyclone was to occur in your area, how likely would it be that each of the following would occur?

	Extremely unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Extremely likely
Your property is damaged	O	O	O	O	O	O	O
Your, or a member of your household's, daily life is disturbed	O	O	O	O	O	O	O
You, or a member of your household, are prevented from going to work or doing their job	O	O	O	O	O	O	O
Your, or a member of your household's, mental health is negatively affected	O	O	O	O	O	O	O
Your, or a member of your household's, physical health is negatively affected	O	O	O	O	O	O	O

How strongly do you agree with the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
I am knowledgeable about cyclone risks (to be able to make informed preparation decisions)	O	O	O	O	O	O	O
I am knowledgeable about the types of property damage that can be caused by a cyclone	O	O	O	O	O	O	O
I am knowledgeable about what I can do to reduce cyclone related property damage	O	O	O	O	O	O	O
I think about the potential negative effects from cyclones regularly	O	O	O	O	O	O	O
Cyclone related issues are discussed regularly in my household	O	O	O	O	O	O	O

How likely do you believe it is that in the next 5 years your region will experience a...

	Extremely unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Extremely likely
Category 1 (or above) cyclone	O	O	O	O	O	O	O
Category 3 (or above) cyclone	O	O	O	O	O	O	O
Category 5 cyclone	O	O	O	O	O	O	O

If the following category cyclone were to occur next week, what level of property damage would you expect to receive? (Assume you will perform your usual amount of preparation)

	Very Low	Low	Somewhat Low	Medium	Somewhat high	High	Very High
Category 1-2 cyclone	O	O	O	O	O	O	O
Category 3-4 cyclone	O	O	O	O	O	O	O
Category 5 cyclone	O	O	O	O	O	O	O

If a cyclone were to occur in the next 5 years, what do you believe is the likelihood that the government would provide financial assistance to homeowners who have received property damage?

- O Very Likely
- O Likely
- O Somewhat Likely
- O Unsure
- O Somewhat Unlikely
- O Unlikely
- O Extremely Unlikely

Since living in your region, have you actively looked for information regarding what you can do to reduce cyclone related property damage?

- O Yes
- O No

Please indicate the frequency in which you have (or intend to) access/contact the following sources for information about an upcoming cyclone event.

	Daily	Every few hours	Hourly	Never	N/A
Phone or visit a family member	O	O	O	O	O
Watch television updates	O	O	O	O	O
Listen to radio updates	O	O	O	O	O
Read the newspaper	O	O	O	O	O
Visit the Bureau of Meteorology website	O	O	O	O	O
Phone or visit neighbours	O	O	O	O	O
Visit the council website	O	O	O	O	O
Phone or visit SES members	O	O	O	O	O
Phone or visit the police	O	O	O	O	O
Social media updates from unofficial group pages (e.g. friends, local groups, etc.)	O	O	O	O	O
If there are any other resources you would regularly access, please list below	O	O	O	O	O

The following questions will ask you about property you own in NQ (if you own more than one, answer the questions with one property in mind). We are primarily interested in the types of building upgrades that are installed on your property.

Do you own property in NQ?

O Yes

O No

Do you live in the property you own?

O Yes

O No

How many years have you lived in/owned your current property?

____

How many more years do you plan on living in/owning your current property?

- O 0-1
- O 2-3
- O 4-5
- O 6+

Property Information

	Yes	No	Unsure
Did you build/use a building contractor to build the house you own in NQ?	O	O	O
Was your house built before 1982?	O	O	O
Was your house built before 2012?	O	O	O
Do you have a shed?	O	O	O

Please indicate which of the following are installed on your property and when the item(s) were installed.

	Not Installed	Installed when the house was built	Already installed when you purchased the house	Installed after the house was built/purchased	Unsure	N/A
Deadlocks on external doors	O	O	O	O	O	O
Metal screens on all glass areas (e.g., windows, sliding doors)	O	O	O	O	O	O
Cyclone shutters (as shown below or similar window protection)	O	O	O	O	O	O

Please indicate the primary reason why the following were installed on your property.

	Security	Cyclone Protection	Other	N/A
Deadlocks on external doors	O	O	O	O
Metal screens on all glass areas (e.g., windows, sliding doors)	O	O	O	O
Cyclone shutters (as shown below or similar window protection)	O	O	O	O

Please indicate which of the following property upgrades were installed on your property and when the upgrades were installed.

	Not Installed	Installed when the house was built	Already installed when you purchased the house	Installed after the house was built/purchased	Unsure	N/A
Roller door bracing (pre 2012 homes)	O	O	O	O	O	O
Shed anchored to concrete slab	O	O	O	O	O	O
Shed designed for high wind rating/reinforced with cyclone kit	O	O	O	O	O	O

Roof upgrades (pre 1982 homes)

	Not Installed	Installed when the house was built	Already installed when you purchased the house	Installed after the house was built/purchased	Unsure	N/A
Complete roof replacement (not only the external cladding but batten to rafter attachments and tie-downs)	O	O	O	O	O	O
Strapping Upgrades	O	O	O	O	O	O
Roof over-batten system	O	O	O	O	O	O
Sarking (layer of protection placed underneath roof tiles or sheeting to help prevent wind driven rain and dust from entering the home)	O	O	O	O	O	O

Have any of your friends, family, or neighbours installed any of these upgrades?

- O Yes
- O No
- O Unsure

Did you implement any of these upgrade items due to incentive from insurers?

- O Yes
- O No
- O N/A

Did you implement any of these upgrade items due to an incentive or encouragement from anyone else?

- O Yes
- O No
- O N/A

How likely is it that you will install the following upgrades in the next 5 years? (If you are not a homeowner, think of how likely you would be to install these upgrades if you were to buy a house without these items already installed) Leave blank if not applicable.

	Extremely unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Extremely likely	Already Installed
Complete roof replacement	O	O	O	O	O	O	O	O
Upgrade roof structural connections during roof replacement (pre-1982 homes)	O	O	O	O	O	O	O	O
Deadlocks on external doors	O	O	O	O	O	O	O	O
Cyclone shutters	O	O	O	O	O	O	O	O
Metal screens on all glass areas (e.g. windows, sliding doors)	O	O	O	O	O	O	O	O
Roller door bracing	O	O	O	O	O	O	O	O
Shed anchored to concrete slab	O	O	O	O	O	O	O	O
shed designed to high wind rating/reinforced with cyclone kit	O	O	O	O	O	O	O	O

Please indicate how strongly you agree with the following statements.

Cyclone shutters...

	Strongly disagree	Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Agree	Strongly agree
are effective for reducing damage and financial consequences of cyclones to my property and belongings	O	O	O	O	O	O	O
are effective for my family's safety during a cyclone	O	O	O	O	O	O	O
are useful for purposes other than preventing cyclone damage	O	O	O	O	O	O	O
increase property value	O	O	O	O	O	O	O
expensive to install considering my income and other expenses	O	O	O	O	O	O	O
require a lot of time and effort to install considering my free time	O	O	O	O	O	O	O
are difficult to get installed considering the knowledge and skill that is required	O	O	O	O	O	O	O
would require a lot of help/cooperation from others (friends, family, neighbours, or government) to be installed	O	O	O	O	O	O	O
can be installed by myself or a family member	O	O	O	O	O	O	O
are visually appealing	O	O	O	O	O	O	O

The next section will ask you about the preparation behaviours you preformed leading up to Cyclone Yasi and/or Cyclone Debbie.

From the list below please indicate which activities you performed at the **start of the cyclone season**. Please indicate the activities you performed before Cyclone Yasi and Cyclone Debbie.

	Cyclone Yasi			Cyclone Debbie		
	Yes	No	N/A	Yes	No	N/A
Trim treetops and branches	O	O	O	O	O	O
Check property for rust, rotten timber, termite infestations, and loose fittings	O	O	O	O	O	O
Check that the walls, roof, and eaves of your home are secure	O	O	O	O	O	O
Check fencing is not loose or damaged	O	O	O	O	O	O
Clean gutters and downpipes	O	O	O	O	O	O

From the list below please indicate which activities you performed **once the cyclone watch/warning had been issued**. Please indicate the activities you performed before Cyclone Yasi and Cyclone Debbie.

	Cyclone Yasi			Cyclone Debbie		
	Yes	No	N/A	Yes	No	N/A
Trim treetops and branches	O	O	O	O	O	O
Check property for rust, rotten timber, termite infestations, and loose fittings	O	O	O	O	O	O
Check that the walls, roof, and eaves of your home are secure	O	O	O	O	O	O
Check fencing is not loose or damaged	O	O	O	O	O	O
Clean gutters and downpipes	O	O	O	O	O	O
Put plywood up on glass windows/doors	O	O	O	O	O	O
Secure outdoor furniture and garden items	O	O	O	O	O	O
Clear yard of any loose items	O	O	O	O	O	O
Remove shade sails	O	O	O	O	O	O

If you experienced both Cyclone Yasi and Cyclone Debbie, did you prepare for the cyclones differently? How so?

____

From the list below please indicate how likely you are to perform the following activities during the next cyclone season/when a cyclone watch/warning is issued.

	Extremely unlikely	Moderately unlikely	Slightly unlikely	Neither likely nor unlikely	Slightly likely	Moderately likely	Extremely likely
Trim treetops and branches	O	O	O	O	O	O	O
Check property for rust, rotten timber, termite infestations, and loose fittings	O	O	O	O	O	O	O
Check that the walls, roof, and eaves of your home are secure	O	O	O	O	O	O	O
Check fencing is not loose or damaged	O	O	O	O	O	O	O
Clean gutters and downpipes	O	O	O	O	O	O	O
Put plywood up on glass windows/doors	O	O	O	O	O	O	O
Secure outdoor furniture and garden items	O	O	O	O	O	O	O
Clear yard of any loose items	O	O	O	O	O	O	O
Remove shade sails	O	O	O	O	O	O	O

Thinking about both structural upgrades and preparedness activities, what factors are most important to you when preparing for a cyclone? Please rank items from most important to least important.

That preparations...
_______ increase my (and my family's) safety during a cyclone
_______ limit the financial impact and damage to my property and belongings
_______ are also useful for events other than cyclones
_______ are affordable
_______ take little time and effort
_______ require little knowledge and skills
_______ require little help and cooperation from others

This section will ask you about your home insurance

What type of home insurance do you have (if any)?

- O Home and contents insurance
- O Only Home insurance
- O Only contents insurance (homeowner)
- O Only contents insurance (renter)
- O No insurance (homeowner)
- O No insurance (renter)

How much does your current insurance premium cost p.a.?

____

When choosing your insurance policy did you look for anything in particular regarding cyclone coverage? If yes, what did you look for?

- O Yes ______________________________
- O No

Are there any particular reasons why you do not have insurance?

____

Assuming it would cost $3000 (including labour) to install cyclone shutters on all of your windows, how much reimbursement would you require to go ahead with the purchase?

As a reduction on your premium p.a. over 5 years (would require some paperwork).

____

As a government rebate (would require some paperwork).

____

Other questions

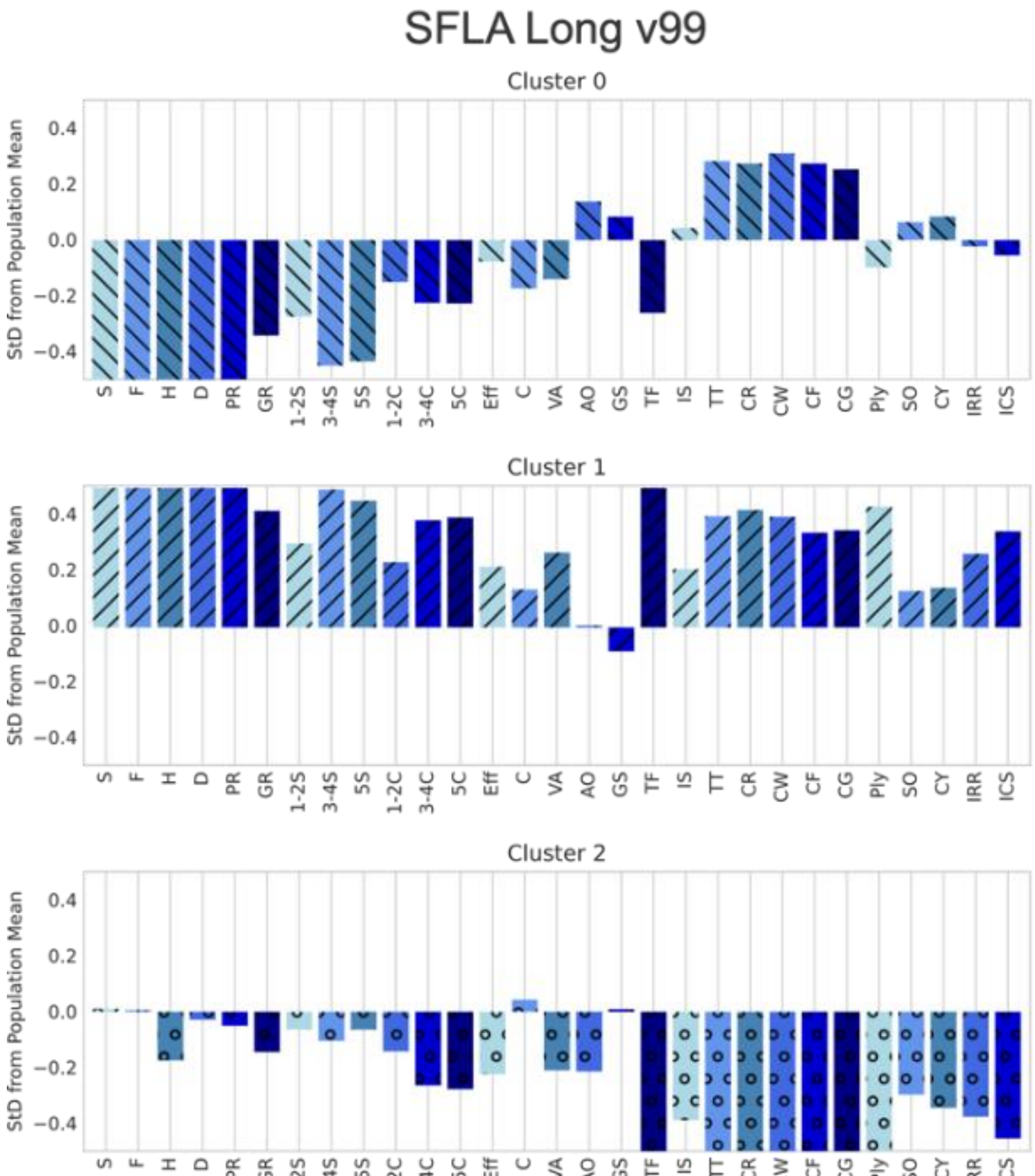
What is the highest level of education you have completed?

- O Grade 9 or Below
- O Grade 10 or 11
- O Grade 12
- O Certificate I-IV
- O Diploma
- O Bachelor's Degree
- O Postgraduate Degree (Masters/PhD)

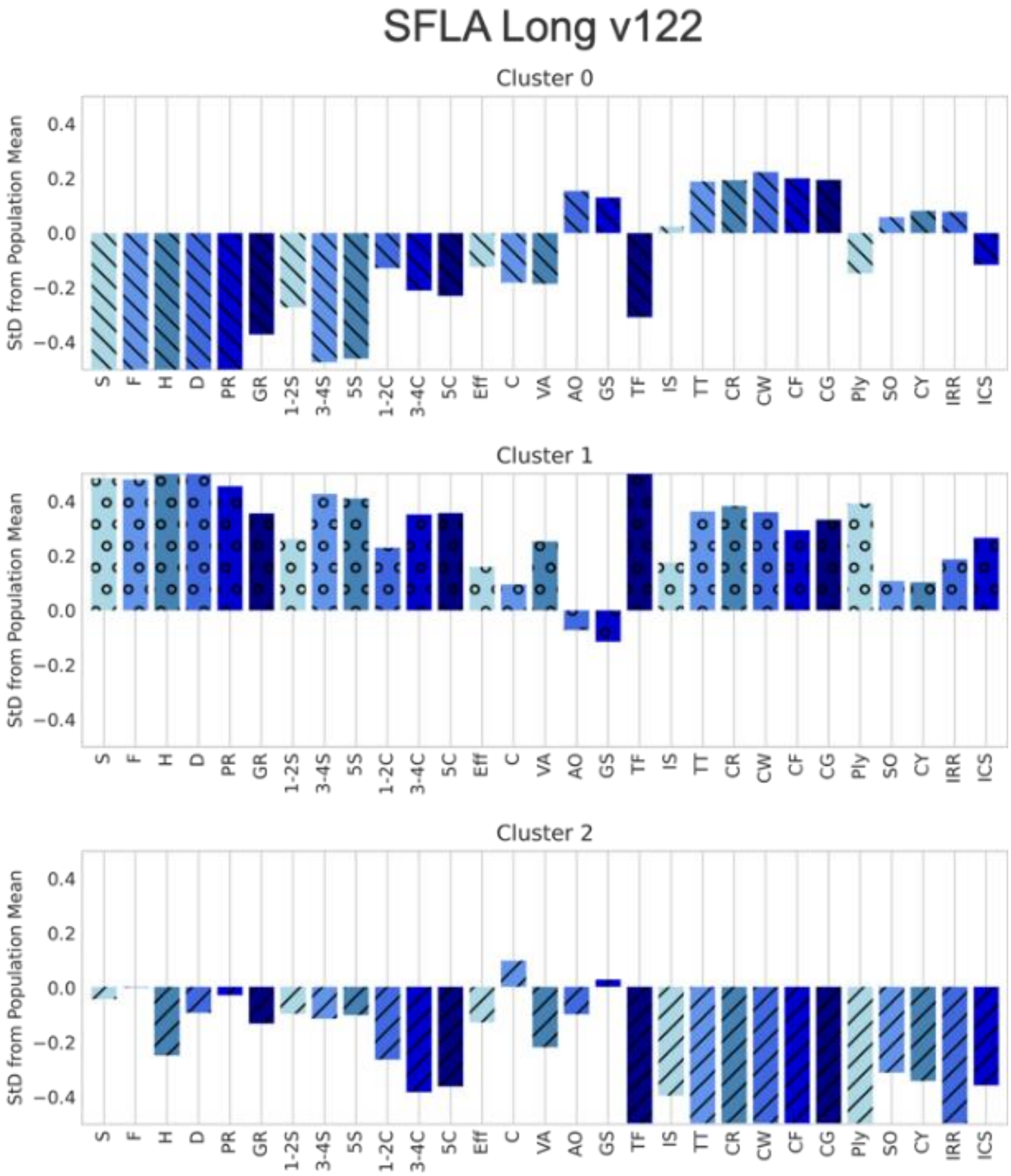
What is your average household income?

- O <$22,000
- O $22,000 - $50,000
- O $50,000 - $80,000
- O $80,000 - $125,000
- O $125,000 - $260,000
- O >$260,000

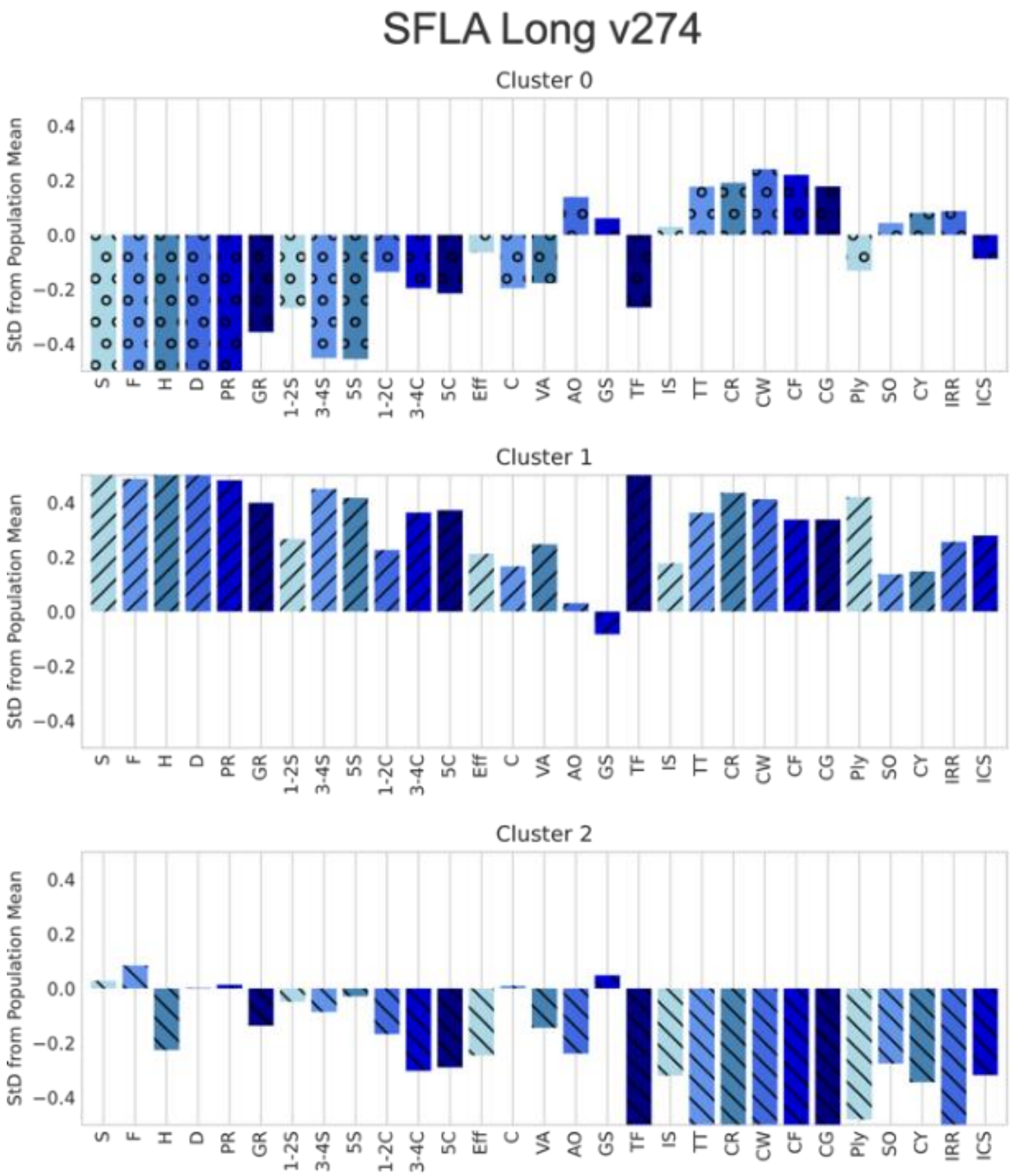
APPENDIX C: GRAPHS OF THE TOP 4 FUNCTIONALLY UNIQUE CLUSTER SETS ON THE 2018 DATA SET



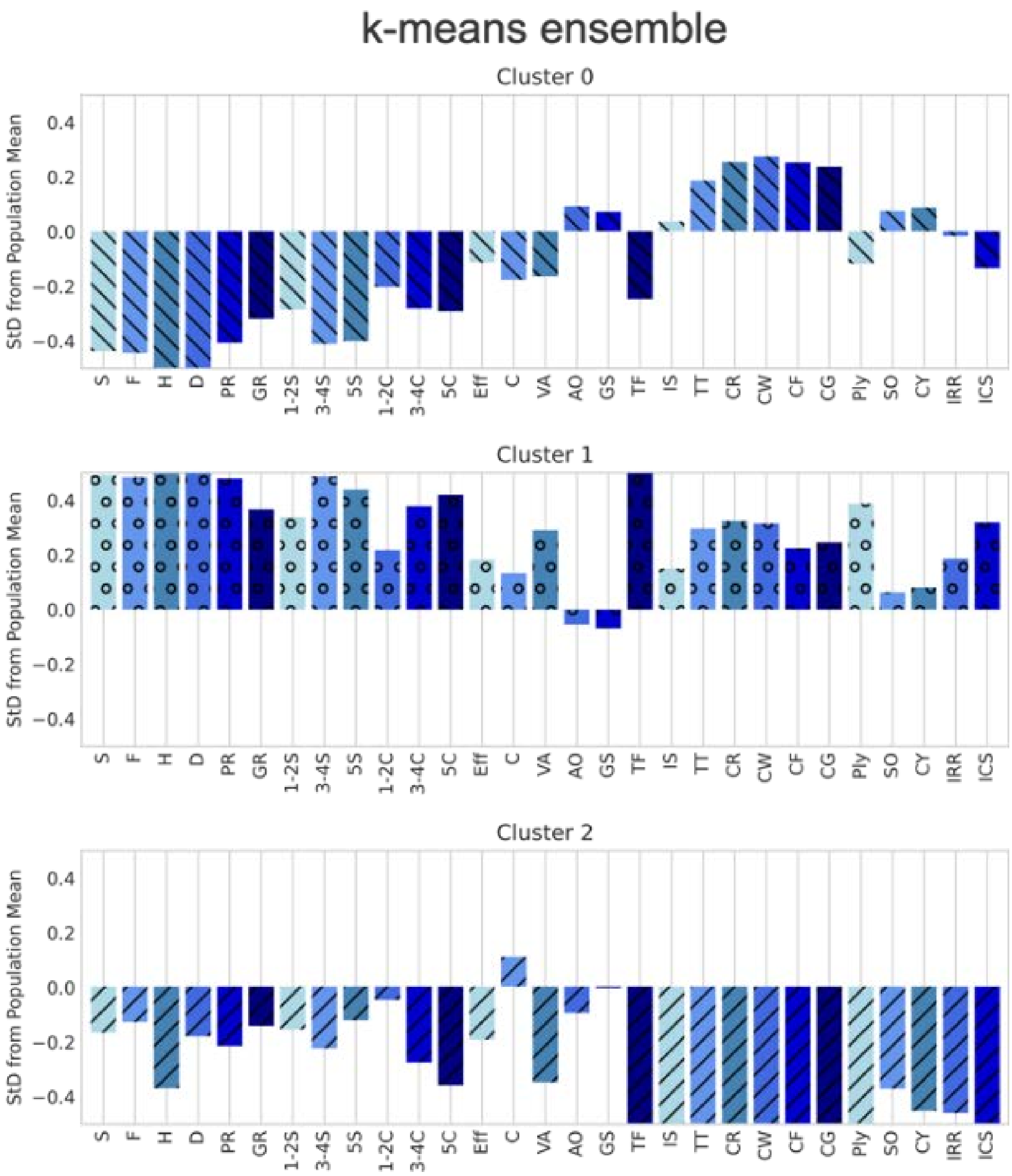
Appendix C: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2018 Data Set



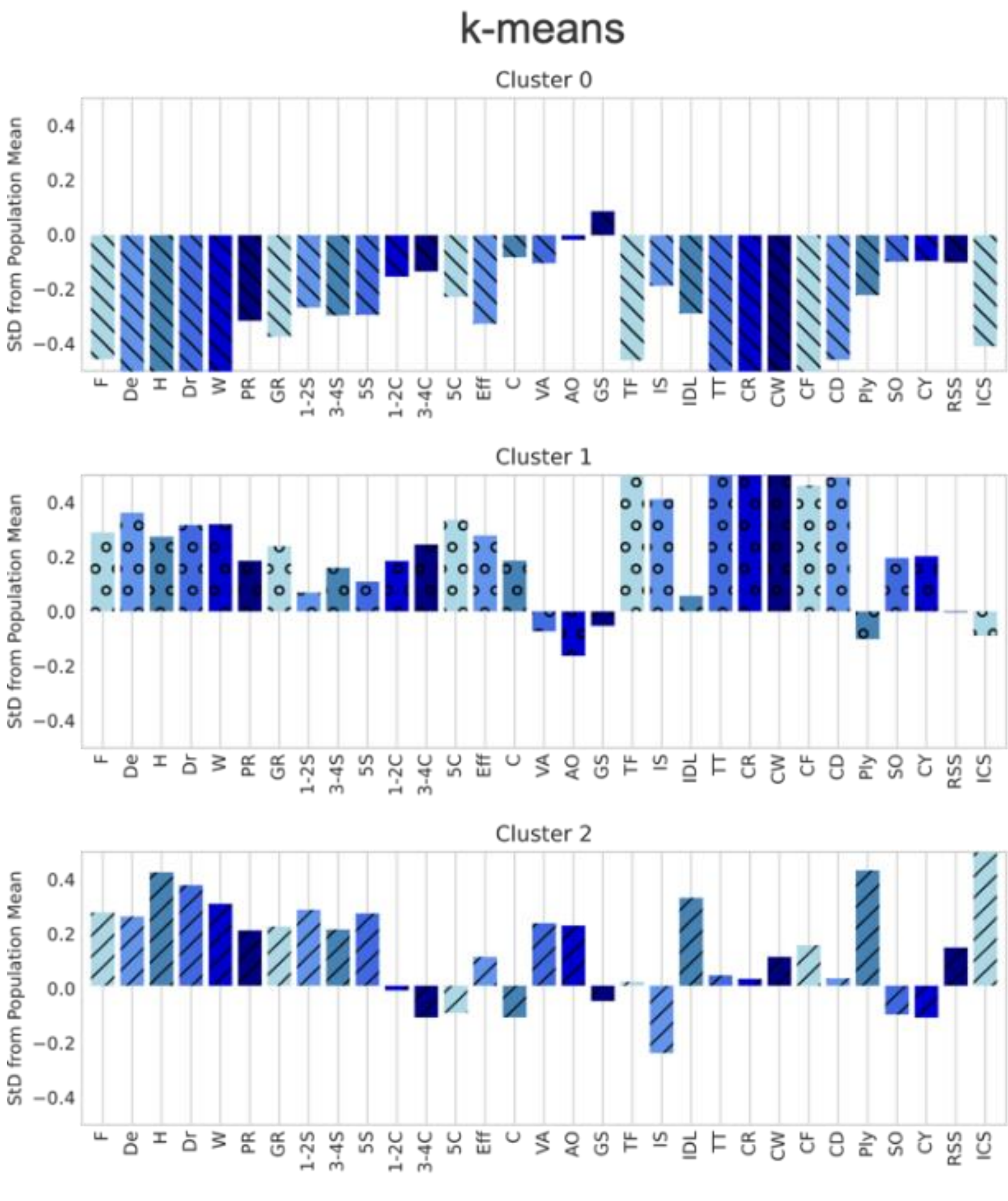
Appendix C: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2018 Data Set



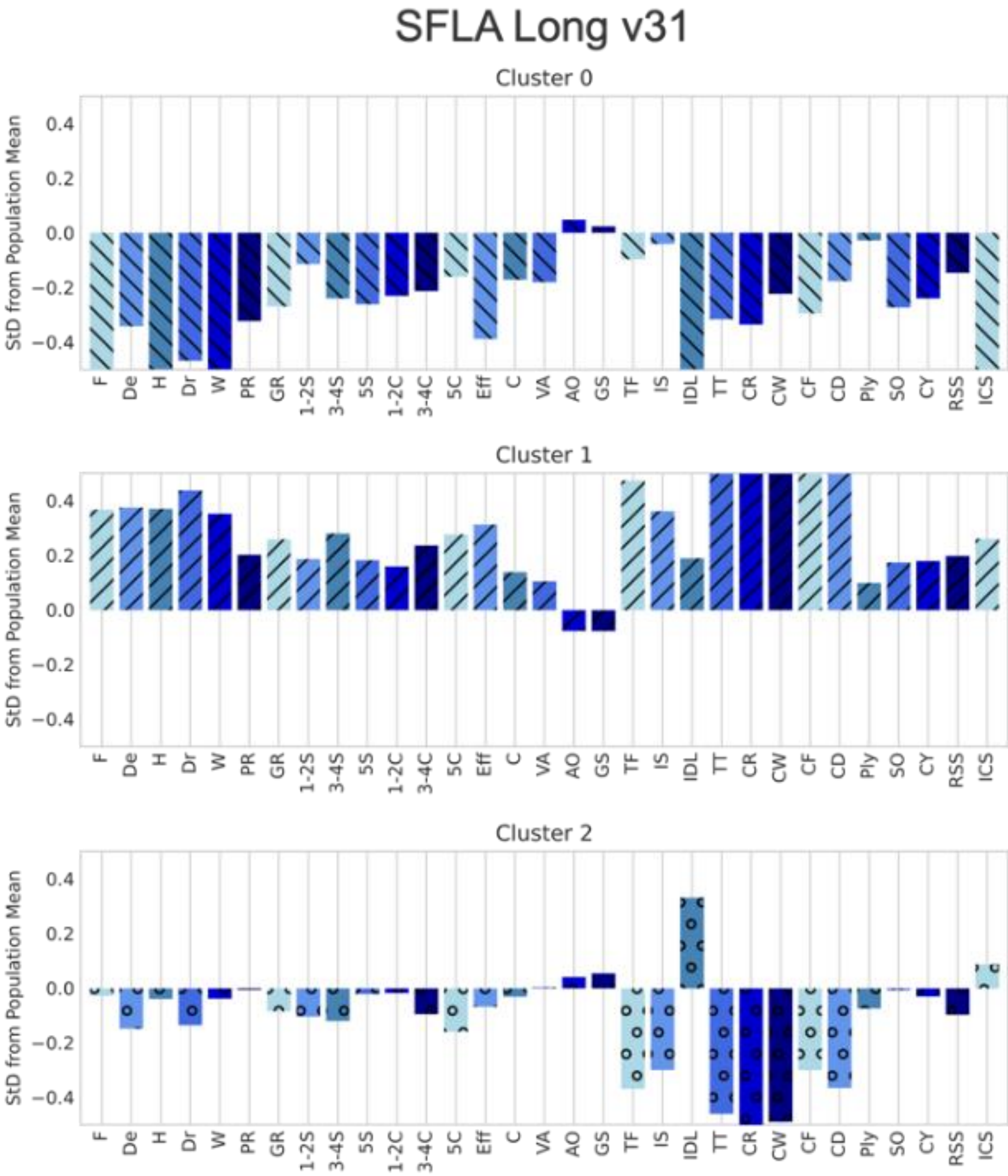
Appendix C: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2018 Data Set



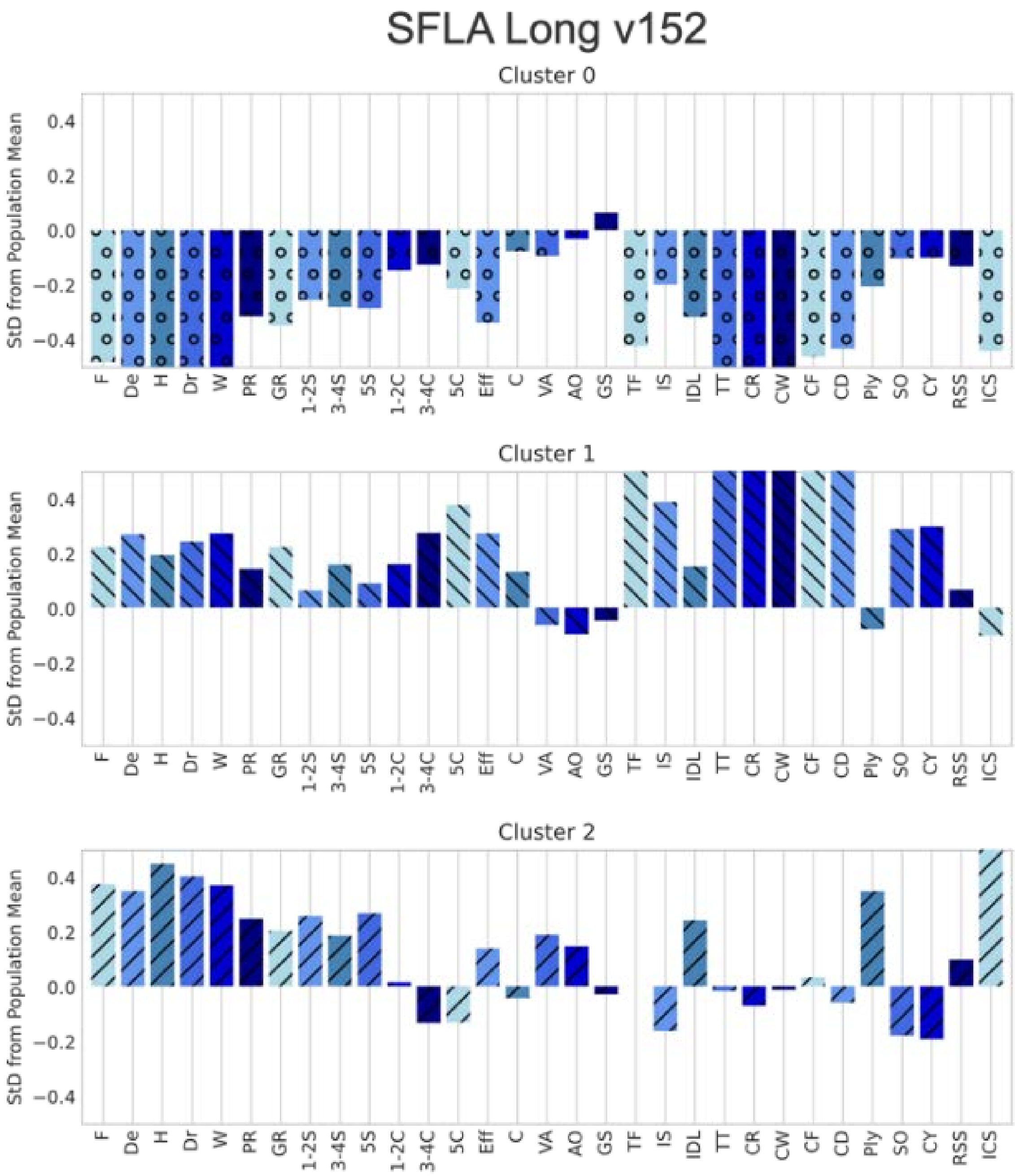
APPENDIX D: GRAPHS OF THE TOP 4 FUNCTIONALLY UNIQUE CLUSTER SETS ON THE 2021 DATA SET



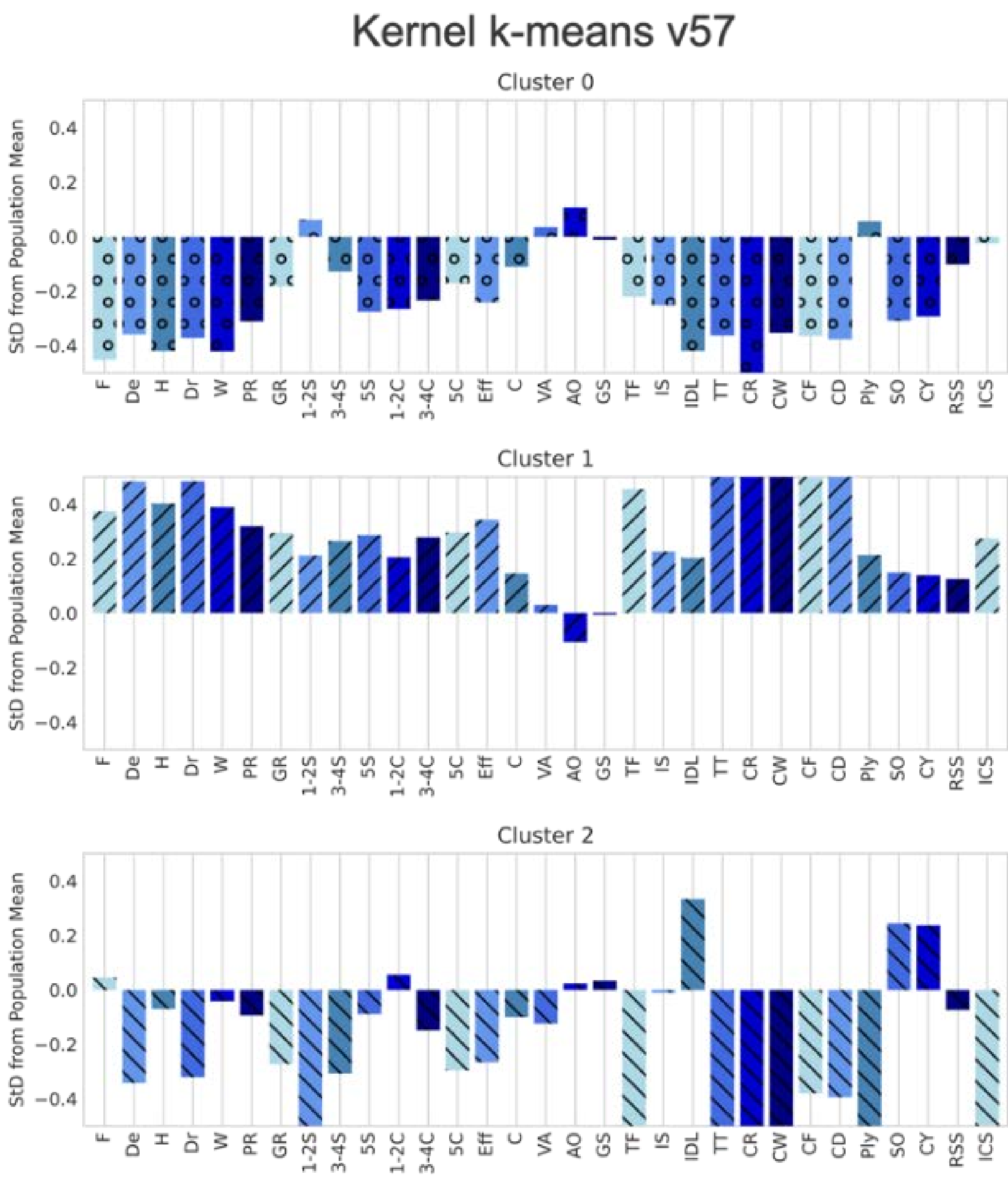
Appendix D: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2021 Data Set



Appendix D: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2021 Data Set



Appendix D: Graphs of the Top 4 Functionally Unique Cluster Sets on the 2021 Data Set



APPENDIX E: 2018 DATA SET PERSONAS

Unconcerned Persona

They don't really consider cyclones as a pressing issue, whether due to lack of experience, the age of their house, or being a renter.

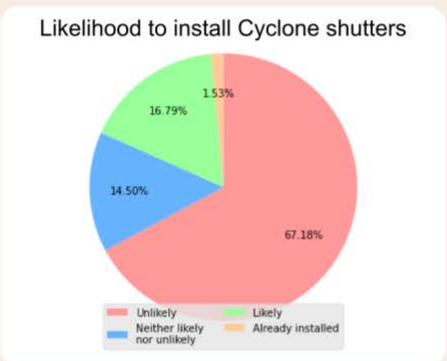
They spend the least amount of time discussing or thinking about cyclones. While they do believe a cyclone occurring would have significant consequences, and they have the average emotional response to the possibility of a cyclone occurring, they have the lowest expectation of a cyclone occurring. They are likely to perform the very simple mitigation behaviours but are unsure about the more difficult behaviours and would not perform behaviours such as putting up plywood or installing cyclone shutters. They are the least likely to have experienced a cyclone and the most likely to be renting and be living in a more recently built property, which may mitigate a level of concern about property damage.

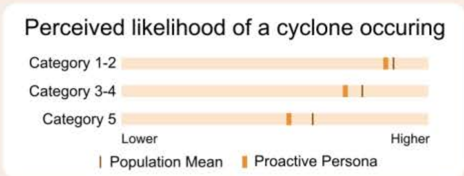
About Them
More formal education
Average income level
Less likely to be a homeowner (60.30%) of a home built before 2012 (47.32%)

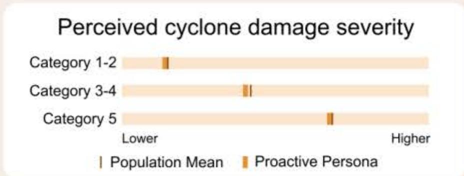
They generally don't discuss and think about cyclones

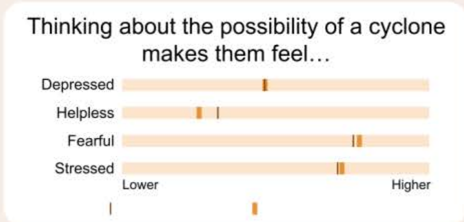
Beliefs about Cyclones:
They believe a cyclone occuring would:
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

And might also:
- Damage their home









Before the next cyclone they are:
Very likely to:
- Secure outdoor furniture and items
- Clear yard of any loose items

Are unsure whether they would:
- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes

But would not:
- Put plywood up on glass windows/doors

Believe cyclone shutters:
- They are unsure whether they are effective
- Are not visually appealing
- Are not particularly expensive

Proactive Persona

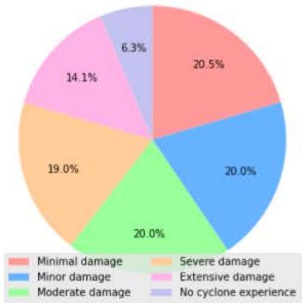
They are actively worried about the possibility of a cyclone occurring and are willing to perform any behaviours to protect themselves, their family, and their property.

They are the most concerned about the possibility of a cyclone, with the highest expectation of both a cyclone occurring and doing considerable damage. They have the strongest emotional response to the possibility of a cyclone, with it causing feelings of depression and helplessness, and are most likely to spend time thinking about or discussing the possibility of a cyclone. They are the most likely to perform preparatory behaviours leading up to the next cyclone, including more difficult behaviours such as installing cyclone shutters. They also believe cyclone shutters to be quite effective and finds them the most visually appealing.

About Them
Less formal education
Average income level
Likely to be a homeowner (81.16%) of a home built before 2012 (67.63%)

They somewhat regularly discuss and think about cyclones

Cyclone Damage Experience



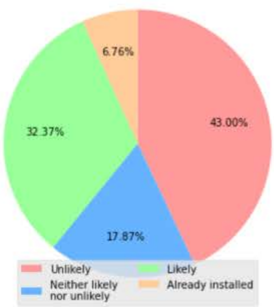
Beliefs about Cyclones:
They believe a cyclone occuring would:

- Cause catastrophic destruction
- Pose a great financial threat
- Damage their home
- Disturb daily life
- Disrupt their ability to work

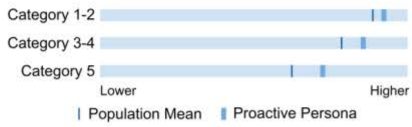
And might also:

- Negatively affect their mental health
- Pose a threat to future generation

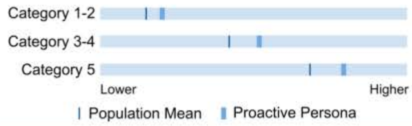
Likelihood to install Cyclone shutters



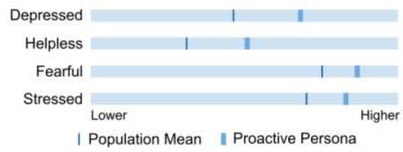
Perceived likelihood of a cyclone occuring



Perceived cyclone damage severity



Thinking about the possibility of a cyclone makes them feel…



Before the next cyclone they are:
Very likely to:

- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes
- Secure outdoor furniture and items
- Clear yard of any loose items

Might:

- Put plywood up on glass windows/doors

Beliefs about cyclone shutters:

- They are moderately effective
- Are somewhat visually appealing
- Somewhat expensive

Confident Persona

They are aware of the risks associated with cyclones but are confident they know what to do to mitigate any damage – as they have done before.

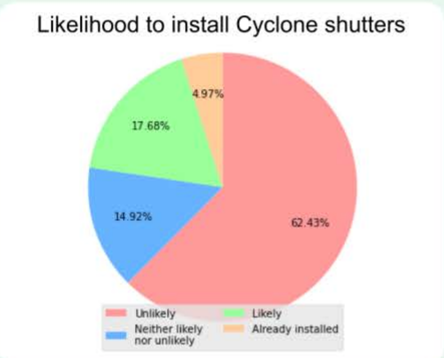
They perceive the least amount of risk associated with a cyclone occurring and have the lowest perception of the damage that would accompany a cyclone. They have weakest emotional response to the potential of a cyclone occurring, feeling particularly less feelings of depression or helplessness. They are most likely to have been through a cyclone before where they have received no damage or only minor damage. They occasionally discuss or think about cyclones and are likely to do simple mitigation behaviours leading up to the next cyclone but are unlikely to perform any of the more difficult or expensive tasks.

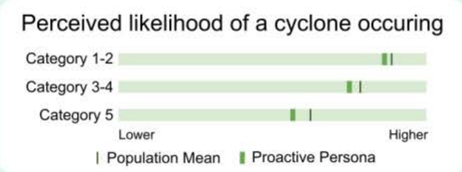
About Them
Average amount of formal education
Higher income level
Likely to be a homeowner (76.24%) of a home built before 2012 (64.08%)

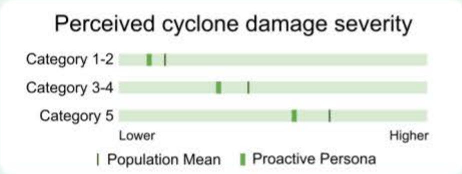
They are somewhat unlikely to discuss or think about cyclones

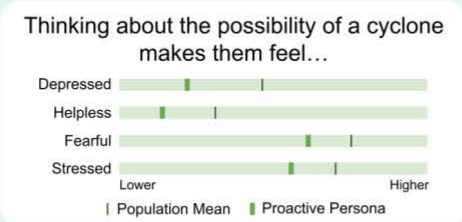
Beliefs about Cyclones:
They believe a cyclone occuring might:
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

But is unlikely to:
- Negatively affect their mental health
- Negatively affect their physical health









Before the next cyclone they are:
Very likely to:
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes
- Secure outdoor furniture and items
- Clear yard of any loose items

Might:
- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings

Are unsure whether they would:
- Put plywood up on glass windows/doors

Beliefs about cyclone shutters:
- are somewhat effective
- aren't particularly visually appealing
- are not particularly expensive

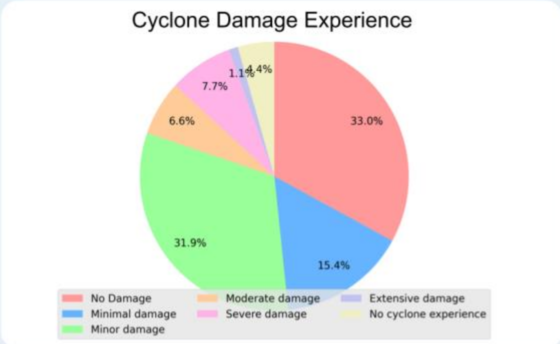
APPENDIX F: 2021 DATA SET COLOURED PERSONAS

Proactive Persona

They are actively worried about the possibility of a cyclone occuring and are willing to perform any behaviours to protect themselves, their family, and their property.

They are the most concerned about the possibility of a cyclone, with the highest expectation of both a cyclone occurring and doing considerable damage. They have the strongest emotional response to the possibility of a cyclone and are most likely to spend time thinking about or discussing the possibility of a cyclone. They are the most likely to perform preparatory behaviours leading up to the next cyclone, including more difficult behaviours such as installing cyclone shutters, and to have looked up ways to prevent cyclone damage. They are the most likely to have previously played a role in cyclone preparation and have the strongest belief that a cyclone occurring could damage their property.

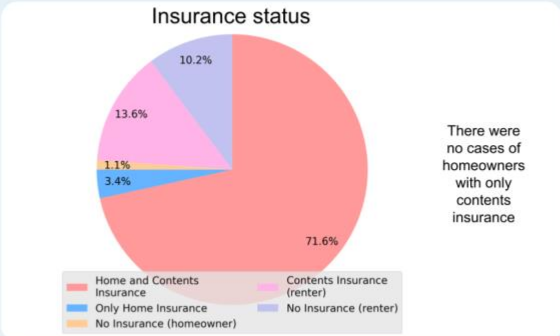
About Them
They are probably a homeowner (73.03%) of a home built before 2012 (62.92%), and intend to live in their house for 3-4 more years.
They are most likely to have played a role in the preparation leading up to previous cyclones (94.38%)

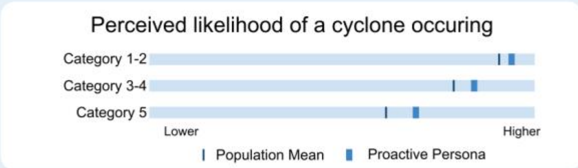


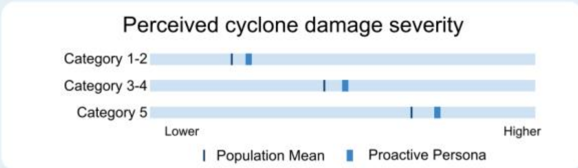
Attitudes towards cyclones
They are likely to discuss and think about cyclones somewhat regularly and to have sought out more information on ways to reduce cyclone damage.
They consider themselves as fairly knowledgeable about cyclones and the related risks.
They think cyclone shutters are likely to be somewhat effective

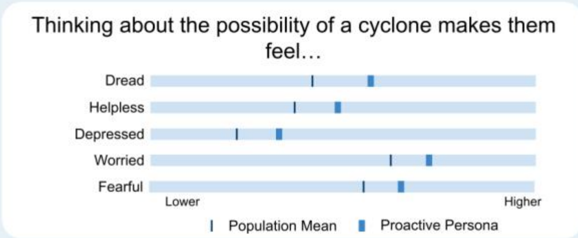
Beliefs about Cyclones:
They believe a cyclone occuring would:
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

And might also:
- Damage their home



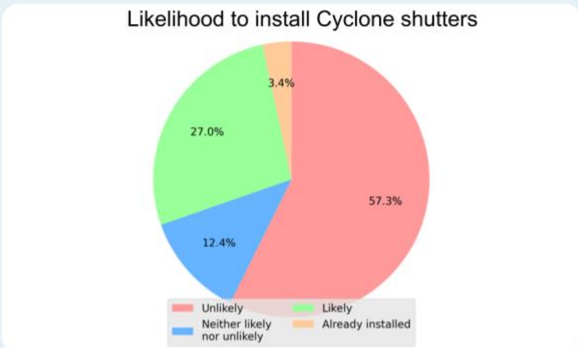






Before the next cyclone they are:
Very likely to:
- Secure outdoor furniture and items
- Clear yard of any loose items
- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes

Are unsure whether they would:
- Put plywood up on glass windows/doors



Unconcerned Persona

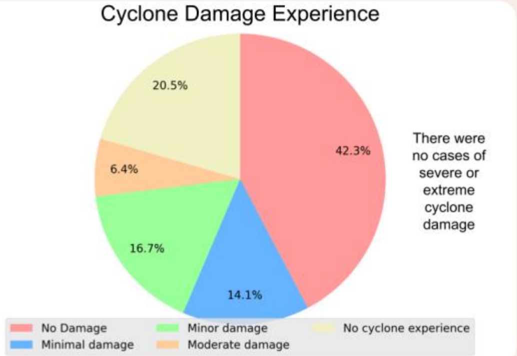
They don't really consider cyclones as a pressing issue, whether due to lack of experience, the age of their house, or being a renter.

They are unlikely to spend time discussing or thinking about cyclones, rate themselves the least knowledgeable about cyclones, and are least likely to have looked up ways to reduce cyclone damage. While they do believe a cyclone occurring would have significant consequences, they have the weakest emotional response to the prospect of a cyclone occuring. They are likely to perform the very simple mitigation behaviours but are unsure about the more difficult behaviours. They are the least likely to have experienced a cyclone, or to have been previously involved in cyclone preparation, and of those with cyclone experience are the most likely to have received no damage. They are most likely to be renting, living in a more recently built property, and to be uninsured.

About Them

They are probably a renter (56.41%) of a home built after 2012 (69.23%), and only intend to live in their house for 2-3 more years.
They are least likely to have played a role in the preparation leading up to previous cyclones (71.79%)

Cyclone Damage Experience



Attitudes towards cyclones

They are less likely to discuss and think about cyclones regularly and to have not sought out information on ways to reduce cyclone damage. They don't consider themselves as knowledgeable about cyclones and the related risks. Are unsure whether cyclone shutters would be effective

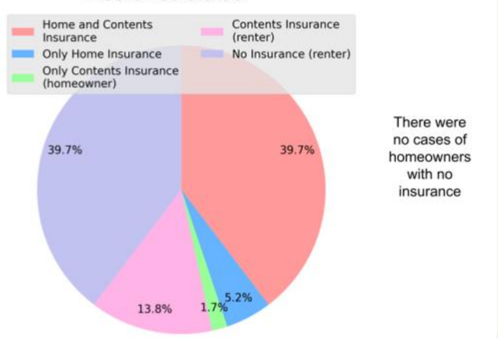
Beliefs about Cyclones:

They believe a cyclone occuring might:

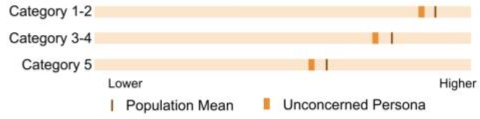
- Cause catastrophic destruction
- Pose a great financial threat
- Disturb daily life
- Disrupt their ability to work

But they are unsure on any other possible effects.

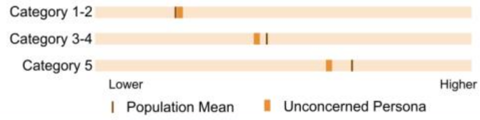
Insurance status



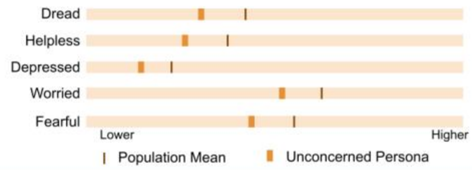
Perceived likelihood of a cyclone occuring



Perceived cyclone damage severity



Thinking about the possibility of a cyclone makes them feel…



Before the next cyclone they are:

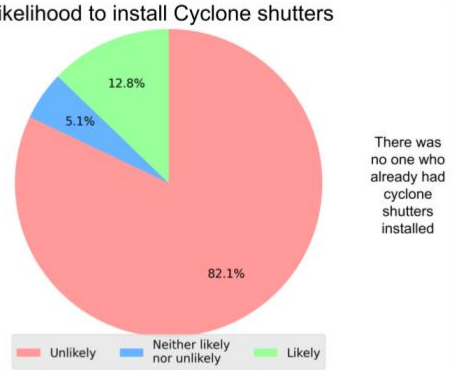
Very likely to:

- Secure outdoor furniture and items
- Clear yard of any loose items

Are unsure whether they would:

- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes
- Put plywood up on glass windows/doors

Likelihood to install Cyclone shutters



Confident Persona

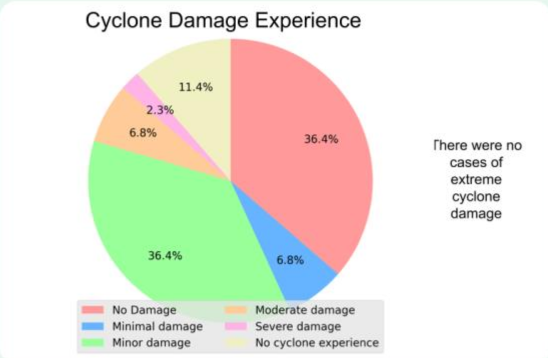
They are aware of the risks associated with cyclones but are confident they know what to do to mitigate any damage – as they have done before.

They perceive the least amount of risk associated with a cyclone occurring, only having a strong belief that a cyclone occuring would disrupt their daily life. They are most likely to have been through a cyclone before where they received minor damage and the most likely to be fully insured. They rate themselves as the most knowledgeable about cyclones, although they are the least likely to think about or discuss cyclones. They are likely to do simple mitigation behaviours leading up to the next cyclone but are unlikely to perform any of the more difficult tasks or install structural upgrades, although they are the most likely to be homeowners and living in a property built before 2012.

About Them

They are most likely to be a homeowner (97.73%) of a home built before 2012 (72.73%), and intend to live in their house for 5+ more years.

They likely played a role in the preparation leading up to previous cyclones (86.36%)



Attitudes towards cyclones

They are unlikely to discuss or think about cyclones regularly but might have sought out information on ways to reduce cyclone damage. They consider themselves very knowledgeable about cyclones and the related risks and are unsure the effectiveness of cyclone shutters.

Beliefs about Cyclones:

They believe a cyclone occuring will:

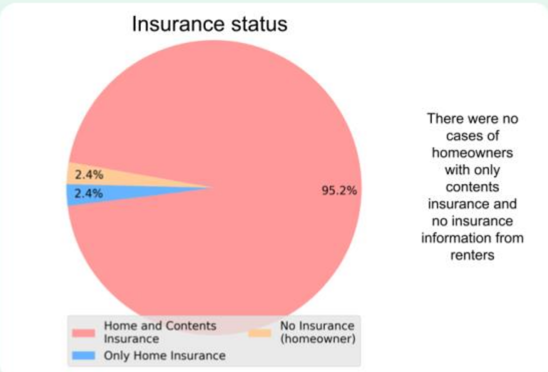
- Disturb daily life

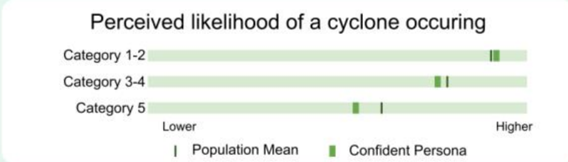
Might also:

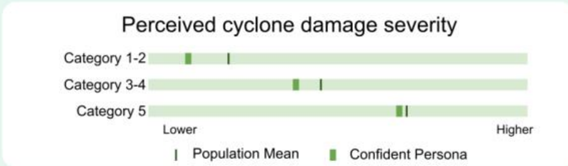
- Cause catastrophic destruction
- Pose a great financial threat
- Disrupt their ability to work

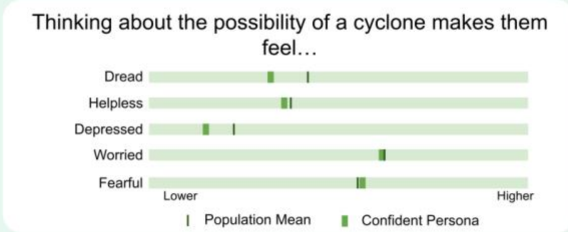
But probably wouldn't:

- Cause widespread death
- Pose a significant threat to future generations
- Negatively affect physical health









Before the next cyclone they are:

Very likely to:

- Secure outdoor furniture and items
- Clear yard of any loose items

Are unsure whether they would:

- Trim treetops and branches
- Check property for rust, rotten timber, termite infestations and loose fittings
- Check that the walls, roof and eaves are secure
- Check fencing is not loose or damaged
- Clean gutters and downpipes

Are unlikely to:

- Put plywood up on glass windows/doors