Faculty of Engineering of University of Porto

Driving innovation through social data: a methodology for building buyer personas

Natália Cavalcanti Carneiro Leão

Master in Innovation and Technological Entrepreneurship 2016-2018

Supervisor: Prof. Doutor Manuel Firmino Torres

2018

Abstract

Knowing your audience is crucial to the success of a communication or product strategy. In this context, buyer-personas are fictional profiles that personify targeted customers and are a powerful tool for providing marketers and managers with strong audience insights and directives. It's originally a demanding and costly task through the use of traditional market research methods. The main focus of this study is to explore the power of social data analysis and develop a methodological procedure for building buyer-personas using social data conversations. This will be achieved through the study of customer segmentation theories and examination of traditional market research methods, followed by the application of the new guidelines in a case studies for building buyer-persona profiles and audience map.

Acknowledgements

v

Index

Figures List

Table List

1. Introduction

1.1 - Motivation

One of the toughest challenges that marketers face is understanding who are their customers, how different they are, and then deciding how to best market to each of them. This dissertation delves into specific methodologies for further exploring customer segmentation and the buyer persona as strategic tools. Persona profiling is originally an extensive and costly task but this has been changing with marketers discovering new ways to use technologies and automation to their benefit.

As a former marketing professional and currently manager of social listening projects, I would like to further explore a methodological procedure for customer profiling based on qualitative and quantitative analysis of available social data. This study aims to create a body of knowledge regarding the value of buyer personas as well as building an execution road map for buyer persona generation that can be applied by marketers and even entrepreneurs for a more affordable cost and within shorter timelines.

Also, writing this thesis is an opportunity for improving current methodologies applied for market segmentation joining the researcher's professional life experience with the academic knowledge acquired from the Master in Innovation and Technological Entrepreneurship, and present solid case studies for buyer persona generation that can be later be referred as an innovation and a competitive advantage in organizations.

1.2 - Basic Framework

A brief conceptual framework is needed to provide the reader with a short overview of the main concepts which will allow a better understanding of the literature review discussion presented later. Those are:

Customer Segmentation - Richard Tedlow identifies segmentation starting in the 1920s. As market size increased, manufacturers could pitch different models of their products and meet specific needs of a number of demographic and psychographic market segments. The concept has evolved together with technological advancements, allowing marketers to narrow

market segments even more and therefore communicate more tailored messages and get the most results from their marketing efforts.

Personas - The purpose of persona generation is to develop a solid representation of your audience segment for reference; using fictional characters to represent behavioural patterns that can help in better understanding the values and attitudes of that group of people. Although the naming "persona" came only a few years later, the overall concept was first suggested by Angus Jenkinson in 1994. Nowadays, it's used by both product designers and developers (user persona) and by marketers (see buyer persona). Traditionally, personas rely on data collect thru interviews and, more recently, Customer-relationship-management database.

Buyer Persona - Considering the definition of personas, buyer personas are semi-fictional profiles representing a segment of buyers for a product or service.

Psychographics – Defines the process of measuring consumers by behaviours and segmenting by personality (Samuel, 2016).

Social Data - Refers to any information that social media users publicly share and that includes metadata such as language, location or biographical details. Tweets on Twitter, posts on Facebook and comments on Forums are a few examples of social data.

Social Listening - It's the process of capturing online conversation around pre-defined terms (search query) to understand what users are saying about a brand or an industry. Those conversations will then be used to discover opportunities or answer specific business questions. In this dissertation, can also be referred as social media research.

2. Literature Review

An analysis of previous literature is critical for fully understanding the topic of the research and helpful to progressively define the objectives of the study. Fink (2005) defines a literature review as a *"systematic, explicit and reproducible method for identifying, evaluating and synthesising the existing body of completed and recorded work produced by researchers, scholars and practitioners"*.

This research looked for articles published in Scopus based on the query ["buyer persona" OR "buyer personas" OR "buyer profile" OR "customer persona" OR "brand persona" OR "persona generation" OR "persona development" OR "user archetypes" OR ((personas OR persona) AND (marketing OR "market research" OR "market researches" OR "market segment" OR "customer segment" OR "segmentation" OR "research method" OR "social media datum" OR "ethnography"))]. A total of 329 documents were obtained from the search from which 150 most relevant were selected based on its titles and keywords of which the abstract of 42 was read to rank these papers based on the relevancy to the topic of this research. Also, additional references were analyzed based on the bibliography of such pre-select papers. Within the academic papers, 18 out of 42 articles had been analysed to find the gap in the literature, as shown in the Table 1 below.

Table 1 — Literature Review

Authors & Year	*Keywords* & Concepts Covered
(Smith, 1956)	Reviewed the concepts of product differentiation and marketing segmentation as emerging alternative marketing strategies.
(Jenkinson, 1994)	Proposed Maslow's hierarchy of values as a useful typology for the ladder of value, representing an evolution of quality thinking. The author discussed the concept of "group" and "alignment" as potentially more powerful than "segment" and "target".
(Cooper, 1998)	The author first features the difficulties for engineers and developers to access the user needs when designing a product and suggests the use of persona during the process.
(Kelly, 2003)	*Data mining, customer segmentation, CRM (Customer Relationship Management), mass customisation, clustering, direct marketing*

	Discussed on the application of data mining and cluster analysis to the basic problem of market segmentation thru a priory acquired database.
(Nielsen, 2004)	On how a user-centred approach influenced the perception of the design process in the e-business group at AstraZeneca. The author used a qualitative approach, performing user inquiries and with a perspective focused on designing. **Future Research:** - Further investigate grouping or segmentation of users - Problematics of disseminating the knowledge of personas to large groups within the enterprise
(Johansson & Messeter, 2006)	*Concept design, ethnography, interaction design, interdisciplinary design, participatory design, process improvement, user-centred design* Reported a persona approach built on a combination of ethnographic exploration, participatory inquiry and collaborative design. The author also suggests a continuous process of constructing personas as the design process evolves.
(McGinn & Kotamraju, 2008)	*Personas, persona development, data gathering, data analysis, factor analysis, user research, survey, interview.* Described a method for creating and validating personas, based on the statistical analysis of data. The researchers performed a survey, a factor analysis to allow the persona groups to emerge and then interviews to make adjustments.
(Portigal, 2008)	Discusses persona generation as a tool and its benefits in the designing process as well as the risks of misuse.
(Jenkinson, 2009)	*Segmentation, collaborative filtering, analytics, strategy, audit, data mining* Discussed about the art of the possible and the effect on business success of segmentation methods, providing a simple "test of effectiveness" self-audit for organisations decision-making.
(Kotler & Kevin, 2009)	Discussed a vast list of marketing concepts, walking through segmentation and targeting and presenting personas as a valuable strategy.

(Siegel, 2010)	*User research, personas, segmentation, quantitative research, qualitative research* Presented an instructive case in which qualitative field research helped contribute to abandoning a "state of the art" quantitative user segmentation. This research was performed through a UCD (User-centred design) perspective. The quantitative raw data was gathered from a large volume of responses from users while the qualitative was verified through interviews. The author discusses questions the value of using statistics blindly.
(Awad, 2011)	*Green marketing; segmentation; environment; green consumerism; cluster analysis, demographics, ECCB (ecologically conscious consumer behaviour), Kingdom of Bahrain* Used questionnaire survey and data analysis to profile green consumers in the kingdom of Bahrain. **Future Research** - Supporting the self-report measures with observational or behavioural measures to strengthen the findings
(Miaskiewicz, 2011)	*Personas, design processes, design methodology, research methods* Examined the benefits of incorporating personas into design processes thru a literature review of the concept. **Future Research**: - Further examination of persona use and methods
(Chroneos-Krasavac, Soldic-Aleksic & Petkovic, 2016)	*Big data, advanced analytics, the impact at scale, marketing, retail, customer data, Networked Readiness Index* Discusses the emergence of big data and its application for business' needs with the specific emphasis on marketing and new abilities to profile the customers, especially related to CRM database. **Future Research:** The role of business analytics techniques for structuring data for organizations use - Investigate the relevance of big data usage in the businesses processes - Explore the connection between big data usage and data-driven decision making

(Dagustan, Buchory, Satya & Diryana, 2016)	*Segmentation, Positioning* Conducted re-profile consumer snacks food brand borondong through surveys and data analysis to come up with six consumer profiles for the brand.
(Félix, 2016)	*Tradition Marketing; Inbound Marketing; Focus; Customers; Profiling; Data; Database Marketing; Data Mining* Performed a series of interviews with experts on management and marketing to collect their insights about the importance of content and profile creation for new marketing strategies. Discussion focused on database and use of CRM. **Future Research:** - The author aims to create in the future an automated customer profile thru the use of database and data mining techniques.
(Jung, An, Kwak & Jansen, 2017)	*Persona; User Experience Research; User Analytics* Developed a methodology for persona generation using real time social media data focused on digital content creators' accounts. The methodology here is focused on NMF (non-negative matrix factorization) and automated generation with data from a large Youtuber account. **Future Research:** - Investigate other approaches besides NMF to identify user segment
(Salminen, Şengün, Kwak, Jansen, An, Jung, Vieweg & Harrell, 2017)	*Social media analytics; personas; Middle East; mixed method research* Conducted a mixed-method study to understand the content consumption patterns of Middle Eastern social media users and explored new ways to present online data by using automatic persona generation. Methodology included analysis of YouTube interactions content and interviews. **Future Research:** - Further comparisons of social media personas across and within cultures - Validation of the usefulness of the automatic personas in real decision-making situations - Encourage multi-disciplinary, mixed-method studies in the field of social media analytics, especially for generating rich data representations such as user personas

It's noticeable that several authors have covered the topic of segmentation and discussed different perspectives and methodologies for persona generation. Thus, the literature discussion was grouped into two pieces: "Persona as a Segmentation Tool" and "Methodologies for persona generation", followed by a conclusion about the call for further research and the gap in literature.

2.1 – Persona as a Segmentation Tool

Wendell R. Smith (1956) was the first one to discuss the marketing segmentation as emerging alternative marketing strategy and an evidence of the consumer's pre-eminence in the contemporary economy, highlighting the benefits of using science to solve marketing problems: *"Market segmentation involves viewing a heterogeneous market as a number of smaller homogeneous markets in response to differing preferences, attributable to the desires of consumers for more precise satisfaction of their varying wants"* (p. 6). In other words, a market segment represents a slice of customers who share similar attributes and needs.

As both the market demand and supply side increase its diversity, theories of perfect competition and monopoly had become insufficient to explain the new business scenarios. Variations in resources and processes of manufactures led to differences in products and the release of mass production approaches, while the diversity in consumer demand prompted the loss of precision when predicting customer's needs. Smith (1956) also suggests that the increase in buying power contributed to intensifying comparison in the buyer cycle and consumers became more willing to pay a little extra for products with an identity and authenticity that they could relate to. As a response to those changes, market segmentation started to be seen by companies as a necessary strategy for growing and not becoming obsolete.

In 1994, Angus Jenkinson questioned if, however, market segmentation was enough to address the individuality needs of the consuming behaviour. He introduced the concept of "group" and "alignment" as a more powerful tool than segmentation, through which enterprises would better understand the attitudes and personalities behind the groups and will then be able to shift its marketing efforts accordingly and, thus, more efficiently. While "segments" entails segregating consumers into fragments based on given factors, very often demographics, "groups" would combine those sharing common characteristics. Jenkinson himself admits the difference may be subtle but has its benefits and highlights that grouping

would imply more efforts to nurture an intimate relationship with the community. Eventually, these groups' archetypes would become known by the term "personas".

Table 2 — Differences between Segments and Groups (future known as "personas"). *Note. Adapted from Beyond Segmentation (p. 64), by A. Jenkinson, 1994.*

SEGMENTS AND GROUPS	
"Segments" are more:	**"Groups" are more:**
Top down	Bottom up
What divides?	What unites?
Population based	People based
Macro-economic	Micro-economic
Class determined values	Individual values shared
(Useful) abstraction	Concrete people
Product driven	Requirements driven
Aiming to generate aspirations	Aiming to recognize aspirations
Targeted solutions	Tailored solutions
Targeted, focused on	Aligned with, focused on
Aimed at "target" audience	Aimed at "listening" audience
Selected by factors in common	Recognized as community of individuals
Transactional potential	Relationship potential

The term "persona" was only first introduced in 1999 by Alan Cooper. He informally developed personas as a practical design tool for engineers and developers to better internalize the mindset of their audience when designing a new software. With the transition from mass marketing to target marketing, the need for more refined segmentation strategies emerged and the term "persona" gained popularity and eventually started to be used by marketers and not only with a User Interface approach.

Jenkinson (2009) came back to discuss the effect on business success of segmentation methods, providing a simple "test of effectiveness" self-audit for organisations decision-making. While Jenkinson most often suggest personas as a general strategy, others broach on employing personas thru a marketing approach to gain insights about a target market to life (Kotler & Kevin, 2009). More recently, João Félix (2016) performed a series of interviews with experts on management and marketing to collect insights on customer data acquisition

leading to profiling. He concluded that customer segmentation is important but refinement to a customer profiling and creating buyer persona profiles can better direct marketing efforts.

On the other hand, the concept is still frequently discussed in academia from a designing perspective. Steve Portigal (2008) follows such approach when discussing persona generation as a tool and highlights its benefits in the designing process as well as the risks of misuse. In addition to it, Tomaz Miaskiewicz (2011) provided a literature review of the benefits of incorporating persona into the user-centered design process.

The persona helps translating the available consumer numbers into humanized data to strengthen the understanding of software developers about users' life and behaviours, hence improving developers' capabilities to invent better solutions and features (Cooper, 1999). Similarly, for marketers, buyer personas are representations of a targeted market, brought to life through fictitious yet realistic profiles of one or more group of similar consumers, in terms of demographics, psychographic, behavioural and objectives towards a product or solution. The final output for a persona may include a photo, a name, a short biography and other embodying attributes that can help to efficiently convey the archetype of a group of thinking alike people (Kotler & Kevin, 2009). Figure 1 shows an example of a persona translated into a visual and informative profile.

Figure 1 – Example of a Persona profile. *Reproduced from Buyer Personas: How to Gain Insight into your Customer's Expectations, Align your Marketing Strategies, and Win More Business (p. 8), by A. Revella, 2015.*

Companies such as Delta Airlines and Unilever employ personas to help their marketing and sales team to visualize customer segments (Buss, 2006). One of the most mentioned cases in literature is "Katie", a 25-year-old single woman persona constructed by Unilever for the promotion of Sunsilk's haircare products. Seeing Katie as a "real person" helped both the company's marketers and its agencies to better understand that female target segment and efficiently tailor any marketing strategies that would better fit Katie's personality and preferences, such as what TV show Katie would watch or how would she react to specific pieces of communication (Wood, 2007).

Buyer personas had clearly demonstrated to improve results for enterprises across a wide range of industries. In 2013, NetProspex - a B2B data management - was looking to provide more relevant content to increase engagement and conversation on their website. According to Banko (2014), their idea was to build different identities for each website visitor, offering more relevant topics and resources:

The site achieved this customization by marketing to NetProspex's different buyer personas with carefully targeted content through progressive profiling, which allows visitors to provide data over time in exchange for resources such as product guides and webinars. The effort would also encompass the alignment of email marketing automation with buyer behavior to build a better relationship with visitors as they move along their purchase journey (Banko, A. 2014, p. 8).

NetProspex had identified three main personas to target based on their purchasing and behaviour as current customers; then the team recruited customers matching each persona prospect to conduct both interviews and focus groups to obtain answers about their key pain points, motivations and what type of content they'd like to consume. The results helped the company to formulate their messaging, highlighting what type of content would work within each targeted group and what needed to be updated to generate rapport. When the new website was launched in April of 2013, NetProspex experienced a 900% increase in visit duration; a 1000% increase in the number of pages visited; 111% increase in email open rate; and 46% increase in conversions.

Similarly, Whirlpool has used persona-based design to raise its site customer loyalty by 68%. From a brand perspective, the company had already a target audience, the "active balancer", those who balance different life demands, from careers to family. In that context, the study of personas helped the company to deep dive into their targeted audience and

understand consumers' different needs. The team conducted a series of interviews with consumers, as well as call centre data and website analytics to create four primary personas: purchaser under duress (those whose refrigerator has just broken and are in an urgent situation such as ice melting over their kitchen floor), planned remodeller, new owner and owner with repair. Each persona was given an age, name, family, occupation and income, along with additional paragraphs explaining their mindset towards the website usage. Those personas helped the development team to create different path and functionalities having in mind each of the identified personas. For example, one of the personas was named "Remodeller" and had a more considered purchasing path and usually explored more the available options. To meet such behaviour pattern, developers included functionalities such as a "showroom" highlighting important areas to consider while remodelling a home, or the ability to email a product, which allows the consumer to send a product to your inbox at home and discuss the options with the family, for instance (Marketing Sherpa, 2005).

Another great buyer persona demonstration is Skytap, a cloud automation solutions provider who had reported a 73% increase in opportunities generated through online marketing after the usage of buyer personas to tailor its content strategy (Sutton, 2013).
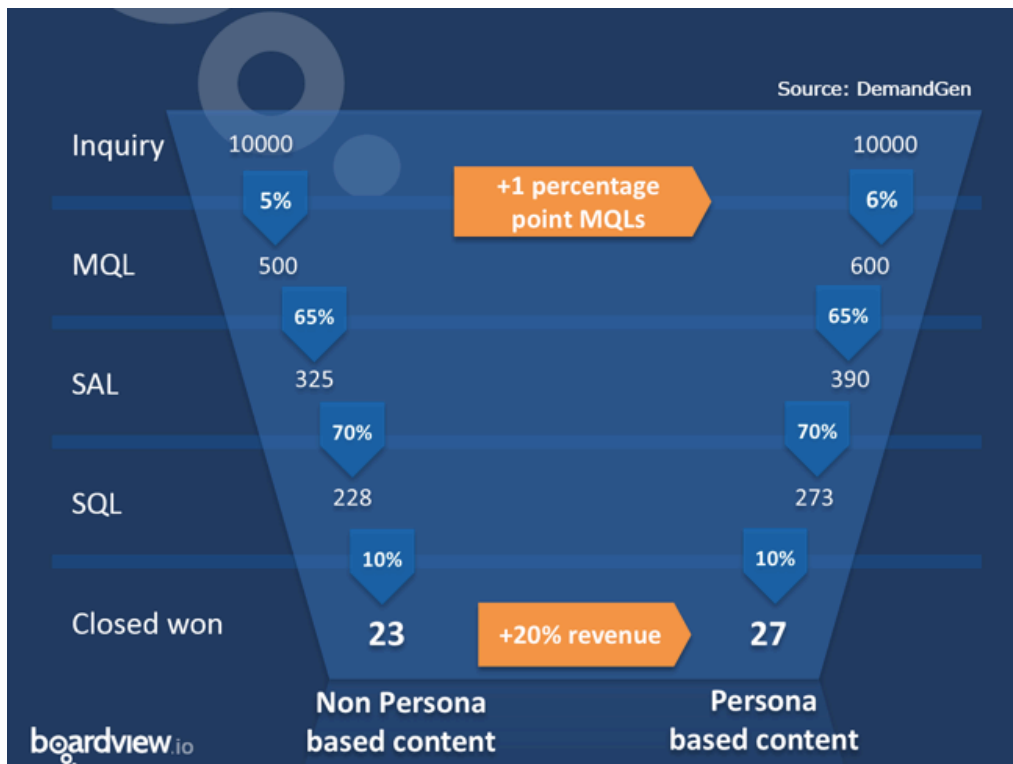
Figure 2 – Illustration of A/B tests comparing persona and non-persona-based content and its impacts in a Software firm's revenue. *Reproduced from Persona ROI Case Study – How Personas Help You Close Deals, by F. Riemersma, 2015. Boardview.io.*

There are many other successful stories of Persona ROI (Return-of-Investment) available to justify the investments of building and updating personas. Figure 2 represents the case of a software firm that investigated the impact of personas by performing A/B tests on their email campaigns of which one of the variables was customized content based on persona insights. The firm found out that persona-based content was able to increase engagement almost six times more than non-persona-based content when targeting cold leads (58% versus 10%). As illustrated by the Figure 2 funnel, the Marketing Qualified Leads (MQL) increased from 5% to 6% its conversion rate; although the small raise of 1% may seem insignificant, it was able to increase 20% in revenue at the bottom of the funnel.

Although it's easier to make assumptions based on one's own knowledge, the current scenario of social and technological rapid developments tends to open more doors to companies with a consumer-centric perspective and able to rapidly adapt their businesses and communication to buyers' needs. By better understanding the buyer decisions, marketers and managers will then be able to become a reliable influencer through the buying decision process (Revella, 2015).

2.2 – Methodologies for Persona Generation

Among the investigated literature, many authors suggested different methodologies for building personas, either for designing or marketing purposes. Alan Cooper (1998) interviewed several colleagues and acquaintances who would likely be among the intended audience of a project he was working on and got to know their behaviours and engage in a play-acting that would allow him to evaluate ideas from the audience perspective. Traditional methods of market research and user evaluation were applied for persona generation such as interviews, workshops with collaborative exercise, field study and surveys as suggested by Cooper (1998), Nielsen (2004), Johansson & Messeter (2006), McGinn & Kotamraju (2008), Siegel (2010) and Awad (2011). Moreover, Seán Kelly discussed the application of data mining and cluster analysis to the basic problem of market segmentation thru a priory acquired database such as data from enterprises' CRM (Customer Relationship Management) tools (2003)

Analysing the Table 3 below, it's evident that big data and social data are emerging as relevant methods to explore insights thru data mining. Krasavac et al (2016) discussed the emergence of big data and its application in marketing and customer profiling, also suggesting the use of a CRM database. More recently articles tested a new methodology using NMF (non-negative

matrix factorization) and automated generation of personas based on data from a popular YouTube account (Jansen, Jung, Salminen, An & Kwak, 2017).

Table 3 – Breakdown of Methodologies by Literature Review's Authors

	Interviews	Prior Database	Workshop	Field Study	Ethnographic	Survey	Big Data	Social Data	Site Analytics	Automatic
(Cooper, 1998)	x									
(Kelly, 2003)		x								x
(Nielsen, 2004)	x		x	x						
(Johansson & Messeter, 2006)			x		x					
(McGinn & Kotamraju, 2008)	x					x				
(Siegel, 2010)	x		x	x						
(Awad, 2011)						x				
(Chroneos-Krasavac, Soldic-Aleksic & Petkovic, 2016)		x					x			x
(Dagustan, Buchory, Satya & Diryana, 2016)						x				
(Jung, An, Kwak & Jansen, 2017)									x	x
(Salminen, Şengün, Kwak, Jansen, An, Jung, Vieweg & Harrell, 2017)	x								x	x

2.3 – Conclusions

Within the reviewed literature, some of the suggested future research was considered in this dissertation while other papers contributed to the body of knowledge regarding the conceptualization as well as the practical application of segmentation and personas. Several authors proposed further examination of persona use and methodologies (Miaskiewicz, 2011; Félix, 2016; Jung, An, Kwak & Jansen, 2017) encouraging multi-disciplinary mixed-method studies in the field of social media analytics (Salminen, Şengün, Kwak, Jansen, An, Jung, Vieweg & Harrell, 2017). Related to that, the problematics of disseminating the knowledge of

personas to large groups within the enterprise was also among the suggested future research (Nielsen, 2004). Therefore, it was identified an unmet demand for:

- Exploring the power of social data as an existing and non-biased conversation database;

- Exploring cost-effective persona generation methods that allow marketers and managers to uncover buyer personas without big market research investments and extensive timelines;

- Providing actionable insights and rich data representation to marketers and managers in an easy-to-digest format to ensure the dissemination of the knowledge among all areas of the enterprise.

3. Research Question & Design

3.1 – Research Question

From the analysis of the literature, it's possible to realize a predominance of traditional research models for market segmentation. Although data-driven methods have surged in the last years, currently solutions yet don't fully use the power of unbiased social data sources and are often not accessible for the average marketer or manager decision-makers due to its complexity and extensive resources requirements. Hence, the following research question emerged from the literature review:

- How to develop buyer personas using social data conversation, thru a cost-effective methodological procedure that can be implemented by marketers and managers and provide easy-to-digest outputs within less money and time efforts than traditional market research methods?

3.2 – Research Design

As the intent of this study is to develop a methodological procedure for building buyer personas using social data interpretation, this research will follow the Design Science approach, as pictured in Figure 3. The Design Science approach is an Information System (IS)-based methodology that aims to build knowledge not only limited to understanding a field but also to develop alternative solutions for real-professional problems (Aken, 2005).

In the design sciences, academic research objectives are of a more pragmatic nature. Research in these disciplines can be seen as a quest for understanding and improving human performance. It is solution-oriented, using the results of description-oriented research from supporting (explanatory) disciplines as well as from its own efforts, but the ultimate objective of academic research in these disciplines is to produce knowledge that can be used in designing solutions to field problems. (Aken, 2005, p.22).

This study unfolds in a United Kingdom technology-based firm that provide business intelligence solutions, seeking also opportunities for continuous improvement in its process and standardization of its methodologies. In 2017, the company was selected as a "Best of British" company by the UK Trade and Investment authority to represent UK innovation at the South by Southwest (SXSW®) Interactive Conference. The author of this research has been working for 2 years in the company as Senior Manager in the Insights team, therefore, ease access and future application of the findings and outcomes of this study were one of the main reasons for choosing this unit of analysis.
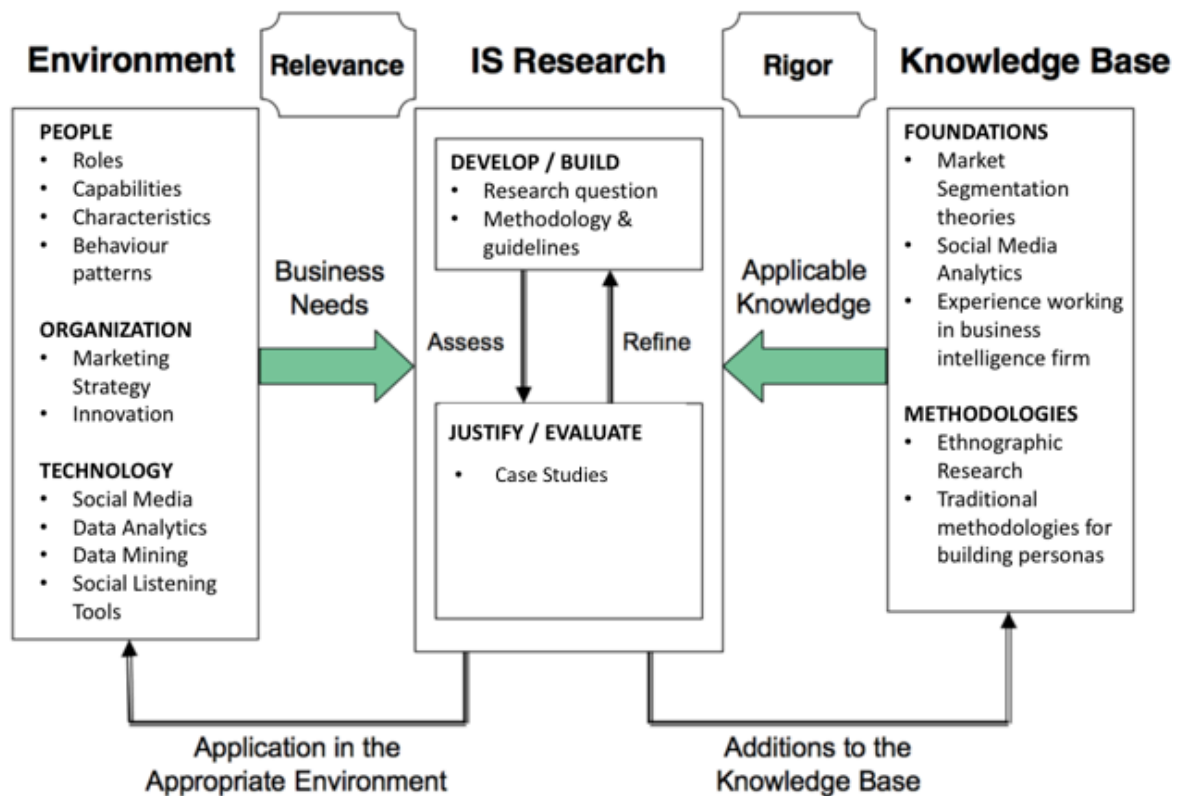
Figure 3 - Framework for Understanding Design Science Approach *Note. Adapted from Design Science in Information Systems Research (p. 80), by A.R. Hevner, S.T. March, P. Jinsoo, and R. Sudha, 2004.*

The Figure 3 above illustrates the framework to be used in this study. The Environment characterize the problem space (Simon, 1996) of the research. It includes the perspective of analysing people (consumers) from the perspective of persona-development, together with the business (organization) needs and while applying existing technologies. The left part of

the framework defines the business needs that will be covered by the activities in the two middle quadrants, while the right side of the framework underlines the Knowledge Base for carrying out the study, which includes foundations and existing methodologies that will serve as guidelines for performing research activities in the two middle quadrants (Hevner et al, 2004).

The research itself will consist of two main phases:

- **Develop / Build** phase which aims to develop a structured methodological procedure for using social data in the byer persona-generation process. The study of usual ethnographic method and traditional methods for building personas will be crucial for elaborating an efficient "virtual ethnographic" version of the current processes, aligned with the available technologies.

- **Justify / Evaluate** phase which aims to validate the new methodological procedure elaborated in the previous step. This will be achieved by ensuring the final deliverables are fulfilling the literature gaps and opportunities previously stated here. The procedure will be empirically applied, and case studies will be presented at the end, including a buyer personas summary within a specified market/product.

4. Developing buyer-personas through social listening

4.1 – Traditional market research methods constraints and limitations

By analysing the literature review, it's noticeable the predominance of traditional market research methods for market segmentation such as ethnography and interviews. These methods, however, are normally very costly and can take several months or even years to produce the desired outcomes. The MSN Explorer Persona effort for Microsoft started in January of 2000 and lasted about 10 months (Pruitt & Grudin, 2003), while the persona development at Cisco took several years (Nieters, Ivaturi & Ahmed, 2007). Although these cases were focused on a design perspective, rather than a marketing perspective, it gives a rough estimate on the great amount of resources a company necessary to allocate.

According to Pruitt & Grudin (2003), one of the key problems that usually take place when developing personas can be non-believable characters, mainly due to the lack of data during the persona conception process or an unclear connection between the data and the persona. Kotamraju (2008) perceives the absence of original research as the potential cause for such unclear connection, highlighting the importance of conducting original study as part of the persona development process. Pruitt & Grudin rationale showed that they had no resources to do original research, hence had to rely on third parties' field studies.

Another potentially rich source of market data are customer databases. To know their customers, the company compiles information and build a database for customer relationships management (CRM). From the literature review, it's possible to observe a few applications of customer database for persona development. However, the process of creating and maintaining a CRM also infer several risks which should also be considered. First, building and maintaining a customer database involves a sizable investment in software, hardware and skilled staff. Secondly, it requires a large effort from all the staff to ensure the database is well managed and used across the company. Furthermore, some customers may not want to release personal information to the company or in some circumstances it may not even be worth to build a database management process such as products that are an once-in-a-lifetime purchase or when sellers have no direct contact with the final buyers (Kotler & Kevin, 2009).

4.2 – Owned channels analytics: first steps for consumer insights and its limitations

To have valuable customer insights, an enterprise might not always have to look much further. Tools such as Facebook Insights and Google Analytics can be great initial sources to uncover customers' behavior patterns online when related to a brand or product, without having to rely on primary market research and potentially simpler to manage.

A less resource-intensive way to get started with buyer personas is by using website and social media analytics. Analytics can quickly and easily provide a wealth of data that will enable you to understand and more effectively engage your audience. Using Google Analytics to build a buyer persona is a great place to start before investing a lot of time and money into more detailed methods (Engel, 2015, p. 20-21).

Through Google Analytics, for instance, one can see what users are searching for when they find the company's or a product's website, as well as more detailed information of what devices are they accessing the website from, what are the most clicked links and how long do they stay there. Similarly, Facebook Insights, and other social media analytics such as Twitter Insights, help to unveil valuable statistics such as what type of content is working, the demographics of your audience (e.g. gender, age and location) or even the audience's main interests.

Statistics provided by analytics can indeed be a near to no cost for unveiling personas to a business. It's an especially resourceful method for online stores, who can benefit from the detailed behavior pattern of users online and use that information to customize and automate communication (e.g. keep showing advertising of iPhone's earbuds for people who have recently bought an iPhone at your online store).

However, this method does not allow an in-depth study of customers' needs, objectives and preferences as an interview or focus-group would, it's primarily quantitative-based and lack enough qualitative insights to trace a proper buyer persona profile. Following the New York traveller example, statistics alone would be unlikely to tell the triggers for purchasing a new phone, the emotional journey behind choosing an iPhone over another smartphone, the pros and cons pondered during the process or even the drivers to the final purchasing decision. Moreover, to take advantage of analytics statistics, it's assumed that a company already owns and maintains a website and social media channels, which not always apply to

new businesses or at least not to a resourceful degree - if a website or Facebook page have low traffic or engagement, there might not be significant statistics to guide future actions.

Monitoring its own websites and owned social media channels is still crucial to any modern company, no matter if big global enterprises or small start-ups. These data becomes even more powerful when analysed in parallel to other methods that allow analysing the consumer voice.

4.3 – Social Listening & Market Intelligence

The introduction of the World Wide Web to the public has been reshaping the world in the last two decades and so has social media, including blogs, microblogs, forums, photo-sharing platforms, web chats and social networks such as Facebook. By 2021, it's expected that one third of the world population to be active social media users. In its well-known annual *Digital Global Overview*, We Are Social (2017) highlighted some milestones for the rapidly growth of internet across the world: more than half the world now uses a smartphone and more than one in five of the world's population has shopped online in the past 30 days. In Europe and in the Americas, the internet population is more than 70% of the total population; 412 million Europeans are active social media users.

While social media becomes largely used, more consumer data is available online. Twitter generated 10 terabytes per day in 2010. Facebook generated 25 terabytes per day in 2009 in log files alone. By monitoring social media conversation, one can gather valuable consumer insights and market intelligence for not only marketing purposes but also for product development.

Social Listening is the process of actively monitoring social media conversation to understand what customers may be saying about a brand, a product or an industry. Social data extracted through Social Listening can provide valuable insights to a wide range of areas within an enterprise, from measuring the brand's online visibility or measuring a marketing campaign effectiveness to identifying new opportunities to engage with consumers or tracking competitors' activities on social media.

For marketers, social listening can be used to guide decisions and strategies. *"For the first time ever, marketers can gain real-time access to a compendium of consumer behaviours, attitudes preferences, ideas, affinities and sentiment"* (Gleansight Benchmark Report, 2013. p. 3). A deep understanding of the consumers can help the company when planning future marketing strategies such knowing what messages will most likely reverberate within that

specific market segment or what channels are consumers using to discuss a determine product.

Social listening also brings many advantages for improving market research effectiveness and enhancing innovation within the development processes by gaining insights on customers' unmet needs and market demands, as well as tracking market response to new products.

1. Social data conversation enables one to listen to what an audience is saying organically. Because the data is collected in its natural ecosystem, there is lower risk for bias and behaviours will be placed into an organic and unsolicited context.

2. Monitoring social data allows one to identify any issues in near real-time and apply any necessary changes to an enterprise's strategies faster than ever. Social listening tools also allow to research past data, which means that it's possible to understand behaviours relative to a specific event or defined point in time and as well as its evolution in a given period.

As the millennials are expected to comprise three-quarters of the global workforce by 2025 (EY, 2015), the world becomes more populated with "digital natives" and both internet and social media penetration increases its ratio in the global population, broaden even more the power of social listening.

4.4 – Filling the Innovation Gap

"At some point many businesses stopped seeking efficient approaches and started clinging to the established way of doing thing. They became suspicious about innovative tech that challenges existing methods. Many became positively precious about sophisticated analytical processes that should never be questioned" says Omiyale (as cited in Raconteur 2017, p.1). The quote was voiced in 2017 by the Vice-President at a market research solutions provider, and explains the transition that the market research industry now faces. According to him, inaccurate polling results and other similar cases started to raise questions and opened opportunities for market research to evolve. *"The future of market research lies in combining existing research skills, technologies and methodologies with new data science and social listening approaches that enable researchers to deliver true business insight"* (Omiyale, as cited in Raconteur, 2017, p. 5).

The key take-away from Omiyale's testimonial is that market research industry should move towards to be an insight-industry instead. Automation and new technologies permit accessing valuable data and therefore empower business owners or decision makers.

The GreenBook Research Industry Trends Report (GRIT, 2013) is a leading vehicle that reports trends in the market research industry. Among their survey, respondents were asked the main issues they faced that have triggered a change in to how they collect data within the past year. As seen in Figure 4 below, budgetary constraints placed the biggest issue prompting changes in data collection, with both clients and suppliers looking for cost-saving alternatives to get the required information. According to the GRIT report, budget constraints was also considered the biggest issue in the previous year and other issues such as "changes in staffing or layoffs" or "the economy - business growth/slowdown" could similarly reflect the same type of budgetary pressure among the market research industry. Technological changes were also a driver of changes in data collection such as "client demands for innovation" or "increased use of mobile communications", reflecting the increase of opportunities brought by technology.
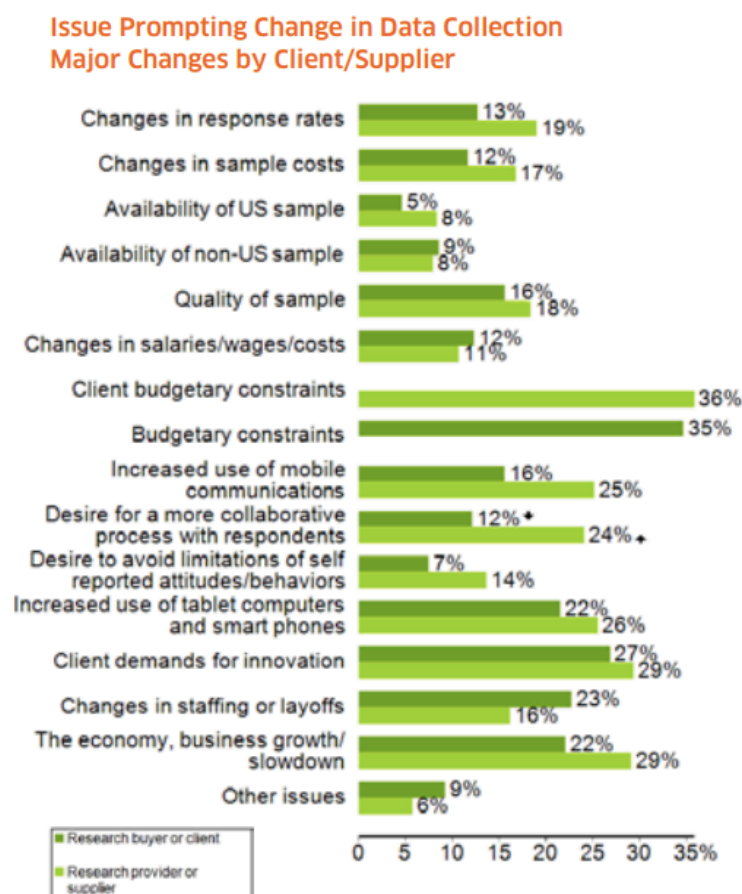
Figure 4 – Issue Prompting Change in Data Collection by Client/Supplier. *Reproduced from GreenBook Research Industry Trends Report (p. 12), by GreenBook, 2013.*

Respondents were also asked to answer what options or solutions they are exploring to address these data collection issues. Figure 5 resumes the most mentioned solutions. Both clients and suppliers were actively exploring innovative solutions by either new methodologies (50%) or new technologies (48%).

	Client	Supplier	Net
Exploring new methodologies	51%	50%	50%
Exploring new technologies	38%	50%	48%
Doing more with same resources	31%	30%	30%
Exploring new sample resources	13%	25%	23%
Close partnering with quality access panels	15%	22%	21%
Investing in more in-house technology	18%	20%	19%
Working longer hours with less staff	13%	14%	14%
Spending/charging less for the same services	4%	16%	14%
Opening direct channels of communications with customers	16%	11%	12%
Negotiating discounts or shorter timelines with vendors	13%	12%	12%
More diligent validation	6%	11%	11%
Changing ratio of insourcing vs. outsourcing of research functions	17%	10%	11%
Recommending higher respondent incentives	7%	11%	10%
Asking for sacrifices from vendors	13%	8%	9%
Asking for sacrifices from employees	8%	8%	8%
Other	7%	6%	6%
NET	99%	97%	98%

Figure 5 – Solutions for Data Collection by Client/Supplier. *Reproduced from GreenBook Research Industry Trends Report (p. 13), by GreenBook, 2013.*

These statistics forecasted the emergence of automation and social data analysis in market research. In this context, social listening can be placed as an effective response to the pressure of doing more with less or the same resources. As discussed in previous chapters, social listening is not only an effective research method, but it can also be a cost-saving alternative. Table 4 below shows an estimative of costs per research method compared to a breakdown of costs for conducting a social listening research.

Table 4 – Estimative of Costs per Research Method

Traditional Market Research Methods	Social Listening Research Methods
Survey	**Paid Social Listening Tool**
- By phone: $5,000 - $15,000 with an average of $40 per person surveyed. - By e-mail: $3,000 - $5,000 (Lee, 2002)	Platforms that help to gather and structure available online data can be free but richer functionalities and capabilities are available from $800 - $2,000 monthly or per dashboard (e.g. Synthesio, Crimson

	Hexagon or Brandwatch)
Focus Groups	**Data Analysis**
- Range from $4,000 - $7,000 per session, with a recommendation of at least two sessions. Compensation to participants may vary from $35 - $50 per consumer; $50 - $100 per professional individuals; $100 - $150 for executives. (Lee, 2002)	A 3 weeks project performed by specialized business intelligence companies versed in social listening can range from $5,000 - $25,000 but can go as low as $3,000 depending on project scope and needs.
In-Depth Interviews	
- $200 - $300 per consumer interview plus incentives; $300 - $500 per business interviews plus incentives. (Vernon Research, 2018)	

Above all, it's important to highlight that each method has its own advantages and disadvantages and both costs and timeline necessary to get to results will depend on the objectives of the research as well as many variants such as the incident rate, whether research is conducted masked or not, geographical coverage, incentives provided and sampling size. This dissertation does not aim to provide an accurate comparison of methods prices and demanded time but rather a high-level estimative to rationale the advantages of social listening flexibility. Figure 6 attempts to visually demonstrate the flexibility and scalability of social listening studies compared to conventional methods. Although all methods are impacted by variables of each project, social listening scalability is rather more vertical when compared to conventional methods. For instance, a social listening research can easily go from one-market project to a 10-markets project requiring a significantly less effort than doing the same on interviews, focuses groups or even surveys, avoiding the burden of running multiple sampling phases for all the different targeted locations. As illustrated in the graph, social listening also holds the advantage of providing both qualitative and quantitative insights together, which is sometimes hard to achieve through conventional research methods. While a simple survey can nowadays be done easily and cheaper, it could hardly provide actionable insights as a standalone method and without having to further explore questions qualitatively.
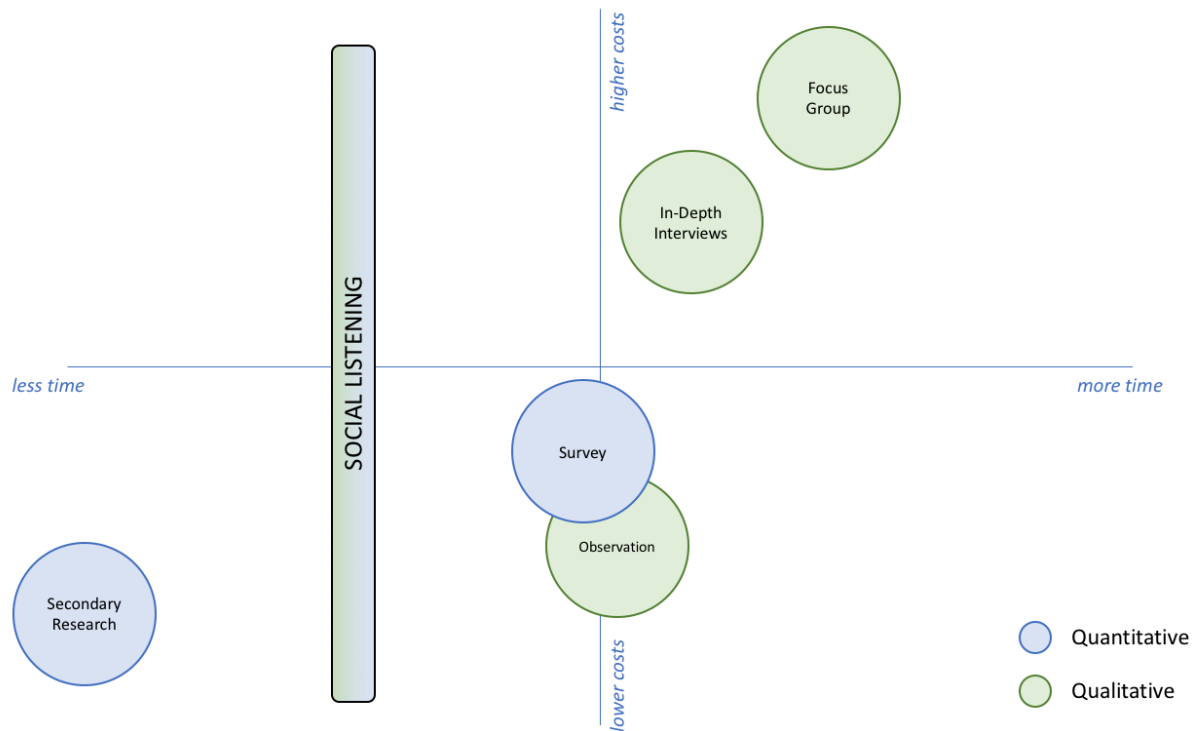
Figure 6 – Visualization of main methods' capabilities, costs and time efforts. Time-efforts are based on *What Goes In to Market Research and How Much Does it Cost?* (Wessell, 2018) and personal experience

With free tools or third-party solutions provided for as low as $3,000 - without having to lose quantitative significance and timelines seldom requiring more than 3 to 4 weeks of work - it certainly open possibilities to small companies and young start-ups who otherwise wouldn't be able to unveil a deeper layer of consumer insights. Moreover, it also allows insights in almost to real-time, essential for start-ups and fast-growing technological companies who need to keep up with the always changing market demands.

The IT leader Cisco Systems is a great example of company that has already tasted the power of social listening. Using Salesforce Radian6, a social listening platform, Cisco found a uniform way to gather, understand, share and act on social media conversation about the company. According to a case study published by Nucleus Research (2013), in less than 1-year the company reduced marketing costs, improved profits and was able to truly understand and build empathy with their customers, achieving a 281 percent Return-on-Investment (ROI). Cisco's social media command centre continuously monitors topics of interest, trends and sentiments but can also address specific events in time such as a new product launching or sponsorship and coverage of events and trade shows. Insights are then shared not only with the marketing team but all company's sectors, inciting data-driven innovation across all areas.

Social listening has also helped Cisco to increase its profits and reduce market research investments:

Because Cisco can mine and analyse customer and partner social mentions in a meaningful way, they can have a better overall view of sentiment and behavior that would be cost-prohibitive to achieve through traditional means such as market research surveys… Insights gained from social listening have resulted in sales opportunities that might otherwise not have been recognized (Ciarallo, 2013, p. 12).

Although Cisco's investment in its social media command centre was likely a high one, this case illustrates the potential impacts of social listening in an innovative culture which can equivalently be applied by Small and Medium Enterprises (SME) and start-ups.

Social intelligence agency Convosphere (2018) has recently published a case study looking at how social listening predicted Trump's win when traditional polling failed. While all traditional polls and prediction models forecasted Clinton's victory on the American 2016 Elections, social listening was able to deep understand patterns in social media conversation and come to a more accurate prediction for the president's run results. Although the 2016 Elections were not the first time that tracking polls have been wrong, it was the scale of "wrong" that surprised researchers and reverberated globally, with media sources all over the world questioning how polls could have missed it to such a scale. Election polls are especially important because of its ability to drive behavior: high margins can determine if American voters will bother to get up and go out to vote or not. Convosphere's CEO, Jackie Cuyvers, explains the importance of running this case study and the benefits of social listening versus traditional polls:

*If traditional methods are failing us, what are we going to do? (…) One of the reasons that we were so confident doing that in social data is because there are some strong advantages that we've already found in other studies about social data over traditional methodologies (*Cuyvers, J. 2018, 1:32)

She then highlights the three important benefits as:

1.   Being able to study both real-time data, to predict near-future outcomes, as well as historic data, to reveal not only what happened but "why";

2.   Organic opinions, as people on social media are less guarded when voicing their opinions and tend to be more honest;

3.   Social engagement could better predict voter turnout as people are already voluntarily going online to discuss their candidates.

The team has retrieved more than 25 million social conversations over the period of the last 3 months prior to the election day. To analyse the data, they have taken an approach of observing the out-layers behavior to then deep-dive into inner-layers drivers for the observable patterns. First, the team looked at daily volume and sentiment for both Hilary Clinton and Donald Trump, to try to identify any high-level discrepancies that could lead to trends. As Figure 7 shows, sentiment was surprisingly very similar for both candidates, independently of the events that occurred during the election time: people expressed positive sentiment for their candidate of preference, but they would also equally express negativity towards the opponent. This mean that sentiment was not an indicative of any difference between the candidates.
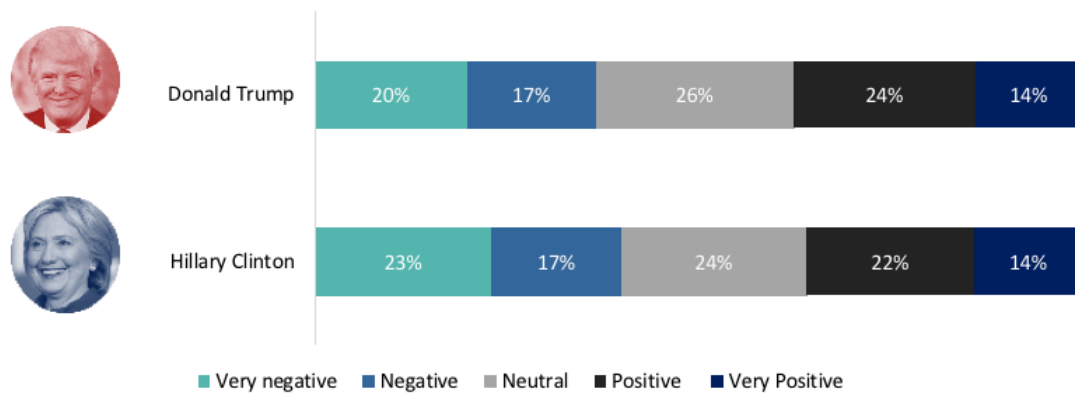
Figure 7 – Sentiment Breakdown for the two main candidates in 2016 American Election (Cuyvers, 2018, 3:50)

Then, the team looked at the daily volume of social media conversation and noticed that there was consistent stronger engagement for Donald Trump; he was often inspiring more social media conversation and responses to debates, as shown in Figure 8.
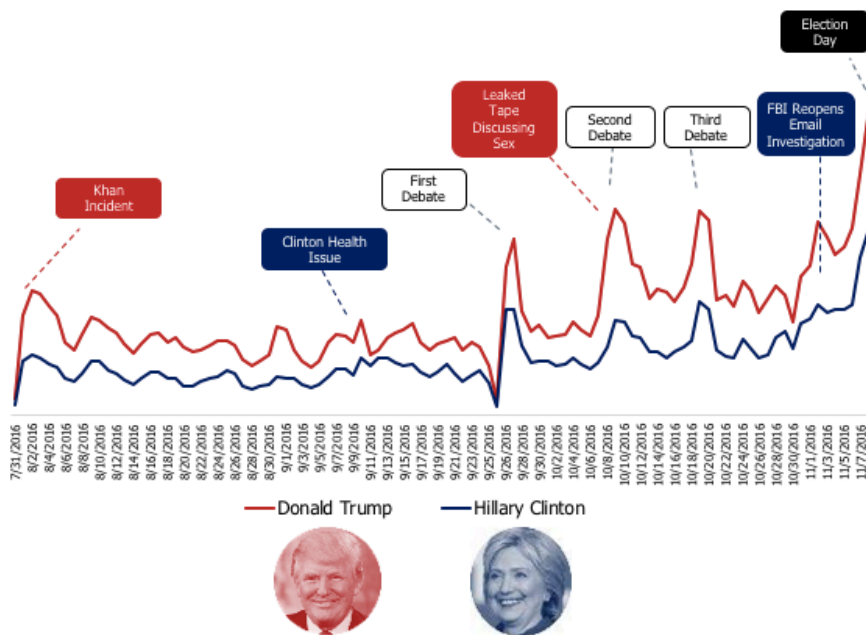
Figure 8 – Timeline of Volume for the election period for the two main candidates. Graph shows that Trump inspired more data volume and social media conversation than Hilary (Cuyvers, 2018, 4:14)

Observing Figure 8 helped the team to understand that what polls could not capture was the level of enthusiasm and motivation a voter had to convince other to go out and vote. This is an especially important behavior in the United States, where voting requires an extra amount of engagement as people have to vote on a weekday and often having to wait in line.

To further understand the levels of engagement and enthusiasm for each candidate and across the elections period, the team identified conversation that indicate voter commitment and pinpoint where they land on an 8-stage journey as described in Figure 9. Although this was a more structured project and not aiming to profile buyer-personas, the Voter Journey can work similar to a "Buyer Journey" and clearly exemplify how different stages of the journey can unveil distinct audience groups and engagement levels, helping to understand and even predict behaviour. From a product or service perspective, "Buyer Journey" tagging can classify audience groups by their commitment to a brand, product or even a new technology, which could later be translated into buyer-persona profiles.

Figure 9 – Voter Journey Map. Displays core emotional states driving commit to a candidate, with Affinity, Influence and Evangelism placing the strongest support and advocacy for or against a candidate (Cuyvers, 2018, 8:19)

In the case of the 2016 American Elections analysis, Affinity, Influence and Evangelism represented the strongest stages, with the highest levels of enthusiasm towards a candidate. This mean that people on these stages were more willing to not only voting for a candidate but also advocate for their preferred option or fight against the opponent.

- Affinity conversation happened where a person communicates positivity towards a candidate and disliking towards the opponent;

- Influence measures the degree of how voter's opinions of candidates influence the opinion of other voters.

- Evangelism conversations happened where a person actively tries to persuade others, either by promoting their preferred candidate or by trying to denigrate the opponent.

When comparing the volumes of conversation per journey stage, the ratio between Trump and Clinton is highest in the final 3 stages, especially in Affinity and Evangelism, as shown in Figure 10. The team has also explored the emotional levels in the conversation for each candidate and found that while 39% of Trump's voters expressed "strong emotions", only 18% of Hilary voters have shown the same intensity in their messages. This confirms that by understanding the different audience groups and their emotional intensity, social listening

could predict the outcome of American Elections in 2016. The same data here described could be then translated into profiles to target the different personas (e.g. Influencers or Evangelists) during the election campaigns.
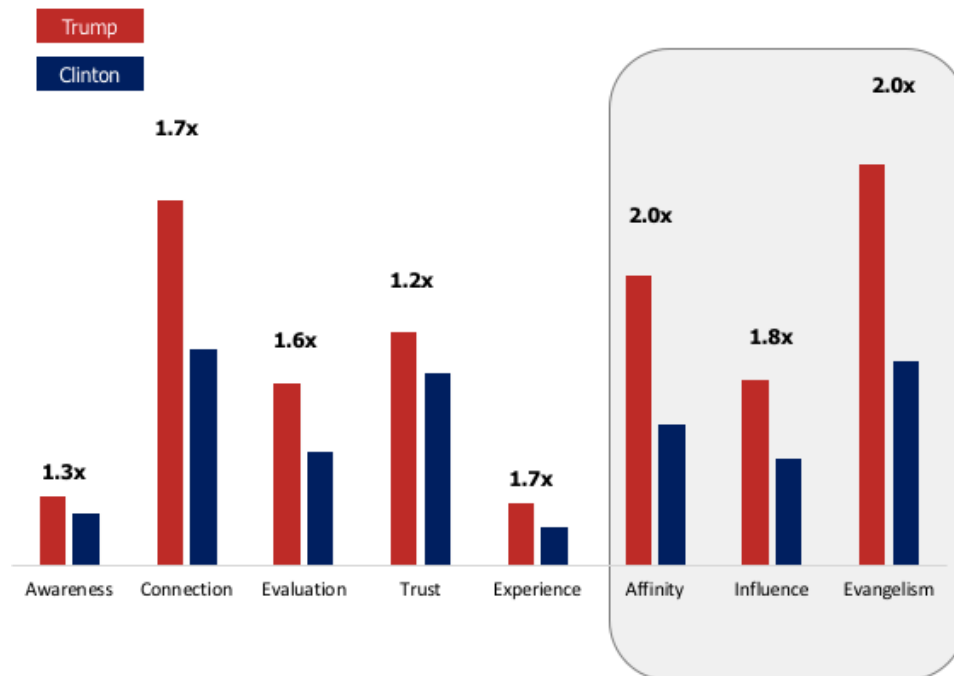
Figure 10 – Journey Stage Volume per candidate (Cuyvers, 2018, 9:08)

4.5 – Challenges and Limitations of Social Listening

During the last years, social listening has evolved from its data privacy concerns to a promising market in technology and market research. At its early years, social listening was often compared to a "Big Brother"; by 2012, 51% respondents from a survey stated they wanted to be able to talk about brands on social media without being listened (NetBase - J.D. Power and Associates, 2012). However, as social media unfolded into more mature communication channels and with more elaborated privacy settings, users started to accept the idea of having brands listening to customers' conversations, sometimes even expecting brands to do so (Davey, 2015). Social media analytics has become a great business, with

MarketsandMarkets (2017) forecasting the market will reach the value of $2.73 billion in 2019. Although there are many benefits taken from social listening, there are also a number of challenges and limitations that can restraint projects' outcomes.

Data privacy is yet a challenge often raised regarding monitoring data online. The late scandal affecting Facebook and Cambridge Analytica was vastly discussed on media channels and symbolizes the huge importance of companies' responsibility when handling users' personal and social data. Data providers and social listening tools are required to follow data privacy laws such as the General Data Protection Regulation (GDPR) in Europe, which means, for example, that personal data may only be collected and used upon consent or when it's public or else anonymized.

Despite the fact that Internet has given immediate access to a huge volume of data, even the most advanced tools and technologies won't be able to automatically uncover insights and deliver valuable conclusions to organizations. While social listening tools can reveal quantitative data, qualitative insights are much more complex to achieve and usually require business analysts fluent in the technology and in analytical skills to collect relevant data and visualize it in an easy manner to detect trends. This study aims to provide a beginner but valuable methodological procedure to enable start-ups and SME entrepreneurs to also benefit from such valuable data and insights themselves.

Limitations of a social listening project will also directly be affected by what industry or product one is looking for. Because social media is just one channel for conversating, there is a chance that frequent posters provide a skewed or incomplete picture of the landscape. For instance, dissatisfied consumers may be more active voicing their opinions, depending on the product category. Lack of data may also become a constraint when dealing with delicate subjects of study or a minor niche. It's important to keep in mind that many people on social media are passive users and rather than actively posting they are merely observing and looking for their peers' opinions.

With the new General Data Protection Regulation (GDPR) taking place in May 25th 2018, companies, data providers and social media channels will need to change the way they seek and handle data. The new rule forces enterprises to be accountable for what the data they use as well as ensuring consumers have clarity, fully understanding and consent of how their information is being processed. The methodological procedure here described relies on social listening tools that will only collect publicly available data accessible to anyone. However, some platforms have already enforced data restriction terms. For example, Instagram has decreased access from 5000 API requests per hour per account to 200 API besides limiting

access to personal data. This can mean a restriction in data access for enterprising processing huge volumes of personal data but most likely won't reflect significant impacts on consumer insights (Bingham, 2018) such as the case of this study.

Similarly, the shift of the Boomers generation from Facebook and Twitter to "invisible" networks such as Snapchat can also represent a challenge for social listening in the future as monitoring tools can't track conversation that are not public and therefore companies may miss what these consumers are saying. Although social listening it's a cost-effective solution for gathering consumer insights on a flexible timeframe, in ideal circumstances, it can be used as a complement, helping to better scope and target further market research efforts.

4.6 – Will social listening substitute traditional market research methods?

The increasing popularity of social media research in the latest years doesn't mean that conventional methods have become obsolete. To get a real understanding of the customer, businesses have to see them from different perspectives. Also, although social listening has many benefits and advantages, it still falls short at certain points and therefore could not entirely replace other more traditional methods. In this context, technology is an enabler, allowing automation and more effective ways to collect and analyse data, and it's arguable that:

The future of market research lies in combining existing research skills, technologies and methodologies with new data science and social listening approaches that enable researchers to deliver true business insight. This means taking a collaborative approach with other business disciplines to deliver understanding of all the data generated, ensuring it can be digested, manipulated and acted upon more quickly to drive strategic change (Omiyale, as cited in Raconteur, 2017, p. 5).

Traditional market research and social media research methods should not be explored by the industry as two opposite or separated things but rather a constant evolution towards greater outcomes. Previous chapters have highlighted the advantages and limitations of social listening; conventional methods also have valuable benefits depending on each project objective:

1. Market research has a long history and so conventional methods were for many years proven effective. The balance of conventional and social listening methods may help to counter balance any bias brought by strong social media research advocators.

2. As already briefly discussed in previous chapters, social media conversation limits the data to an online audience; traditional methods may allow gathering insights from a wider percentage of population, which can be especially important in third world markets where internet access is only available to a very small percentage of the country.

In sum, both methods have its own relevancy and applications in the market research industry. Businesses need to ponder the strong points of each and integrate different methods in the best way possible for each specific case.

5. Defining a methodological procedure for buyer persona development using social data

To structure a methodological procedure for buyer persona development through the usage of social data, this section will briefly explain the approach taken by traditional market research methods and construct a parallel between the current methodologies and similar alternative steps using social listening, narrowing it down to buyer persona development.

William G. Zikmund and Barry J. Babin (2010) well summarized the general pattern of market research in the following stages: 1. Research objectives; 2. Research Design; 3. Planning a sample; 4. Collecting the data; 5. Analysing the data; and finally, 6. Formulating the conclusions and delivering the results. Figure 11 portrays a flowchart of market research process, breaking down those six stages into smaller steps.
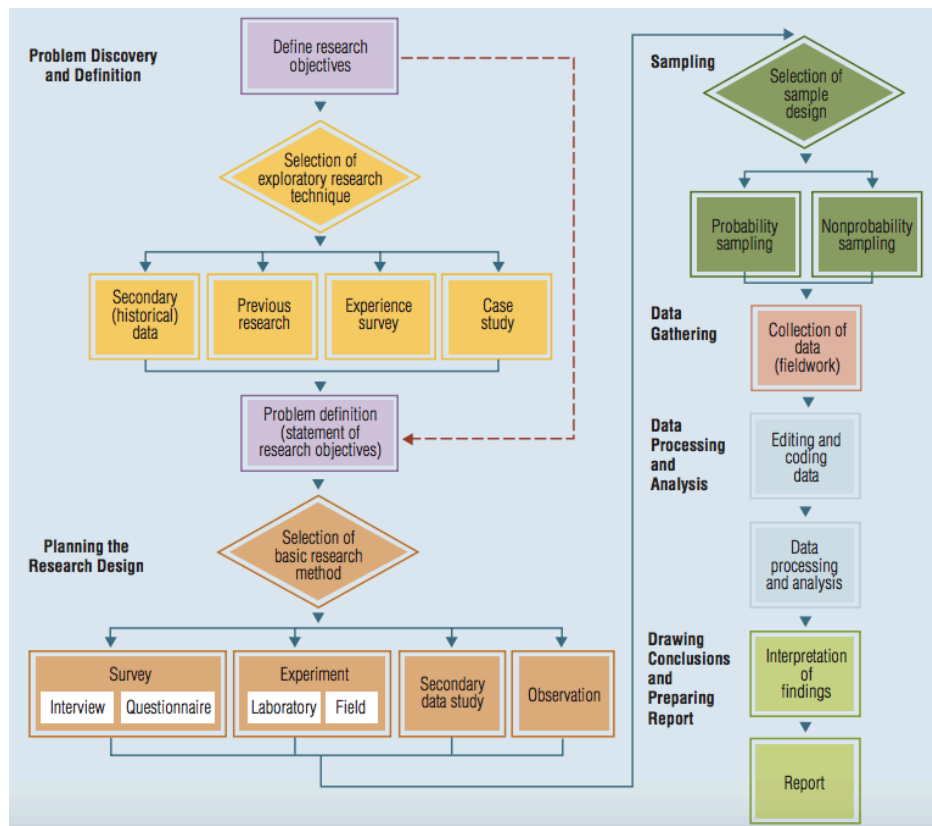
Figure 11 - Flowchart of the Marketing Research Process. Note: Diamond-shaped boxes indicate stages in the research process in which a choice of one or more techniques must be

made. The dotted line indicates an alternative path that skips exploratory research. *Essentials of Marketing Research (p. 52), by W. G. Zikmund and B. J. Babin, 2010*

5.1 – Problem Discovery and Definition

The first step in any market research is to formulate the problem one is trying to solve. In a business scenario, problems often come from an issue that managers are facing. However, one of the big challenges of market research is to be able to understand and translate the management problem into a research question that is possible to study. For example, "sales are stagnant" is a management problem which would need to be translated to a research problem such as what are the expectations and experiences of different groups of customers. While management problems focus on business action, research problems aim to provide the necessary knowledge to solve the management problem (Smith, 2012). In the specific case of buyer personas development, the management problem relies on: *who are our customers and how do I engage with each of them?* The research problem would be then translated into identifying, understanding and organizing archetypes to represent similar groups of customers based on their behaviours, unmet needs and expectations.

Whenever possible, the market research process can be improved by employing scientific methods. In a buyer persona research project, the organization could start by developing hypothesis based on any prior research or preconception of their target segments. In addition to bringing to light knowledge gaps, these hypotheses will portray as a starting point and help planning and scoping questions for internal interviews. For social listening, hypothesis can be very useful for narrowing down the volume of data in order to answer or to validate specific questions or preconceptions.

It's important to notice here that social listening can be used either as a stand-alone research method or an exploratory research to help narrowing down the scope for future more traditional methods. For example, a global organization may analyse the online conversation in 10 different countries to identify consumer groups' behaviours, concerns and terminologies, which can later guide questions for more in-depth interviews or focus group for testing marketing campaigns for each persona type.

5.2 – Planning the Research Design

The research design serves as an overall plan of the methods for collecting and analysing the data. *"(...) the researcher must consider the type of data, the design technique (survey, observation, experiment, etc.), the sampling methodology and procedures, the schedule, and the budget. Although every research problem is unique, most research objectives can be met by using one of three types of research designs: exploratory, descriptive, causal"* (Hair et al., 2002 p.41). Descriptive research designs are the most appropriate for the objectives of developing buyer personas as the objectives include finding the how, who, what, when and where answers for the different segment groups of consumers. Figure 11 shows four basic design techniques for both descriptive and causal research: surveys, experiments, secondary data, and observation. The choice of techniques will be directly impacted by the objectives, the availability of data, as well as by the timeline and budget of the project, bearing in mind the benefits of social listening regarding time efforts and cost-wise.

Based on the Literature Review, interview is often the technique of choice for developing personas because it allows the researcher to synthesize and prioritize the key elements of the consumer group's narrative. Revella (2015) says that *"the most effective way to build buyer personas is to interview buyers who have previously weighed their options, considered or rejected solutions, and made a decision similar to the one you want to influence"* (p.08). She criticizes marketers who don't understand that hearing and connecting their customer's stories is the key for understanding them as buyers.

Although social listening can't rely on interviews, it can be interpreted as an adaptation of the classical ethnographic techniques in the online environment, therefore enabling the exploitation of its flexibility in time and space. *"For all the years that anthropologists have claimed that 'being there' is an essential part of our practice, what does this really mean when the vast majority of the people we want to understand are physically present but digitally somewhere else entirely?"* (Mott, 2015, p. 1). While classical ethnography faces the challenges of the digital age, social listening poses a comprehensive way to capture a completely organic record of the social interaction happening online around a particular subject. Despite not being able to individually address questions as in an interview, the virtual ethnography method enables access to non-biased (or at least fewer chances of bias) and non-requested opinions from customers. Boolean searches and data mining tools help researchers to explore the available data and get detailed answers with often as rich details as the

information provided on interviews, however without the ability to take the discussion further.

5.3 – Sampling

One of the biggest perks of social listening is the ability to monitor a significant sample of the target audience at low cost and being able to gather new samples of data over time to update any marketing and businesses' strategies (Jackson, 2017). In social listening, the total population of people conversating online about a topic is often considered as the total population for a determine search; in these terms, it can be said that it's Census Data. Nevertheless, as discussed on topic 4.3, the distinct nature of sharing and the importance of social media in the construction of personal identity may still mean that social data may has a bias that needs to be considered.

Although no sampling process is required in social listening, the main measure of accuracy relies on the relevancy percentage, namely the fraction of the retrieved data that is relevant. To generate the search terms for a social listening study, it's important to test data volumes and search terms' results in a snowball sampling process. For example, one can start with a set of relevant keywords such as "compact automotive" and then retrieve a sample of data to evaluate relevancy and identify additional keywords to be used for the next rounds of searches or exclusion terms for narrowing down the conversation to a more relevant dataset according to the objectives. This process ensures the quality of the search query and therefore the project's dataset, as well it helps to confirm feasibility of the research in terms of data volume.

Also, the persona development process requires a more in-depth analysis which is not always possible to perform by analysing the whole dataset. For the proposed methodological procedure, a representative sample set is recommended, to get an extended degree of accuracy and context from conversations. Sample analysis will be further explained when examining Data Processing and Analysis.

5.4 – Data Gathering

In traditional research methods, this step may require considerable human resources and a big piece of one's budget to perform the selection, training, coordination and assessment of necessary fieldwork such as interviewing (either focus-group format or in-place interception

or telephone-assisted interviews). Although online surveys require fewer personnel and implies lower costs, it won't reveal anything that it's not already known by the enterprise as it requires pre-scripted questions based on prior knowledge and high-level assumptions. Furthermore, even if the sample is relevant to the total population, the answers are given voluntarily and may never represent the real trend as it's impossible to guarantee that the non-respondents would follow the same line of answers as the respondents.

As previously discussed on this chapter, interviews are usually the method of choice when researching customer for buyer persona development. In social listening, on the other hand, virtual ethnography consists of an unobtrusive method for data gathering and allow researchers to collect the data without having to disturb the subjects.

In social listening, data gathering process is a bit like finding a needle in a big haystack that is online data. For established brands or depending on the industry, there are likely to be hundreds or thousands interesting mentions every day, but the challenge is that they are often hidden among other thousands of less interesting or entirely irrelevant mentions. Most social listening platforms use search queries that are based on Boolean logic, also known as Boolean Search. Boolean operators may vary in different platforms but the fundamental logic (AND, OR, NEAR, NOT) often applies across different tools and also in native social networks search such as Twitter or Google itself to find relevant forums or blogs conversation, for instance. Boolean search query provides flexibility to the search as well as the necessary precision to get the most accurate results, aligned with the project's goals.

Revella (2015) highlights the preparation process prior to each buyer persona interview, which includes getting familiar with the terminology the buyer is likely to use. Similarly, in social listening, healthcare professionals and patients may not always use the same terminologies when conversating about cancer. Knowing the stakeholders' terminologies can have a substantial impact in how effective a Boolean search is in finding relevant data. For buyer personas development, there are essentially two ways the research setup can go:

1. Narrow down the Boolean search for the enterprises' brands or specific products (e.g. "Toyota Prius"). The data is most likely to be relevant though lower in volume.

2. Setup the Boolean search to cover broad topics that often drive more conversation than brand queries. Although buyer personas are more commonly focused on a specific product or products' line, this might be particularly useful for start-ups or smaller enterprises trying to understand the buyers' characteristics within a determine market/industry. For example, one might look for conversations around electric automotive models to uncover who are the personas purchasing and commenting about this type of products. Alternatively,

tracking competitors' products can also offer a landscape of buyer personas playing in the current market.

For either the case, one will still need to create a targeted query to avoid noise in the data. When looking for buyer personas, writing a query that includes personal pronouns may help one to have a cleaner and more relevant dataset in terms of consumers' first-person conversation and experiences.

Based on professional experience, many social listening tools provide a step-by-step guide to facilitate the setup of Boolean queries even without prior knowledge of Boolean logic nor operators. There is a wide variety of free alternative tools in the market that besides some limitations in terms of features and volume of data, it can provide great insights for beginners or small entrepreneurs who are not yet ready to invest in more robust mechanisms. However, these free options are not a direct substitute to Enterprise-level tools, such as Brandwatch, Synthesio or Crimson Hexagon, which brings social listening to a new level and are able to provide more functionalities and deeper insights. Some well-recognized free tools in the market include Hootsuite, Mention, Twazzup and Buzzsumo.

This study does not aim to include an in-depth comparison of social listening platform alternatives in the market. For the purpose of demonstration, Crimson Hexagon will sometimes be used as an example. Enterprises should be aware that data volumes and quality will vary according to each tool and depending on the market, which means that not all social listening platforms will have the same data capabilities to find conversation in the United States, in English, and in China, in Chinese. Available options should be weighted according to each projects' requirements and objectives.

5.5 – Data Processing & Analysis

Data cleaning refers to the process of identifying and correcting irrelevant or incorrect parts of the data with the goal of having a more accurate dataset. In traditional methods such as interview or survey, this can mean ensuring if respondents or interviewers have entered the fields correctly or check for any skip patterns. In social listening, data cleaning can help to filter out any irrelevant or less interesting results that may have not be filtered out during the Boolean query process. As an example, a project tracking awareness around cardiovascular diseases may want to clean out the noise of people mentioning to "have a heart attack" while watching a scary movie, as such verbatims would be obviously out of context.

After reaching the most accurate dataset possible, coding the data is usually the step that follows. Codes are the rules for interpreting and categorizing the data into meaningful groups of responses and facilitates both computer or hand tabulation (Zikmund & Babin, 2010).

Revella introduces the concept of *The 5 Rings of Buying Insights*, five types of insights to guide a better understanding of what happens and who is involved in the buyer journey. These five insights include: Priority Initiative, which explains the reason why buyers decided to purchase a certain solution or chose to stay with the status quo; Success Factors, which describes the expected results by purchasing a solution; Perceived Barriers, referring to what prevents consumers to decide for one solution over another; Buyer's Journey, which reveal the details of the buyer's journey, including other stakeholders and factors involved in the decision-making process; and, finally, Decision Criteria, referring to the specific attributes that buyers evaluate when choosing a solution. Based on *The 5 Rings of Buying Insights*, Revella underlines three main steps for mining interviews for buying insights:

1.   Mark up interviews' transcript: It's recommended reading the interview transcript from the beginning and marking up quotations that answers one of the 5 Rings of Buying Insight's questions, tagging it to the related insight (Priority Initiatives, Success Factor, Perceived Barriers, Buyer's Journey or Decision Criteria). Codes such as "SF" for Success Factor can be used during the process.

2.   Organize the story based on buying insights: then, it's suggested using a spreadsheet to create what she calls the "Insights Aggregator", which list the quotes that best exemplify each of the different five insights' groups.

3.   Write a headline for each key insight: The next steps would be to identify the key insights based on interviews' quotations and which are the most important ones for using in the buyer persona profiles.

Although social listening will most likely result in a much higher volume of data than interviews' transcripts, a similar approach can be placed for analysing social datasets. While automated tools can be useful for gathering and summarizing high level insights, identifying buyer personas requires a balance between the human and the machine. As briefly mentioned in chapter 5.3., it's recommended a manageable sample which needs to be representative enough but yet feasible to be analysed in detail. Considering a confidence level of 95% and a confidence interval of five, researchers will rarely need to code a sample larger than 400 verbatims. Most monitoring platforms allow their users to export a random sample for coding.

As both the dataset and the attributes of the subject may vary, so will the coding template. This phase is rather an exploratory step where the researcher will hunt down insights by grouping themes and topics. To follow a similar approach to the one suggested by Revella, the data can be coded by *The 5 Rings of Buying Insights* and then more in details regarding different attributes of the product such as "Value per money" or "Guarantee". The researcher may start by creating initial categories that he or she believe it may appear based on past experiences or company's prior market research results. However, new topics may and will emerge as one work through the data.

Figure 12 represents an example of coding sheet analysing attributes of Decision Criteria, one of *The 5 Rings of Buying Insights.* The quantification of Importance of each attribute and the tagging of accurate sentiment will help to further visualize insights through a spider chart for each buyer persona.

		DECISION CRITERIA			
		Price		Guarantee	
Author name	Content	Importance	Sentiment	Importance	Sentiment
		2	Negative	4	Neutral
		3	Negative	1	Positive

Figure 12 – Example of coding sheet for attributes within the Decision Criteria insight.

Analysing the coded sample will enable the researcher to identify key trends and map out the main audience groups talking about the determine product; then, deep-diving into each audience group to find out their characteristics, similarities and differences, both in terms of attributes and buying journey as well as online behaviour such as their channels of preferences and who are they influenced by.

Revella summarizes well the process of mining interviews data, which can be easily related to mining social data for buyer insights: *"It's a familiar scene from countless investigative police procedural dramas. The chief detective assembles the team and presents the evidence on a bulletin board. There are forensic photographs and images of witnesses and suspects. Pinned next to them is a detailed map of the crime scene, a timeline, photocopies of key evidence, and selected witness statements. This visual image is a powerful dramatic device that helps all observers understand the relevant personalities and clues (as well as red herrings) in the story. (..) We will combine all of the stories to create a single narrative that represents the mind-set of a group of buyers who think alike. When we are done, we will have a factual description of our person (or persons) of interest, and a story that details expectations, thinking, and decision-making process as that person approaches the decision you want to influence."* (Revella, 2015, p. 97).

5.6 – Drawing Conclusions and Preparing Report

The last phase of the research process consists of interpreting the results, drawing the conclusions and presenting the findings in a comprehensible format, providing the deliverables promised at the objectives stage to help future managerial decisions.

The structure for building and presenting buyer persona can be very similar no matter if research was conducted using interviews, social listening or other methods; structure may vary, however, according to each project's objectives and needs, focusing more in certain information and details over another. Buyer profiles help to get across the consumers' grouped during the analysis phase as well as specific insights and what strategies may work for each buyer persona. As also suggested by Pruitt and Adlin (2006), marketers usually give their persona profile a photo and a real name to enhance the "human" character of the persona to stakeholders across the company, as well as it serves as a recall of all the persona attributes associated to that name. Persona profiles often also include job titles, demographics (not only age and gender but also location, household income or family size), goals and challenges (and how the product in subject help this person to reach these goals) as well as the main barriers for purchase or what type of communication may be most appealing for each buyer group.

5.7 – Proposed methodological procedure flowchart

Considering the preceding chapters, the following flowchart is proposed for researching buyer personas through the usage of social data:
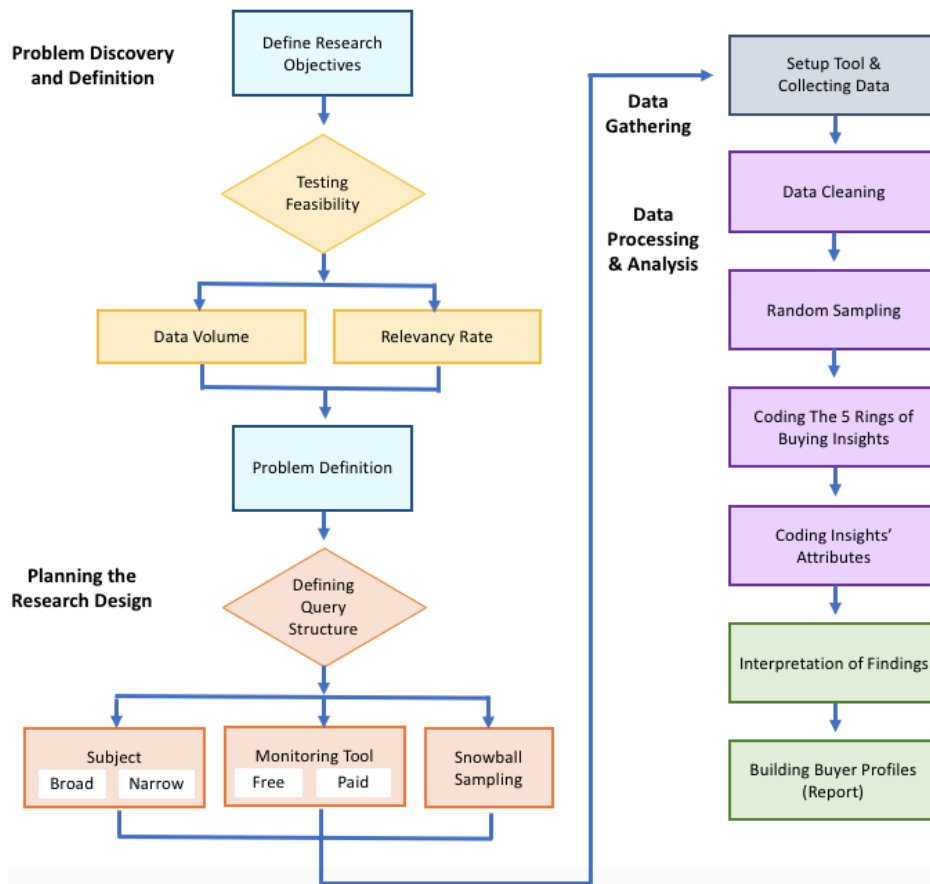
Figure 13 - Flowchart of the Buyer Persona Research Process through Social Listening.

With the objectives highlighted, a phase for testing the feasibility of the project is required. As discussed in previous chapters, social listening may not be able to provide the necessary data volume for certain research subjects and testing the volumes and relevancy rates may avoid wasting further efforts and resources.

The Problem Definition will serve as a guide for delineating the query structure, deciding if the subject of research will be broad (topic area) or narrowed down to a product or its competitors; the choice of the best monitoring tools, which may take into account not only the allocated budget but also the markets to be covered and necessary capabilities, as well as the fluency of the researcher with social media; and a snowball sampling process to ensure the best query to provide accurate results to the determined objectives.

After setup the monitoring platform and the data collection is completed, the data processing starts by data cleaning, which can be completed either in a dashboard or manually by removing irrelevant verbatim from an exported spreadsheet. When the data has reached a reasonable level of accuracy (usually 80%), the coding process will be necessary to provide

more in-depth findings which will serve as the basis for grouping buyers and building profiles.

As described in chapter 5.5, a simple spreadsheet allows the researcher to tag verbatim accordingly to each of *The 5 Rings of Buying Insights*, as suggested by Revella (2015). This will help to identify different sides and considerations of the consumer within the buyer journey, which will later be taken to the next step by coding for specific attributes to identify importance and sentiment around points such as Price, Value per Money, High-End or Guarantee, for example.

Finally, by interpreting findings, the researcher is then able to build buyer profiles: realistic archetypes with proper names and pictures which will help to communicate consumers' characteristics across the company.

6. Case Studies

6.1 – Unveiling a Buyer-Persona in the Home Appliance Industry

This chapter will briefly underline the steps and results based on the process illustrated on Figure 13 with the objective to enlighten and exemplify further the stages of previous proposed methodological procedure for building buyer personas through social listening.

Based on professional experience it's possible to say that building buyer personas through social listening tend to be a more enriching process especially for products and brands related to a complex buying process, namely consumer journeys defined by high involvement from the consumer when making the decision; complex buying processes are often associated to more expensive products and less frequently purchases. Due to the characterization of the purchasing scenario, consumers will tend to consider and ponder their options for longer and enablers and barriers are likely to have a bigger impact in the final decision. As a simple example, the buying cycle of a detergent won't require as much as thought as buying a car. For that reason, social listening can be even more powerful for innovative enterprises and start-ups dealing with high technological solutions.

For the purpose of demonstration, the case study will consider the technological kitchen appliance brand named Instant Pot. According to the company's website, Instant Pot is the biggest seller kitchen device in Amazon, designed to combine different cooking features in one single product. Prices may vary from $80 to $400 and it's not an everyday purchasing, characterizing it as a complex buying process.

The procedure described here was performed using Crimson Hexagon due to existing access to the tool but, as described in previous chapters, the same process can be easily reproduced in different platforms and acquired data. The case study will consider last 12 months of data for Instant Pot within the United States for publicly available verbatims on Twitter, Facebook, Instagram, blogs, forums and online reviews' websites.

Following the objectives of this research, this case study aims to underline consumers' characteristics for building buyer personas. The first step was to run a feasibility test for checking on data volume estimation and data relevancy rate. A simple query containing only variations of the product's name ("instant pot" OR instantpot OR #instantpot) was tested and returned an average of 854 monthly data within the last quarter (January to March 2018).

Data relevancy was confirmed by tagging as relevant or irrelevant a sample of 200 which resulted in a relevancy rate higher than 80%. To improve data quality, exclusion terms were added to the query to exclude spammy promotions and giveaways which wouldn't offer any interesting insights to the study's objectives. Should the relevancy rate had been lower than 80%, it'd recommended to revise the queries further in a snowball sampling process previously described, to avoid misleading the data and unnecessary efforts while going through the verbatims to find consumers' opinions.

A total of 4,933 posts were collected by the tool using the specified query for Instant Pot of which the majority was pulled from Twitter (38%), Instagram (31%) and Forums (22%). A random sample of 356 verbatims was retrieved, considering the population as the total volume and at a confidence level of 95% and interval of 5.

A spreadsheet was created including The 5 Rings of Insights adapted from Revella (2015) and its subcategories. A first group of subcategories were pre-determined as hypothesis and a second group was brought up while analysing verbatims.

- **Priority Initiative:** represent the triggers to start the buyer journey. Examples: Needs to cook healthy, Needs to cook for the family or Needs to cook to save money.

- **Success Factors:** represents the benefits and results expected from the usage of the product. Examples: Healthier lifestyle, Avoid restaurant food, Being Fit, Having fresh food every day or Keeping up with special dietary needs. Success Factors were also coded in a scale of Importance from 1 to 4, with 4 being very important and 1 being low importance. The classification helps to understand any nuances between different groups of audience.

- **Decision Criteria:** as the name itself says, it represents the criteria pondered by consumers during the decision-making process. Examples: Price, Value for Money, Easier to use or Better Client Support. Decision Criteria were coded both in an importance scale of 1 to 4 as well as for sentiment (negative, neutral and positive) which helps to identify the impact of that criteria for that particular consumer.

- **Buyer's Journey:** identifies where in the buyer journey the consumer is at. Examples: switching brands, considering options, already owns the product or recommends the products to peers.

- **Perceived Barriers:** finally, whenever a barrier was identified, a subcategory was allocated. Example: Affordability or Availability of the said product.

With the coded spreadsheet, it's possible to start retrieving some valuable insights which will then serve as a base for building the buyer persona. For the purpose of demonstration, this case study will only analyse results for people looking to cook in order to be embrace a healthier lifestyle. Those are consumers placed under the "Need to cook healthy" category as their trigger to the usage of Instant Pot.

As explained in Chapter 5.5., attributes for "Successful Factors" (benefits) and "Decision Criteria" were coded on an importance scale of 1 to 4, when mentioned by the verbatim. The easiest approach to start analysing the data is to start filtering by Priority Needs" (purchasing triggers), one at a time, to unveil detailed insights for the audience associated to each trigger. Once the data is filtered by one of the Priority Needs, one can sum up the coded values for importance of each "Success Factor" to highlight potential trends within that segment. Totals for each Success Factor can then be calculated into averages to allow by dividing the sum by the total n sample. Figure 14 shows that "Healthier Life" and "Reduce Time & Effort", for instance, place as strong (in green) benefits perceived by those looking to cook healthy.

Priority Initiative	Success Factors					
Needs to cook healthy	Reduces Time & Effort	Healthier Lifestyle	Avoid Restaurant Food	Meal Prep	Avoid Restaurant Food to save mone	Tastes good
1	4	4				
1	4	2				
1	3	4		3		
1	3	4				
1	4					
	84	104	0	34	8	62
average importance	2.470588235	3.058823529	0	1	0.235294118	1.823529412

Figure 14 - Example of coding sheet for attributes in Success Factors filtered by "Needs to cook healthy" as an trigger. Coloured cells at the bottom are a sum of "Success Factors" codes and can represent potential trends within that segment.

This mean that filtering verbatims by consumers under "Need to cook healthy", quantitative can already disclose a few trends while the quotes will lead insights regarding the in-depth behaviour and drivers for that specific audience group. The table below displays the key themes of conversation, ranked by the average of importance (with 4 being the highest) within the coded sample, along with a key insight and an illustrative example quote.

Table 5 – Breakdown of Insights for "Need to cook healthy" audience group

	Topic	Key Insight	Example Quote
Success Factors	Healthier Lifestyle (3.0)	Consumers consider Instant Pot a tool to accomplish healthier meals due to features such as the lack of frying.	*"Tonight's oil free vegan black bean tomato barley potato mushroom soup with kale, chard, bok choy, Napa cabbage, butternut squash, green oinions, mustard greens and parsley..yum..I love my new instapot..a few ingredients made an energizing and nourishing meal for a small army. #oilfreevegan #vegan #plantbasedcooking #drmcdougall #eattolive #instapot"*- Instagram
	Ease to Use (2.7)	Instant Pot is often associated with its smart and fast cooking capabilities. Consumers were happy to be able to add frozen or raw ingredients in a pot and get back a delicious meal.	*"I love my new insta-pot. I have made several meals in it now and have loved them all. I have tried some that were frozen solid and the meal was ready in 30 minutes. I used it the most for steamed vegetables. Usually about 5 minutes and my vegetables are perfect. This is very easy to use."* – Online Review
	Reduces Time & Effort (2.5)	Consumers appreciated time saving due and often referred to how many minutes took them to cook a recipe on an Instant Pot.	*"I love roasted garlic! I make it in my instant pot now! SO FAST! https://t.co/tspX3IunUa #instantpot #instapot #recipes #glutenfree"* - Twitter
	Tastes Good (2.0)	Tasteful food is also considerate by consumers. They celebrate meals accomplished by the Instant Pot, sometimes with an implicit surprise that the tool helped them to accomplish that in such a short time.	*"Lemon Pepper Duck w/ Sauteed Vegz! Swipe 4! Hubs doesn't even like duck but he was lovin on this yummiliciousness! Pressure cooked semi-frozen duck for 20 mins then pan seared to a nice golden brown! Veggies only took a few mins while duck was being seared. Lovin my @instantpotofficial!!!"* - Instagram
	Special Nutrition Needs (1.8)	Healthy eating is commonly associated with special diets or nutrition needs such as veganism, keto diet or gluten free.	*"My #instapot really got the workout today. Cauliflower and greens beens for lunches (3 min), chicken for dinner tonight (frozen 20 min) (...) I'm ready to do this #lowcarb #keto way of eating to help me move in the right direction. #foodprep"* - Facebook
Decision Criteria	Great Results (4)	Consumers were pleased with the overall great results produced by Instant Pot, usually associated with great taste versus lower effort.	*"Thank you @instantpotofficial for helping me make yummy coconut curry rice ~ even though I burnt my naan, it was still a super yummy and healthy dinner. #foodblogger #foodie #foodgram #homecooking #homecook"* - Instagram

	Simplicity (2.8)	The key motivations were reflected in consumers' statements of why they love or bought their Instant Pot. Simplicity in the kitchen (being able to use the pot for different tasks) and being able to cook faster were raised as important factors.	*"Do you have an instapot or crockpot? I haven't found them (specifically the instapot, despite its name) to be significantly faster, but I use them a lot because they are hands off cooking. And they make a lot of food so I have leftovers for lunches the next couple of days."* - Forum
	Time Saving (2.3)		
Perceived Barriers	Hard to Learn (26% of total coded sample)	Consumers were sometimes overwhelmed by the many features available and how robust is the product, making them slightly afraid to try using it.	*"the bhb convinced me that i needed an insta pot. last Prime Day i ordered the mother of all insta pots - the 8 quart one. i opened the box - saw the size of that mama jama and closed the lid. my goal is to open it and try it out before thanksgiving. must get back to healthier eating and meals!"* - Forum

Because of the large scale of social media possibilities, the researcher can go as granular as it's feasible, gathering demographic data when necessary (and available) and exploring channels preferences. For instance, on this case study, healthy enthusiasts tended to "show-off" their meals on Instagram more than other audience groups. Such granularity of information can help to better tailor communication strategies for each persona.

Each of The Five Rings of Insights can provide valuable characteristics to our persona. For instance, we are specifically analysing those looking to cook health (main Priority Need), which is a trigger for them to own or to consider buying an Instant Pot. As we code, it's important to consider that attributes can often overlap, depending on the subject of the study. In this case, this mean that a person with the "Need to cook healthy" can also fall into other Priority of Needs categories such as "Need to cook for family" or "Needs to cook to save money". We calculated the percentages of overlapping across the secondary attributes to find out the hierarchy of motivations for this persona. Besides being healthier, Family was the biggest driver (73% overlap), followed by Hobby or the enjoyment of cooking (47%) and finally by Save Money (13%). From a business perspective, this may mean that when communicating to this persona, appealing to family moments will be more likely to result than budgetary rationale. Similarly, "Decision Criteria" quantitative helped to build the persona's goals and objectives and her challenges, by showing that this audience group was highly concerned on obtaining great results with their kitchen appliance, in an easier and

faster way. Paired with qualitative findings from verbatims, this was translated into a busy mother who is also managing many activities at the same time while still trying to impress and provide the best to her loved ones. She wants things faster, but she is not willing to give hand of quality results.

Decision Criteria							
Technology		Great Results		Easier to use		Work Faster	
		4	Positive				
		4	Positive	3	Positive		
				3	Neutral		
						4	Positive
		4	Positive	4	Positive	4	Positive
		4	Positive			4	Positive
						4	Neutral
3	Positive					4	Positive
6		120		84		76	
0.176470588		3.529411765		2.470588235		2.235294118	

Figure 15 - Example of coding sheet for attributes in "Decision Criteria" filtered by "Needs to cook healthy" as an trigger. Coloured cells at the bottom are a sum of "Decision Criteria" codes and can represent potential trends within that segment.

With the key insights in mind and backed by the coded spreadsheet, a buyer persona was drafted for better representing this audience group, as shown in Figure 16. The full-project to unveil the different personas within the Instant Pot online conversation demanded between 3 to 4 weeks of work, while traditional interview methods would require months to be completed, as referenced in previous chapters. For this specific case Crimson Hexagon was the tool of choice for retrieving and exporting the data; although the usage of professional tools guarantees better data availability and analysis functionalities, free platforms could have reached similar results with enough outcomes to provide guidance for SME and new start-ups with no other methods available for gathering consumer insights.

Figure 16 – Example of buyer persona for Instant Pot

6.2 – Identifying Buyer-Personas for a Beauty Technology Brand

Another case study we performed helped a beauty technology brand to identify potential buyer personas for their haircare appliance and understand how to better engage with them or how to spend their marketing budget towards specific audience groups. More than 600,000 messages in English were identified within the 12 months of period range for the research in the United States. The virtual ethnography process allowed the team to evaluate the drivers of preferences and consumption behaviour. Unstructured data was brought to life by identifying commonalities to detect different conversation patterns across audience groups.

First, the team has manually coded a random sample for purchasing intent, demographics (age group, occupation, gender), barriers (price, availability, negative reviews from peers), purchasing behaviour (benefits, design, product features, to offer as a gift, etc) and other attributes found during the process. Analysing the coded sample and reading through verbatims, it was possible to identify the main audience groups that would later lead to the buyer personas, usually driven by distinct Priority of Needs which in this case were closely related to different hair type needs such as curly, straight or stylish. For instance, Figure 17 shows a deep-dive for "Curling hair" audience type, including sentiment breakdown by the client's products and its competitors and what platforms were experiencing most conversation around this topic. The list of top terms in conversations helps the unveiling

process of benefits and barriers for this audience such the constant worry about damaging the hair with heat or the need for an easier routine for curly hair customers.
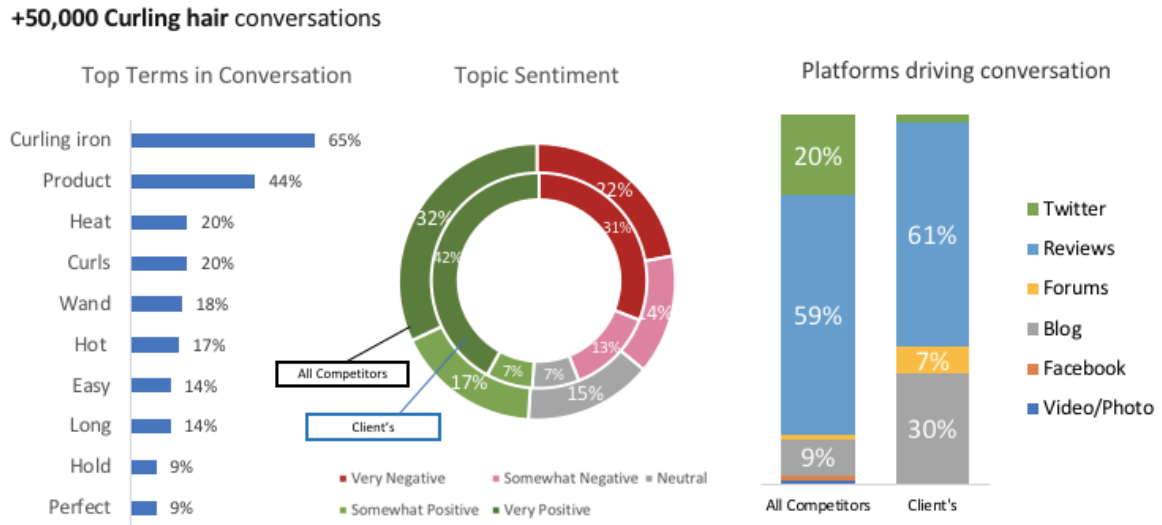
Figure 17 – Deep-Dive for "Curling hair" audience type

With the coded sheet filled for a significant sample, it becomes easier to prepare visualizations and compare attributes importance between buyer personas, as shown in Figure 18. At the end of the project, the team was able to identify 9 buyer personas, of which 4 were recommended to the client as their targeted audience groups.
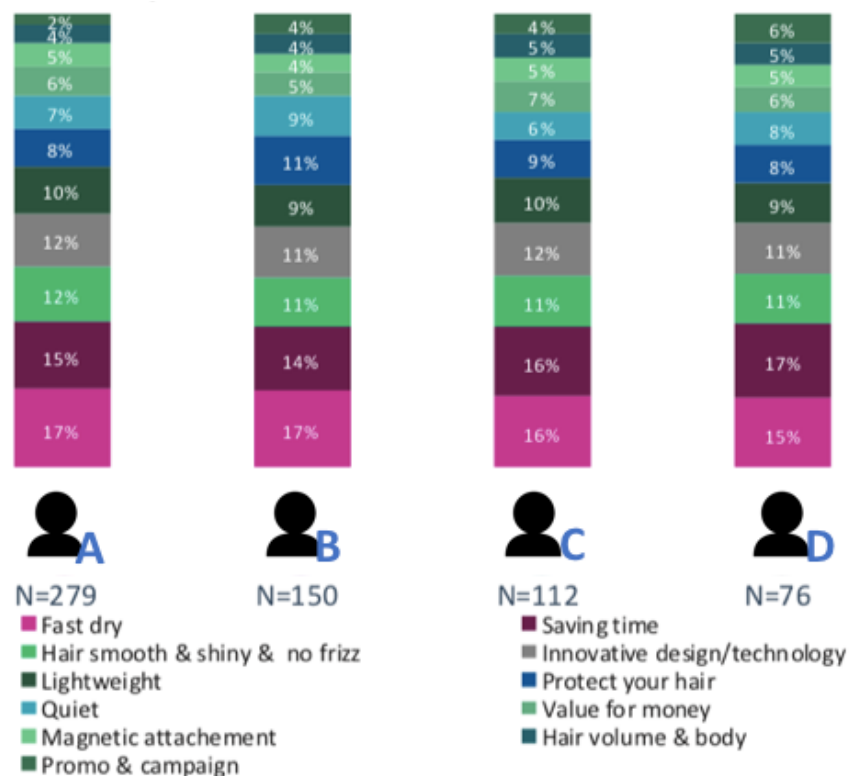
Figure 18 – Breakdown of perceived Success Factors (benefits) by buyer persona.

7. Conclusions

7.1 – Summary

This work highlights the growing usage and the importance of social listening for consumer insights and, more specifically, for building buyer personas. Several authors have covered the topic of market segmentation and discussed different perspectives and methodologies for persona generation. As the diversity of both market demand and supply side increase through the history, new business scenarios were raised with changes in consumers' needs and a higher buying power. The revolution on the consumption cycle brought the need of market segmentations to better target the diversity in consumer demand, which would later lead and evolve to personas.

While market segments evoked a segregation of consumers based on given factors such as demographics, personas would combine audiences by their common characteristics. In simple words, buyer-personas are a representation of a targeted market brought to life through a realistic profile, to illustrate an audience group based on not only demographics but also psychographic, behaviours, needs and barriers towards a product or service. Companies such as Delta Airlines, Unilever and Whirlpool already make usage and have demonstrated the power of buyer personas, either to bring more impact to marketing campaigns or to improve the brand's channels functionalities according to the identified archetypes.

From the literature review, it was possible to notice a predominance of traditional research methods for market segmentation or building personas, either for marketing or design purposes. Although big data and social data sources have been emerging in the last few years, those are not often accessible for the average marketer or manager decision-makers either due to lack of knowledge about social listening, exclusive reliability on conventional methods or to the fear of complexity and resources necessary of available technological solutions.

Traditional persona development relied mostly on interviews and ethnography. These methods were normally referred to be costly and taking several months or even years to produce the desired findings. Budgetary constraints placed as a big constraint for data collection, both from enterprises' and researchers' perspective. They look for cost-saving alternatives that can allow them to get more information with less or the same resources.

In a world where social media has become largely used, social listening emerges as an important alternative to such constraints, enabling to actively monitor a big volume of social

media conversation to understand consumers talking about a topic, a product or an industry. Social data is extracted through social listening and can provide valuable insights not only to marketing teams but also for product development and towards innovation. It also brings clear advantages within the market research process:

1. Enables listening to what the audience is saying organically. Different from other methods, social data provides unbiased and unsolicited context;

2. Enables identifying trends and potential changes in the market in near real-time, allowing business to act and adjust strategies almost immediately.

Paired with the advantages of cost-effectiveness and time-saving, the benefits of social listening could be promising for SME and start-ups who first, not always have high budgets for market research, especially at the beginning of their business course; secondly, require fast information and fast change of actions to keep up with bigger enterprises and technological developments.

7.2 – Discussion

To structure a procedure for buyer persona development through the usage of social data, it was traced a parallel between traditional market research methods and similar alternative steps using social listening, resulting in the following stages:

1. Defining the research objectives, including a phase for testing the study's feasibility in terms of online volume of conversation and data relevancy;

2. Planning the research design and therefore the query structure which will determine the data available for the analysis. This stage also requires deciding on any specific monitoring tools;

3. Setup tool and collecting the necessary data;

4. Performing data cleaning where necessary and retrieve a sample for coding for The 5 Rings of Buying Insights, unveiling the importance of different drives, benefits and barriers as part of the data analysis process. This phase is crucial to identify different sides and considerations of the consumer within the buyer journey;

5. Finally, formulating conclusions and building the persona profiles accordingly.

The case studies helped to exemplify the process of building buyer personas by using social media data available online in just 3-5 weeks of work while conventional methods would require additional timeline and resources until reaching an usable outcome. It also presented different applications and how the same procedure can help unveiling further

insights such as levels of engagement or where the conversation about a topic is happening online, all valuable information for marketers and product developers looking to keep their enterprises at the front of the technological run.

7.3 – Conclusions

Personas can be an important asset for any business but it's usually an expensive and time-taking process which traditionally involves ethnography studies and can be easily outdated within a short period of time. Social listening places as a cost-effective method that allows marketers and enterprises to explore non-biased and non-requested data with quicker and most up-to-date results. These advantages are especially important for technological and innovative enterprises who require rapid decision-making and pivot moments for a reasonable budget, which is the case for many self-funded start-ups and SME. Nevertheless, data is not always easily accessible and manageable, limiting smaller players to gather valuable market intelligence due to the lack of budget or for not knowing from where to start.

Guided by the research question, this study demonstrates the viability of exploiting social listening methods for generating buyer personas. It has answered the unmet needs determined on chapter 2.3 as follows:

1. Exploring the power of social data as an existing and non-biased conversation database;

2. Exploring cost-effective persona generation methods that allow marketers and managers to uncover buyer personas without big market research investments and extensive timelines;

3. Providing actionable insights and rich data representation to marketers and managers in an easy-to-digest format to ensure the dissemination of the knowledge among all areas of the enterprise.

In those regard, this dissertation has systematised information about social listening as an efficient research method to be used by marketers and decision makers. Its outcomes contribute to enrich academic knowledge around the relatively new method that is social listening; especially when related to buyer personas generation, for which we found there was a lack of studies exploring the usage of social data for that purpose. Through evidences from business cases and examples of usage, the work here presented intends to incite further adoption of this solution by enterprises, in particular by SME and start-ups.

This dissertation has proposed and validated an innovative yet simplified methodological procedure for building buyer personas through the usage of social data, highlighting its advantages regarding other available methods in place. Originally used in interviews, The 5 Rings of Insights (Revella, 2015) were applied to a social listening context to establish a methodological process that allows entrepreneurs to transform available information into customer insights, test hypothesis and tailor their strategies based on buyer persona profiles, not only assumptions.

While social listening can be a benefit to consumer research methods and buyer persona development, it still falls short at certain points and experience a few potential constraints such as data privacy concerns, limitations of data availability according to the subject or the demography of the study as well as the challenges to structure and transform data points into actionable insights. It's certainly an economic and time-saving alternative but it is still important to consider it as part of a more complete framework to guide or be enriched by other methods whenever possible and according to each project's objectives. Both social listening and other more conventional methods have its own relevancy and applications; businesses need to evaluate strong and weak points of each to integrate different methods and achieve the best results possible.

Hoping that this work will inspire other research on social listening in order to support possible improvements and to facilitate its use more and more, further studies could explore, for instance, the Return-On-Investment (ROI) of social data-driven buyer personas in marketing campaigns, increasing reliability for persona-driven communication strategies. Additionally, would be interesting to trace a standardized quantitative comparison of different persona development methods in terms of time-effort and budget to determine the cost-effectiveness of each approach. Results could guide companies from different sizes and backgrounds to decide what better methodologies apply to their business needs. Moreover, the methodological procedure used here could be further applied for different areas and buying cycle complexity levels to validate the framework in a wider range of context not yet considered by the case studies here in this dissertation. In addition, future studies can also consider other authors and methodologies to enrich the "5 Rings of Insights" categorization system, both from a conventional market research and social listening perspective.

References

Aken, J.E. (2005). Management Research as a Design Science: Articulating the Research Products of Mode 2 Knowledge Production in Management. *British Journal of Management, Vol. 16,* 19–36

Awad, T. A. (2011, March). Environmental segmentation alternatives: Buyers' profiles and implications. *Journal of Islamic Marketing 2(1):55-73.* 10.1108/17590831111115240

Banko, A. (2014). *Persona Marketing: NetProspex increases website visit duration 900%, lifts marketing-generated revenue 171%.* Accessed 10th June 2018: https://www.marketingsherpa.com/article/case-study/netprospex-increases-website-visit-duration

Bingham, C. (2018, April). *Understanding Instagram's API Changes: How Instagram's new data restrictions affect data partners like Crimson Hexagon.* Accessed 25th May 2018: https://www.crimsonhexagon.com/blog/understanding-instagrams-api-changes

Chroneos-Krasavac, B.; Soldic-Aleksic, J. & Petkovic, G. (2016). The big data phenomenon: The business and public impact. *Industrija.* 44. 117-144.

Chui, M., Manyika, J., Bughin, J., Dobbs, R., Roxburgh, C., Sarrazin, H., Sands, G. & Westergren, M. (2012) The social economy: Unlocking value and productivity through social technologies. *McKinsey Global Institute.* Accessed 14th March 2018: https://www.mckinsey.com/industries/high-tech/our-insights/the-social-economy

Ciarallo, J. (2013) *How Cisco Achieved 281% ROI with Social Listening and Salesforce Radian6.* Accessed 10th June 2018: https://www.salesforce.com/blog/2013/08/cisco-social-listening.html

Cooper, A. (1999). *The Inmates Are Running the Asylum*: *Why High Tech Products Drive Us Crazy and How to Restore the Sanity*. Indianapolis, Indiana. Sams Publishing.

Cuyvers, J. (2018, February 9) How social listening wins, where political polling fails [Webinar]. *In Convosphere Webinar.* Accessed 10th June 2018: https://convosphere.com/social-listening-wins-political-polling-fails/

Dagustani, D,; Buchory, H.A.; Satya, M.T. & Diryana, I. (2016). The internal branding process: Customer reprofiling and brand positioning (empirical study on traditional snack food in Badung-Indonesia). *International Journal of Business, Economics and Management, 2016, 3(2): 18-30*

Davey, N. (2015) *Social listening and privacy: How can brands maintain consumer confidence?* Accessed 11th April 2018: https://www.mycustomer.com/experience/voice-of-the-customer/social-listening-and-privacy-how-can-brands-maintain-consumer

Engel, K. (2015) *How to Develop Buyer Personas From Social Media Analytics.* Accessed 10th June 2018: https://klear.com/blog/buyer-personas-social-analytics

Félix, J. (2016). *Application of automatic customer profiling in small companies.* Accessed 23th January 2018: https://researchgate.net/publication/304540953_Application_of_automatic_customer_profiling_in_small_companies

Fink, A. (2005) *Conducting Research Literature Reviews*: *From the Internet to Paper*, 2nd edn. London: Sage Publications.

GreenBook (2013) *GreenBook Research Industry Trends Report.* Accessed 10th June 2018: https://www.greenbook.org/PDFs/GRIT-W13.pdf

Hair, J.F.; Bush, R.P. & Ortinau, J.D. (2002) *Marketing Research: Within a Changing Information Environment.* New York: The McGraw-Hill Irwin Companies, 2002.

Hevner, A.; March, S.; Park, J. & Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly (28:1)*, 75-81.

Jackson, D. (2017) *What Is Social Listening & Why Is It Important?* Accessed 23th April

2018: https://sproutsocial.com/insights/social-listening/

Jenkinson, A. (1994). Beyond Segmentation. *Journal of Targeting, Measurement and*

Analysis for Marketing, Volume Three. Number One. London: Henry Stewart Publications.

Jenkinson, A. (2009). What happened to strategic segmentation? *Journal of Direct, Data and*

Digital Marketing Practice 11: 124. 10.1057/dddmp.2009.27

Johansson, M. & Messeter, J. (2005). Present-ing the user: constructing the persona. *Digital*

Creativity. 16. 231-243. 10.1080/14626260500476606.

Jung, S.; An, J.; Kwak, H.; Ahmad, M.; Nielsen, L. & Jansen, J. (2017). *Persona Generation*

from Aggregated Social Media Data. 1748-1755. 10.1145/3027063.3053120.

Kelly, S. (2003). Mining data to discover customer segments. *Interactive Marketing,*

4(3):235-242.

Kemp, S. (2017) *Digital in 2017: Global overview. We Are Social.* Accessed 16th March

2018: https://wearesocial.com/special-reports/digital-in-2017-global-overview

Kotler, P. & Keller, K. L. (2009). *Analyzing Consumer Markets, Marketing Management*

(150-181). New Jersey: Prentice Hall.

Lee, Y. (2002) *Conducting Surveys and Focus Groups.* Accessed 10th June 2018:

https://www.entrepreneur.com/article/55680

MarketingSherpa (2005). *How Whirlpool Raised Customer Loyalty by 68% by Revamping its*

Site Using Persona-Based Design. Accessed 10th June 2018:

https://www.marketingsherpa.com/article/case-study/how-whirlpool-raised-customer-

loyalty

MarketsandMarkets (2017) *Social Media Analytics Market by Application (Sales and*

Marketing Management, Customer Experience Management, Competitive Intelligence),

Component, Analytics Type, Deployment, Organization Size, Industry Vertical, and

Region - Global Forecast to 2022. Accessed 22th April 2018:

https://www.marketsandmarkets.com/Market-Reports/social-media-analytics-market-

96768946.html

Mcginn, J. & Kotamraju, N. (2008). *Data-driven persona development.*

10.1145/1357054.1357292.

Miaskiewicz, T. & Kozar, K.A. (2011). *Personas and user-centered design: How can*

personas benefit product design processes?. Design Studies. 32. 417-430.

10.1016/j.destud.2011.03.003.

Mott, J. (2015) *What Room is There for Ethnography in a Digital World?* Accessed 22th

April 2018: https://bammglobal.com/room-ethnography-digital-world/

Neiters, J. E., Ivaturi, S., & Ahmed, I. (2007) Making Personas Memorable. Proceedings of

CHI 2007, *ACM Press (2007)*, pp. 1817 – 1823.

Netbase (2013) *Social Listening vs. Digital Privacy.* Accessed 22th April 2018:

https://www.slideshare.net/mosofsky/net-base-jdpalisteningprivacyfinal-16493659/1

Nielsen, L. (2004). *Engaging Personas and Narrative Scenarios.* Vol. 17, PhD Series.

Denmark: Samfundslitteratur.

Piller, F. T. & Tseng, M. M. (2010) *Handbook of research in mass customization and*

personalization. Volume 1: Strategies and Concepts, (pp. 611). Singapore: World

Scientific Publishing.

Portigal, S. (2008, January-February). *Persona Non Grata. True Tales*, 15(1), 72-73.

Accessed 22th January 2018: http://www.portigal.com/wp-

content/uploads/2008/01/Portigal-Consulting-White-Paper-Persona-Non-Grata.pdf

Pruitt, J. & Adlin, T. (2005*) The Persona Lifecycle: Keeping People in Mind Throughout*

Product Design. San Francisco: Morgan Kaufman.

Pruitt, J. & Grudin, J. (2003) Personas: Practice and Theory. Proceedings of the 2003
Conference on Designing for User Experiences (DUX'03), *ACM Press (2003)*.

Racounter (2017) *At last the reinvention of market research.* Accessed 10th June 2018:
https://www.raconteur.net/sponsored/at-last-the-reinvention-of-market-research

Revella, A. (2015) *Buyer Personas: How to Gain Insight into your Customer's Expectations,
Align your Marketing Strategies, and Win More Business*. New Jersey: John Wiley & Sons

Riemersma, F. (2015) Persona ROI Case Study - How Personas Help You Close Deals.
Accessed 10th June 2018: http://boardview.io/blog/roi-case-study-how-personas-help-you-
close-deals

Salminen, J.; Şengün, S.; Kwak, H.; Jansen, J.; An, J.; Jung, S.; Vieweg, S. & Fox Harrell, D.
(2017). *Generating Cultural Personas from Social Data: A Perspective of Middle Eastern
Users*. 10.1109/FiCloudW.2017.97.

Samuel, A. (2016) Psychographics Are Just as Important for Marketers as Demographics.
Accessed 18th June 2018: https://hbr.org/2016/03/psychographics-are-just-as-important-
for-marketers-as-demographics

Siegel, D.A. (2010). *The mystique of numbers: belief in quantitative approaches to
segmentation and persona development*. 4721-4732. 10.1145/1753846.1754221.

Simon, H. A. (1996). *The Sciences of the Artificial*, 3rd edn. MIT Press, Cambridge MA.

Smith, S. (2012) 9 Key Stages in the Marketing Research Process. Accessed 22th April 2018:
https://www.qualtrics.com/blog/marketing-research-process

Smith, W. 1956. Product Differentiation and Market Segmentation as Alternative Marketing
Strategies. *Journal of Marketing 21 (1): 3-8.* doi:10.2307/ 1247695.

Tedlow, R. (1990) *New and Improved: The Story of Mass Marketing in America*, (pp. 4–12).
New York: Basic Books.

Vernon Research (2018) *How Much Does Market Research Cost?* Accessed 10th June 2018:

http://www.vernonresearch.com/wp-

content/uploads/2018/01/HowMuchDoesMarketResearchCost_ebook.pdf

Wasserman, T. (2006) Unilever, Whirlpool get personal with personas. *Brandweek*,

September 18, 2006 pp. 13

Wessel, R. (2018) *What Goes In to Market Research and How Much Does it Cost?* Accessed

10th June 2018: https://startupnation.com/grow-your-business/goes-market-research-

much-cost/

Wood, M. (2007) *Essential Guide to Marketing Planning*, (pp.75). England: Pearson

Education Limited.

Zikmund, W. & Babin, B. (2010) Essentials of marketing Research, (pp. 42-60). United

States: Cengage Learning.