DEVELOPMENT OF A FRAMEWORK FOR THE ANALYSIS AND ASSESSMENT OF DAILY AIRPORT OPERATIONS

A Dissertation
Presented to
The Academic Faculty

By

Eugene Mangortey

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Aerospace Engineering

Georgia Institute of Technology

December 2020

DEVELOPMENT OF A FRAMEWORK FOR THE ANALYSIS AND ASSESSMENT OF DAILY AIRPORT OPERATIONS

Thesis committee:

Dr. Dimitri Mavris, Advisor
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Olivia Pinon-Fischer
School of Aerospace Engineering
Georgia Institute of Technology

Dr. Daniel Schragge
School of Aerospace Engineering
Georgia Institute of Technology

Mr. Tom Tessitore
William J. Hughes Technical Center
Federal Aviation Administration

Mr. Mike Paglione
William J. Hughes Technical Center
Federal Aviation Administration

Date Approved: December 1, 2020

Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do.

Pele

I dedicate this thesis to my family. Thank you for your love and support!

ACKNOWLEDGEMENTS

When I joined Georgia Institute of Technology as a college freshman in the fall of 2013, I never expected to defend my doctoral dissertation at the same institution. The past seven years have been filled with numerous memories and opportunities, and I am extremely grateful for the love, support and encouragement of many people.

First and foremost, I would like to express my deepest gratitude to my family. I am the man I am today because of your unconditional love and support.

Second, I would like to thank my advisor, Dr. Dimitri Mavris. You and the Aerospace Systems Design Laboratory provided me with numerous opportunities to grow and learn, not just as an engineer but also as an individual. Research opportunities in the lab provided me with the opportunity to learn about Big Data analytics and its numerous applications in the aerospace domain, leading to this dissertation.

I would also like to thank Dr. Olivia Pinon Fischer. You are one of the most hardworking people I have met, and have been a great source of inspiration, encouragement and guidance since my first year of graduate school. Thank you for supporting me throughout this entire process.

Special thanks goes to Tom Tessitore, my manager at the FAA's William J. Hughes Technical Center. The opportunity to intern with you over the past year provided me with the opportunity to learn about Big Data architectures and how they can be leveraged to improve operations in the National Airspace System. Special thanks also goes to Dr. Daniel Schragge and Mr. Mike Paglione for agreeing to be on my dissertation committee and providing invaluable feedback. I would also like to thank my colleagues at the FAA's William J. Hughes Technical Center for providing technical assistance when needed throughout this entire process.

I would be remiss if I do not thank Anya Berges and Warren Strickland at the FAA's William J. Hughes Technical Center. Thank you for helping shape this dissertation, and providing invaluable feedback and advice throughout this process.

Finally, I would like to thank my mentor, John Nussear, as well as Dr. Tejas Puranik, and my colleagues and friends: Ali Erktin, Akua Asare, Marc-Henri Bleu-Laine, Gabriel Achour, Andrew Melissas, Victor Zhang, Dylan Monteiro, Ghislain Dard, Federico Bonalumi, Petros Eskinder, Kunal Gangolli, Matthew Oliver, Paola Zanella, Lydia Manigwa, Maddie Pike, Bradley Jenkins, Pablo Gonzalez, Lorena Carapaica, Richard Hunter, Osman Ghandour, Shareef Ali, Philip Amankwah, Artur Cabral, James Sharpley, Nana Entsiwaa Adenu-Mensah, Roslynne Amoah, Joseph Brown, Sanggyu Min, Wentiirim Annankra, and Joe Brucculeri. Thanks for being a great support system!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Tremendous progress has been made over the last two decades towards modernizing the National Airspace System (NAS) by way of technological advancements, and the introduction of procedures and policies that have maintained the safety of the United States airspace. However, as with any other system, there is a need to continuously address evolving challenges pertaining to the sustainment and resiliency of the NAS. One of these challenges involves efficiently analyzing and assessing daily airport operations for the identification of trends and patterns to inform better decision making so as to improve the efficiency and safety of airport operations. Efforts have been undertaken by stakeholders in the aviation industry to categorize airports as a means to facilitate the analysis of their operations. However, a comprehensive, repeatable, and robust approach for this very purpose is lacking. In addition, these efforts have not provided a means for stakeholders to assess the impacts and effectiveness of traffic management decisions and procedures on daily airport operations. Furthermore, an efficient and secure framework for extracting, processing, and storing the data needed for the analysis and assessment of daily airport operations is needed, as the current process employed by FAA analysts is manual, time-consuming, and prone to human error.

Consequently, this dissertation addresses these gaps through a set of methodologies that 1) leverage unsupervised Machine Learning algorithms to categorize daily airport operations, 2) leverage a supervised Machine Learning algorithm to determine the category that subsequent daily airport operations belong to, 3) facilitate the comparison of similar and different daily airport operations for the identification of trends and patterns, 4) enable stakeholders to analyze and assess the impacts and effectiveness of traffic management decisions and procedures on daily airport operations, and 5) develop a framework to facilitate the efficient and secure extraction, processing and storage of data needed for the analysis and assessment of daily airport operations.

The developed framework automates the flow of data from extraction through storage, and enables users to track the flow of data in real time. It also facilitates data provenance by logging

the history of all processes and is equipped with the capability to log errors and their causes, and to notify analysts via email whenever they occur. In addition, it has the capacity to automatically extract, process, and store the data needed for the analysis and assessment of the daily operations of all airports in the NAS. Indeed, this framework will be one of the first of its kind to be deployed into the FAA's Enterprise Information Management platform and will serve as a template for leveraging cloud-based services and technologies to improve operations in the NAS. Finally, this framework will enable FAA analysts to analyze and assess daily airport operations in an efficient manner to facilitate the identification of trends and patterns for better decision making, which will lead to improved airport operational performance.

CHAPTER 1
INTRODUCTION

The International Air Transportation Association (IATA) estimates that in an average year, the aviation industry supports over 50 million jobs and transports 3 billion people and 50 million tons of cargo worldwide [1]. Indeed, the aviation industry is seen as an important driver of the economical and social development of countries [2, 3]. This has been particularly evident in the United States, as increased economic activity in the aviation industry has contributed greatly to the nation's economy, particularly after the economic recession of 2008 [4]. Between 2012 and 2014, the aviation industry accounted for approximately 5% of the Gross Domestic Product (GDP) of the U.S. economy, over $1 billion in U.S. economic activity and over 9 million jobs [5]. The aviation industry was also found to be the 7^{th} leading contributor to overall productivity in the United States in 2014 [6], despite its ranking as the 41^{st} largest industry out of 63 industries [7].

Economic activity in the aviation industry also outpaced economic growth in the United States between 2012 and 2014 [6]. This can be attributed to increased investments in aviation-related research and development. Indeed, increased investments in aviation-related research and development after 2011, as seen in Figure 1.1, overlapped with the aforementioned increased economic activity in the aviation industry. Some of these investments have led to the development and implementation of new procedures and policies, and technological advancements which have modernized the National Airspace System.

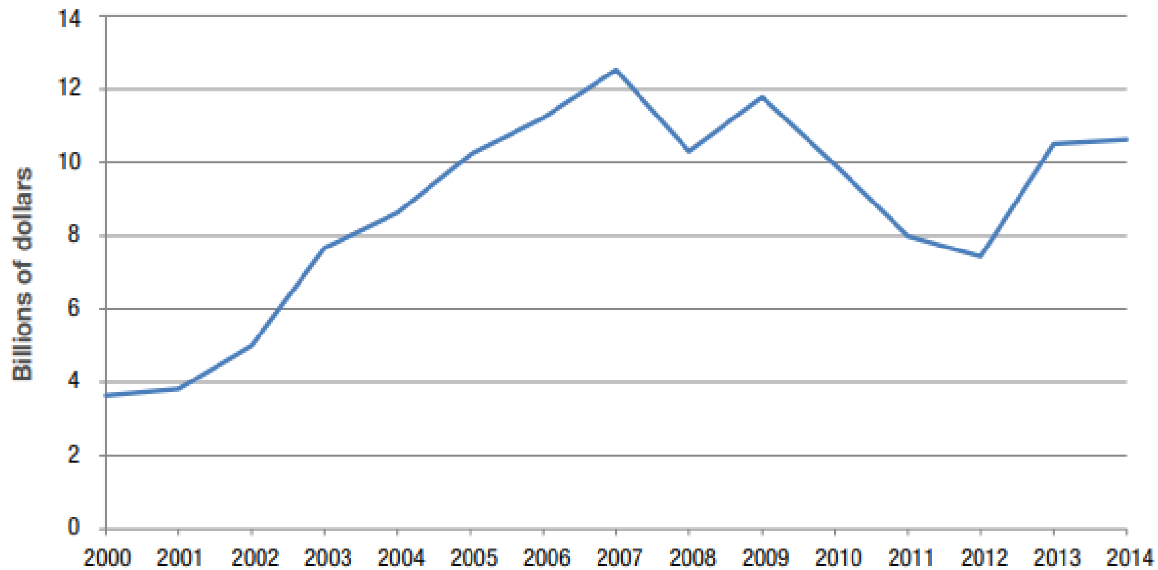

Figure 1.1: Investments in aviation-related research and development (2000-2014) [6]

1.1 National Airspace System

The National Airspace System (NAS) was created to ensure that the U.S. airspace remains safe and efficient for general, commercial, and military aviation [8–11]. It is comprised of air navigation facilities, airports, equipment, regulations, technologies, etc. that support the largest, busiest, and most complex airspace in the world [12]. Indeed, it handles an average of 44,000 daily flights, and is comprised of over 19,000 public and private airports [13]. The NAS, which is supported by over 77,000 pieces of equipment and systems that operate all year round, handles approximately 1 billion passengers annually [13].

The NAS is managed by the Air Traffic Control System Command Center (ATCSCC) [14, 15]. The ATCSCC's primary responsibility is to monitor and manage air traffic flow in the NAS to ensure that traffic flow remains safe, orderly, and expeditious, while minimizing delays. This is achieved by monitoring and analyzing system components, demand, capacity, and weather patterns to assess their impacts on operations in the NAS. The ATCSCC also continually monitors operations in the NAS in order to identify and address constraints in a timely manner [16]. Operations in the NAS are managed by approximately 14,695 Air Traffic Controllers stationed at the ATCSCC, 518 Airport Traffic Control Towers, 154 Terminal Radar Approach Control (TRACON) facilities, and 22 Air Route Traffic Control Centers (ARTCC) [13]. The NAS supports aircraft throughout the different phases of flight from taxiing at their departure airports through takeoff, to operations around airports as well as en-route to their destinations, and during final approaches and landings at their destinations. Figure 1.2 provides an overview of which facilities handle aircraft during the different phases of flight.

Figure 1.2: Air Traffic Flow Chart [17]

Airport Towers control aircraft within 5 miles of airports and on the ground, and clear aircraft to depart and land [18]. Terminal Radar Approach Control (TRACON) facilities on the other hand, control aircraft in terminal spaces [18, 19], which are typically 5 - 50 nautical miles around airports [17].

The 22 Air Route Traffic Control Centers (ARTCC) manage en-route flights operating under Instrument Flight Rules (IFR) [20], and support flights operating under Visual Flight Rules (VFR), [21] with VFR traffic advisories [18, 22]. 20 ARTCCs are located in the contiguous United States [19], as seen in Figure 1.3, while the other two are located in Anchorage and Honolulu. Table A.1 in Appendix A provides a list of ARTCC and their locations in the United States.

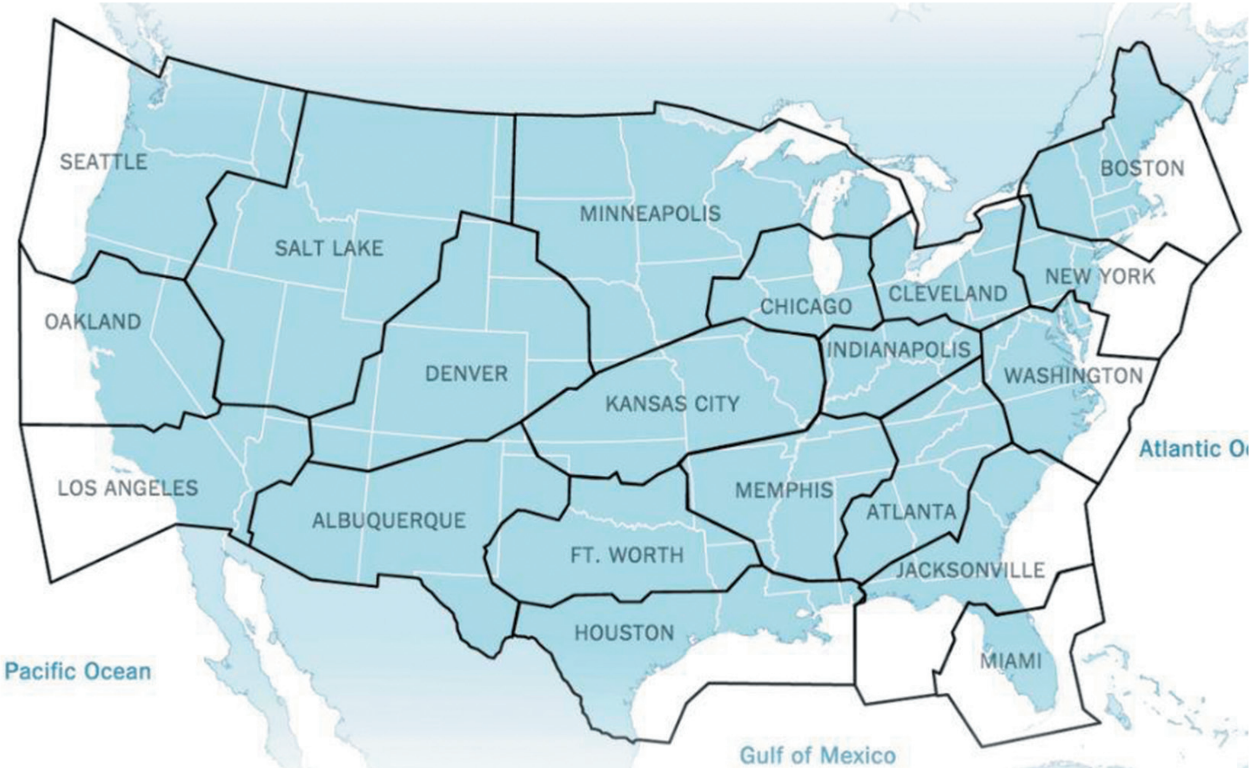

Figure 1.3: Air Route Traffic Control Centers (ARTCC) except Anchorage and Honolulu [23]

As previously mentioned, the NAS is the largest, busiest, and most complex airspace in the world. However, as with any other system, it is continually constrained by evolving challenges that need to be addressed. One of these challenges involves improving the efficiency of its operations so as to reduce the number, duration, and impact of flight delays. Table 1.1 shows the proportion of on-time flights compared to the different types of delays that occurred from June, 2003 to April, 2020. In particular, it shows that 78% of flights arrived on time while late arriving aircraft and NAS-related delays were the highest causes of flight delays.

Table 1.1: Flight Delay Statistics (June, 2003 - April, 2020) [24]

Metric	**Contributing value (%)**
On Time	78
Late Arriving Aircraft	7
National Airspace System	7
Air Carrier	5
Cancelled	2
Inclement Weather	1
Diverted	<1
Security Delays	<1

Flight delays have significant impacts on airlines, passengers, and the United States economy, as seen in Figure 1.4, which provides a breakdown of the direct costs of air transportation delays in terms of passengers, airlines, lost demand, and impact on the Gross Demand Product (GDP) of the United States in 2007 dollars. The $16.7 billion passenger component is comprised of the passenger time lost due to schedule buffers, delayed flights, flight cancellations, and missed connections, which lead to loss in productivity and opportunities for business travelers as well as an opportunity cost of time for leisure passengers [25]. The $8.3 billion airline component is comprised of increased operating costs for crew, fuel, maintenance, etc. The $3.9 billion cost associated with lost demand represents an estimate of the time or productivity loss incurred by passengers who avoid air travel as a result of delays. The effects of flight delays on airlines and passengers also have indirect impacts on the economy. Delays may lead to increased fuel costs, which may lead to increased airfares. Increased airfares as well as delays in general may lead to changes in consumer spending on travel and tourism-related goods and services [26], which eventually impacts the economy.

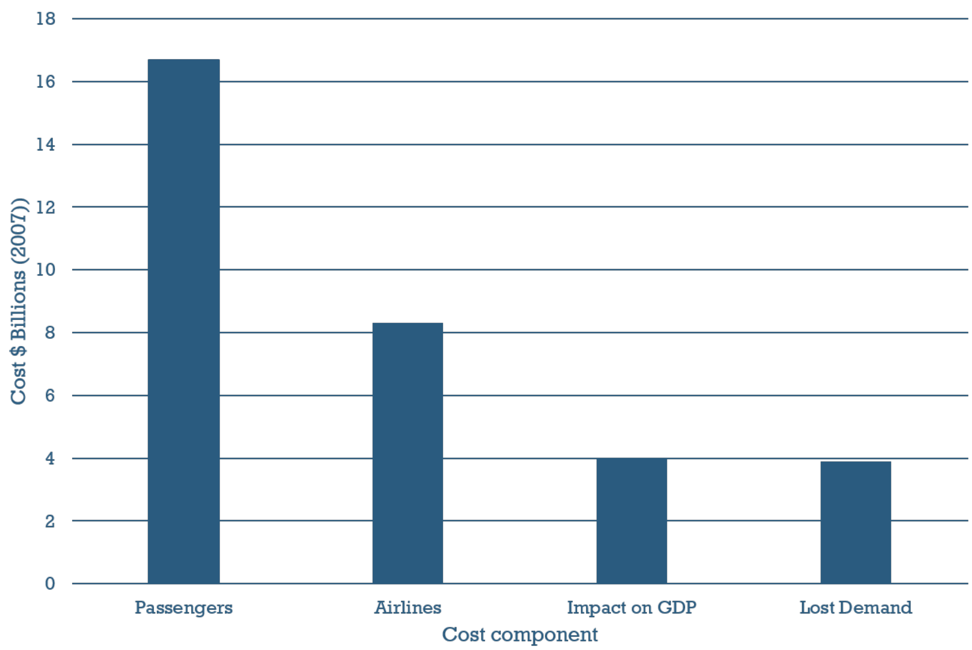

Figure 1.4: Direct cost of air transportation delays (2007 dollars) [27]

As illustrated, flight delays are costly to passengers, airlines, and the U.S. economy. As such, there is a need to improve the efficiency of operations in the NAS, particularly at airports as flight delays that originate at airports often propagate across the NAS. Unfortunately, the efficient management and operation of airports in the NAS has been hindered by factors such as airport capacity constraints and the implementation of various traffic management decisions and procedures.

1.1.1 Airport Capacity Constraints

Scheduled demand was often close to, or exceeded airport capacity during inclement weather at a majority of major U.S. airports prior to the economic downturn caused by the Novel Coronavirus (COVID-19) pandemic [28–31] in 2020. Scheduled demand was also often close to, and during certain hours of the day exceeded airport capacity even in good weather at airports such as LaGuardia, Newark Liberty, and John F. Kennedy International Airports [32], causing them to be capacity-constrained. Indeed, these three airports in addition to the Philadelphia and Hartsfield-

Jackson Atlanta International Airports were projected by the FAA to be capacity constrained by 2020 in a report published in 2015 [33]. The FAA also projected that these airports, in addition to the four airports represented by yellow circles in Figure 1.5, will be significantly capacity constrained by 2030 unless significant efforts are made to improve airport infrastructure and increase airport capacity. It is worth noting that a majority of these airports are the hubs to major U.S. airlines, some of which were formed as a result of the consolidation of airlines through mergers over the last couple of decades. As such, the operations of major U.S. airlines have been concentrated at fewer hub airports [33], leading to increased capacity constraints.

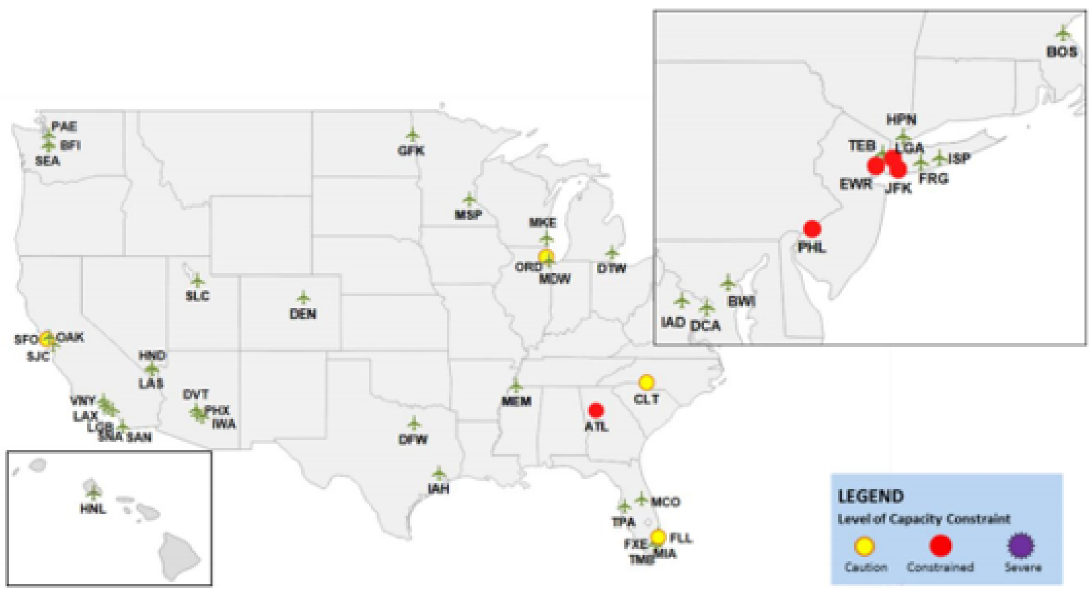

Figure 1.5: Assessment of Airport Capacity (2015) [33]

The aviation industry witnessed a steady increase in the volume and frequency of global air traffic prior to the economic downturn caused by the Novel Coronavirus (COVID-19) pandemic. This increase was largely influenced by technological advancements and increased demand as a result of global Gross Domestic Product (GDP) [34–37] and Revenue Passenger Kilometers (RPKs) growth [38]. This trend was evident in the United States, and demand for air transportation was projected to further increase over the next decade, as seen in Figure 1.6, which shows actual and forecasted annual airport operations from studies conducted in 2005 and 2012 at 30 major U.S. airports.

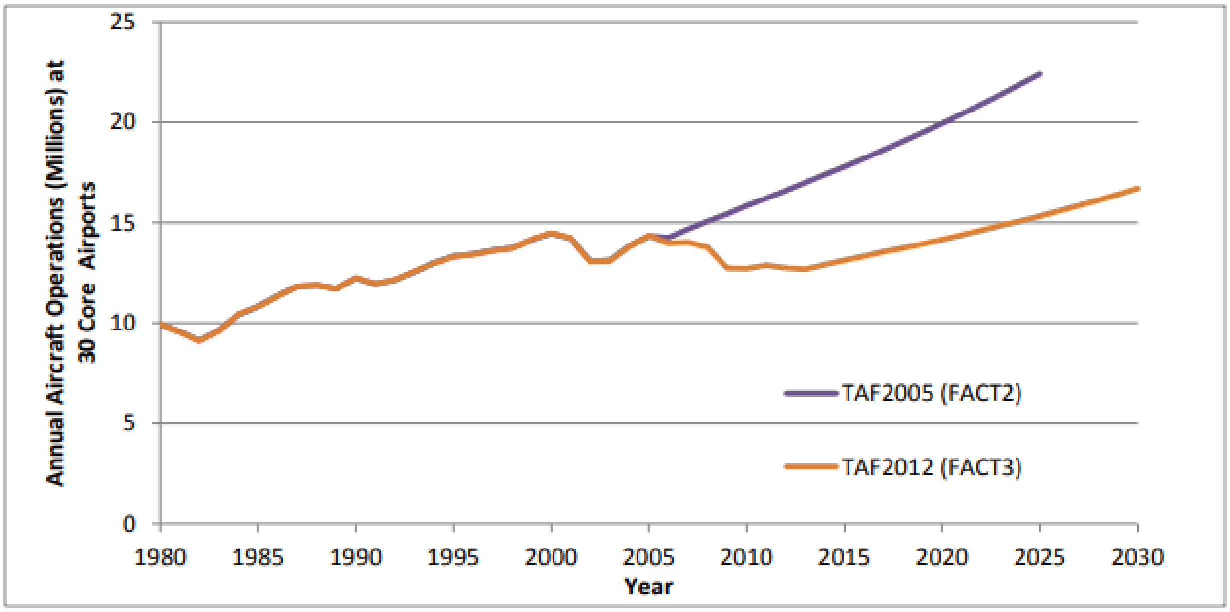

Figure 1.6: Actual and Forecasted Annual Operations at 30 U.S. Airports [33]

This trend as well as the projected increase in demand for air transportation can be attributed to a variety of reasons. Air transportation is regarded as safer compared to other modes of transportation [3] as air-related accidents had the lowest frequency among the four major modes of transportation in the United States from 1998 to 2017 [39]. Air transportation is also less time consuming compared to other modes of transportation [40], particularly for intercontinental and transcontinental travel, and the opportunity cost of time is much lower with air travel, which is particularly important for business travelers [41]. It is thus expected that major U.S. airports will continue to be capacity-constrained unless significant efforts are made to improve airport infrastructure and increase airport capacity to match the projected demand for air transportation.

1.1.2 Traffic Management Decisions and Procedures

The NAS is often constrained by inclement weather, volume constraints, equipment failures, closed runways, etc. Whenever this happens, traffic management personnel plan and implement Traffic Management Initiatives (TMI) to ensure that there is a balance between demand and capacity in constrained areas of the NAS [42–45]. Traffic Management Initiatives are also implemented to ensure that en-route and terminal areas of the NAS remain safe and operate optimally in spite of constraints. En-route Traffic Management Initiatives such as Reroutes [46], Airspace Flow Programs [47–50], and Miles-in Trails [51–53] are implemented to manage air traffic at constrained areas of the NAS during the en-route phase of flight [17]. Terminal Traffic Management Initiatives such as Ground Delay Programs [54–57] and Ground Stops [45, 58], on the other hand, are implemented at airports to ensure that their Airport Acceptance Rates (AAR) exceed or match aircraft demand.

Ideally, Terminal Traffic Management Initiatives should be planned well ahead of time to enable traffic management personnel and flight operators to better manage airport and flight operations, in spite of constraints at airports. However, this is often not the case due to rapidly evolving conditions at airports. Rapidly changing conditions at airports also often lead to changes in the scope and duration of Traffic Management Initiatives prior to and/or during their implementation, which further increases the number, duration, and impact of flight delays. Furthermore, the dynamic nature of airport and flight operations occasionally results in the initiation of a Traffic Management Initiative during the implementation of another Traffic Management Initiative [59]. This coincidence often occurs due to rapid changes in conditions, which leaves traffic management personnel with limited time to efficiently plan and implement the Traffic Management Initiatives separately [58]. As such, efforts need to be made to ensure that Traffic Management Initiatives are planned and implemented in an efficient manner so as to reduce the number, duration, and impact of flight delays caused by their implementation.

1.1.3 Current Efforts Towards Improving the Efficiency of Airport Operations

As illustrated, the efficient management and operation of U.S. airports is hindered by airport capacity constraints and the implementation of various traffic management decisions and procedures. As such, various efforts have been pursued by stakeholders in the aviation industry to address this evolving challenge by way of improving airport infrastructure, increasing airport capacity, etc. These efforts, primarily through the FAA's Next Generation Air Transportation System (NextGen) initiatives, have aimed to modernize operations in the NAS through the planning and implementation of new technologies and airspace procedures [60–65]. Indeed, the development and implementation of programs such as the Automatic Dependent Surveillance – Broadcast (ADS-B) [66–73], Data Communications (Data Comm) [74–76], En Route Automation Modernization (ERAM) [77–81], Terminal Automation Modernization and Replacement (TAMR) [81–87], System Wide Information Management (SWIM) [81, 88–91], and other efforts [92–102], have led to the avoidance of over 15,000 hours of delays and nearly 25,000 hours of communication time saved due to improved data sharing processes [65]. In addition, average delay duration in 2018 matched that of 2017 despite an increase in the number of constraints due to inclement weather and air traffic congestion [65]. Even though these efforts have improved operations at airports and consequently, the NAS, much more needs to be done to address this evolving challenge as inefficient airport operations will most likely impede the growth of the aviation sector in the United States. One approach towards addressing this challenge involves efficiently analyzing and assessing airport operational performance to facilitate the identification of trends and patterns for better decision making.

1.1.4 Analysis and Assessment of Airport Operational Performance

Traffic management personnel regularly analyze projected airport demand, forecasted weather conditions, and the statuses of airport systems, equipment and infrastructure in order to plan daily airport operations. Ideally, the impact and effectiveness of the implementation of traffic management decisions and procedures on daily airport operations should be analyzed and assessed in an efficient manner, so as to identify trends and patterns to better inform decision making, as summarized in Figure 1.7. As such, efforts such as the development of efficiency metrics for assessing TRACON and airport operational performance, and the Operational Service Performance Criteria have been pursued by stakeholders.

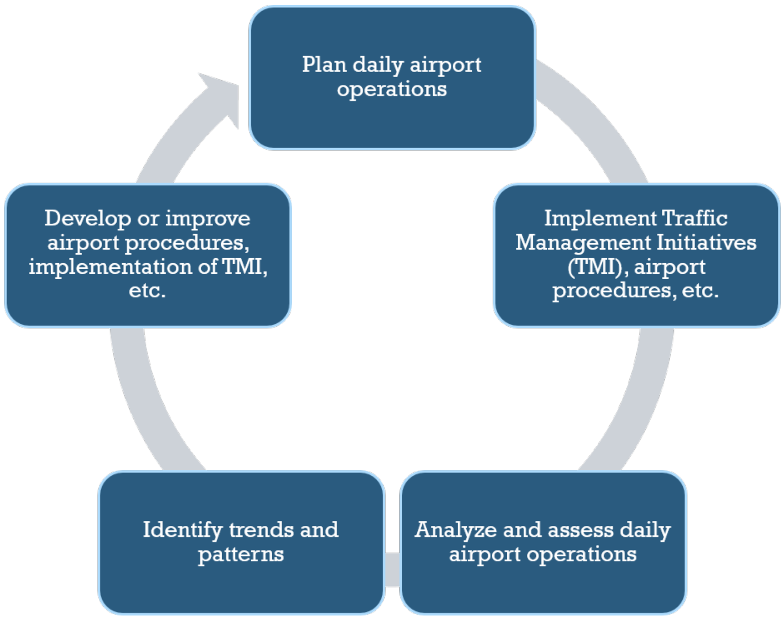

Figure 1.7: Overview of need to analyze and assess daily airport operations

Efficiency Metrics for TRACON and Airport Operations

Metrics such as Terminal Arrival Efficiency Rate, Departure Efficiency Rate, Arrival Efficiency Rate, and System Airport Efficiency Rate [103, 104] have been developed and implemented to measure the efficiency of TRACON and airport operations.

The Terminal Arrival Efficiency Rate (TAER) "measures the arrival efficiency of flights from 100 miles out to Wheels On for a given time period. It is calculated by dividing the actual number of arrivals by the lesser of the facility set arrival rate or the number of demand units and is reported as a percentage not to exceed 100". This metric was designed to measure "TRACON performance and the impact of Traffic Management Initiatives within 100 miles of airports" [105]. The Departure Efficiency Rate (DER) measures system efficiency from the time of take-off up to top of climb [103], whereas the Arrival Efficiency Rate (AER) measures "the extent to which an airport handled the number of aircraft it indicated it could accommodate, and how well the demand was met" [104]. The System Airport Efficiency Rate (SAER) is the weighted average of Departure Efficiency Rate and Arrival Efficiency Rate, and "measures the extent to which an airport handled the number of aircraft it indicated it could accommodate and how well demand was met" [104].

These metrics provide a measure of how TRACON facilities and airports handle demand. However, they do not provide insights into how traffic management decisions and procedures impact daily airport operations, and do not facilitate the identification of trends and patterns for better decision making.

Operational Service Performance Criteria

The Operational Service Performance Criteria (OSPC) was developed by FAA analysts to categorize daily airport operations as a means to facilitate their assessment. This effort categorizes daily operations of eight U.S. airports using the following metrics that capture the impact of traffic management decisions on daily airport operations:

- **TMI To Airport Delays**: These are delays to airports caused by the implementation of Traffic Management Initiatives (TMI) [106]

- **Departure Delays**: Departure delays in excess of 15 minutes attributed to conditions at the departure airport [106]

- **GDP Revisions**: Ground Delay Programs are Traffic Management Initiatives implemented when aircraft demand is projected to exceed airport capacity over a long period of time [107–109]. This parameter refers to the number of times that Ground Delay Programs are updated

- **GDP Lead-in Time (Minutes)**: The time between the proposal of a Ground Delay Program and its implementation

- **Ground Stops**: These are Traffic Management Initiatives implemented when aircraft demand is projected to exceed airport capacity over a short period of time [110]. This parameter refers to the number of unique Ground Stops implemented at airports

- **Number of Aircraft Affected by Airborne Holdings**: An Airborne Holding occurs when an en-route aircraft is issued a clearance in excess of 15 minutes for a predetermined maneuver to keep the aircraft within a specified airspace while awaiting further clearance from Air Traffic Controllers [111]

- **Total Duration of Airborne Holdings (Minutes)**: The summation of durations of all Airborne Holdings over 15 minutes

- **Diversions**: The number of flights that were diverted from their originally intended arrival airport

- **Completion Rate**: The percentage of scheduled and/or planned air carrier arrivals that were not cancelled [111]

Each of these metrics is classified as green (good), yellow (average), or red (bad) using predefined ranges of values determined by Subject Matter Experts, as seen in Figure 1.8. Table 1.2 shows the criteria for good performance for each of the metrics. Each daily operation is then classified as a "Good day", "Average day", or "Bad day" by identifying the predominant class of metrics (green, yellow, red) for the airport, as seen in Figure 1.8, where EWR, for example, was classified as a "Good day" because green (good) was the predominant class of metrics. This effort is currently conducted each weekday morning using the previous day's data, whereas daily operations for each Friday and Saturday are categorized on Monday mornings. This effort currently categorizes daily operations of the Boston Logan (BOS), Newark Liberty (EWR), LaGuardia (LGA), John F. Kennedy (JFK), Philadelphia (PHL), Dulles (IAD), Baltimore/Washington International Thurgood Marshall (BWI), and Reagan National (DCA) airports, as seen in Figure 1.8.

August XX, 20XX	Green	Yellow	Red	BOS	EWR	LGA	JFK	PHL	IAD	BWI	DCA
TMI To (including GS, GDPs, Other)	0-75	76-200	+201	0	211	260	88	134	0	0	60
Departure Delays	0-25	26-75	+75	0	90	262	60	152	36	30	22
GDP Revisions (*Manual Entry from NE Recap*)	0-1	2-3	+4	0	2	1	1	0	0	0	0
GDP Lead-in Time (Minutes)	+120	45-119	-45	n/a	146	0	143	105	n/a	n/a	0
Ground Stops (*Manual Entry from NE Recap*)	0-1	2-4	+5	0	0	1	1	1	0	0	1
Airborne Holding (Minutes)	0-75	76-200	+201	0	0	33	0	279	0	114	595
Airborne Holding (# of aircraft)	0-7	8-20	+21	0	0	2	0	10	0	5	25
Diversions	0-4	5-10	+11	0	1	1	1	1	0	1	9
Completion Rate	+90	80-90	-80	98.91	96.94	97.89	93.54	97.78	98.76	98.11	96.76
				G	G	G	G	Y	G	G	G

Figure 1.8: Operational Service Performance Criteria for a particular day

Table 1.2: Criteria for assessing OSPC metrics

Metric	Criteria for good performance
TMI To Airport Delays	Minimize
Departure Delays	Minimize
GDP Revisions	Minimize
GDP Lead-In Time	Maximize
Ground Stops	Minimize
Airborne Holdings (Minutes)	Minimize
Airborne Holdings (Aircraft)	Minimize
Diversions	Minimize
Completion Rate	Maximize

As with any other system, OSPC has a number of limitations that need to be addressed. First, this process is time consuming as analysts manually extract and process data into day-specific documents, as seen in Figure 1.8, from reports provided by the FAA's Aviation Systems Performance Metrics (ASPM) [112] platform, which provides FAA operations and performance data such as traffic counts, forecasts of aviation activity, delay statistics, etc. over different time periods (daily, monthly, etc). Indeed, FAA analysts spend more time extracting and processing the data for each airport from multiple ASPM reports compared to the amount of time spent analyzing the data. Metrics such as Ground Delay Program lead-in time and number of revisions, number of Ground Stops, and Completion Rate are calculated and compiled with the other metrics into a document on a daily basis. As such, the current process is also prone to errors as the calculation and compilation of data is done manually. The current manual process of data extraction, compilation, and storage on a local machine also limits the ability of FAA analysts to expand the scope of their work to include additional airports, and highlights a challenge pertaining to the efficient extraction, processing, fusion, and analysis of aviation data that researchers and analysts continually face.

The current process classifies daily operations of the eight airports into three categories as means to assess operational performance. However, additional categories may provide more insights into airport operations. As such, there is a need to determine if classifying daily airport operations into three categories is the best suited approach for assessing airport operational performance, and if and/or how this varies across airports. The current process also assumes a broad set of predetermined ranges for metrics across the eight airports. These predetermined ranges are based on the opinions of Subject Matter Experts and may not be necessarily accurate for all eight airports.

The current approach of using the predominant class (green, yellow, red) of metrics to categorize daily airport operations assumes that each parameter is weighted equally. However, the impact of each parameter may vary on an airport-by-airport basis. As such, robust and repeatable methodologies are needed to better categorize and determine the category that daily airport operations belong to, instead of using predefined ranges of metrics and the predominant class of metrics, respectively. Furthermore, OSPC currently does not account for any impacts that the time of year may have on the categorization of daily airport operations.

OSPC currently classifies daily airport operations into three categories without quantifying the degree to which a daily operation belongs to its category. As such, an approach for efficiently comparing, analyzing, and assessing daily airport operations across different days, months, and/or years for the identification of trends and patterns is lacking. Addressing this gap will enable stakeholders to analyze and compare daily operations in similar and different airport categories, which will enable them to identify traffic management decisions that lead to “very good” or “barely good” operational performance, for example. Doing so will also lead to the identification of trends and patterns for the improved planning and implementation of various traffic management decisions and airport procedures.

Finally, even though OSPC provides an assessment of airport operational performance, it does not provide insights into how traffic management decisions and procedures impact daily airport operations. Addressing this gap will enable stakeholders to make better decisions to ensure safe and efficient airport operations.

1.2 Motivation

Given the aforementioned limitations and gaps of current efforts pursued by stakeholders to analyze and assess airport operational performance, this present work presents a set of methodologies encapsulated in a framework that facilitate the analysis and assessment of daily airport operations. Given this, the overarching objective of this research is formulated as follows:

Develop a framework to facilitate the analysis and assessment of daily airport operations to improve airport operational performance

The remainder of this dissertation consists of the following chapters:

- Chapter 2 presents a review of relevant background material collected from literature review on the topics of analyzing airports and/or their operations
- Chapter 3 presents the formulation of the problem in terms of the research objective, questions, hypotheses, and experiments
- Chapter 4 presents the development and testing of a methodology for categorizing daily airport operations

- Chapter 5 presents the development and testing of methodologies for determining the category that daily airport operations belong to, and analyzing and assessing daily airport operations

- Chapter 6 presents the development and testing of a framework to facilitates the efficient extraction, processing, and storage of data for the analysis and assessment of daily airport operations

- Chapter 7 presents a summary of the work done in addition to contributions and recommendations for future work

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter presents a review of relevant background material collected from a survey of literature regarding the assessment and/or analysis of airports and/or their operations.

2.1 Survey of Literature

Bieblich et al. [113] developed a methodology for generating generic flight schedules so as to optimize airport and flight operations. This methodology involved categorizing airports using the minimum-variance-linkage/Ward-linkage clustering algorithm and using the clusters or categories as inputs for determining generic flight schedules. The airports were placed into seven categories that correlate to the nature of their operations (cargo hub, small regional airports, etc.) using metrics such as number of passengers and runways, revenue generated, distance to city center, etc. Even though the analysis of results showed that airports with similar characteristics (number of passengers, etc.) had quite comparable flight operations, the authors indicated that the accuracy of modeling weekly flight distributions could be further improved. In addition, even though this work highlighted the use of a clustering algorithm to successfully categorize airport operations, it did not provide insights into airport operations that were hitherto unknown.

Azzam [114] leveraged the Centroid Linkage clustering algorithm to categorize airports as a means to determine the roles that airports play in global air transportation. This effort was developed and implemented using data from the Official Airline Guide (OAG) [115] flight schedules database from 1979 to 2007. The airports were placed into twelve categories using the following metrics from network theory [116]: "sum of all seats on scheduled direct flights to and from the airport", "sum of all inbound and outbound flights", "sum of directly connected inbound and outbound origin-destination pairs", "degree to which an airport is connected to rest of the network", "number of shortest paths between any pair of airports in the network that connects via the airport", and "average great circle distance of all directly connected inbound and outbound origin-destination pairs". The airport categories were then analyzed to identify their different functions in the global air transport network. They were also used to create evolution graphs inspired by a first order Markov chain [117] to identify evolutionary patterns. Even though this effort provided a means for analyzing airports from a network perspective and established a new airport taxonomy, it did not provide a means for analyzing how traffic management decisions impact airport operations.

Ottl et al. [118] developed a methodology to categorize airports for air traffic simulation scenarios using data obtained from the Official Airline Guide (OAG) [115]. This effort involved using the Single Linkage clustering algorithm and 10 air traffic schedule parameters to initially identify and remove outliers from a set of 22 airports. The remaining group of airports were then used to calculate boundaries of each category and to determine parameter values for air traffic scenarios. Even though the scenarios were independent of specific airports, they captured the characteristics of similar airports within the same category. This effort involved grouping airports into categories. However, airport operations are impacted by a variety of factors (weather, volume etc.) and grouping several airports into categories may not be the most appropriate approach as their underlying characteristics may be lost.

Zambochova [119] grouped 838 airports into clusters based on the number of handled passengers as a means to better understand the popularity and intensity of the use of global air travel. This was achieved by leveraging the K-means clustering algorithm and monthly data from January 2000 to April 2014 to group the airports into four categories. The data used from this work was obtained from the French Institute of Civil Aviation, French Ministry for ecology, sustainable development and energy, and previous efforts by Darda P [120]. This effort outlined a repeatable approach for assessing the impact of events on air transportation. However, it involved grouping airports into categories, which may not be the best approach as their underlying characteristics may be lost.

Gorripaty et al. [121] developed a methodology for identifying similar days of air traffic management using quarter-hour arrival demand and capacity data obtained from Aviation System Performance Metrics (ASPM). This involved leveraging Principal Component Analysis to reduce the dimensionality of capacity data and leveraging clustering analysis to determine if inherent clusters exist in capacity and demand data of four U.S. airports. The outcome of this effort was the development of similarity measures based on capacity and demand data for Newark, San Francisco, Chicago O'Hare and John F. Kennedy International Airports. Even though this effort provided a methodology for identifying similar days of air traffic management, it did not facilitate the analysis and assessment of airport operational performance due to the metrics that were leveraged for this work.

Grabbe et al. [122] used a clustering algorithm to identify hours for which the probability of imposing a Ground Delay Program were similar at Chicago O'Hare International Airport and Newark Liberty International Airport. Ground Delay Programs (GDP) are utilized by controllers to manage air traffic whenever the number of anticipated aircraft is projected to exceed an airport's acceptance rate over a long period of time [54]. This effort was developed and implemented using weather, Ground Delay Program, and airport arrival demand and capacity data obtained from the Localized Aviation MOS (Model Output Statistics) Program (LAMP), National Traffic Management Log (NTML), and Aviation System Performance Metrics (ASPM), respectively. An analysis of the clusters was also conducted to identify the underlying weather conditions in each of these clusters. While this effort demonstrated how the fusion of historical weather and air traffic data, and the application of a clustering algorithm can be leveraged to provide guidance on the types of traffic management restrictions to implement in response to weather and traffic conditions impacting an airport, it did not provide insights into how traffic management decisions impact airport operations.

Grabbe et al. [123] leveraged a modified version of the K-means clustering algorithm to identify similar daily airport operations in the National Airspace System based on the locations and causes of Ground Delay Programs. This effort was implemented using weather, hourly scheduled arrival rates, and Ground Delay Program data obtained from Meteorological Aerodrome Reports (METAR), the FAA's Aviation System Performance Metrics (ASPM), and the National Traffic Management Log (NTML). Grabbe et al. [124] also developed a similar methodology that leveraged Ground Delay Program, weather, airport arrival delay, and total NAS delay data from the National Traffic Management Log (NTML), Rapid Refresh, Aviation System Performance Metrics (ASPM), and the FAA's Operations Network (OPSNET), respectively. Both efforts used Ground Delay Program-related variables to identify unique daily categories across the NAS which were then analyzed to verify the causes of the Ground Delay Programs. It was observed from these efforts that similar daily operations in the NAS can be identified, and that clustering algorithms can be leveraged to identify underlying causes of Ground Delay Programs. However, these efforts did not provide insights into how traffic management decisions impact airport operations.

Hoffman et al. [125] leveraged clustering algorithms to transform an initial set of 65 NAS-related variables to 8 key variables that constituted NAS feature vectors, one for each day from Jan. 1, 2000 through Sept. 10, 2001. The 65 variables were obtained from the FAA's Operations Network Database (OPSNET), FAA's Aviation System Performance Metrics (ASPM), Bureau of Transportation Statistics (BTS), and Air Traffic Control System Command Center (ATCSCC) quality assurance data sources. The variables captured delay statistics, traffic counts, Traffic Management Initiatives, and weather conditions. Clustering analysis was then performed with the feature vectors to rank daily NAS operations by "levels of normality". Insights gained from this work served as recommendations on how similar daily operations can be used to validate NAS simulations. This effort was implemented on a NAS-wide level. However, its implementation on an airport-specific level may provide insights into airport operational performance that were hitherto unknown.

Finally, Gano et al. [126] leveraged the K-means clustering algorithm to group 517 days of NAS delay data into 10 categories using daily total delay time in minutes as the distance metric.

The NAS delay data was obtained from the FAA's Operations Network (OPSNET). This effort was developed and implemented to identify days with similar air traffic flow patterns and operational characteristics so as to evaluate and validate air traffic management concepts and simulations, respectively. Similar to other efforts, implementing this methodology on an airport-specific level may provide insights into airport operations that were hitherto unknown.

2.2 Observations from Literature

This section outlines key observations made from the body of research identified from the literature that motivate this work. First, it was observed that several efforts have been made to categorize airports and/or their operations in order to characterize and/or analyze them. Even though these efforts did not facilitate the analysis and assessment of daily airport operations for the identification of trends and patterns, they provided insights into airports and their operations that were hitherto unknown.

Several of these efforts leveraged clustering algorithms to categorize data for airport operations research. Clustering algorithms perform differently depending on the type and amount of data, as well as the algorithms' methodology and predefined number of clusters. However, it was observed that a rigorous benchmarking of different clustering algorithms to identify the best suited one for various tasks was lacking.

Previous efforts also involved grouping airports and/or their operations into categories. However, airport operations are characterized and impacted differently by a variety of factors (geographic, runway configurations, etc.). As such, grouping several airports into categories may not be the most appropriate approach, as their underlying characteristics may be lost.

It was also observed from the survey of literature that an approach for analyzing and assessing the impact and effectiveness of traffic management decisions on airport operations is lacking, primarily due to the metrics used in various efforts. Addressing this gap will enable stakeholders to make beteer decisions to ensure safe and secure airport operations. In addition, an approach for quantifying the degree to which airports and/or their operations belong to various categories was lacking. Addressing this gap will enable stakeholders to analyze and compare operations across in similar and different airport categories. Doing so will lead to the identification of trends and patterns for the improved planning and implementation of various traffic management decisions and airport procedures.

As previously discussed, traffic management personnel regularly analyze projected airport demand, forecasted weather conditions, and the statuses of airport systems, equipment and infrastructure in order to plan daily airport operations. Ideally, the impact and effectiveness of the implementation of traffic management decisions and procedures on daily airport operations should be analyzed and assessed in an efficient manner, so as to identify trends and patterns to better inform decision making, and develop or improve the implementation of airport procedures, Traffic Management Initiatives, etc. However, it was observed that a comprehensive and robust approach is lacking, and that there are several gaps in literature that need to be addressed. As discussed in this chapter and in the review of the Operational Service Performance Criteria in Section 1.1.4, the categorization of airports and their operations has been successfully leveraged to analyze and gain insights into airports and/or their operations that were hitherto unknown. As such, there is a need to identify, extract, process, and store metrics that capture the impact of traffic management decisions on daily airport operations, and to leverage these metrics for the categorization of daily airport operations, which will facilitate their analysis and assessment for the identification of trends and patterns, as summarized in Figure 2.1.

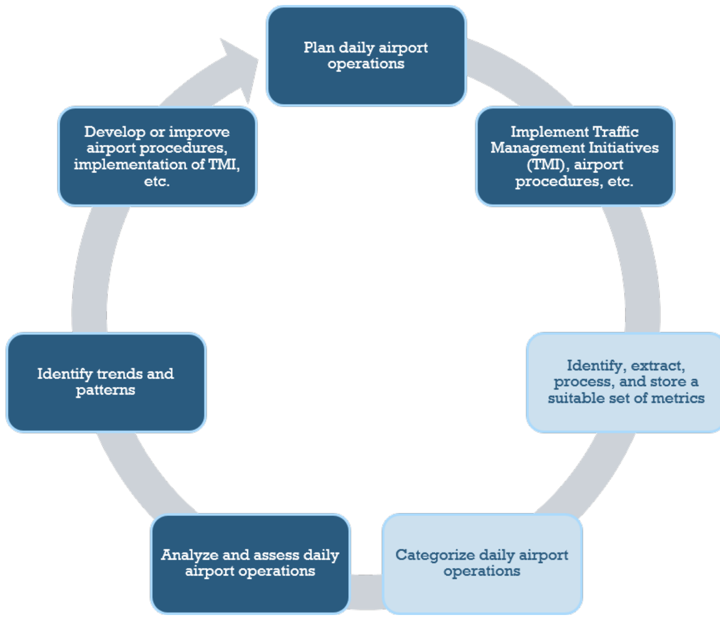

Figure 2.1: Proposed approach for the analysis and assessment of daily airport operations

The aforementioned observations led to the formulation of research questions, and their associated hypotheses and experiments, which are discussed in Chapter 3.

CHAPTER 3
PROBLEM FORMULATION

Chapter 1 discussed the need to efficiently analyze and assess daily airport operations to facilitate the identification of trends and patterns, which will inform better decision making to improve airport operational performance. However, the survey of literature in Chapter 2 revealed that a comprehensive and robust approach is lacking, and that there are several gaps in literature that need to be addressed. These observations led to the formulation of the overall objective of this dissertation presented in Section 3.1. In pursuit of this objective, several research questions were identified, and are presented in Section 3.2.

3.1 Research Objective

In order to address the need for a comprehensive, robust, and repeatable approach for analyzing and assessing daily airport operations, the objective of this dissertation is to:

Develop a framework to facilitate the analysis and assessment of daily airport operations to improve airport operational performance

With this research objective in mind, the overarching hypothesis of this work is stated as follows:

A framework that automates the extraction and processing of airport data, and facilitates the analysis and assessment of daily airport operations in a comprehensive, robust, and repeatable manner will enable stakeholders to identify trends and patterns for better decision making and as a consequence lead to improved airport operational performance

This overarching hypothesis is associated with methodologies that herein categorize daily airport operations, determine the category that subsequent daily airport operations belong to, provide a means for analyzing and assessing daily airport operations, and develop a framework to automate the extraction, processing, analysis, and storage of airport data. An overview of the overall methodology comprised of four main components is given in Figure 3.1. First, a repeatable methodology was developed to categorize daily airport operations, instead of using predefined ranges of metrics. A methodology for determining the category that daily airport operations belong to was then developed as an alternative to identifying the predominant class of metrics, as is currently done with OSPC. Outcomes of these two methodologies were then used to facilitate the analysis and assessment of daily airport operations in order to better understand how traffic management decisions and procedures impact airport operations, and to identify trends and patterns. Finally, a framework was developed to encapsulate the outcomes of the aforementioned methodologies and to automate the extraction, processing, and storage of data needed to efficiently analyze and assess daily airport operations.

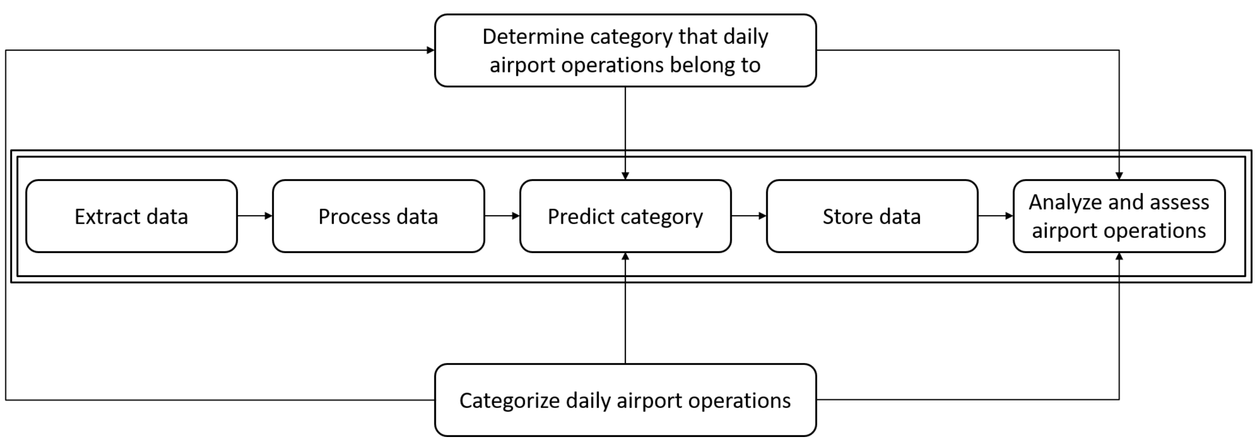

Figure 3.1: Overview of overall methodology

3.2 Research Questions, Hypotheses Development, and Experiments

Pursuant to the overarching objective of this work, a series of research questions were developed. The first question relates to the development of a methodology for categorizing daily airport operations. Then, a subsequent set of questions were posed which relate to determining the categories that subsequent daily airport operations belong to, and analyzing and assessing daily airport operations. The final research question probed the development of a framework to automate the extraction, processing, and storage of data needed for the efficient analysis and assessment of daily airport operations.

3.2.1 Research Question, Hypothesis, and Experiment 1

Research Question 1

It was observed from the review of the literature in Chapter 2 that various efforts analyzed airports and/or their operations by categorizing airport data using various approaches. However, these approaches have limitations that need to be addressed. The Operational Service Performance Criteria, for example, uses a broad set of predefined ranges of metrics to categorize daily operations of eight U.S. airports as a means to assess their operational performance. However, these predetermined ranges are based on the opinions of Subject Matter Experts and may not be necessarily accurate for all eight airports or may only work for these eight airports.

Other efforts that leveraged clustering algorithms to categorize airports and/or their operations also lacked benchmarking exercises to determine the best suited algorithms for various tasks. Benchmarking exercises are important as clustering algorithms perform differently depending on the type and amount of data as well as their methodologies and the predefined number of clusters. There is also a need to determine the optimal number of clusters needed to categorize airports and/or their operations, and how this varies by airport.

Finally, the developed methodology also needs to be systematic and repeatable so that it can be easily replicated for any airport. Given this, the first research question is formulated as follows:

Research Question 1

How can daily airport operations be categorized in a systematic, robust and repeatable manner?

Hypothesis 1

Within the literature, it was observed that clustering, which is a category of unsupervised Machine Learning, has been leveraged to categorize airports and/or their operations. Unsupervised Machine Learning algorithms develop models that benefit from the insights gained from summarizing data in new and interesting ways. These models are commonly referred to as descriptive models where no single feature is more important than the others. In unsupervised Machine Learning, the data is unlabeled and the algorithms learn to understand the structure of the data. Unsupervised Machine Learning algorithms can be used for tasks such as pattern discovery [127, 128] and clustering [127, 129–131]. Clustering algorithms categorize data into clusters, without any prior training or understanding of the data [127, 132, 133]. They are widely used to explore data in order to identify hidden patterns and can categorize data into a predefined number of clusters even if no meaningful clusters exist. This is done using different measures of similarity defined by metrics such as probabilistic distance [127, 134–136], which differ based on the methodologies of the algorithms. The methodologies of clustering algorithms as well as the predefined number of clusters, and the type and amount of data influences the consistency of clustering results and the quality of clusters. As such, there is often a need to determine the best suited combination of algorithm(s) and number of clusters for various tasks [137]. Given this, the following hypothesis is formed:

Hypothesis 1

If clustering algorithms are benchmarked while varying the number of clusters, then daily airport operations will be categorized in a systematic, robust and repeatable manner

Experiment 1

It was observed in the survey of literature that previous efforts do not facilitate the analysis and assessment of the impact of traffic management decisions on airport operations due to the metrics (number of arrivals, departures, etc.) used in these efforts. As such, Research Question 1 and Hypothesis 1 are answered and validated, respectively, by initially identifying a set of metrics that capture the impacts of traffic management decisions and procedures on daily airport operations. Consequently, the following metrics currently used in the Operational Service Performance Criteria are used for this work: delays to airports due to the implementation of Traffic Management Initiatives, departure delays, number of aircraft affected by Airborne Holdings, total duration of Airborne Holdings, number of diversions, Completion Rate, number of Ground Stops, and Ground Delay Program lead-in time and number of revisions.

Figure 3.2 provides an overview of the methodology for Experiment 1 which is implemented in R [138] with data from May 1, 2016 - December 31, 2019 for the following airports: Boston Logan (BOS), Baltimore/Washington International Thurgood Marshall (BWI), Reagan National (DCA), Newark Liberty (EWR), Dulles (IAD), John F. Kennedy (JFK), LaGuardia (LGA), and Philadelphia (PHL) airports.

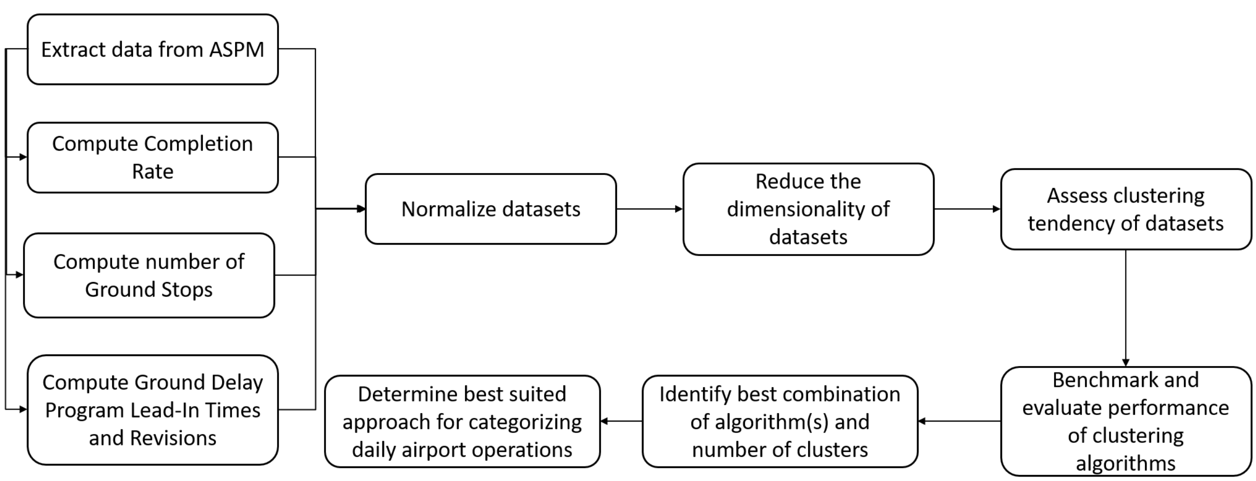

Figure 3.2: Methodology for Experiment 1

Experiment 1, as described in detail in Chapter 4, is carried out by extracting data from the FAA's Aviation Systems Performance Metrics (ASPM) platform. Completion Rate is computed using airport arrival and departure counts, whereas the number of Ground Stops, and Ground Delay Program lead-in times and number of revisions are computed using Ground Stop and Ground Delay Program data, respectively. Steps are then taken to normalize, reduce the dimensionality, and assess the clustering tendency of each airport's dataset. The datasets are then clustered by benchmarking the performance of ten clustering algorithms while varying the number of clusters from three to eight for each airport. This range of clusters was determined to be appropriate by FAA Subject Matter Experts. The performance of the algorithms is evaluated with seven metrics to assess the consistency of clustering results and the quality of clusters. The combination of best suited clustering algorithm and number of clusters for each airport are then determined by identifying the most common combination of algorithm(s) and number of clusters. Finally, the results from clustering are compared to results from the Operational Service Performance Criteria and reviewed by FAA Subject Matter Experts to determine which approach better categorizes daily airport operations.

3.2.2 Research Questions, Hypotheses, and Experiments 2

Research Questions 2

Within the literature, it was observed that a robust and repeatable methodology for determining the category that subsequent daily airport operations belong to is lacking. It was also observed that an approach for comparing daily operations at airports is lacking. Addressing this gap will enable stakeholders to analyze and compare daily operations in similar and different airport categories, which will allow them to identify traffic management decisions that lead to "very good" or "barely good" operational performance, for example. Doing so will also lead to the identification of trends and patterns for the improved planning and implementation of various traffic management decisions and airport procedures.

The review of the literature also showed that an approach for analyzing and assessing the impacts and effectiveness of traffic management decisions on airport operations is lacking. Addressing this gap will enable stakeholders to identify trends and patterns, which will inform better decision making thereby ensuring safe and efficient airport operations. Consequently, the second research question is threefold:

Research Question 2.1

How can the category that a daily airport operation belongs to be better determined?

Research Question 2.2

How can daily airport operations in similar and different categories be compared for the identification of trends and patterns?

Research Question 2.3

How can the impacts and effectiveness of traffic management decisions on daily airport operations be analyzed and assessed?

Hypotheses 2

Models used in supervised learning are known as predictive models where the algorithms attempt to discover and model the relationships between the value being predicted (target) and other values (predictors). Predictive models can be used to predict not only events in the future, but can also be used to predict previous and real-time events [139]. Unlike with unsupervised Machine Learning algorithms, the data is labeled in supervised learning so that the algorithms can learn to predict the target from the predictors. Supervised Machine Learning algorithms can be used for two tasks: regression and classification [127].

Regression is used to predict a numeric target from a set of predictors. The targets are continuous because there are no discontinuities or gaps in the values that they can take [127]. Examples of regression tasks include predicting the duration of flight delays, cancellations, etc. Classification is often used to predict which class an instance belongs to. This involves mapping predictors to a target by learning how predictors are related to the target [127]. Examples of classification tasks include predicting the occurrence of certain Traffic Management Initiatives, flight delays, etc [140–145]. The performance of regression and classification models are often evaluated using a set of metrics so as to assess how well the models will predict new observations. Therefore, the hypothesis for Research Question 2.1 is:

Hypothesis 2.1

If the performance of classification models developed with a supervised Machine Learning algorithm is observed to be excellent, then the category that a daily airport operation belongs to will be determined with a robust approach

Certain supervised Machine Learning algorithms compute the posterior probability or degree of support of predictions, which measures how likely each observation belongs to the different categories of a dataset [127]. As such, it is hypothesized that this feature will provide a means for quantifying the degree to which a daily operation belongs to a category. In particular, this will enable analysts to compare daily operations in similar and different airport categories as a means to identify trends and patterns. Consequently, the hypothesis for Research Question 2.2 is:

Hypothesis 2.2

If the posterior probability or degree of support of predictions is computed, then the comparison of daily operations in similar and different airport categories will be facilitated

Machine Learning models are often referred to as "black boxes" because their underlying processes are not explicitly clear to users. This challenge is often remedied by leveraging Decision Trees which allow for a global interpretation of Machine Learning models developed with tree-based algorithms. They also provide a means to identify underlying relationships between predictors and targets of prediction models [146]. Certain supervised Machine Learning algorithms also provide a ranking of predictor importance which can be used to gain insights into the underlying causes of events. It is thus hypothesized that Decision Trees and the ranking of predictor importance of each airport will provide a means for determining how various traffic management decisions impact airport operations and how this varies by airport. Consequently, the hypothesis for Research Question 2.3 is:

Hypothesis 2.3

If the ranking of predictor importance and Decision Trees of classification models are analyzed, then the impact and effectiveness of traffic management decisions on airport operations will be analyzed and assessed for the identification of trends and patterns

Experiment 2

Research Questions and Hypotheses 2 are answered and validated, respectively, by developing prediction models using the methodology shown in Figure 3.3 which is implemented with Python [147, 148]. The target of each model is the category of daily airport operation determined in Experiment 1, and the predictors are the nine airport metrics used in Experiment 1. As discussed in Section 1.1.4, the Operational Service Performance Criteria does not account for the impacts that the time of year may have on the categorization, analysis, and assessment of daily airport operations. As such, the month of year is also included as a predictor of the prediction models.

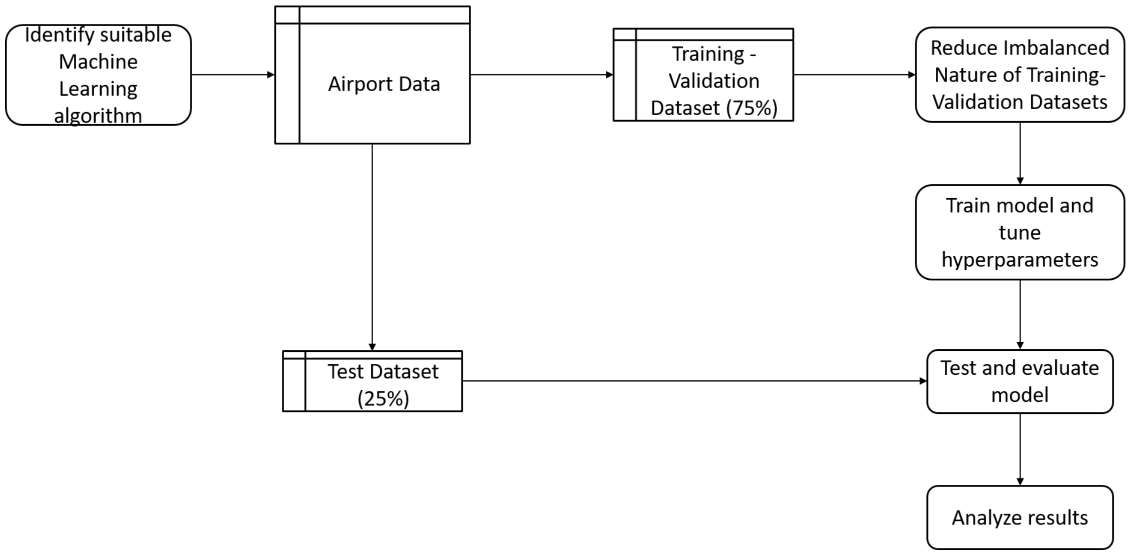

Figure 3.3: Methodology for Experiment 2

First, a suitable supervised Machine Learning algorithm is identified and the datasets from Experiment 1 are randomly split into two sets: training-validation and testing. The training-validation set is used to train and tune the algorithm's hyperparameters to ensure optimal model performance. The performance of the models are then assessed with the testing set and evaluation metrics in order to answer and validate Research Question and Hypothesis 2.1, respectively.

The posterior probability or degree of support of predictions made by the prediction models is generated as outputs of the models to facilitate the comparison of daily operations in similar and different airport categories. These outcomes are analyzed in order to answer and validate Research Question and Hypothesis 2.2, respectively.

Finally, Research Question and Hypothesis 2.3 are answered and validated, respectively, by analyzing the ranking of predictor importance and Decision Trees of the prediction models. The analysis of the ranking of predictor importance of the prediction models facilitates the identification of how each metric influences the categorization of daily airport operations, and how this varies across airport. The Decision Trees of the prediction models on the other hand, are leveraged to provide insights into how various traffic management decisions and airport procedures impact daily airport operations.

3.2.3 Research Question, Hypothesis, and Experiment 3

Research Question 3

As previously discussed in Section 1.1.4, FAA analysts manually extract airport needed for the Operational Service Performance Criteria from multiple ASPM reports each weekday morning. Indeed, FAA analysts spend more time extracting and processing data needed compared to the amount of time spent assessing airport operational performance. The current manual process of data extraction, processing, and storage on a local machine also limits the ability of FAA analysts to expand the scope of their work to include additional airports. In addition, an efficient means for comparing, analyzing, and assessing airport operations across different days, months, and/or years for the identification of trends and patterns is lacking. Consequently, the final research question is formulated as follows:

Research Question 3

How can the efficient analysis and assessment of daily airport operations be automated from data extraction, through processing, analysis, and storage?

Hypothesis 3

The manual and time-consuming nature of the process currently used by FAA analysts in OSPC highlights a challenge pertaining to the efficient extraction, processing, fusion, and analysis of aviation data that stakeholders continually face. This challenge is further exacerbated by the large volumes of data, commonly referred to as Big Data [149–152], are continually generated and utilized from different data sources within the NAS. This data is used by stakeholders to improve flight safety and fuel efficiency [153], ensure airline profitability [154–156], plan maintenance schedules [157], and proactively detect maintenance issues [158]. Aviation Big Data is also leveraged to "make accurate forecasts, streamline processes, identify operational patterns, and inform the development of new concepts and methods aimed at ensuring that operations in the NAS remain as safe and efficient as possible" [43]. However, challenges associated with the volume, veracity, velocity, and variety of the data prevents stakeholders from efficiently utilizing the data. In particular, stakeholders in the aviation industry are often unable to access integrated data in a timely manner, and often lack tools needed to efficiently analyze aviation Big Data. As such, various efforts have been pursued by stakeholders to address this challenge. One of these involves the development of the Enterprise Information Management (EIM) platform by the FAA. EIM is a cloud-based environment that is expected to provide reusable data, information management services, and Big Data processing capabilities for broad cross-agency use, once completed. This environment, represented in Figure 3.4 will facilitate continuous innovation by providing stakeholders with advanced tools and capabilities as a way to keep up with the fast-changing data and analytics landscape in the aviation domain.

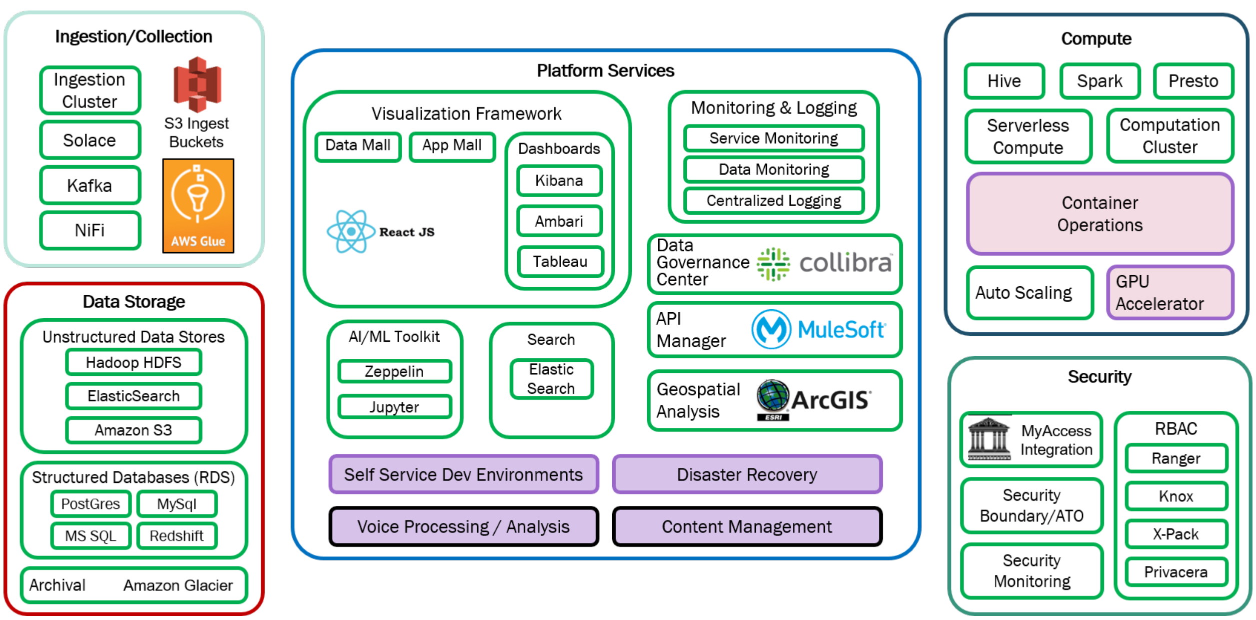

Figure 3.4: Overview of the FAA's Enterprise Information Management (EIM) platform

EIM will be equipped to rapidly ingest internal FAA datasets and external datasets in real-time, and store them in cloud-native formats for efficiency. It will also have the capability to transform datasets into formats suitable for consumption and data analytics purposes. These will be achieved by leveraging cloud computing services and scalable architectures such as Amazon Web Services [159–161], Elastic Search [162–164], Apache NiFi [165–169], PostgreSQL [170–172], and Hadoop and the Hadoop Distributed File System (HDFS) [173–175]. Indeed, leveraging these technologies and software will lead to increased cost efficiency, provision speed and enhanced information sharing, and will support innovation, research and sustainability in the aviation domain [176, 177]. As such, it is hypothesized that leveraging cloud-based technologies and software to develop a framework that encapsulates the methodologies and outcomes of Experiments 1 and 2, and automates the extraction, processing and storage of airport data will facilitate the efficient analysis and assessment of daily airport operations. Given this, the final hypothesis is:

Hypothesis 3

If a framework is developed to automate the extraction, processing, analysis and storage of airport data, then daily airport operations will be analyzed and assessed in an efficient manner

Experiment 3

Figure 3.5 provides a broad overview of the methodology for Experiment 3 which is used to answer and validate Research Question 3 and Hypothesis 3, respectively. In particular, it shows that the first step focuses on identifying technologies and software that are compatible with the FAA's EIM platform as the framework will be deployed into the EIM platform for the real-time analysis and assessment of daily airport operations. Subsequent steps focus on the deployment of various software and scripts needed for the framework, the initiation of the framework, extraction of data needed, processing of data, temporary storage of data, execution of Machine Learning scripts, permanent storage of data, and analysis of data.

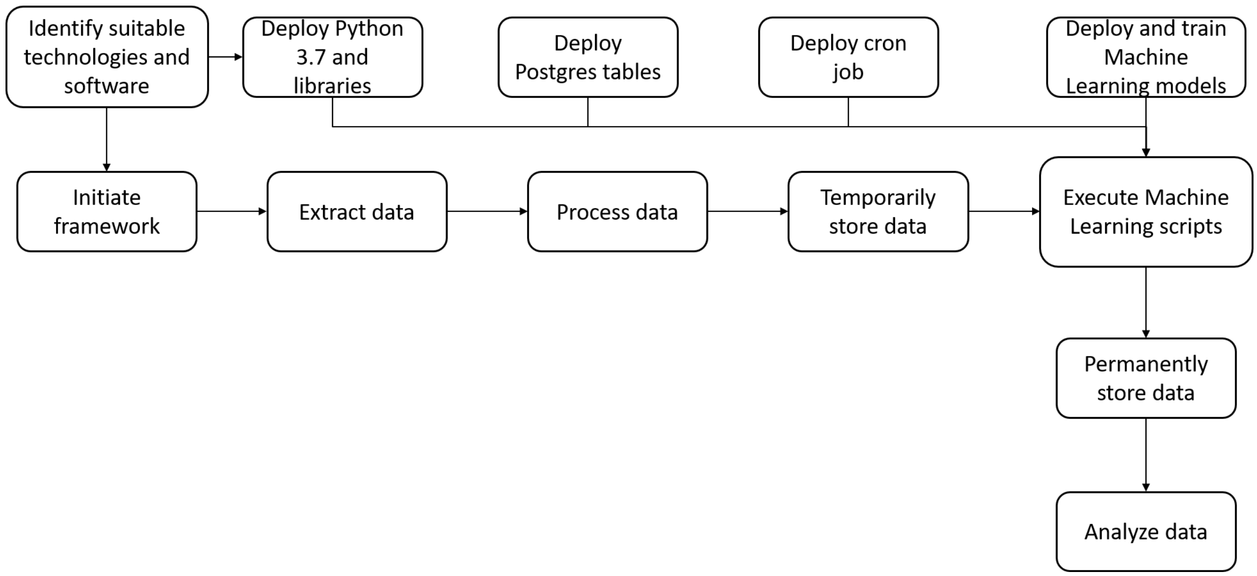

Figure 3.5: Methodology for Experiment 3

The framework is developed and tested in the FAA's Computing and Analytics Shared Services Integrated Environment (CASSIE), which is a collaborative and flexible environment for conducting research by bringing FAA organizations, partners and sponsors together in a shared services environment consisting of data, computing power, and analytical tools. The developed framework will eventually be deployed into the FAA's EIM platform, as CASSIE will be phased out of operation once EIM is operational. The technologies and software that are leveraged to develop the framework are deployed into CASSIE with Ansible scripts which automate the deployment of system configurations, software, codes, files, etc. into computing environments [178–180]. This replaces the manual and time consuming effort required to transfer, install and/or configure software or files in new environments. The technologies and software are then configured to automatically trigger the initiation of the framework each weekday morning, and ingest, process and store the data needed for this work. The outcomes of Experiments 1 and 2 are also incorporated into the framework to facilitate the analysis and assessment of daily airport operations. The completed framework is then tested over a four-month period to evaluate and validate its performance so as to answer and validate Research Question and Hypothesis 3, respectively.

CHAPTER 4

CATEGORIZATION OF DAILY AIRPORT OPERATIONS

In this chapter, the development, application, and testing of a methodology for categorizing daily airport operations is discussed so as to successfully answer Research Question 1:

Research Question 1

How can daily airport operations be categorized in a systematic, robust, and repeatable manner?

4.1 Methodology Overview

Figure 4.1 provides an overview of the methodology for Experiment 1 which was implemented with a set of metrics that capture the impacts of traffic management decisions and procedures on daily airport operations. Data from May 1, 2016 - December 31, 2019 was extracted from ASPM for the following airports: Boston Logan (BOS), Baltimore/Washington International Thurgood Marshall (BWI), Reagan National (DCA), Newark Liberty (EWR), Dulles (IAD), John F. Kennedy (JFK), LaGuardia (LGA), and Philadelphia (PHL) airports. The remainder of this section discusses each of these steps in detail.

4.1.1 Step #1: Extract Data from ASPM

The developed methodology for Experiment 1 was implemented using metrics identified by Subject Matter Experts that capture the impacts of traffic management decisions and airport procedures on daily airport operations. As such, the following metrics currently used in the Operational Service Performance Criteria (OSPC) were used for this work: delays to airports due to the implementation of Traffic Management Initiatives, departure delays, number of aircraft affected by Airborne Holdings, total duration of Airborne Holdings, number of diversions, Completion Rate, number of

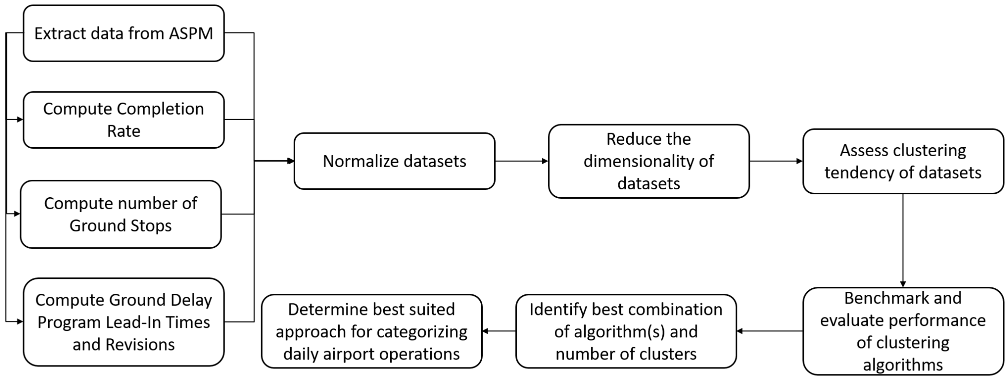

Figure 4.1: Overview of methodology for Experiment 1

Ground Stops, and Ground Delay Program lead-in time and number of revisions. These metrics were also used to facilitate the comparison of the outcomes of this work and OSPC in order to identify the better approach for categorizing daily airport operations to facilitate their analysis and assessment. Metrics such as delays to airports due to the implementation of Traffic Management Initiatives, departure delays, number of aircraft affected by Airborne Holdings, total duration of Airborne Holdings, and number of diversions were directly extracted from ASPM [112], whereas Completion Rate was calculated by extracting the number of actual arrivals and flight cancellations from ASPM. The number of Ground Stops, and Ground Delay Program revisions and lead-in times were calculated by extracting Ground Stop and Ground Delay Program data from ASPM.

4.1.2 Step #2: Compute Completion Rate

Completion Rate was calculated by extracting the number of actual arrivals and flight cancellations, and using Equation 4.1:

$$Completion\ Rate = \frac{Arrivals}{Arrivals\ +\ Cancellations} \tag{4.1}$$

4.1.3 Step #3: Compute the Number of Ground Stops

The number of Grounds Stops at each airport was computed using the following process:

1. The first step involved sorting the Ground Stop advisories by time, and removing duplicate advisories. Duplicate advisories exist because ASPM occasionally stores the same Ground Stop advisory multiple times

2. The duration and scope of an ongoing Ground Stop may be modified whenever conditions change. This leads to overlapping Ground Stop advisories, which is inaccurate. In order to address this inconsistency, the end time of an initial Ground Stop advisory was set as the start time of a new Ground Stop advisory as seen in Figure 4.2. In particular it shows that the start time of advisory number **0117** is prior to the end time of advisory number **0071**. Thus, the end time of advisory number **0071** was set as the start time of advisory number **0117**

Advisory Number	Start Time	End Time
0071	2017-04-20T16:00:00Z	2017-04-21T03:00:00Z
0117	2017-04-20T18:15:00Z	2017-04-21T03:00:00Z

Advisory Number	Start Time	End Time
0071	2017-04-20T16:00:00Z	2017-04-20T18:15:00Z
0117	2017-04-20T18:15:00Z	2017-04-21T03:00:00Z

Figure 4.2: Updating the end dates and times of updated active Ground Stop advisories

3. The number of Ground Stops for each day was then computed by identifying original Ground Stop advisories and ignoring subsequent updated advisories. This was achieved by identifying the number of times that start times of advisories did not match the end times of previous advisories. Finally, zero was indicated whenever a Ground Stop was not implemented on a specific day

4.1.4 Step #4: Compute Ground Delay Program Revisions and Lead-In Times

Ground Delay Program revisions and lead-in times were calculated by implementing the following steps:

1. The first step involved sorting the Ground Delay Program advisories by time, and removing duplicate advisories. Duplicate advisories exist because ASPM occasionally stores the same Ground Delay Program advisory multiple times

2. The duration and scope of an ongoing Ground Delay Program may be modified whenever conditions change. This leads to overlapping Ground Delay Program advisories which is inaccurate. In order to address this inconsistency, the end time of an initial Ground Delay Program advisory was set as the start time of the new Ground Delay Program advisory, similar to the process implemented with Ground Stop data, as seen in Figure 4.2. A Ground Delay Program revision was recorded whenever this step was implemented. A value of -1 was indicated whenever a Ground Delay Program was not implemented at an airport on a specific day, whereas zero was indicated whenever a Ground Delay Program was implemented with no revisions. This was done to distinguish between when a Ground Delay Program was implemented with no revisions, and when a Ground Delay Program was not implemented at an airport on a specific day

3. Ground Delay Program lead-in time was computed as the time between the proposal of original Ground Delay Programs and their implementation, in minutes. A value of 500 minutes was indicated whenever a Ground Delay Program was not implemented at an airport on a specific day. 500 minutes was used because a Ground Delay Program will never be proposed that much in advance. This was done to distinguish between when a Ground Delay Program was implemented with zero lead-in time, and when a Ground Delay Program was not implemented at an airport on a specific day

4.1.5 Step #5: Normalize Datasets

To ensure that metrics with larger ranges of values do not skew the clustering process, there is a need to normalize the parameters. Z-score standardization was thus leveraged to scale each metric to ensure that they had a mean of zero and a standard deviation of one [181, 182]. This was achieved with the mean and standard deviation of each metric, and was calculated using Equation 4.2.

$$Z = \frac{Value - Mean}{Standard\ Deviation} \tag{4.2}$$

4.1.6 Step #6: Reduce the Dimensionality of Datasets

It is usually difficult to explore and visualize the relationships between features in highly dimensional datasets. Thus, techniques such as Principal Component Analysis (PCA) have been widely used to reduce the dimensionality of datasets. This is achieved by orthogonally transforming a set of variables into principal components. Principal components are linear combinations of original variables of a dataset that capture the variance of the dataset. The transformation is done to ensure that the first principal component captures the maximum variance of the dataset, while each subsequent principal component captures the remaining variance of the dataset [183–186]. Orthogonal components that capture the maximum variance of the dataset are then identified using a scree plot, as seen in Figure 4.3. The scree plot, for this example, shows that two principal components captured over 90% of the variance of the dataset. This means that the dimensionality of the dataset was reduced from over 40 variables to two principal components without significantly reducing the variance of the dataset. Principal Component Analysis was implemented in this work by leveraging R's *prcomp* library [187, 188].

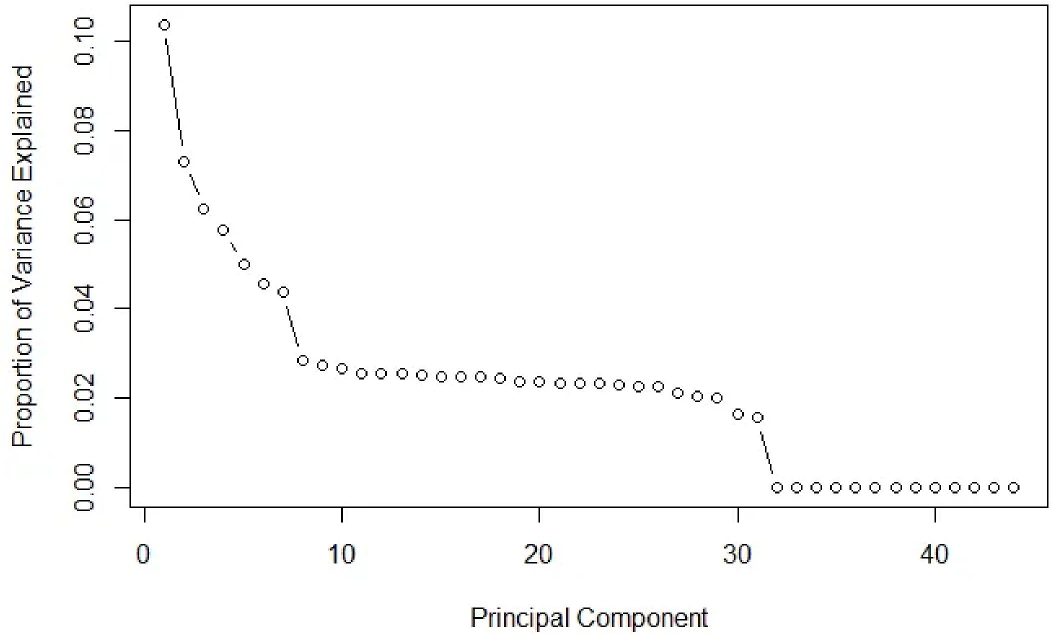

Figure 4.3: Sample scree plot [189]

4.1.7 Step #7: Assess the Clustering Tendency of Datasets

The majority of clustering algorithms split up datasets into predefined number of clusters, even if no meaningful clusters exist. It is thus important to assess the clustering tendency of a dataset to determine if meaningful clusters can be created [190–193]. Hopkins statistic [194] and Visual Assessment of cluster Tendency (VAT) [190–193, 195] are two methods that are usually used to determine if a dataset has a non-random structure and will produce useful clusters [195].

Hopkins Statistic

Hopkins Statistic assesses the clustering tendency of datasets by measuring the probability that a given data set is generated by a uniform data distribution [196]. The Hopkins Statistic for each airport was obtained using Algorithm 1 [197].

Algorithm 1: Hopkins Statistic

for each airport **do**
  Uniformly sample n points (p_1,.....p_n) from data set, T
  Find the nearest neighbor, p_j of each point, $p_i \in T$
  Compute the distance from each real point (p_i) to each nearest neighbor (p_j) and denote it as x_i = dist(p_i,p_j)
  Generate a simulated data set ($random_T$) drawn from a random uniform distribution with n points (q_1,.....q_n) and the same variation as the original real data set T
  Find the nearest neighbor, q_j of each point, $q_i \in random_T$
  Compute the distance from each artificial point (q_i) to the nearest neighbor (q_j) and denote it as y_i = dist(q_i,q_j)
  Calculate the Hopkins statistic (H) as the mean nearest neighbor distance in the random data set divided by the sum of the mean nearest neighbor distances in the real and across the simulated data set, as seen in Equation 4.3

$$H = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i + \sum_{i=1}^{n} x_i} \tag{4.3}$$

  return Hopkins Statistic
end for

If the data set T is uniformly distributed, then $\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n} x_i$ will be close to each other, and the Hopkins Statistic (H) will be approximately 0.5. However, if clusters are present in T, then the distances for points from the simulated data set ($\sum_{i=1}^{n} y_i$) will be substantially larger than for the ones from the uniformly distributed data set T ($\sum_{i=1}^{n} x_i$), and Hopkins Statistic (H) will be greater than 0.5. A value for H higher than 0.75 indicates a clustering tendency at the 90% confidence level.

Consequently, the null hypothesis for the Hopkins Statistic is defined as the dataset being uniformly distributed. The alternative hypothesis on the other hand is defined as the dataset not being uniformly distributed. Thus, a Hopkins Statistic close to one means that the null hypothesis is rejected and the dataset has a very high clustering tendency [198].

Visual Assessment of cluster Tendency (VAT)

The Visual Assessment of cluster Tendency (VAT) is an image that indicates the presence of meaningful and well separated clusters, represented by dark boxes along the main diagonal of the image, as seen in Figure 4.4. The VAT is implemented by computing the Dissimilarity Matrix (DM) between objects in the dataset using the Euclidean distance measure [199, 200]. An Ordered Dissimilarity Matrix (ODM) is then created by reordering the original Dissimilarity Matrix so that similar objects are close to one other. The Ordered Dissimilarity Matrix is then displayed as the VAT plot for the dataset [190–193]. The values indicated in Figure 4.4 are a measure of dissimilarity in a cluster or dataset. As such, lower dissimilarity values and darker boxes correspond to the presence of very similar objects in clusters. On the other hand, higher dissimilarity values and lighter shaded boxes correspond to dissimilar objects in clusters. Indeed, the size and number of the dark boxes along the diagonal provide a means for visually assessing the potential size and number of meaningful clusters in a dataset. This step was implemented by leveraging R's *factoextra* library [201, 202]

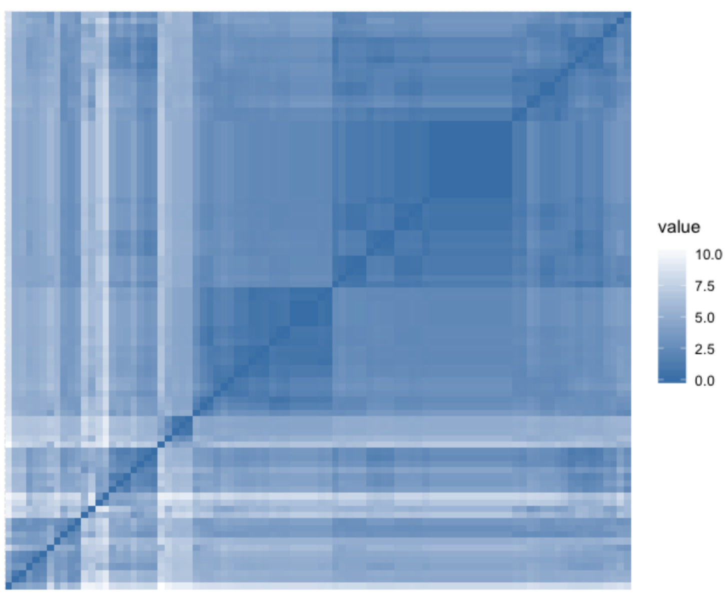

Figure 4.4: Sample Visual Assessment of cluster Tendency (VAT)

4.1.8 Step #8: Benchmark and Evaluate the Performance of Clustering Algorithms

A. Clustering Algorithms

Clustering algorithms have different methodologies and assumptions. It is thus important to benchmark different clustering algorithms in order to identify the best suited clustering algorithm for a dataset. Some clustering algorithms also require users to arbitrarily select the number of clusters to be used. In order to ensure that the optimal number of clusters is selected, there is a need to vary the number of clusters used while benchmarking the different algorithms to identify a combination of best suited algorithm(s) and number of clusters to be used for each airport. Consequently, the following clustering algorithms were benchmarked while varying the number of clusters for each airport using R's *clValid* library [203] due to their applications across multiple research domains [204–212]:

1. Divisive Hierarchical Clustering Algorithms

The divisive hierarchical clustering algorithms group similar objects in multidimensional spaces into categories by repeatedly dividing the largest cluster into at least two clusters [213]. The divisive hierarchical clustering algorithms utilized for this research were:

- Divisive Analysis (DIANA): This algorithm initially places all objects into the same cluster. At each point in time, the algorithm then splits the largest available cluster into two smaller clusters until each cluster contains at least one object [204, 205, 214, 215]. The individual clusters are then merged based on their similarity to each other

- Self Organizing Tree Algorithm (SOTA) Clustering Algorithm: This algorithm is an unsupervised network with a binary tree topology that combines advantages of hierarchical clustering and Self-Organizing Maps (SOM). It is implemented by splitting the most diverse node into two nodes, called cells. This process is repeated until a desired hierarchical level is reached or a predefined criteria based on the approximate distribution of probability is obtained by randomisation of the original data set [207, 208, 216]

2. Agglomerative Hierarchical Clustering Algorithms

The agglomerative hierarchical clustering algorithms group similar objects in multidimensional spaces into categories by assigning each object into a cluster, and then merging similar clusters by their proximity to each other [213, 217–220]. These algorithms differ based on their method for measuring distances between objects. The agglomerative hierarchical clustering algorithms utilized for this research were:

- Complete Linkage: The distance between two clusters is defined as the longest distance between two objects in each cluster [221]
- Average Linkage: The distance between two clusters is defined as the average distance between each object in one cluster to every object in the other cluster [221]
- Single Linkage: The distance between two clusters is defined as the shortest distance between two objects in each cluster [221]
- Ward: The distance between two clusters is defined as how much the sum of squares will increase when the clusters are merged [222]

3. Kmeans Clustering Algorithm

The Kmeans algorithm assigns objects to a predetermined number of clusters, where the differences between objects in each cluster are minimized, and the differences between objects in different clusters are maximized [127]. The algorithm is implemented by randomly selecting centroids which serve as the beginning points of each cluster. Iterative calculations are then carried out to optimize the positions of the centroids. The iterations are suspended when there is no change in centroid values or when a defined number of iterations is reached [223–226]

4. Partitioning Around Medoids (PAM) Clustering Algorithm

The Partitioning Around Medoids (PAM) or k-medoids clustering algorithm is similar to the Kmeans clustering algorithm. However, the PAM algorithm clusters objects into a predetermined number of clusters around medoids or centers [211, 212, 214]. The medoids are computed such that the total dissimilarity of all objects to their nearest medoid is minimal

5. Clustering for Large Applications (CLARA) Clustering Algorithm

This algorithm works similarly to the Partitioning Around Medoids (PAM) algorithm, where objects are clustered around centers or medoids. The PAM algorithm stores entire dissimilarity matrices in central memory, which greatly increases computation time. This is particularly costly with large datasets. The CLARA algorithm does not store entire dissimilarity matrices in central memory. Instead, it only clusters a sample of the large dataset using the PAM algorithm's methodology, and then assigns the remaining objects in the dataset to the clusters obtained from the sample [214, 227, 228]

6. Model-based Clustering Algorithm

This algorithm is a statistical model made up of a combination of Gaussian distributions that are used to fit the data, where each combination of Gaussian distributions represents a cluster [229–232]

B. Evaluation Metrics for Clustering Algorithms

This subsection outlines a set of metrics that were leveraged to evaluate the consistency of clustering results [233, 234] and assess the quality of clusters [203, 234, 235]. The following metrics were weighted equally for the purpose of this research and implemented with R's *clValid* library due to their application across multiple research domains [203]:

1. Connectivity

Connectivity measures "the extent to which items are placed in the same cluster as their nearest neighbors in the data space" [203, 234, 236]. Connectivity is defined as [234]:

$$Connectivity(C) = \sum_{i=1}^{N} \sum_{j=1}^{R} x_{i,n_{ij}} \tag{4.4}$$

Where N is the number of observations in a dataset, and $x_{i,n_{ij}}$ equals 0 if i and j belong to the same cluster, and equals 1/j if i and j do not belong to the same cluster. n_{ij} is the j_{th} nearest neighbor of observation i. C has k disjoint clusters, (C_1, ..., C_k), and R refers to the number of nearest neighbors that contribute to the connectivity measure.

2. Dunn Index

This measures the ratio between the smallest distance between items in different clusters and the largest distance between items in the same cluster [234, 236, 237]. The Dunn Index is defined as [234]:

$$Dunn\ Index = \frac{min(O)}{max(I)} \tag{4.5}$$

Where O is the distance between observations in different clusters and I is the distance between observations in the same cluster.

3. Silhouette

This measures the average distance between different clusters [234, 236, 238]. The Silhouette width of observation i is defined as [234]:

$$Silhouette\ Width = \frac{b(i)\ -\ a(i)}{max(a(i), b(i))} \tag{4.6}$$

Where a(i) is the average distance between i and all other observations in cluster A and b(i) is the average distance of i to the observations in the nearest neighbor cluster.

4. Average Proportion of Non-overlap (APN)

"This measures the ratio of items placed in different clusters by clustering using the entire dataset and clustering using the dataset with one excluded column" [203].

Let $C^{i,o}$ represent the cluster containing observation i using the original clustering (based on all available data), and $C^{i,l}$ represent the cluster containing observation i where the clustering is based on the dataset with column l removed. Then, APN is defined as [203]:

$$APN(L) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}(1 - \frac{n(C^{i,l} \cap C^{i,0})}{n(C^{i,0})}) \tag{4.7}$$

Where N is the total number of observations (rows) in a dataset and M is the total number of columns.

5. Average Distance (AD)

"This measures the average distance between items placed in the same cluster when the entire dataset is clustered, and when the dataset is clustered without one column" [203, 234]. AD is defined as:

$$AD(L) = \frac{1}{MN}\sum_{i=1}^{N}\sum_{l=1}^{M}\frac{1}{n(C^{i,l})n(C^{i,0})}[\sum_{i \in C^{i,0}, j \in C^{i,l}} dist(i,j)] \tag{4.8}$$

Where N is the total number of observations (rows) in a dataset and M is the total number of columns.

6. Average Distance between Means (ADM)

"This measures the average distance between cluster centers for items in the same cluster when the entire dataset is clustered, and when the dataset is clustered without one column" [203]. ADM is defined as:

$$ADM(L) = \frac{1}{MN} \sum_{i=1}^{N} \sum_{l=1}^{M} dist(\bar{x}_{C^{i,l}}, \bar{x}_{C^{i,0}}) \tag{4.9}$$

Where $\bar{x}_{C^{i,0}}$ is the mean of the observations in the cluster which contain observation i, when clustering is based on the full data, and $\bar{x}_{C^{i,l}}$ is similarly defined.

7. Figures of Merit (FOM)

"This measures the average intra-cluster variance of the deleted column, where the clustering is based on the remaining (undeleted) columns" [203, 234]. FOM is calculated using the following formula with deleted column l:

$$FOM(l, L) = \sqrt{\frac{1}{N} \sum_{k=1}^{K} \sum_{i \in C_k(l)} dist(x_{i,l}, \bar{x}_{C_k(l)})} \tag{4.10}$$

where $x_{i,l}$ is the value of the i^{th} observation in the l^{th} column, $\bar{x}_{C_k(l)}$ is the average of cluster $C_k(l)$.

Table 4.1 shows the criteria for good clustering of each evaluation metric.

Table 4.1: Criteria for clustering evaluation metrics

Evaluation metric	**Range of values**	**Criteria for good clustering**
Connectivity	$(0,\infty)$	Minimize
Silhouette	$(-1,1)$	Near 1
Dunn Index	$(0,\infty)$	Maximize
APN	$(0,1)$	Close to 0
AD	$(0,\infty)$	Minimize
ADM	$(0,\infty)$	Minimize
FOM	$(0,\infty)$	Minimize

4.1.9 Step #9: Identify Best Combination of Algorithm(s) and Number of Clusters

Algorithm 2 provides an overview of how the clustering algorithms were benchmarked to identify the best combination of algorithm(s) and number of clusters for each airport.

Algorithm 2: Algorithm for identifying best combination of clustering algorithm(s) and number of clusters

```
Let j be the number of clusters
for each airport do
    for each algorithm do
        vary j from 3 to 8
        for each j do
            evaluate algorithm performance using each metric
        end for
    end for
    return number of clusters identified as the best suited by each metric
end for
return combination of algorithm(s) and number of clusters identified as the best suited by
each metric
return combination of algorithm(s) and number of clusters identified as the best suited by a
majority of metrics
```

4.2 Implementation and Testing of Methodology for the Categorization of Daily Airport Operations

This section discusses the implementation and testing of Experiment 1 with normalized data from each of the 8 U.S. airports used for this work.

4.2.1 Boston Logan International Airport (BOS)

Application of Principal Component Analysis (BOS)

Figure 4.5 shows the scree plot for Boston Logan International Airport (BOS). In particular, it shows that 4 principal components captured about 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 4, and used for this work.

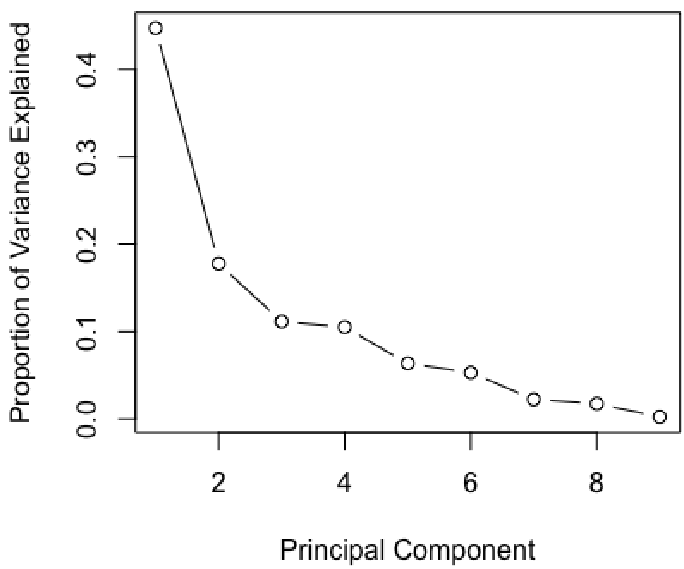

Figure 4.5: Scree plot for Boston Logan International Airport (BOS)

Assessment of the Clustering Tendency of the Dataset (BOS)

The Hopkins Statistic value of **0.957** obtained for this dataset indicates that the clustering tendency of the dataset is very high. The presence of three distinct boxes separated by white vertical and horizontal lines, and comprised of smaller boxes along the diagonal of the VAT plot in Figure 4.6 indicates that multiple clusters can be generated from the dataset. The size and dark shade of the largest blue box, as well as its very low dissimilarity value indicates that a majority of objects in the dataset are very similar. Finally, the high dissimilarity values and light blue shades of smaller boxes along the diagonal of the VAT plot, as seen in Figure 4.6, indicates the presence of outliers and/or very dissimilar objects in the dataset.

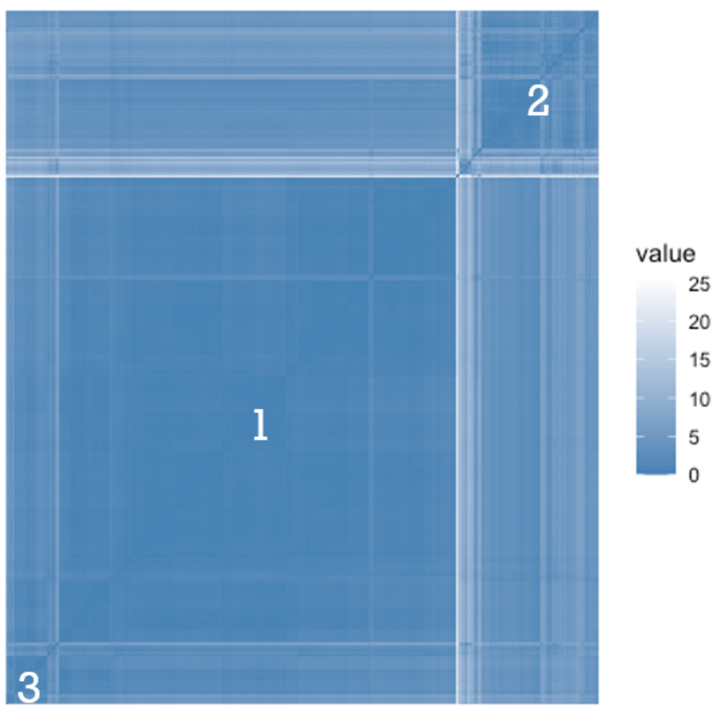

Figure 4.6: Visual Assessment of clustering Tendency (VAT) plot for Boston Logan International Airport (BOS)

Benchmarking and Evaluation of Clustering Algorithms (BOS)

Figures 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, and 4.13 show how each of the clustering algorithms performed while varying the number of clusters. In particular, the non-linear performance certain algorithms across clusters emphasizes the need to identify the optimal combination of clustering algorithm and number of clusters. They also show that the Hierarchical clustering algorithms on average, performed better than the other algorithms.

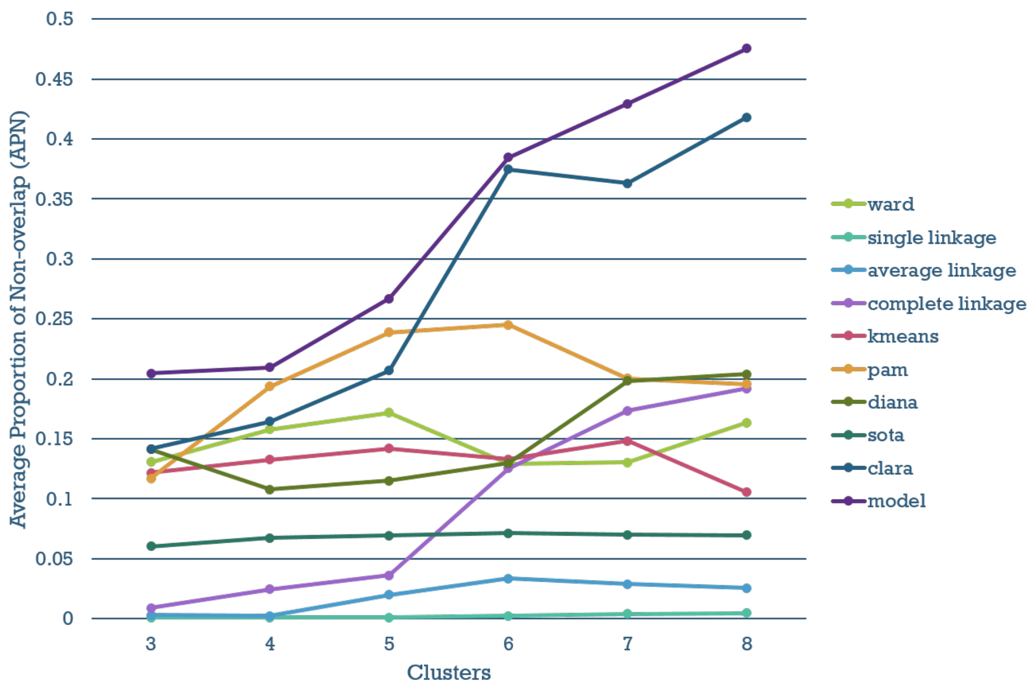

Figure 4.7: Average Proportion of Non-overlap (APN) for Boston Logan International Airport (BOS)

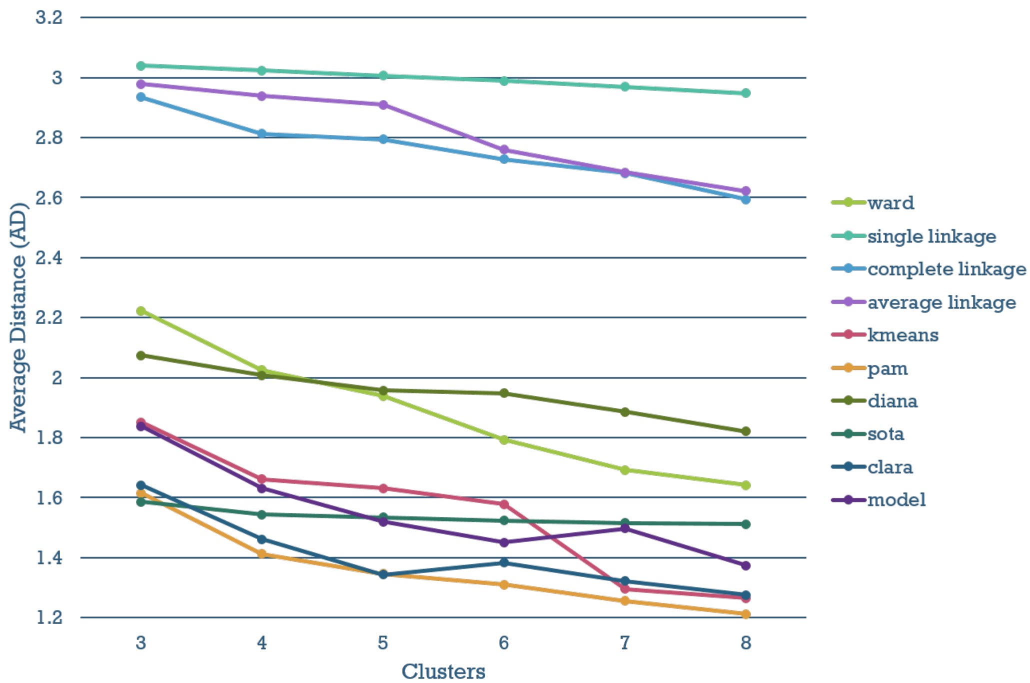

Figure 4.8: Average Distance (AD) for Boston Logan International Airport (BOS)

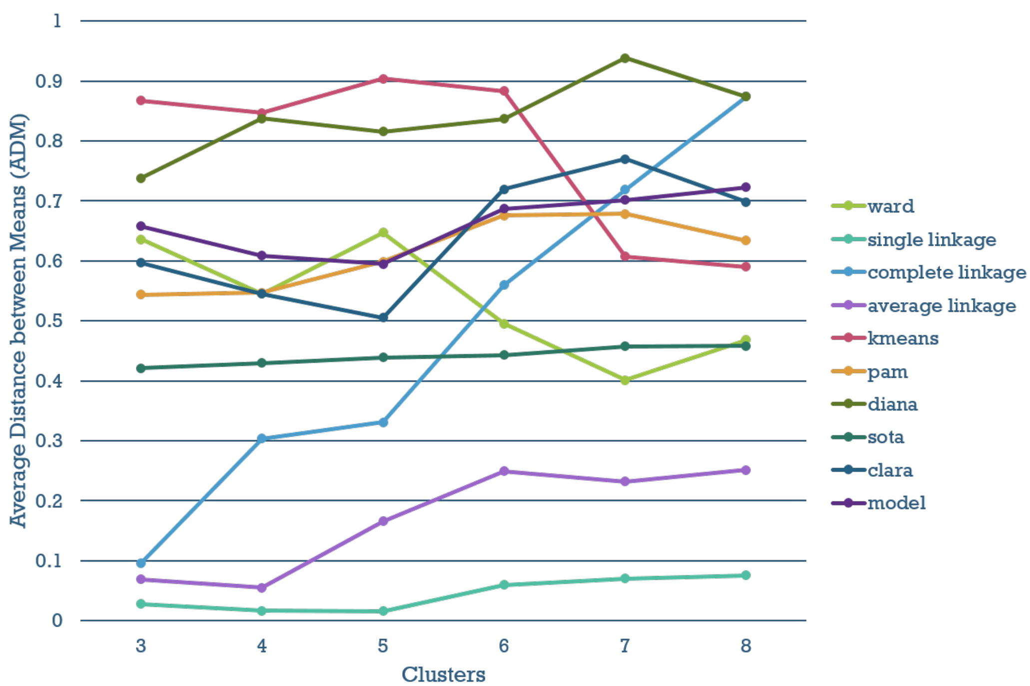

Figure 4.9: Average Distance between Means (ADM) for Boston Logan International Airport (BOS)

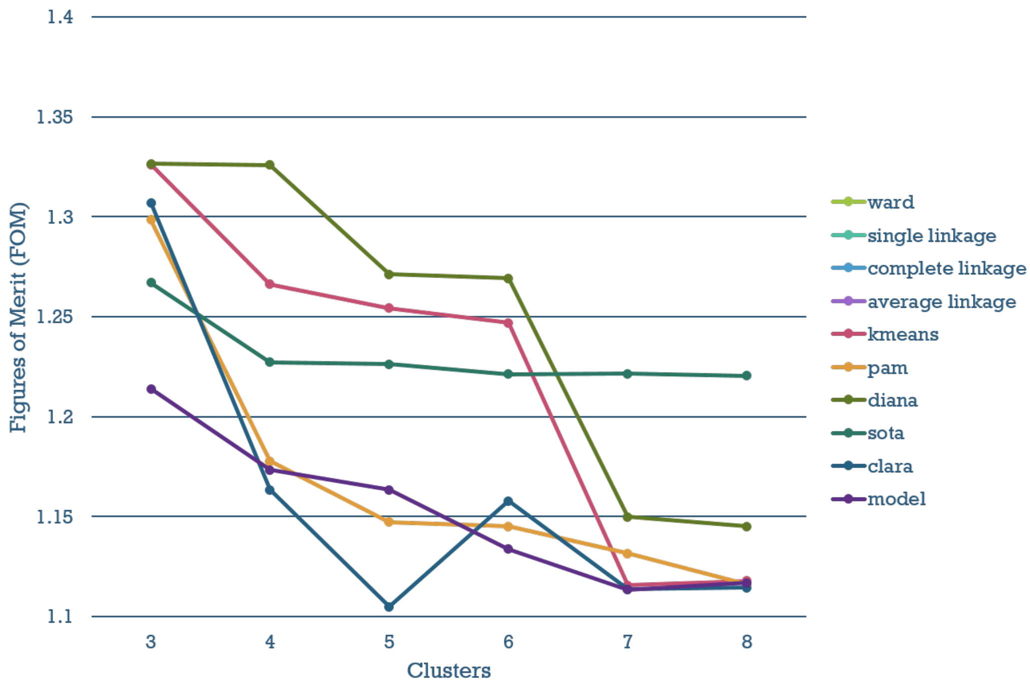

Figure 4.10: Figures of Merit (FOM) for Boston Logan International Airport (BOS)

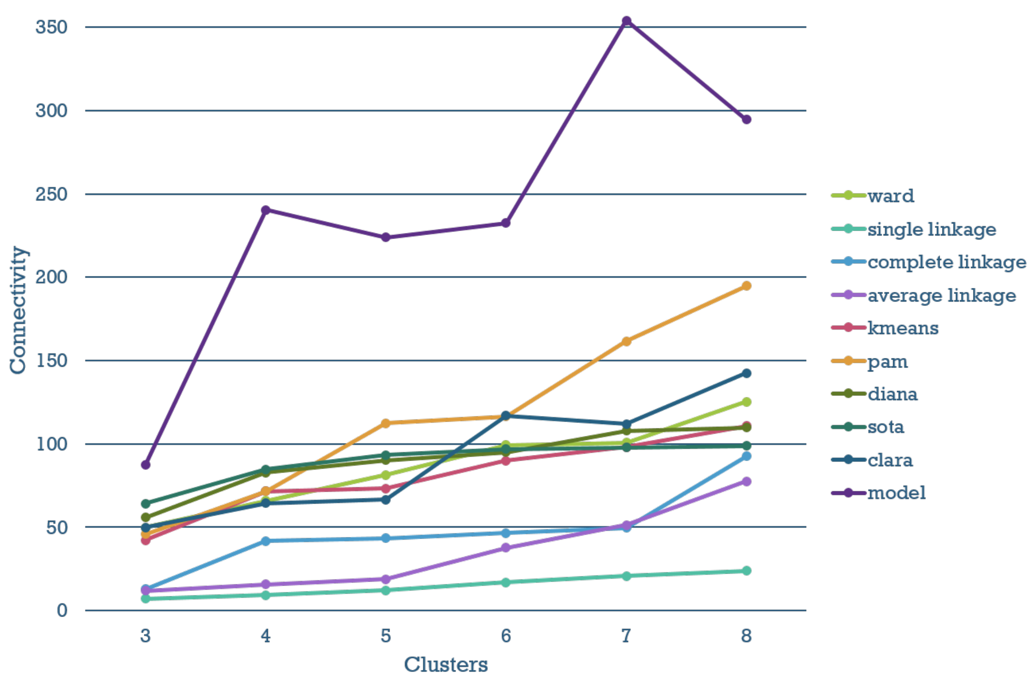

Figure 4.11: Connectivity for Boston Logan International Airport (BOS)

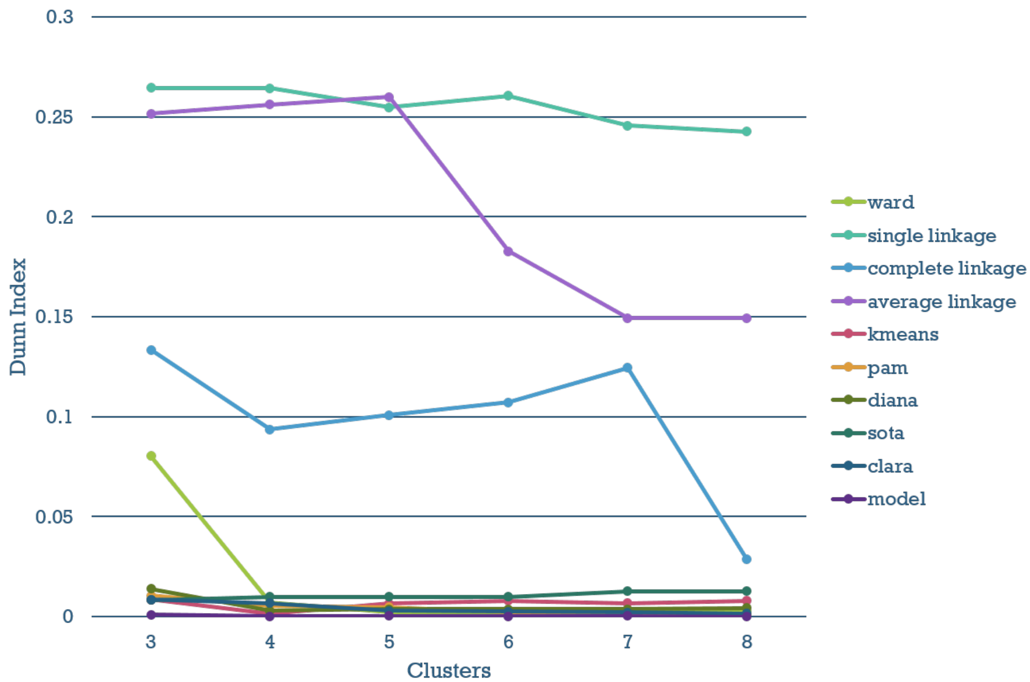

Figure 4.12: Dunn Index for Boston Logan International Airport (BOS)

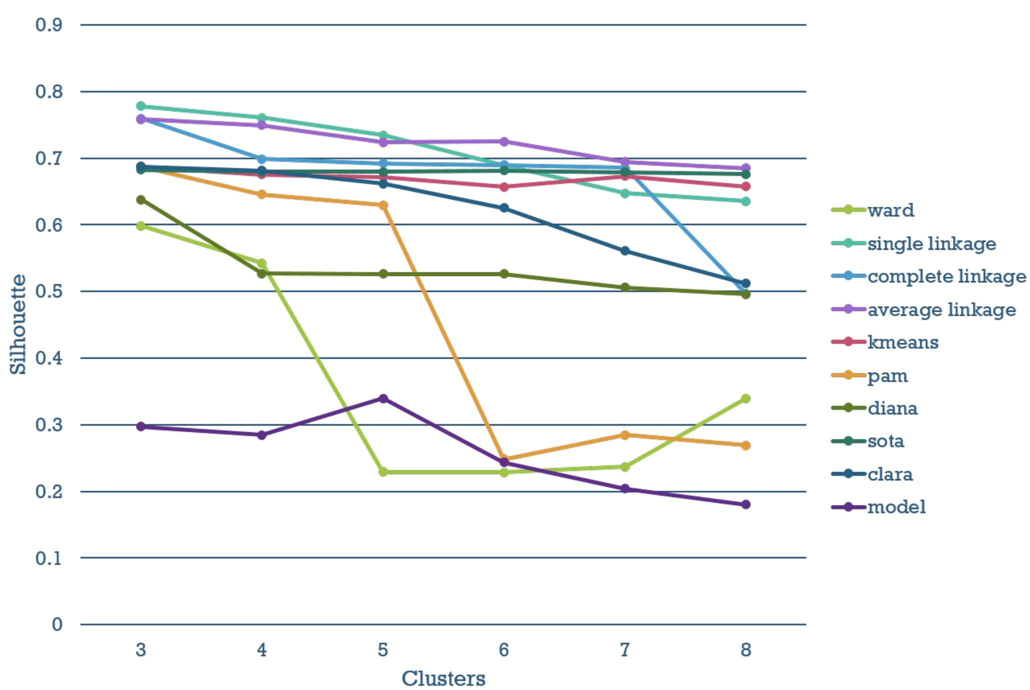

Figure 4.13: Silhouette for Boston Logan International Airport (BOS)

Table 4.2 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for BOS. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** was identified as the best suited combination for BOS by a majority of metrics.

Table 4.2: Optimal combination of algorithm and number of clusters per evaluation metric for BOS

Evaluation Metric	**Value**	**Clustering Algorithm**	**Clusters**
APN	0.0008	Single Linkage	4
AD	1.2117	PAM	8
ADM	0.0158	Single Linkage	4
FOM	0.7205	Ward	8
Connectivity	***7.1619***	***Single Linkage***	***3***
Dunn Index	***0.2646***	***Single Linkage***	***3***
Silhouette	***0.7779***	***Single Linkage***	***3***

Table 4.3 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. The identification of 3 as the optimal number of clusters and the breakdown of daily airport operations in Table 4.3 is validated by observations made from the assessment of clustering tendency of the dataset with the VAT plot. Indeed, Figure 4.6 shows one large dark blue box which corresponds to cluster 1, a small blue box at the top right end of the diagonal corresponding to cluster 2, and a smaller box at the bottom left end of the diagonal which corresponds to cluster 3.

Table 4.3: Breakdown of daily operations of BOS by cluster

Cluster	**Number of Daily Operations**
1	1018
2	315
3	7

Analysis of Clusters (BOS)

Figure 4.14 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by low to high number of departure delays, airborne holdings (minutes and aircraft), and diversions. It is also characterized by low to moderate TMI to airport delays, and high completion rates. Furthermore, a majority of daily operations in the first cluster are not characterized by Ground Stops and Ground Delay Programs, as seen in Figure 4.14 where the majority of daily airport operations have Ground Stops, and Ground Delay Program lead-in time and revision values of 0, 500, and -1, respectively. Ground Stops and Ground Delay Programs significantly impact airport operations. As such, a lack of their implementation bodes well for airport operations. Table 4.4 shows that the first cluster has high mean and median completion rates, and low mean and median airborne holdings (aircraft and minutes), diversions, GDP revisions, Ground Stops, and TMI to airport delays. Table 4.4 also shows that the mode of airborne holdings (aircraft and minutes), diversions, departure delays, GDP revisions, Ground Stops, and TMI to airport delays of this cluster is zero, while that of completion rate is 100%. As such, overall, the characteristics of daily airport operations in the first cluster correspond to good operational performance.

Figure 4.14 shows that the second cluster is characterized by a wider range of TMI to airport delays, airborne holdings (aircraft and minutes), diversions, departure delays, Ground Stops, and GDP revisions compared to the first cluster. It is also characterized by moderate to high completion rates. Unlike the daily operations in the first cluster, a majority of operations in the second cluster are characterized by a wide range of GDP lead-in times. Table 4.4 shows that this cluster has the highest mean airborne holdings (minutes and aircraft), diversions, departure delays, and TMI to airport delays out of the three clusters which indicates poor operational performance. However, this cluster's mean and median values for completion rates and GDP lead-in times are high which indicates good operational performance. Overall, these characteristics correspond to varying operational performance where one or more metrics cause sub-optimal to poor operational performance on a specific day. This observation is validated by the high dissimilarity value of objects in the blue box at the top right end of the diagonal in Figure 4.6, which was previously identified to correspond to the second cluster. The high dissimilarity value indicates that daily operations in this cluster are not necessarily similar to each other, as evidenced by the wide ranges of metrics.

Table 4.4: Mean, Median, and Mode of Airport Metrics across clusters for Boston Logan International Airport (BOS)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ Holdings\ (aircraft)$	**0.6**	5.5	1.1	**0**	1	**0**	0	0	0
$Airborne\ Holdings\ (minutes)$	**11.6**	123.4	25.7	**0**	15	0	0	0	0
$Completion\ rate\ (\%)$	**99.1**	96.1	27.6	**99.4**	98.3	20	**100**	**100**	M[1]
$Diversions$	**0.8**	2.4	2.1	1	1	1	**0**	M[1]	M[1]
$Departure\ delays$	8	9	**0**	0	0	0	0	0	0
$GDP\ lead-in\ time\ (minutes)$	13.5	64.2	**96**	-2.5	60	**96**	-3	0	**96**
$GDP\ revisions$	0.2	0.6	**0**	0	0	0	0	0	0
$Ground\ Stops$	0.1	0.6	**0**	0	0	0	0	0	0
$TMI\ to\ airport\ delays$	0.7	126.5	**0.6**	**0**	120	**0**	0	0	0

The third cluster comprised of seven daily airport operations is characterized by low departure delays, TMI to airport delays, Ground Stops, airborne holdings (aircraft and minutes), and diversions. Similar to cluster 1, a majority of daily operations in this cluster are not characterized by Ground Delay Programs. As such, all but one of daily operations in the cluster have GDP lead-in time and revision values of 500 and -1, respectively. Even though these observations highlight good operational performance, the low average and median completion rates, as seen in Table 4.4, as well as their poor distribution presented in 4.14 indicates that a majority of scheduled air carrier arrivals did not arrive as planned on days assigned to this cluster. As such, the daily operations in this cluster can be observed to have poor operational performance.

[1]Multiple values exist

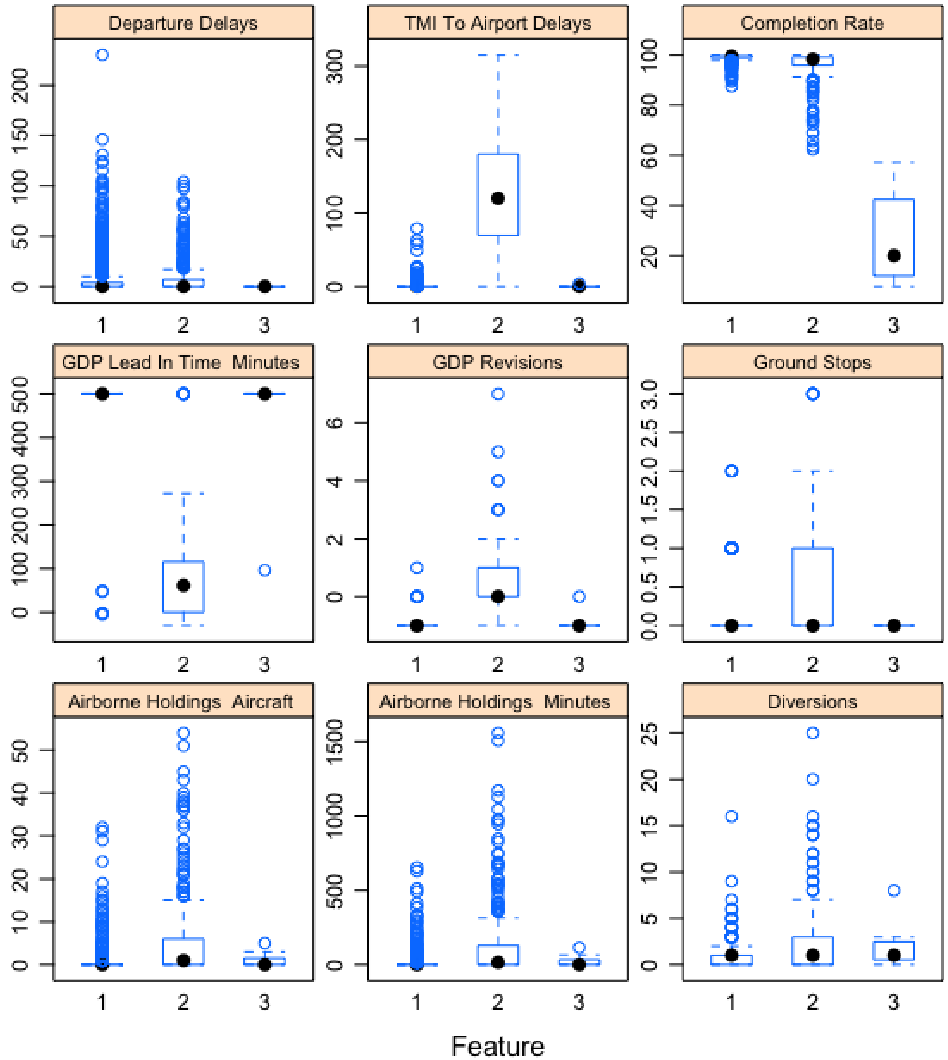

Figure 4.14: Box plots showing the distribution of airport metrics across clusters for Boston Logan International Airport (BOS)

Comparison of results from clustering and using predefined ranges of metrics (BOS)

Experiment 1 validates the categorization of daily BOS operations into 3 categories, as is currently done with OSPC. Indeed, the analysis of the clusters revealed that each cluster can be characterized as having good, varying, or poor operational performance, which is similar to OSPC's approach. As such, a one-to-one comparison of the categorization of daily operations of BOS with clustering and OSPC was conducted and reviewed by FAA Subject Matter Experts so as to determine the better approach for categorizing daily operations of BOS. Table 4.5 shows that 1017, 296, and 7 days categorized as "Good days" by OSPC were placed in the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 1 day categorized as an "Average day" by OSPC was placed in the cluster characterized by good operational performance, while 9 days categorized as "Average days" and 10 days characterized as "Bad days" by OSPC were determined to have varying operational performance by the clustering algorithm.

Table 4.5: Composition of categories from clustering and Operational Service Performance Criteria (BOS)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	1017	296	7
Average day	1	9	0
Bad day	0	10	0

Table C.1 in Appendix C provides a comparison of a subset of daily BOS operations that were categorized differently by both approaches. In particular, it shows that days 1, 2, 7, 8, 9, 10, 11, 12, 13, and 14 were categorized as "Good days" by OSPC because a majority of the metrics reflected good operational performance. However, these daily operations were characterized by extremely poor completion rates, which indicates that a majority of planned and/or scheduled flights did not arrive as scheduled due to severe constraints at BOS. Ignoring the poor completion rate and designating these daily operations as "Good days" may prevent stakeholders from conducting an in-depth assessment of those operations in order to identify the underlying causes of the poor completion rates.

Days 5 and 6 were characterized by very poor Ground Delay Program lead-in times, which indicates that the Traffic management Initiatives were implemented prior to their announcement. This hinders the efficient operation of airports as various stakeholders such as airlines and passengers in particular, are affected by this phenomenon. The clustering algorithm assigned both days to the second cluster indicating varying operational performance unlike OSPC, which classified them as “Good days”.

Even though a majority of daily operations were classified similarly by both approaches, it can be seen that using predefined ranges of values to classify the daily operations of BOS is not the best suited approach for the task at hand as analysts would continually have to update these ranges based on prior knowledge and experience, instead of using a systematic approach such as clustering. As such, it can be concluded that **clustering is a better approach for categorizing daily operations of Boston Logan International Airport (BOS)**.

4.2.2 Baltimore/Washington International Thurgood Marshall Airport (BWI)

Application of Principal Component Analysis

Figure 4.15 shows the scree plot for Baltimore/Washington International Thurgood Marshall Airport (BWI). In particular, it shows that 3 principal components captured about 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3.

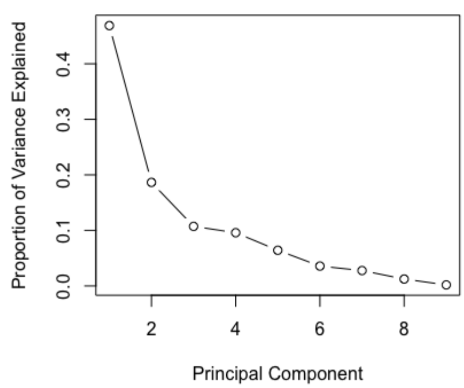

Figure 4.15: Scree plot for Baltimore/Washington International Thurgood Marshall Airport (BWI)

Assessment of the Clustering Tendency of the Dataset (BWI)

The clustering tendency of the dataset was observed to be very high based on its Hopkins Statistic score of **0.973** and the presence of a very large dark blue box as well as smaller blue boxes along the diagonal of the VAT plot in Figure 4.16. The size of the large dark blue box and its low dissimilarity value indicates that the majority of objects in the dataset are very similar. The smaller boxes along the diagonal indicate the presence of a few outliers and/or dissimilar objects, which further highlights the clustering tendency of the dataset.

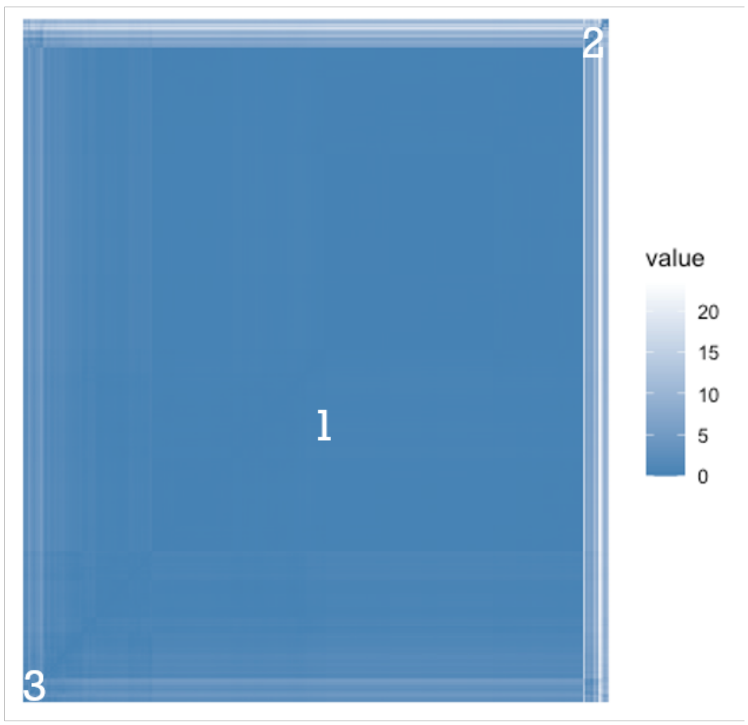

Figure 4.16: Visual Assessment of clustering Tendency (VAT) plot for Baltimore/Washington International Thurgood Marshall Airport (BWI)

Benchmarking and Evaluation of Clustering Algorithms (BWI)

Figures B.1, B.2, B.3, B.4, B.5, B.6, and B.7 in Appendix B show how each clustering algorithm performed with varying numbers of clusters. Similar to BOS, the non-linear performance of certain algorithms across clusters emphasizes the need to identify the optimal combination of clustering algorithm and number of clusters. They also show that the Hierarchical clustering algorithms on average, performed better than the other algorithms. Table 4.6 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for BWI. In particular, it shows that the most common combination was identified to be the **Single Linkage Hierarchical clustering algorithm with 3 clusters**.

Table 4.6: Optimal combination of algorithm and number of clusters per evaluation metric for BWI

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	0.0001	Single Linkage	4
AD	1.288	PAM	8
ADM	0.0626	Single Linkage	6
FOM	0.6686	Kmeans	8
Connectivity	***4.0829***	***Single Linkage***	***3***
Dunn Index	***0.3189***	***Single Linkage***	***3***
Silhouette	0.857	Complete Linkage	4

Table 4.7 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Similarly to BOS, the breakdown of daily operations of BWI in Table 4.7 validates observations made from the assessment of clustering tendency of the dataset with the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.16 shows one large dark blue box which corresponds to cluster 1, and smaller boxes along the diagonal which correspond to clusters 2 and 3.

Table 4.7: Breakdown of daily operations of BWI by cluster

Cluster	Number of Daily Operations
1	1263
2	45
3	32

Analysis of Clusters (BWI)

Figure 4.17 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by a wide range of departure delays, and low to moderate TMI to airport delays, airborne holdings (minutes and aircraft), and diversions. The first cluster is also characterized by high completion rates. Furthermore, none of the daily operations in the first cluster are characterized by the implementation of Ground Delay Programs, as seen in Figure 4.17, where all of the daily airport operations had GDP lead-in times and number of revision values of 500 and -1, respectively. In addition, a majority of daily operations in the first cluster are also not characterized by the implementation of Ground Stops. Table 4.8 shows that this cluster has the lowest mean and median airborne holdings (aircraft and minutes), diversions, departure delays, Ground Stops, and TMI to airport delays. Similar to BOS, these characteristics correspond to good operational performance.

Table 4.8: Mean, Median, and Mode of Airport Metrics across clusters for Baltimore/Washington International Thurgood Marshall Airport (BWI)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ Holdings\ (aircraft)$	**0.3**	14.6	6.3	**0**	16	1	0	0	0
$Airborne\ Holdings\ (minutes)$	**5.8**	407.2	181.5	**0**	415	18	0	0	0
$Completion\ rate\ (\%)$	**98.6**	92.9	83.3	**99.4**	95.3	86.8	**100**	M[1]	M[1]
$Diversions$	**0.5**	9.1	4.9	**0**	7	2	**0**	**0**	1
$Departure\ delays$	**3.5**	16.3	11.6	**1**	16	11	0	0	0
$GDP\ lead-in\ time\ (minutes)$	-	-	75	-	-	72.5	-	-	0
$GDP\ revisions$	-	-	0.5	-	-	0	-	-	0
$Ground\ Stops$	**0.1**	0.9	0.7	**0**	1	0.5	**0**	1	**0**
$TMI\ to\ airport\ delays$	**0.4**	10.6	25.9	**0**	7	8.5	0	0	0

[1] Multiple values exist

The second cluster is characterized by low to moderate TMI to airport delays, departure delays and Ground Stops, as seen in Figure 4.17. It is also characterized by a wide range of airborne holdings (minutes and aircraft) and high completion rates. As with the first cluster, daily airport operations in the second cluster are not characterized by the implementation of Ground Delay Programs which bodes well for airport operations. Even though this cluster is characterized by the highest mean and median airborne holdings (minutes and aircraft), diversions, departure delays, and Ground Stops, the mode of airborne holdings (aircraft and minutes), diversions, departure delays, and TMI to airport delays is zero. This indicates that the cluster is characterized by varying operational performance where one or more metrics cause sub-optimal to poor operational performance on a specific day.

The third cluster is characterized by low to moderate departure delays, and a wide range of TMI to airport delays, completion rates, airborne holdings (minutes and aircraft), and diversions. In addition, unlike the other clusters, which are not characterized by Ground Delay Programs, Ground Delay Programs were implemented on a majority of days placed in this cluster. These observations correspond to poor operational performance.

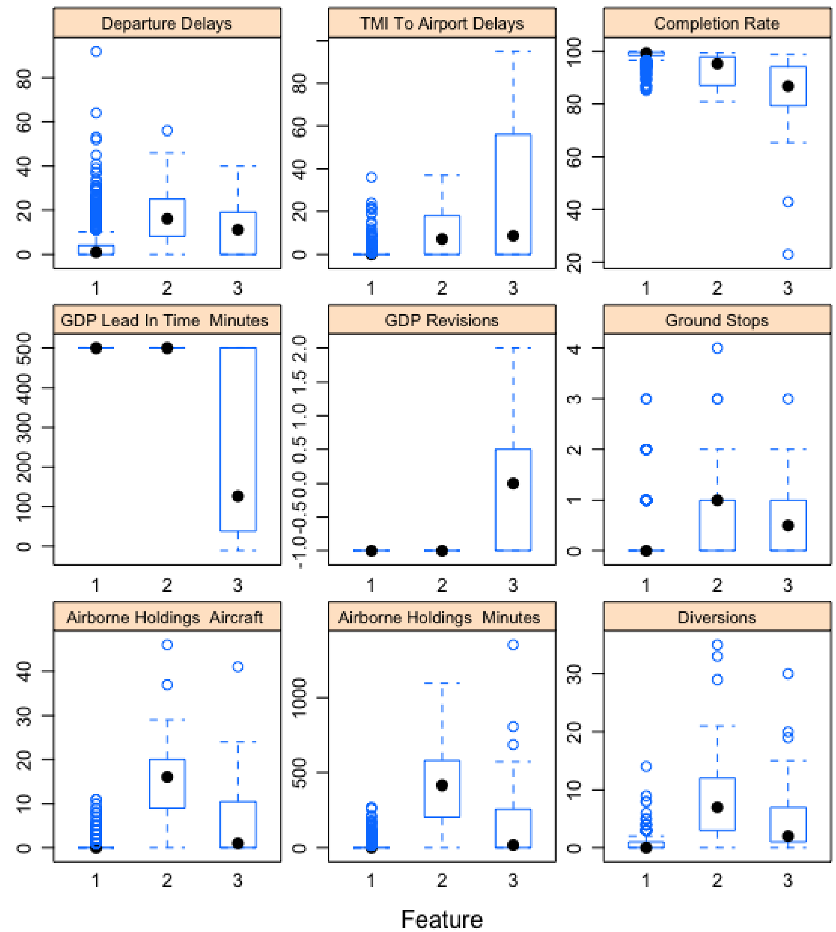

Figure 4.17: Box plots showing the distribution of airport metrics across clusters for Baltimore/Washington International Thurgood Marshall Airport (BWI)

Comparison of results from clustering and using predefined ranges of metrics (BWI)

As with BOS, Experiment 1 validates the categorization of daily BWI operations into 3 categories, as is currently done with OSPC. Similarly, each cluster is observed to either have good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of daily BWI operations with clustering and OSPC was conducted and reviewed by FAA Subject Matter Experts so as to determine which approach better categorizes the daily operations of BWI.

Table 4.9 shows that 1263, 44, and 31 days categorized as “Good days” by OSPC were placed in the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 1 day categorized as an "Average day" by OSPC was placed in the cluster characterized by varying operational performance, while another day categorized as a "Bad day" by OSPC was identified to have poor operational performance.

Table 4.9: Composition of categories from clustering and Operational Service Performance Criteria (BWI)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	1263	44	31
Average day	0	1	0
Bad day	0	0	1

Table C.2 in Appendix C provides a comparison of a subset of daily operations of BWI that were categorized differently by the clustering algorithm and OSPC. In particular, it shows that days 6 through 10 were categorized as "Good days" by OSPC because a majority of the metrics reflected good operational performance. However, similar to BOS, these daily operations had poor completion rates which are ignored by OSPC. Days 1 through 3 were also categorized as "Good days" by OSPC even though they were characterized by very high airborne holdings (minutes and aircraft) and diversions. Days 11 through 16 were also classified as "Good days" by OSPC, even though they were characterized by poor Ground Delay Program lead-in times and/or airborne holdings (minutes and aircraft). Each of these daily airport operations, as well as several others, were assigned to the second or third clusters by the clustering algorithm due to their sub-optimal and poor operational characteristics, respectively. OSPC also classified all but two daily operations as "Good days". As observed, this is clearly not the case as several daily operations were characterized by sub-optimal to poor operational performances. Consequently, similar to BOS, it can be observed that **clustering is a better approach for categorizing daily operations of Baltimore/Washington International Thurgood Marshall Airport (BWI)**.

4.2.3 Reagan National Airport (DCA)

Application of Principal Component Analysis (DCA)

Figure 4.18 shows the scree plot for Reagan National Airport (DCA). In particular, it shows that 3 principal components captured about 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3.

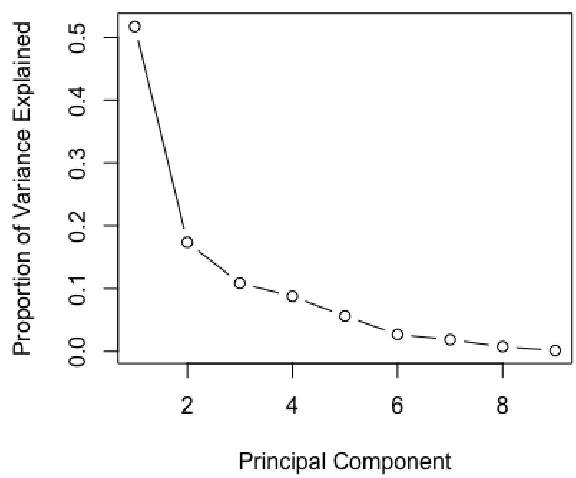

Figure 4.18: Scree plot for Reagan National Airport (DCA)

Assessment of the Clustering Tendency of the Dataset (DCA)

The Hopkins Statistic value of **0.957** obtained for this dataset and the presence of a large dark blue box comprised of several smaller boxes along the diagonal in Figure 4.19 indicates that the clustering tendency of the dataset is high. The smaller boxes located in the top right corner of the diagonal indicate the presence of outliers and/or dissimilar objects which further highlights the clustering tendency of the dataset.

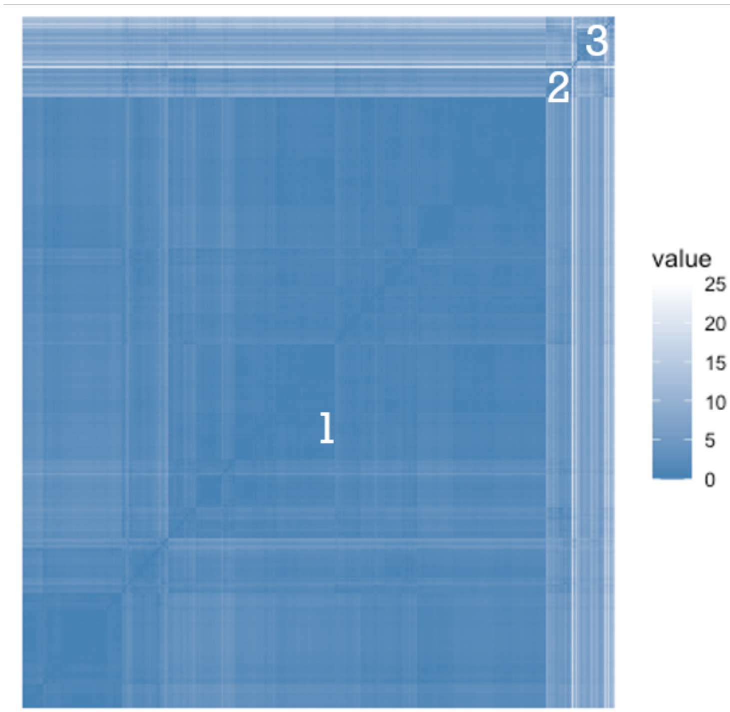

Figure 4.19: Visual Assessment of clustering Tendency (VAT) plot for Reagan National Airport (DCA)

Benchmarking and Evaluation of Clustering Algorithms (DCA)

Figures B.8, B.9, B.10, B.11, B.12, B.13, and B.14 in Appendix B show how each clustering algorithm performed while varying the number of clusters. In particular, they show that the Hierarchical clustering algorithms on average, performed better than the other algorithms. Table 4.10 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for DCA. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** is the best suited combination for DCA.

Table 4.10: Optimal combination of algorithm and number of clusters per evaluation metric for DCA

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	***0.0008***	***Single Linkage***	***3***
AD	1.5316	PAM	8
ADM	***0.0159***	***Single Linkage***	***3***
FOM	0.6064	SOTA	8
Connectivity	***6.2079***	***Single Linkage***	***3***
Dunn Index	***0.2906***	***Single Linkage***	***3***
Silhouette	0.7855	Average Linkage	3

Table 4.11 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Similarly to BWI, the breakdown of daily operations of DCA in Table 4.11 validates observations made from the assessment of clustering tendency of the dataset with the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.19 shows a large dark blue box which corresponds to cluster 1, and smaller boxes along the diagonals which correspond to clusters 2 and 3.

Table 4.11: Breakdown of daily operations of DCA by cluster

Cluster	Number of Daily Operations
1	1175
2	70
3	95

Analysis of Clusters (DCA)

Figure 4.20 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by high completion rates, a wide range of departure delays, and low TMI to airport delays, airborne holdings (minutes and aircraft), Ground Stops and diversions. Similar to BWI, none of the daily operations in the first cluster are characterized by the implementation of Ground Delay Programs, as seen in Figure 4.20 where all of the daily airport operations have GDP lead-in time and revision values of 500 and -1, respectively. Overall, this cluster is characterized by good operational performance as evidenced by the high mean, median, and modal value of completion rate, as well as the low mean, median, and modal values of the airborne holdings (aircraft and minutes), diversions, Ground Stops and TMI to airport delays of this cluster, as shown in Table 4.12.

The second cluster is characterized by low to moderate departure delays, low TMI to airport delays, and low to high airborne holdings (aircraft and minutes) and diversions, as seen in Figure 4.20. As with the first cluster, daily airport operations in the second cluster are not characterized by the implementation of Ground Delay Programs which bodes well for airport operations. However, this cluster is characterized by low to high number of Ground Stops which negatively impact airport operations. This cluster is also characterized by higher mean and median airborne holdings (minutes and aircraft), and departure delays which corresponds to sub-optimal to poor operational performance, compared to the other clusters. However, the absence of Ground Delay Programs, and high completion rates indicates good operational performance. As such, overall, this cluster is characterized by varying operational performance.

Table 4.12: Mean, Median, and Mode of Airport Metrics across clusters for Reagan National Airport (DCA)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
Airborne holdings (aircraft)	**1.6**	20.3	16.5	**0**	21	10	0	0	0
Airborne holdings (minutes)	**30.9**	508.4	438.1	**0**	555	222	0	0	0
Completion rate (%)	**98.4**	92.7	89.3	**99**	93.7	93.8	**100**	M[1]	97.9
Diversions	**0.5**	7.2	7.9	**0**	6	3	0	0	0
Departure delays	**10.2**	21.3	20.4	**7**	20	19	**0**	M[1]	**0**
GDP lead − in time (minutes)	-	-	20.5	0	0	-1	-	-	0
GDP revisions	-	-	0.3	0	0	0	-	-	0
Ground Stops	**0.1**	1.2	1.2	**0**	1	1	**0**	1	1
TMI to airport delays	**0.7**	11.8	81.7	**0**	10.5	78	0	0	0

The third cluster is characterized by a wide range of departure delays, TMI to airport delays, completion rates, airborne holdings (minutes and number of aircraft), and diversions. This cluster also has the highest mean TMI to airport delays and diversions, and the lowest mean completion rate compared to the other clusters. Table 4.12 and Figure 4.20 show that all of the daily operations characterized by the implementation of Ground Delay Programs were assigned to this cluster. In addition, the modal Ground Delay Program lead-in time of this cluster was zero which indicates that there was no time between the announcement and implementation of Ground Delay Program on a majority of days which does not bode well for operational performance. Overall, these characteristics correspond to poor operational performance.

[1]Multiple values exist

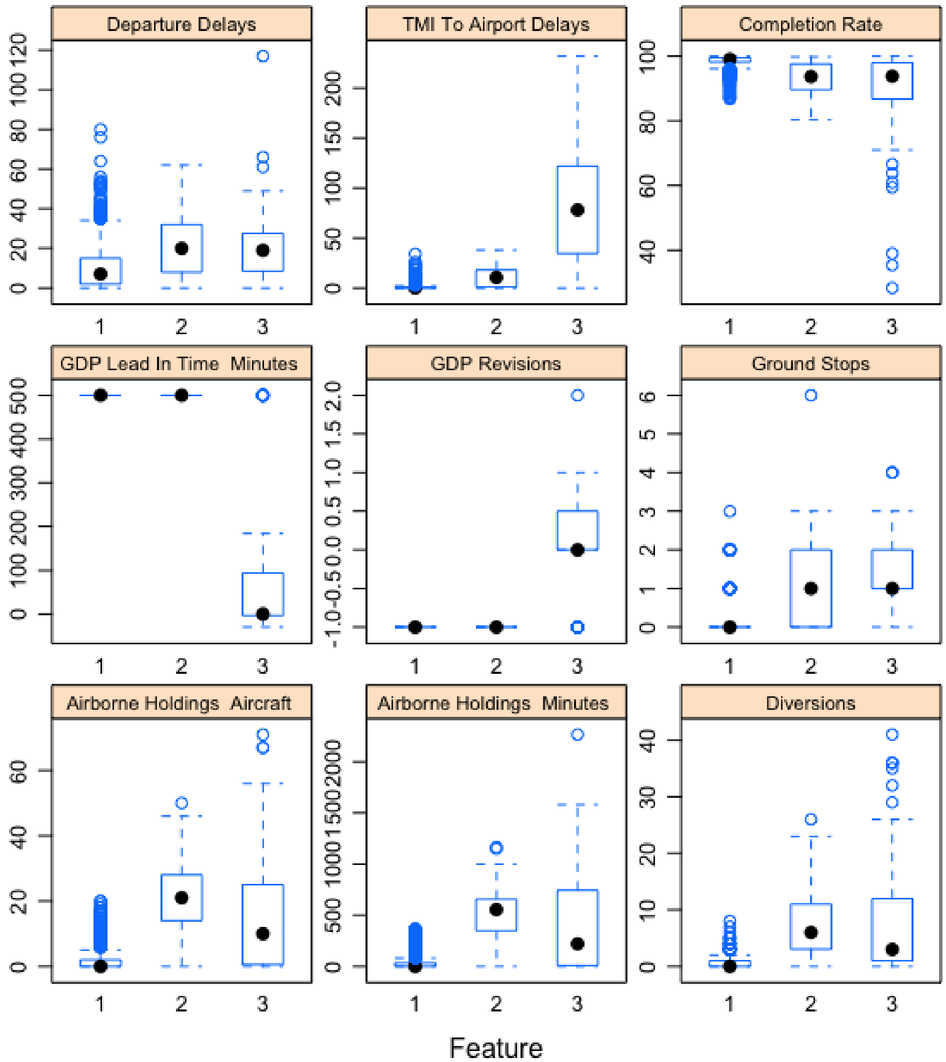

Figure 4.20: Box plots showing the distribution of airport metrics across clusters for Reagan National Airport (DCA)

Comparison of results from clustering and using predefined ranges of metrics (DCA)

As with BOS and BWI, Experiment 1 validates the categorization of the daily operations of DCA into 3 categories, as is currently done with the OSPC. Similarly, each cluster is characterized by either good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of daily operations of DCA with clustering and predefined ranges of metrics was conducted and reviewed by FAA Subject Matter Experts so as to determine which approach better categorizes daily DCA operations.

Table 4.13 shows that 1175, 59, and 77 days categorized as “Good days” by OSPC were assigned to the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 1 day categorized as an “Average day” by OSPC was placed in the cluster characterized by varying operational performance, while 4 days categorized as “Average days” by OSPC were placed in the third cluster. Table 4.13 also shows that 14 days categorized as “Bad days” by OSPC were assigned to the third cluster which is characterized by poor operational performance.

Table 4.13: Composition of categories from clustering and Operational Service Performance Criteria (DCA)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	1175	59	77
Average day	0	1	4
Bad day	0	0	14

Table C.3 in Appendix C provides a comparison of a subset of daily operations of DCA that were categorized differently by the clustering algorithm and OSPC. In particular, it shows that days 1 and 2 which were characterized by high TMI to airport delays and departure delays were classified as having poor operational performances unlike OSPC which classified them as “Good days”. It also shows that even though days 3, 4, and 5 had sub-optimal operational performance due to a high number of diversions and airborne holdings (minutes and aircraft), they were classified as “Good days” by OSPC. Table C.3 also shows that days 6 and 7 which were characterized by poor Ground Delay Program lead-in times and high duration of airborne holdings were classified differently by both approaches. Similar to BOS and BWI, OSPC also classified daily operations with very low completion rates as “Good days”. These observations clearly highlight the limitations of OSPC and how **clustering is better suited for categorizing daily operations of Reagan**

National Airport (DCA).

4.2.4 Newark Liberty International Airport (EWR)

Application of Principal Component Analysis (EWR)

Figure 4.21 shows the scree plot for Newark Liberty International Airport (EWR). In particular, it shows that 3 principal components captured about 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3.

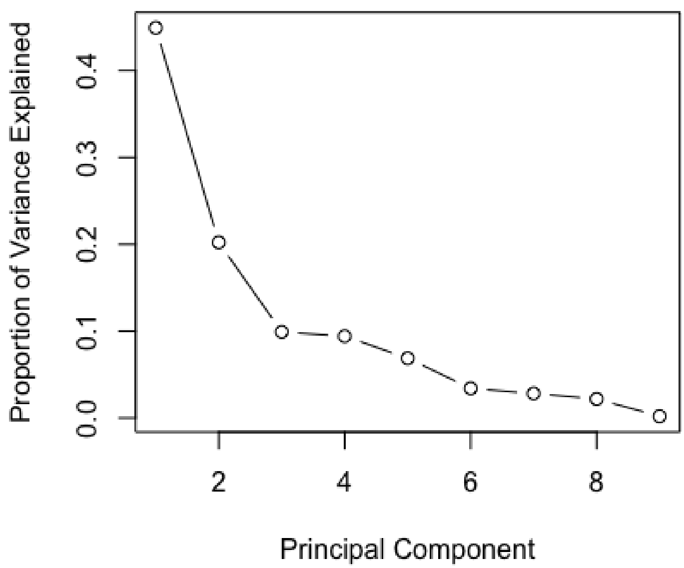

Figure 4.21: Scree plot for Newark Liberty International Airport (EWR)

Assessment of the Clustering Tendency of the Dataset (EWR)

The Hopkins Statistic value of **0.956** for this dataset and the presence of three blue boxes separated by light blue vertical and horizontal lines, and comprised of several smaller boxes along the diagonal in Figure 4.22 indicates that the clustering tendency of the dataset is high. The presence of a small blue box at the top of the diagonal shows that there are objects in the dataset that are different from the others. The dark shade of blue of the three distinct boxes also indicates that objects in each cluster (box) are fairly similar to each other.

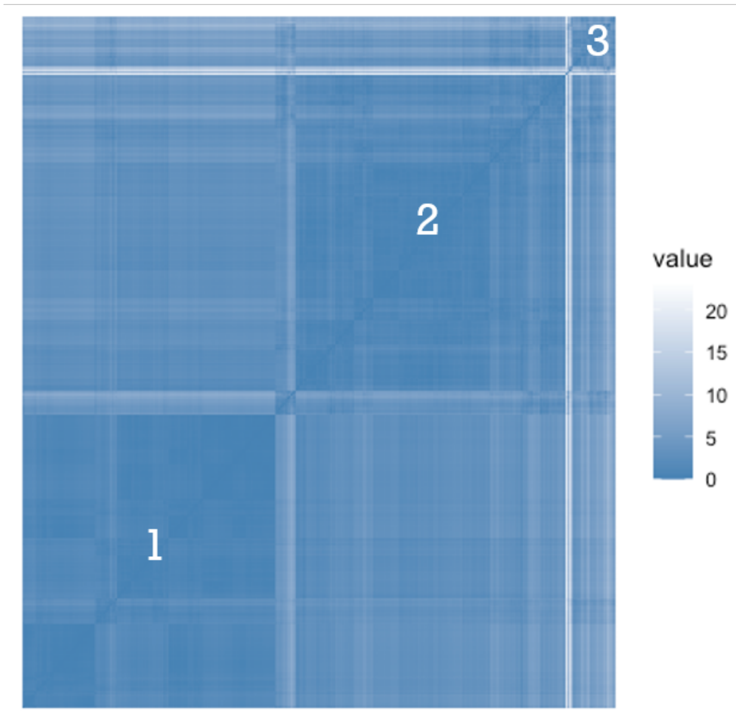

Figure 4.22: Visual Assessment of clustering Tendency (VAT) plot for Newark Liberty International Airport (EWR)

Benchmarking and Evaluation of Clustering Algorithms (EWR)

Figures B.15, B.16, B.17, B.18, B.19, B.20, and B.21 in Appendix B show how each clustering algorithm performed while varying the number of clusters. Table 4.14 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for EWR. In particular, it shows that **Single Linkage Hierarchical clustering algorithm and 3 clusters** is the best suited combination for the EWR dataset.

Table 4.14: Optimal combination of algorithm and number of clusters per evaluation metric for EWR

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	***0.0007***	***Single Linkage***	***3***
AD	1.9049	PAM	8
ADM	***0.0265***	***Single Linkage***	***3***
FOM	0.6821	Kmeans	8
Connectivity	***8.1028***	***Single Linkage***	***3***
Dunn Index	0.3137	Single Linkage	7
Silhouette	0.7241	Average Linkage	3

Table 4.15 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Table 4.15 validates observations made from the assessment of clustering tendency of the dataset with the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.22 shows two large blue boxes and a smaller blue box along the diagonal which correspond to the first two clusters, and the third cluster, respectively.

Table 4.15: Breakdown of daily operations of EWR by cluster

Cluster	Number of Daily Operations
1	570
2	657
3	113

Analysis of Clusters (EWR)

Figure 4.23 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by high completion rates, low to high departure delays, TMI to airport delays, airborne holdings (minutes and aircraft), Ground Stops and diversions. None of the daily operations in the first cluster are characterized by Ground Delay Programs, and a majority of daily operations in the first cluster are not characterized by Ground Stops. Table 4.16 shows that this cluster has the lowest mean and median airborne holdings (aircraft and minutes), diversions, departure delays, Ground Stops, and TMI to airport delays. Overall, these characteristics correspond to good operational performance.

Figure 4.23 shows that the distribution of departure delays, airborne holdings (aircraft and minutes), and diversions in the first and second cluster are similar. However, the second cluster is characterized by a wider range of TMI to airport delays, Ground Stops, and completion rates. The second cluster is also characterized by the implementation of Ground Delay Programs. Overall, these characteristics correspond to varying operational performance.

Table 4.16: Mean, Median, and Mode of Airport Metrics across clusters for Newark Liberty International Airport (EWR)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ holdings\ (aircraft)$	**2.4**	4.5	38.1	**0**	1	36	**0**	**0**	M[1]
$Airborne\ holdings\ (minutes)$	**47.9**	92.4	1024.5	**0**	19	859	**0**	**0**	M[1]
$Completion\ rate\ (\%)$	**98.8**	97.0	87.7	**99.2**	98.8	92.1	**100**	**100**	M[1]
$Diversions$	**0.78**	1.15	17.6	**0**	1	13	**0**	7	**0**
$Departure\ delays$	**14**	31.4	83.6	**0**	11	61	0	0	0
$GDP\ lead-in\ time\ (minutes)$	-	113.5	94.4	-	127	114	-	0	0
$GDP\ revisions$	-	0.7	1.4	-	0	1	-	0	1
$Ground\ Stops$	**0.2**	0.6	1.7	**0**	**0**	2	**0**	**0**	2
$TMI\ to\ airport\ delays$	**5.9**	172.6	157.1	**5**	185	168	**5**	186	M[1]

[1]Multiple values exist

The third cluster is characterized by low to very high departure delays, TMI to airport delays, completion rates, airborne holdings (aircraft and minutes), Ground Stops and diversions. Table 4.16 also shows that this cluster has significantly higher mean and median airborne holdings (aircraft and minutes), diversions, departure delays, and Ground Stops compared to the other clusters. Overall, these characteristics correspond to poor operational performance.

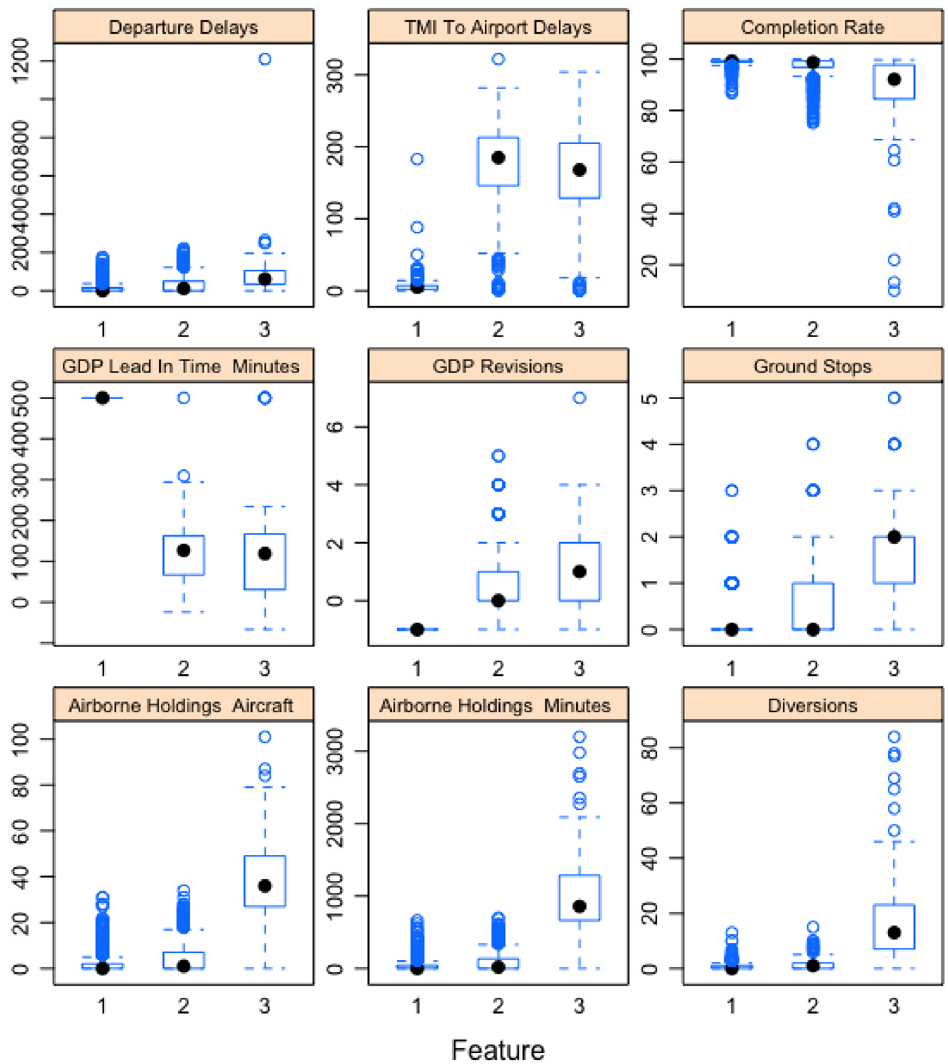

Figure 4.23: Box plots showing the distribution of airport metrics across clusters for Newark Liberty International Airport (EWR)

Comparison of results from clustering and using predefined ranges of metrics (EWR)

As with the other airports, Experiment 1 validates the categorization of daily operations of EWR into 3 categories, as is currently done with the OSPC. Similarly, each cluster is characterized by either good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of daily operations of EWR with clustering and OSPC was conducted and reviewed by Subject Matter Experts so as to determine which approach better categorizes the daily operations of EWR.

Table 4.17 shows that 570, 607, and 38 daily operations categorized as “Good days” by OSPC were assigned to the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 34 and 17 days categorized as “Average days” by OSPC were assigned to the clusters characterized by varying and poor operational performance, respectively. Table 4.17 also shows that 16 and 58 days classified as “Bad days” by OSPC were assigned to clusters characterized by varying and poor operational performance, respectively.

Table 4.17: Composition of categories from clustering and Operational Service Performance Criteria (EWR)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	570	607	38
Average day	0	34	17
Bad day	0	16	58

Table C.4 in Appendix C provides a comparison of a subset of daily EWR operations that were categorized differently by clustering and OSPC. In particular, it shows that days 1 through 3 were classified as “Good days” by OSPC, even though they were characterized by very negative Ground Delay Program lead-in times and high TMI to Airport delays. Days 6 and 7 were also classified as “Good days” by OSPC even though their very low completion rates suggests poor operational performance. Days 4, 5, and 9 are daily operations that were categorized differently by both approaches. Analysis of these three daily operations revealed that one or more metrics were close to the threshold between OSPC categories. Indeed, their classification based on the predefined ranges could change if the ranges were slightly adjusted. As previously discussed, these ranges are based on Subject Matter Expert inputs and slight adjustments will impact how airport operations are classified. As such, it can be observed that **clustering is a better suited approach for categorizing daily operations of Newark Liberty International Airport (EWR).**

4.2.5 Dulles International Airport (IAD)

Application of Principal Component Analysis (IAD)

Figure 4.24 shows the scree plot for Dulles International Airport (IAD). In particular, it shows that 4 principal components captured over 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 4.

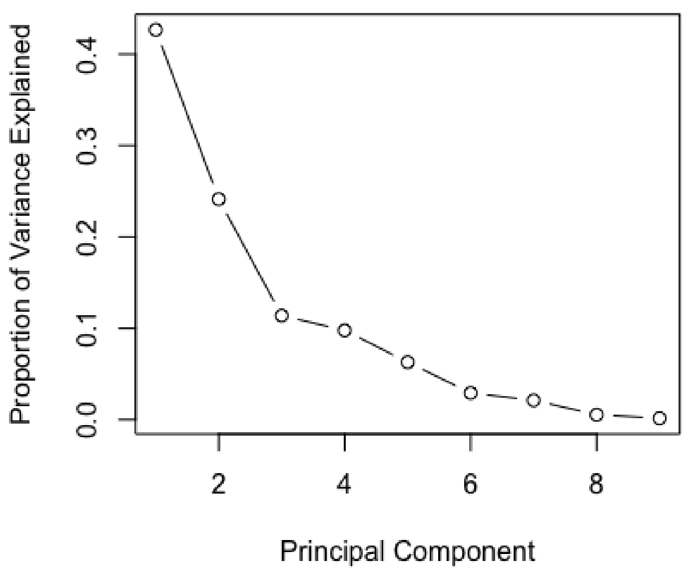

Figure 4.24: Scree plot for Dulles International Airport (IAD)

Assessment of the Clustering Tendency of the Dataset (IAD)

The Hopkins Statistic value of **0.974** for this dataset and the presence of a large dark blue box comprised of smaller boxes along the diagonal in Figure 4.25 indicates that multiple clusters can be generated from the dataset. The size of the large dark blue box also indicates that majority of objects in the dataset are fairly similar to each other. However, they can be further clustered as evidenced by smaller boxes along the diagonal.

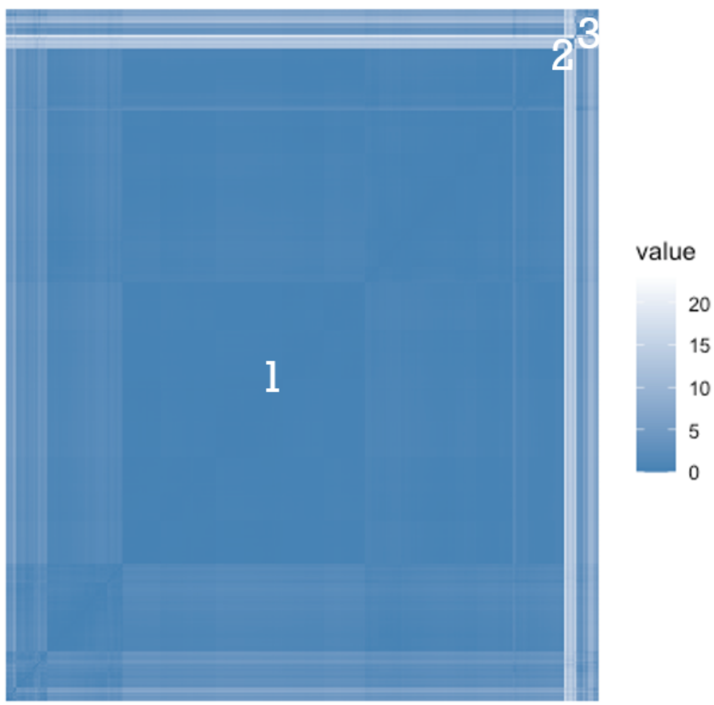

Figure 4.25: Visual Assessment of clustering Tendency (VAT) plot for Dulles International Airport (IAD)

Benchmarking and Evaluation of Clustering Algorithms (IAD)

Figures B.22, B.23, B.24, B.25, B.26, B.27, and B.28 in Appendix B show how each clustering algorithm performed while varying the number of clusters. Table 4.18 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for IAD. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** was identified as the best suited combination for IAD by a majority of metrics.

Table 4.18: Optimal combination of algorithm and number of clusters per evaluation metric for IAD

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	0.0008	Single Linkage	6,7
AD	0.8141	PAM	8
ADM	0.0209	Single Linkage	6
FOM	0.315	Ward	8
Connectivity	***3.6123***	***Single Linkage***	***3***
Dunn Index	***0.3585***	***Single Linkage***	***3***
Silhouette	0.8825	Average Linkage	3

Table 4.19 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Table 4.19 validates observations made from the assessment of clustering tendency of the dataset with the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.25 shows one large dark blue box along the diagonal corresponding to the first cluster, and two smaller boxes at the top right end of the diagonal corresponding to the two smaller clusters.

Table 4.19: Breakdown of daily operations of IAD by cluster

Cluster	Number of Daily Operations
1	1258
2	20
3	62

Analysis of Clusters (IAD)

Figure 4.26 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by high completion rates, a wide range of departure delays, and low to moderate TMI to airport delays, airborne holdings (minutes and aircraft), and diversions. None of the daily operations in this cluster are characterized by Ground Delay Programs. These observations as well as the cluster's low mean and median airborne holdings (minutes and aircraft), diversions, departure delays, Ground Stops, and TMI to airport delays, as seen in Table 4.20, corresponds to good operational performance.

Table 4.20: Mean, Median, and Mode of Airport Metrics across clusters for Dulles International Airport (IAD)

Airport Metric/Cluster	Mean			Median			Mode		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
$Airborne\ holdings\ (aircraft)$	**0.2**	6.5	17.3	**0**	6	18	**0**	**0**	12
$Airborne\ holdings\ (minutes)$	**4.9**	135.3	441.6	**0**	118	348	**0**	**0**	M[1]
$Completion\ rate\ (\%)$	**98.8**	97.1	92.9	**99.2**	97.6	97.	**100**	M[1]	M[1]
$Diversions$	**0.7**	2.3	9.9	**1**	2	9	**0**	1	M[1]
$Departure\ delays$	**9.4**	11.8	23.5	4	**3.5**	24	**0**	1	**0**
$GDP\ lead-in\ time\ (minutes)$	-	90.3	75	-	95.5	75	-	92	75
$GDP\ revisions$	-	0.1	1	-	0	1	-	0	1
$Ground\ Stops$	**0.07**	0.7	0.9	**0**	0.5	1	**0**	**0**	1
$TMI\ to\ airport\ delays$	**0.7**	57.9	12.5	**0**	60.5	8	**0**	M[1]	**0**

The second cluster is characterized by low to moderate departure delays, airborne holdings (minutes and aircraft), and diversions, as seen in Figure 4.26. It is also characterized by a wide range of TMI to airport delays and GDP lead-in times, and high completion rates. In addition, a majority of daily operations in this cluster did not have any GDP revisions which bodes well for operational performance. Table 4.20 shows that the mean and median airborne holdings (aircraft and minutes), diversions, departure delays, and Ground Stops of this cluster are higher than those

[1]Multiple values exist

of the first cluster but lower than those of the third cluster. However, this cluster has higher mean and median TMI to Airport delays values compared to the other clusters. As such, overall, the second cluster can be observed to have varying operational performance.

The third cluster is characterized by a wide range of completion rates, Ground Stops, airborne holdings (minutes and aircraft), and diversions, as seen in Figure 4.26. Table 4.20 shows that this cluster has the highest mean and median airborne holdings (aircraft and minutes), diversions, departure delays, and Ground Stops values. Overall, these characteristics correspond to poor operational performance.

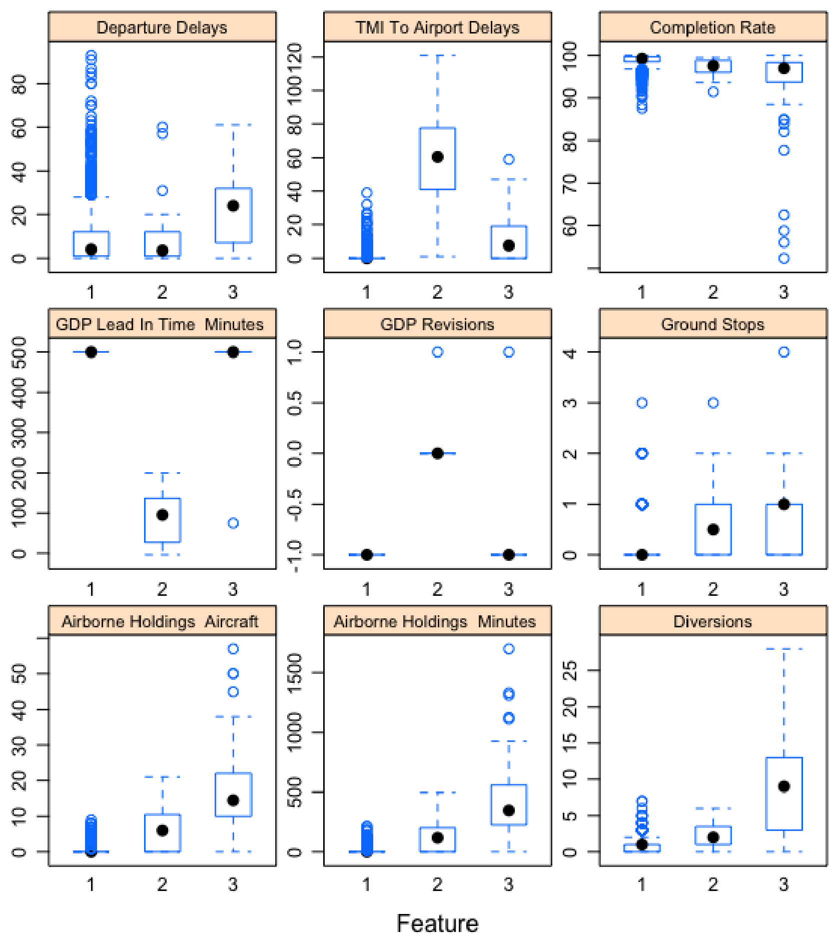

Figure 4.26: Box plots showing the distribution of airport metrics across clusters for Dulles International Airport (IAD)

Comparison of results from clustering and using predefined ranges of metrics (IAD)

As with the other airports, Experiment 1 validates the categorization of daily operations of IAD into 3 categories, as is currently done with OSPC. Similarly, each cluster is characterized by either good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of daily operations of IAD with clustering and predefined ranges of metrics was conducted and reviewed by FAA Subject Matter Experts so as to determine the approach that better categorizes the daily operations of IAD.

Table 4.21 shows that 1258, 19, and 60 days categorized as "Good days" by OSPC were assigned to the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 1 and 2 days categorized as "Average days" by OSPC were assigned to the clusters characterized by varying and poor operational performance, respectively.

Table 4.21: Composition of categories from clustering and Operational Service Performance Criteria (IAD)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	1258	19	60
Average day	0	1	2
Bad day	0	0	0

Table C.5 in Appendix C provides a comparison of a subset of daily operations of IAD that were categorized differently by both approaches. As observed with the other airports, OSPC classified days with poor completion rates as "Good days", as seen with days 1 through 4 in Table C.5. Days 5 through 9 were also classified as "Good days" even though the high duration of airborne holdings and number of diversions do not suggest good operational performance. Based on this analysis, **clustering is a better suited approach for categorizing daily operations of Dulles International Airport (IAD)**

4.2.6 John F. Kennedy International Airport (JFK)

Application of Principal Component Analysis (JFK)

Figure 4.27 shows the scree plot for John F. Kennedy International Airport (JFK). In particular, it shows that 3 principal components captured 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3.

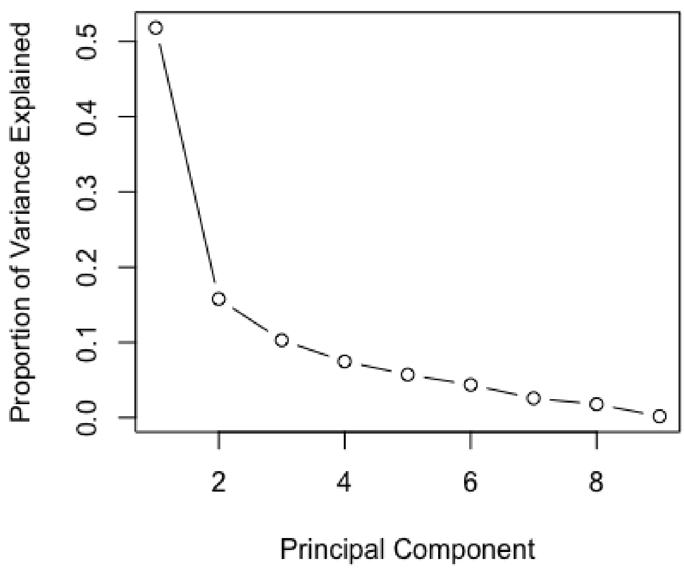

Figure 4.27: Scree plot for John F. Kennedy International Airport (JFK)

Assessment of the Clustering Tendency of the Dataset (JFK)

A Hopkins Statistic value of **0.978** and the presence of multiple blue boxes along the diagonal in Figure 4.28 indicates that the clustering tendency of the dataset is very high. The varying shades of the blue boxes in Figure 4.28 and their accompanying dissimilarity values also indicates that the dataset is characterized by dissimilar objects that can be clustered.

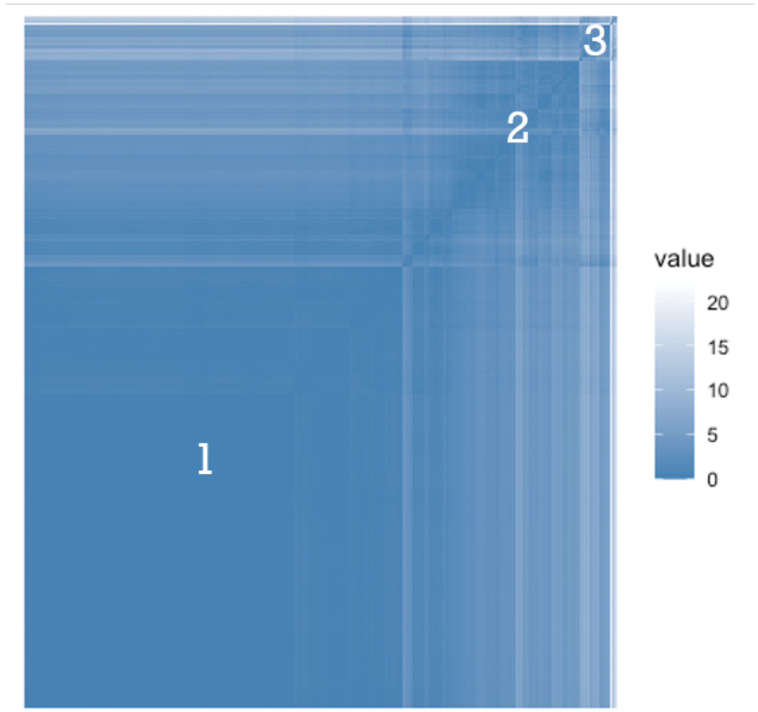

Figure 4.28: Visual Assessment of clustering Tendency (VAT) plot for John F. Kennedy International Airport (JFK)

Benchmarking and Evaluation of Clustering Algorithms (JFK)

Figures B.29, B.30, B.31, B.32, B.33, B.34, and B.35 in Appendix B show how each clustering algorithm performed while varying the number of clusters. Table 4.22 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for JFK. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** was identified as the best suited combination for JFK by a majority of metrics.

Table 4.22: Optimal combination of algorithm and number of clusters per evaluation metric for JFK

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	***0.0012***	***Single Linkage***	***3***
AD	0.6414	Clara	8
ADM	0.0236	Single Linkage	4
FOM	0.6764	Kmeans	8
Connectivity	***8.1313***	***Single Linkage***	***3***
Dunn Index	0.4024	Single Linkage	7
Silhouette	0.87877	Average Linkage	3

Table 4.23 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Table 4.23 validates observations made from assessment of the clustering tendency of the dataset with the Visual Assessment of clustering Tendency (VAT) plot. Indeed, Figure 4.28 shows a large dark blue box along the diagonal corresponding to the first cluster, and smaller boxes along the diagonal which correspond to the smaller clusters.

Table 4.23: Breakdown of daily operations of JFK by cluster

Cluster	Number of Daily Operations
1	967
2	287
3	86

Analysis of Clusters (JFK)

Figure 4.29 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by high completion rates, and low to high departure delays and airborne holdings (minutes and aircraft). It is also characterized by low to moderate TMI to Airport delays and diversions. Majority of the daily operations in the first cluster are also not characterized by the implementation of Ground Delay Programs and Ground Stops which bodes well for airport operations as they typically lead to flight delays. These observations, in addition to lower mean and median values of airborne holdings (aircraft and minutes), diversions, departure delays, GDP revisions, Ground Stops and TMI to airport delays, and the high mean and median completion rate and GDP lead-in times of this cluster compared to the other clusters, as seen in Table 4.24 indicates good operational performance.

Table 4.24: Mean, Median, and Mode of Airport Metrics across clusters for John F. Kennedy International Airport (JFK)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ holdings\ (aircraft)$	**2.5**	7.8	60.1	**0**	4	55	**0**	**0**	61
$Airborne\ holdings\ (minutes)$	**47.4**	153	1468	**0**	71	1362	**0**	**0**	M[1]
$Completion\ rate\ (\%)$	**99.1**	97.5	88.3	**100**	99	93	**100**	**100**	85
$Diversions$	**0.6**	1.3	13.9	**0**	1	9	**0**	**0**	2
$Departure\ delays$	**24.4**	90.5	116	**4**	61	110	**0**	**0**	**0**
$GDP\ lead-in\ time\ (minutes)$	**316**	94.7	99.8	**316**	91	109	M[1]	**0**	**0**
$GDP\ revisions$	**0**	0.6	1.1	**0**	**0**	1	**0**	**0**	1
$Ground\ Stops$	**0.1**	0.6	1.5	**0**	**0**	1	**0**	**0**	1
$TMI\ to\ airport\ delays$	**2.9**	89.3	106.2	**2**	79	113	**1**	2	134

The second cluster is characterized by low to high departure delays, TMI to airport delays, airborne holdings, and GDP lead-in times and revisions, as seen in Figure 4.29. It is also characterized by high completion rates and low diversions. Furthermore, Ground Stops were not implemented on a majority of days in this cluster. Even though these observations are similar to those from the first cluster, Table 4.24 shows that the mean and median values of airborne holdings (minutes and aircraft), diversions, departure delays, Ground Stops, and TMI to airport delays of this cluster are higher than those of the first cluster but lower than those of the third cluster. The mean and median values of completion rate are also observed to be lower than those of the first cluster but higher than those of the third cluster. As such, it is observed that the second cluster exhibits varying

[1]Multiple values exist

operational performance.

The third cluster is characterized by a wide range of departure delays, TMI to airport delays, completion rates, GDP lead-in time and revisions, Ground Stops, airborne holdings (aircraft and minutes), and diversions, as seen in Figure 4.29. This observation as well as the very high mean and median values of airborne holdings, diversions, departure delays, and TMI to airport delays indicate that this cluster exhibits poor operational performance.

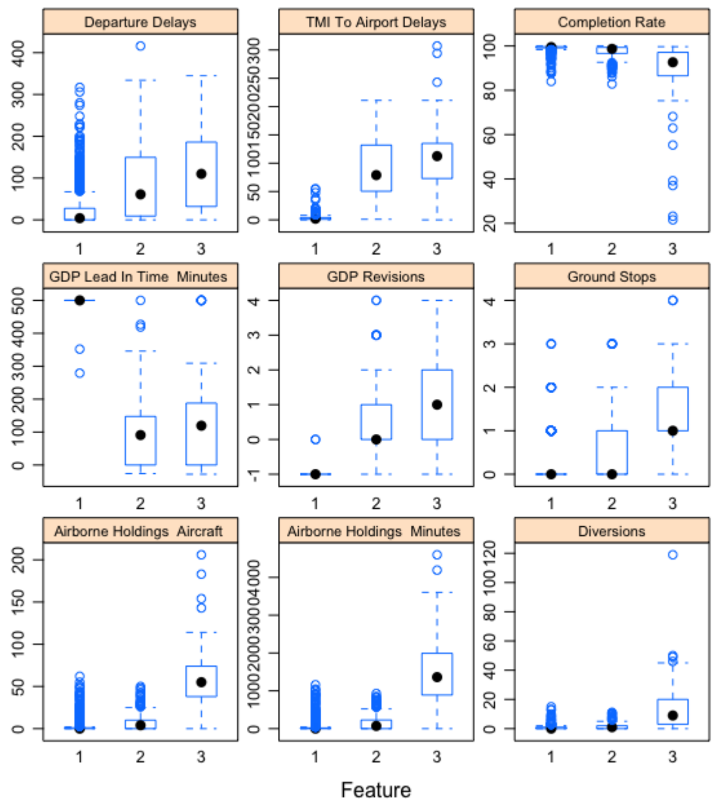

Figure 4.29: Box plots showing the distribution of airport metrics across clusters for John F. Kennedy International Airport (JFK)

Comparison of results from clustering and using predefined ranges of metrics (JFK))

Experiment 1 validates the categorization of the daily operations of JFK into 3 categories, as is currently done with OSPC. Similar to the other airports, each cluster is characterized by either good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of the daily operations of JFK with clustering and OSPC was conducted so as to determine the best approach for categorizing the daily operations of JFK.

Table 4.25 shows that 965, 267, and 28 days categorized as "Good days" by OSPC were placed in the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 15 and 8 days categorized as "Average days" by OSPC were assigned to the clusters characterized by varying and poor operational performance, respectively. Table 4.25 also shows that 2, 5, and 50 days classified as "Bad days" by OSPC were placed in clusters characterized by good, varying and poor operational performance, respectively.

Table 4.25: Composition of categories from clustering and Operational Service Performance Criteria (JFK)

OSPC/Clustering	Cluster 1	Cluster 2	Cluster 3
Good day	965	267	28
Average day	0	15	8
Bad day	2	5	50

Table C.6 in Appendix C provides a comparison of a subset of daily operations of JFK that were categorized differently by clustering and OSPC. As observed with other airports, OSPC classified days with low completion rates as "Good days", as seen with days 1 through 3 in Table C.6. Days 4 through 7 were also classified as "Good days" even though high airborne holdings (minutes and aircraft) do not correspond to good operational performance. Days 8 through 11 were also classified differently by both approaches due to variations in their underlying methodologies. Based on this analysis, it is concluded the **clustering is a better suited approach for categorizing daily**

operations of John F. Kennedy International Airport (JFK).

4.2.7 LaGuardia Airport (LGA)

Application of Principal Component Analysis (LGA)

Figure 4.30 shows the scree plot for LaGuardia Airport (LGA). In particular, it shows that 3 principal components captured 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3 and used for this work.

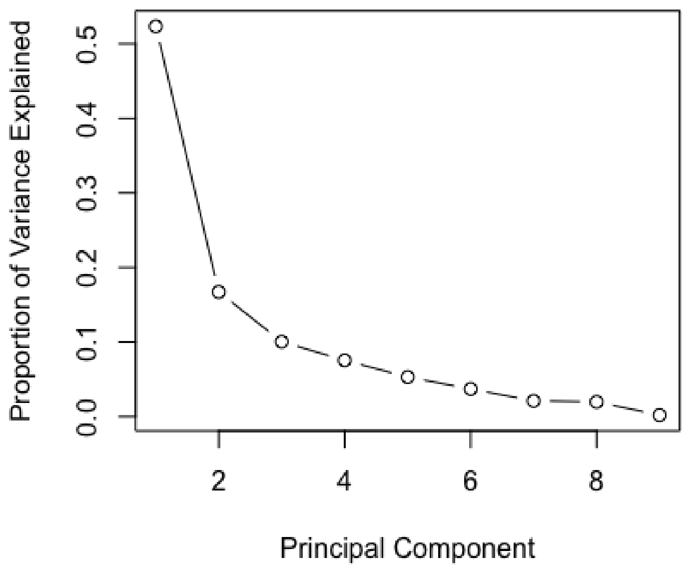

Figure 4.30: Scree plot for LaGuardia Airport (LGA)

Assessment of the Clustering Tendency of the Dataset (LGA)

The Hopkins Statistic value of **0.932** obtained for this dataset and the presence of two distinct blue boxes comprised of smaller boxes along the diagonal in Figure 4.31 indicates that the clustering tendency of the dataset is high. The varying shades of the blue boxes and their accompanying dissimilarity values also indicates that multiple clusters can be formed from the dataset.

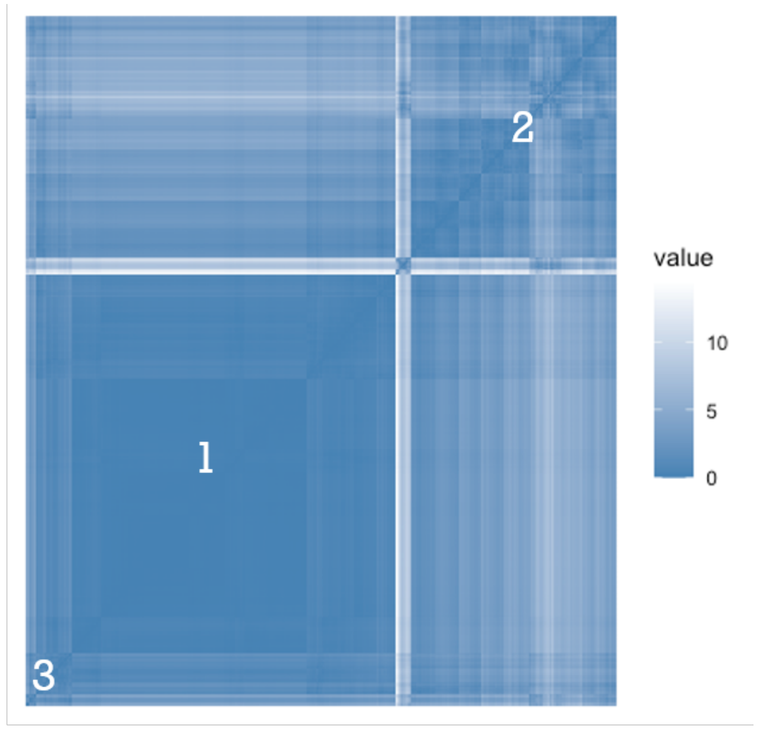

Figure 4.31: Visual Assessment of clustering Tendency (VAT) plot for LaGuardia Airport (LGA)

Benchmarking and Evaluation of Clustering Algorithms (LGA)

Figures B.36, B.37, B.38, B.39, B.40, B.41, and B.42 in Appendix B show how each clustering algorithm performed while varying the number of clusters. Table 4.26 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for LGA. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** was identified as the best suited combination for LGA by a majority of metrics.

Table 4.26: Optimal combination of algorithm and number of clusters per evaluation metric for LGA

Evaluation Metric	Value	Clustering Algorithm	Clusters
APN	***0.0006***	***Single Linkage***	***3***
AD	1.8385	PAM	8
ADM	0.0256	Single Linkage	5
FOM	0.6956	Kmeans	8
Connectivity	***6.7607***	***Single Linkage***	***3***
Dunn Index	0.3181	Single Linkage	4
Silhouette	0.6959	***Single Linkage***	***3***

Table 4.27 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Table 4.27 validates observations made from the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.31 shows a dark blue box along the diagonal corresponding to the first cluster, and smaller boxes in the top and bottom right corners of the diagonal which correspond to the smaller clusters.

Table 4.27: Breakdown of daily operations of LGA by cluster

Cluster	Number of Daily Operations
1	828
2	453
3	59

Analysis of Clusters (LGA)

Figure 4.32 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by high completion rates, low to moderate TMI to airport delays, and low to high departure delays and airborne holdings (aircraft and minutes). It can also be seen that Ground Delay Programs and Ground Stops were not implemented on a majority of days in this cluster. Table 4.28 shows that the mean and median values of a majority of metrics were significantly better for the first cluster compared to the other clusters. As such, it can be observed that overall, this cluster is characterized by good operational performance.

The second cluster is characterized by low to high departure delays, TMI to airport delays, GDP lead-in times and revisions, Ground Stops, airborne holdings, and diversions, as seen in Figure 4.32. Table 4.28 shows that the mean and median values of a majority of metrics are better for this cluster compared to the third cluster but worse than the first cluster. As such, it can be concluded that the second cluster generally exhibits varying operational performance.

Table 4.28: Mean, Median, and Mode of Airport Metrics across clusters for LaGuardia Airport (LGA)

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ holdings\ (aircraft)$	**2.8**	16	46	**0**	9	50	0	0	0
$Airborne\ holdings\ (minutes)$	**60**	354	1301	**0**	179	1464	0	0	0
$Completion\ rate\ (\%)$	**99**	96	71	**99**	98	76	**100**	**100**	M[1]
$Diversions$	**0.8**	2.7	21	**0**	1	17	0	0	0
$Departure\ delays$	**24**	101	52	**0**	81	49	0	0	0
$GDP\ lead-in\ time\ (minutes)$	**176**	51	92	**165**	32	103	M[1]	0	M[1]
$GDP\ revisions$	**0**	0.6	1.1	**0**	1	2	**0**	**0**	2
$Ground\ Stops$	**0.2**	1.3	2	**0**	1	2	**0**	1	1
$TMI\ to\ airport\ delays$	**7.3**	189	185	**6**	187	186	0	0	0

The third cluster is characterized by a wide range of departure delays, TMI to airport delays, completion rate, GDP lead-in times and revisions, airborne holdings, Ground Stops and diversions. Table 4.28 also shows that the third cluster has the worst median and mean values of the metrics out of all three clusters. As such, this cluster is observed to exhibit poor operational performance.

[1] Multiple values exist

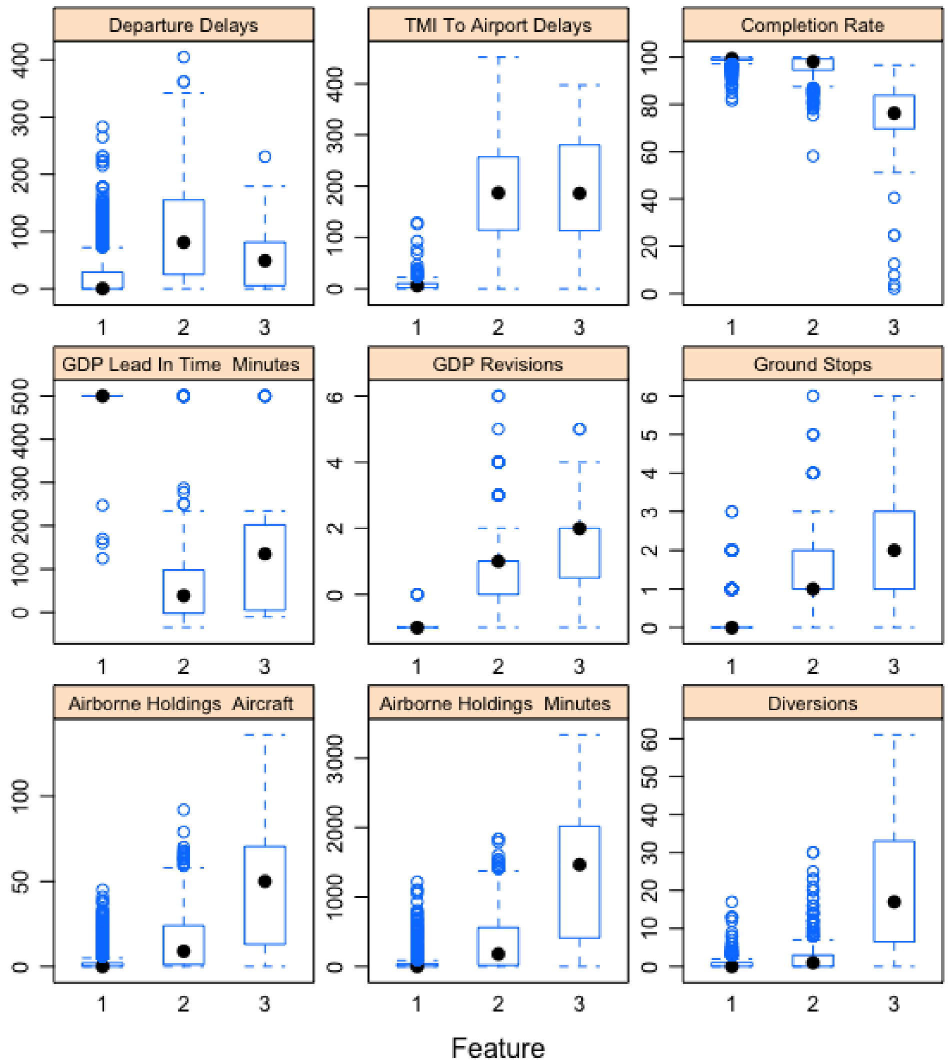

Figure 4.32: Box plots showing the distribution of airport metrics across clusters for LaGuardia Airport (LGA))

Comparison of results from clustering and using predefined ranges of metrics (LGA)

As observed with the other airports, Experiment 1 validates the categorization of the daily operations of LGA into 3 categories, as is currently done with OSPC. Similarly, each cluster is characterized by either good, varying, or poor operational performance. As such, a one-to-one comparison of the categorization of daily operations of LGA with clustering and predefined ranges of metrics was conducted so as to determine a suitable approach for categorizing the daily operations of LGA.

Table 4.29 shows that 828, 310, and 15 days categorized as "Good days" by OSPC were placed in the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 26 and 7 days categorized as "Average days" by OSPC were placed in clusters characterized by varying and poor operational performance, respectively. Table 4.29 also shows that 117 and 37 days classified as "Bad days" by OSPC were placed in clusters characterized by varying and poor operational performance, respectively.

Table 4.29: Composition of categories from clustering and Operational Service Performance Criteria (LGA)

OSPC/Clustering	Cluster 1	Cluster 2	Cluster 3
Good day	828	310	15
Average day	0	26	7
Bad day	0	117	37

Table C.7 in Appendix C provides a comparison of a subset of daily operations of LGA that were categorized differently by clustering and OSPC. Indeed, it shows trends similar to those observed with the other airports, where OSPC classified days with low Completion Rates as "Good days", etc. Similar to the other airports, it is concluded that **clustering is a better approach for categorizing the daily operations of LaGuardia Airport (LGA), instead of using a broad range of predefined values of metrics.**

4.2.8 Philadelphia International Airport (PHL)

Application of Principal Component Analysis (PHL)

Figure 4.33 shows the scree plot for Philadelphia International Airport (PHL). In particular, it shows that 3 principal components captured 90% of the variance of the dataset. As such, the dimensionality of the dataset was reduced from 9 to 3.

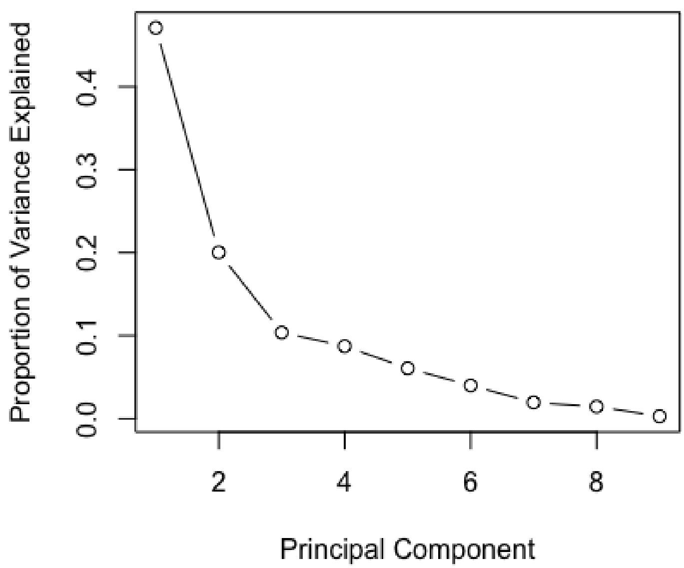

Figure 4.33: Scree plot for Philadelphia International Airport (PHL)

Assessment of the Clustering Tendency of the Dataset (PHL)

The presence of multiple blue boxes along the diagonal in Figure 4.34 and the Hopkins Statistic value of **0.957** for this dataset indicates that the clustering tendency of the dataset is high.

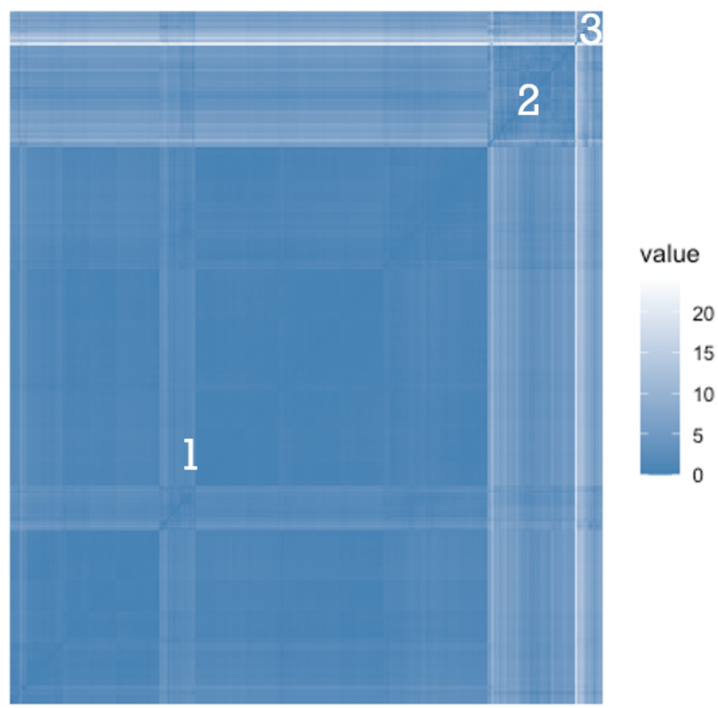

Figure 4.34: Visual Assessment of clustering Tendency (VAT) plot for Philadelphia International Airport (PHL)

Benchmarking and Evaluation of Clustering Algorithms (PHL)

Figures B.43, B.44, B.45, B.46, B.47, B.48, and B.49 in Appendix B show how each clustering algorithm performed while varying the number of clusters. Table 4.30 provides a summary of the best suited combination of algorithm and number of clusters per evaluation metric for PHL. In particular, it shows that the **Single Linkage Hierarchical clustering algorithm and 3 clusters** was identified as the best suited combination for PHL by a majority of metrics.

Table 4.30: Optimal combination of algorithm and number of clusters per evaluation metric for PHL

Evaluation Metric	**Value**	**Clustering Algorithm**	**Clusters**
APN	***0.0001***	***Single Linkage***	***3***
AD	1.5289	PAM	8
ADM	***0.0018***	***Single Linkage***	***3***
FOM	0.7381	Ward	8
Connectivity	***8.1448***	***Single Linkage***	***3***
Dunn Index	***0.5246***	***Single Linkage***	***3***
Silhouette	***0.8097***	***Single Linkage***	***3***

Table 4.31 provides a breakdown of the number of daily airport operations in each cluster, as designated by the Single Linkage Hierarchical clustering algorithm. Table 4.31 validates observations made from the assessment of clustering tendency of the dataset with the Visual Assessment of clustering Tendency plot. Indeed, Figure 4.34 shows a large dark blue box as well as smaller light blue boxes with high dissimilarity values along the diagonal of Figure 4.34.

Table 4.31: Breakdown of daily operations of LGA by cluster

Cluster	**Number of Daily Operations**
1	1078
2	193
3	69

Analysis of Clusters (PHL)

Figure 4.35 shows the distribution of the airport metrics with box plots across the three clusters. In particular, it shows that the first cluster is characterized by moderate to high completion rates, and low to high departure delays, TMI to airport delays, airborne holdings (aircraft and minutes) and diversions. Figure 4.35 also shows that Ground Delay Programs and Ground Stops were not implemented on a majority of days in this cluster. Based on these observations as well as the mean and median values of the metrics for this cluster compared to the other clusters, it can be concluded that overall, the first cluster is characterized by good operational performance.

Figure 4.35 shows that other than TMI to airport delays, and GDP lead-in time and revisions, the distribution of metrics in the second cluster is similar to that of the first cluster. However, Table 4.32 shows that the mean and median values of a majority of metrics are worse than those of the first cluster but better than those of the third cluster. As such, it is concluded that overall, the second cluster is characterized by varying operational performance.

Table 4.32: Mean, Median, and Mode of Airport Metrics across clusters for Philadelphia International Airport (PHL)

.

Airport Metric/Cluster	Mean			Median			Mode		
	1	2	3	1	2	3	1	2	3
$Airborne\ holdings\ (aircraft)$	**0.7**	1.8	18.7	**0**	**0**	18	**0**	**0**	M[1]
$Airborne\ holdings\ (minutes)$	**11.7**	33.4	462	**0**	**0**	442	0	0	0
$Completion\ rate\ (\%)$	**98.2**	94.3	86.2	**98.7**	95.5	92	**100**	M[1]	[1]
$Diversions$	**0.7**	1.1	8.9	**0**	1	6	**0**	**0**	M[1]
$Departure\ delays$	**20.8**	48	72	**5**	25	60	0	0	0
$GDP\ lead-in\ time\ (minutes)$	47	77.1	**97**	47	71	**102**	M[1]	**0**	M[1]
$GDP\ revisions$	**0**	0.5	0.7	0	0	0	0	0	0
$Ground\ Stops$	**0.2**	1.1	1.7	**0**	1	1	**0**	1	1
$TMI\ to\ airport\ delays$	**3.7**	119.4	86.7	**2**	89	116	0	0	0

The third cluster is characterized by a wide range of departure delays, TMI to airport delays, completion rates, GDP lead-in time and revisions, airborne holdings (aircraft and minutes), Ground Stops, and diversions. These observations coupled with the cluster's poor mean and median values of the metrics shown in Table 4.32 indicate that it is characterized by poor operational performance.

[1] Multiple values exist

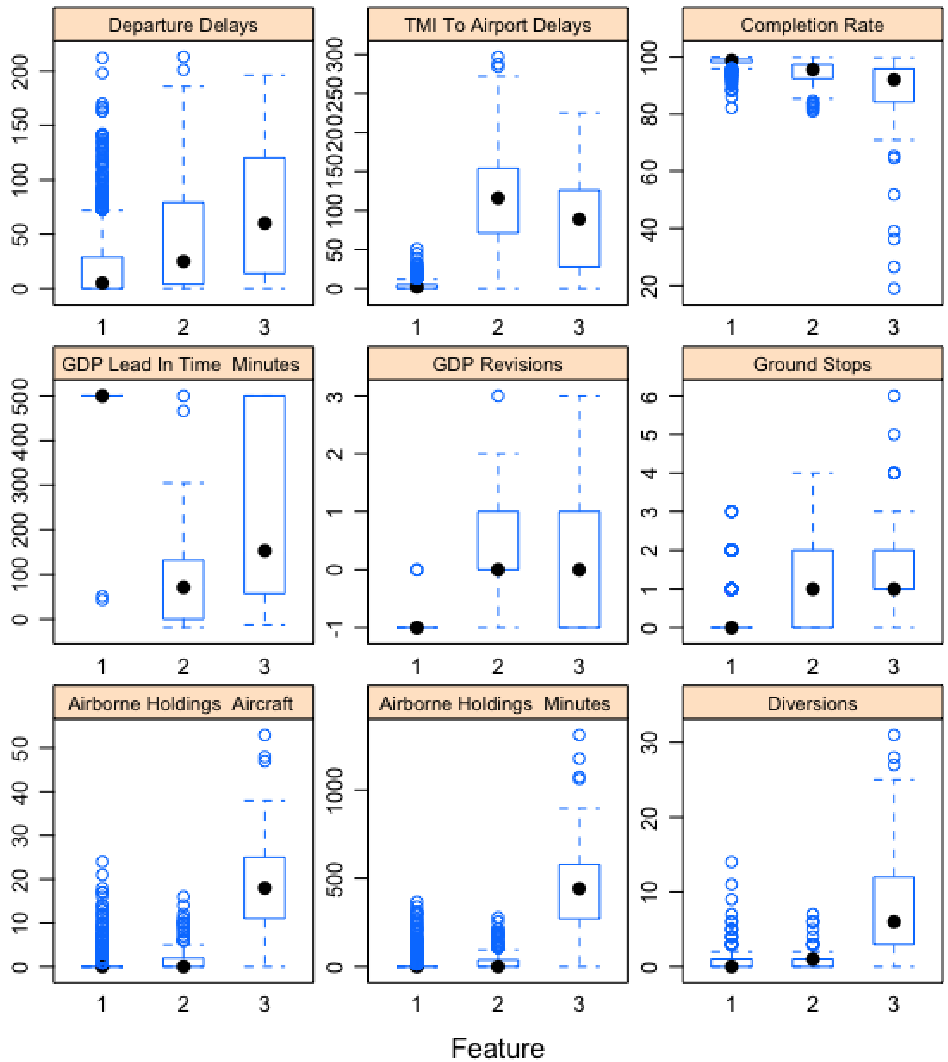

Figure 4.35: Box plots showing the distribution of airport metrics across clusters for Philadelphia International Airport (PHL)

Comparison of results from clustering and using predefined ranges of metrics (PHL)

Experiment 1 validates the categorization of the daily operations of PHL into 3 categories, as is currently done with OSPC. Similar to the other airports, each cluster is either characterized by good, varying, or poor operational performance.

Table 4.33 shows that 1078, 185, and 44 days categorized as "Good days" by OSPC were placed in the clusters characterized by good, varying, and poor operational performance, respectively. It also shows that 8 and 9 days categorized as "Average days" by OSPC were placed in clusters characterized by varying and poor operational performance, respectively. Table 4.33 also shows that 16 days classified as "Bad days" by OSPC were placed in the cluster characterized by poor operational performance.

Table 4.33: Composition of categories from clustering and Operational Service Performance Criteria (PHL)

OSPC/Clustering	**Cluster 1**	**Cluster 2**	**Cluster 3**
Good day	1078	185	44
Average day	0	8	9
Bad day	0	0	16

Table C.8 in Appendix C provides a comparison of a subset of daily operations of PHL that were categorized differently by clustering and OSPC. Indeed, it shows trends similar to those observed with the other airports, where OSPC classified days with low Completion Rates as "Good days", etc. Similar to the other airports, it is concluded that **clustering is a better suited approach for categorizing the daily operations of Philadelphia International Airport (PHL) instead of predefined ranges of metrics.**

4.3 Summary of Findings from Experiment 1

The first research question posed in Chapter 3 examines the capability of the methodology discussed herein to categorize daily airport operations instead of using predefined ranges of metrics. While the existing literature outlined various efforts pursued to categorize airports to analyze their operations, it was observed that a systematic, robust, and repeatable approach was lacking. As such, Experiment 1 was developed, implemented, and tested with data from 8 U.S. airports. This involved extracting and computing the necessary metrics, normalizing, reducing the dimensionality and assessing the clustering tendency of the datasets. The performance of different clustering algorithms was then benchmarked and evaluated using a set of metrics to identify the best combination of algorithm(s) and number of clusters for each airport. The clustering results and the categories developed by OSPC were then compared to determine which approach better categorizes the daily operations of each airport.

The dimensionalities of all but two of the datasets were reduced from 9 to 3 using Principal Component Analysis. The dimensionality of the Boston Logan and Dulles International Airport datasets were reduced from 9 to 4. Hopkins Statistic and Visual Assessment of clustering Tendency (VAT) plots were then leveraged to assess the clustering tendency of the datasets. For all 8 airports, the clustering tendency of the datasets was observed to be very high, as seen in Table 4.34 and Figures 4.6, 4.16, 4.19, 4.22, 4.25, 4.28, 4.31, and 4.34. These observations indicate that there are distinct differences between daily operations of each airport. As such, clustering can be leveraged to categorize daily airport operations.

Table 4.34: Summary of Hopkins Statistic values for the airports

Airport	**Hopkins Statistic**
BOS	0.957
BWI	0.973
DCA	0.957
EWR	0.956
IAD	0.974
JFK	0.978
LGA	0.932
PHL	0.957

The best combination of clustering algorithm and number of clusters was also determined to be the **Single Linkage Hierarchical algorithm and 3 clusters** for each airport by a majority of metrics. The three clusters of each airport were analyzed and observed to either exhibit good, varying, or poor operational performance, as seen in Table 4.35.

Table 4.35: Characteristics of Clusters

Cluster	Characteristic
1	Good Operational Performance
2	Varying Operational Performance
3	Poor Operational Performance

The number of clusters and their characteristics is consistent with the approach currently employed by OSPC which classifies daily airport operations into "Good days", "Average days", and "Bad days". As such, the outcomes of clustering and OSPC were compared and reviewed by FAA Subject Matter Experts to determine which approach better categorizes the daily operations of each airport. The comparison revealed that OSPC classified a majority of daily airport operations as "Good days", even though several of them exhibited sub-optimal to poor operational performance due to very low completion rates and Ground Delay Program lead-in times, and high airborne holdings (minutes and aircraft). As such, FAA analysts may have to manually validate these classifications each day and/or regularly update the predefined ranges of metrics. The clustering algorithm on the other hand, correctly classified a majority of these as either having varying or poor operational performance. Variations in the means, medians, and modes of the airport clusters also indicates that the same predefined set of ranges of metrics should not be applied across the eight airports, as is currently done in OSPC. This observation in addition to the distributions of metrics across clusters, as observed in Figures 4.14, 4.17, 4.20, 4.23, 4.26, 4.29, 4.32, and 4.35 shows that the metrics cannot be weighted equally, as is currently assumed with OSPC.

Based upon this comparison and the review of the developed clusters by FAA Subject Matter Experts, it is concluded that the conditions of Hypothesis 1 are satisfied, namely that benchmarking clustering algorithms while varying the number of clusters will facilitate the categorization of daily airport operations in a systematic, robust, and repeatable manner.

Therefore, the hypothesis for Research Question 1 is verified.

CHAPTER 5

DETERMINING THE CATEGORY THAT DAILY AIRPORT OPERATIONS BELONG TO

In this chapter, the development and application of a methodology for determining the category that subsequent daily airport operations belong to is discussed in order to successfully answer Research Question 2.1:

Research Question 2.1

How can the category that a daily airport operation belongs to be better determined?

This chapter also presents methodologies for comparing daily operations in similar and different airport categories, and analyzing and assessing how traffic management decisions impact daily airport operations. These efforts relate to the subjects of the following research questions:

Research Question 2.2

How can daily airport operations in similar and different categories be compared for the identification of trends and patterns?

Research Question 2.3

How can the impact of traffic management decisions on airport operations be analyzed and assessed?

5.1 Methodology Overview

Figure 5.1 provides an overview of the methodology for Experiment 2 which was implemented with Python [147, 148]. The methodology highlighted in Figure 5.1 was implemented for each airport using data and results from Chapter 4. The target of each model is the category of daily airport operation determined in Experiment 1, and the predictors are the nine metrics, as well as the month of year. The remainder of this section discusses each step of the methodology in detail.

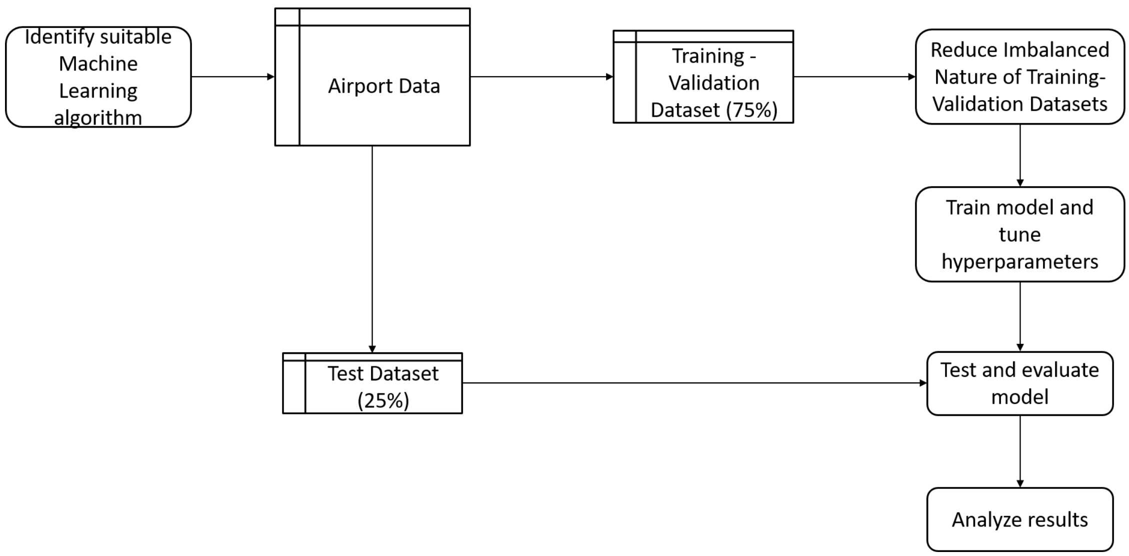

Figure 5.1: Overview of methodology for Experiment 2

5.1.1 Step #1: Identify Suitable Machine Learning Algorithm

As previously discussed, an objective of this work is to leverage the ranking of predictor importance, Decision Trees, and the posterior probability of predictions of a supervised Machine Learning algorithm in order to answer Research Questions 2.2 and 2.3. As such, there was a need to identify a suitable classification tree-based algorithm to undertake these tasks.

The Random Forests algorithm was identified to be a suitable algorithm for this work. Breiman [239] defines Random Forests as “a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest". Random Forests are an ensemble-based method that leverage random feature selection to add additional diversity to Decision Tree models. Random Forests are also widely known for combining versatility and power into a single Machine Learning approach. As the ensemble only uses a small, random portion of the full dataset, Random Forests can handle extremely large datasets where dimensionality might cause other models to fail [127, 239–241]. The algorithm outputs Decision Trees which are often leveraged to interpret how models are developed, and to identify underlying relationships between predictors and targets of prediction models [146]. Random Forests also provide a ranking of predictor importance which can be leveraged to gain insights into underlying causes of events. The algorithm also outputs the posterior probability or degree of support of predictions as a means to measure the confidence of the model's prediction(s).

5.1.2 Step #2: Split Datasets

The data from each airport was divided into two datasets: training-validation and test. As shown in Figure 5.1, three-fourths of the data was assigned to the training-validation dataset which was used to train and tune the models, and one-fourth of the data was assigned to the test dataset to generate predictions for evaluations. The data for each airport was randomly divided to ensure that the training-validation and test datasets do not have systematic differences.

5.1.3 Step #3: Reduce Imbalanced Nature of Training-Validation Datasets

As observed in Chapter 4, the datasets for each airport are heavily imbalanced, as a majority of daily operations belong to one or two clusters. Imbalanced datasets often lead to poorly performing prediction models. As such, the Synthetic Minority Over-sampling Technique (SMOTE) algorithm [242, 243] was leveraged to balance the training-validation datasets of each airport by increasing the minority class(es) of datasets. This is achieved by randomly selecting a k-nearest-neighbor of each member of the minority class. Implementing the SMOTE algorithm instead of naively oversampling the minority class ensures that over-fitting is avoided.

5.1.4 Step #4: Train Models and Tune Hyperparameters

The next step in the methodology focuses on training the prediction models. This is achieved by using the Random Forests Classifier from Python's Scikit library [244] and k-fold cross validation. K-fold cross validation is a resampling method that is used to tune Machine Learning model hyperparameters to ensure optimal performance, where k is a predefined number of groups or folds that the data is split into [245–247]. The following Random Forests hyperparameters were tuned using a grid search in order to identify the best combination(s) of hyperparameters needed for optimal model performance:

Number of trees

The Random Forests algorithm occasionally overfits if the number of trees is too large. As such, there is a need to identify the appropriate number of trees needed for optimal model performance [248].

Maximum depth

Maximum depth refers to the maximum number of splits in each tree, which dictates the complexity of the Random Forests and influences model performance. As such, there is a need to identify the appropriate maximum depth needed for optimal model performance.

These hyperparameters were varied using the range of values shown in Table 5.1.

Table 5.1: Random Forests hyperparameters and their range of values

Hyperparameter	**Range of values**
Number of trees	[100,200,300,400,500,600,700,800,900,1000]
Maximum depth	[1-30]

5.1.5 Step #5: Test and Evaluate Models

The performance of the trained models were then tested and evaluated using a set of metrics. Evaluating the performance of learners is vital as it indicates how a learner will perform on future/unseen data. The type of evaluation metric used depends on whether the task involved classifications or numeric predictions, as well as on how "balanced" the dataset the models are being trained on is. Classification learners are typically evaluated using results obtained from a confusion matrix. A confusion matrix, as seen in Table 5.2, is a table that categorizes predictions according to whether they match the actual value. For classification tasks such as this one, confusion matrices are leveraged to measure performance using metrics such as Sensitivity, Specificity, and Balanced accuracy [127].

Table 5.2: Confusion Matrix

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive (TP)	False Positive (FP)
Predicted Negative	False Negative (FN)	True Negative (TN)

True Positive (TP) refers to the correct classification of the class of interest. True Negative (TN) refers to the correct classification of the class that is not of interest. False Positive (FP) refers to the incorrect classification of the class of interest. False Negative (FN) refers to the incorrect classification of the class that is not of interest [127].

Sensitivity

This refers to the proportion of true positives that were correctly classified, and is specified as [127]:

$$Sensitivity = \frac{TP}{TP + FN}$$

Specificity

This refers to the proportion of negative examples that were correctly classified, and is specified as [127]:

$$Specificity = \frac{TN}{FP + TN}$$

Balanced Accuracy

A model might have high accuracy because it correctly predicts the most frequent class, particularly when the dataset is imbalanced. Balanced accuracy adjusts accuracy by calculating the average of accurate predictions in each class [249, 250] and is specified as:

$$Balanced\ Accuracy = \frac{Sensitivity + Specificity}{2}$$

Table 5.3 shows the range and criteria for good performance of each evaluation metric.

Table 5.3: Range and criteria for prediction model metrics

Evaluation metric	**Range of values**	**Criteria for good performance**
Specificity	(0,1)	Maximize
Sensitivity	(0,1)	Maximize
Balanced Accuracy	(0,1)	Maximize

5.1.6 Step #6: Analyze Results

This section discusses features of the Random Forests algorithm that facilitate the analysis of the outcomes of prediction models. The following features of the algorithm were generated using Python's Scikit library [244]:

Ranking of Predictor Importance

As previously mentioned, the Random Forests Machine Learning algorithm provides a ranking of predictor importance that quantifies the impacts that predictors have on prediction models. This work thus leveraged the ranking of predictor importance to determine how each metric influences the categorization of each airport's daily operations. This will provide insights into similarities and differences between airports, and will facilitate the identification of trends and patterns for better decision making. It will also enable FAA analysts to identify how various traffic management decisions impact airport operations.

Decision Trees

The Random Forests algorithm also outputs a Decision Tree which allows for a global interpretation of prediction models and provide a means to identify underlying relationships between predictors and targets of prediction models [146]. Decision Trees are composed of branches, maximum depth and nodes. Each branch is a path taken from the first node at the top of the tree and corresponds to whether the model agrees with the designated class of that node, as seen in Figure 5.2. Maximum depth refers to the longest path from the first node (root) to the last node of the tree (leaf). Each node other than the leaf node is characterized by five parameters: feature value, gini impurity, samples, value, and class. The *feature value* dictates how each node is split. A data point moves to one of two subsequent nodes based on the feature's value at a particular node. *Gini impurity* refers to the likelihood of incorrectly classifying a randomly chosen data point in a node based on the distribution of data points in the node. Lower gini values indicate less likelihood of incorrectly classifying a randomly chosen data point. *Samples* refers to the number of data points used to determine an outcome, whereas *values* provides a breakdown of the number of data points of each class. *Class* refers to the most likely outcome of that node. The color (orange and blue) of each node corresponds to the assigned class at that node. A white node indicates that there is an equal number of observations of each class at that node. As such, a class is not assigned to that node.

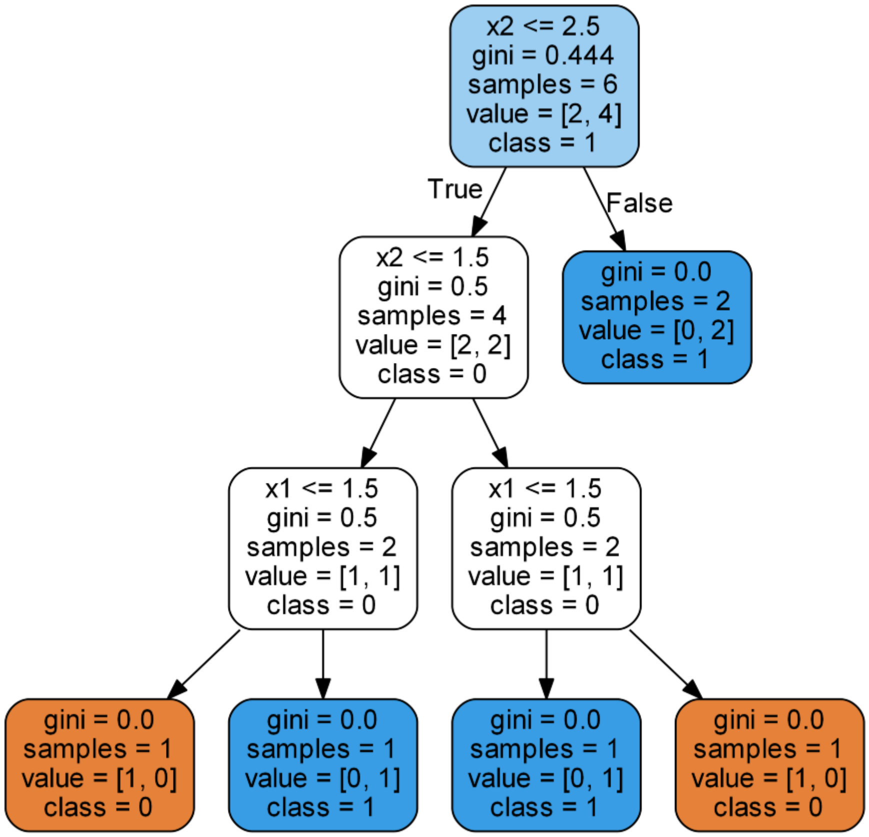

Figure 5.2: Sample Decision Tree [251]

This work thus leveraged Decision Trees to identify underlying relationships between the metrics and airport categories. This will provide a means for analysts to better understand the impact and assess the effectiveness of traffic management decisions on daily airport operations.

Degree of Support of Predictions

The Random Forests algorithm also outputs a posterior probability or degree of support of predictions. This work thus leveraged this feature of the algorithm to facilitate the comparison of daily operations within and across airport categories. Doing so will provide a means for analysts to quantify how well a daily operation fits into an airport category determined in Chapter 4 and will enable them to identify trends and patterns.

Steps 1 through 6 of Experiment 2 were implemented using Algorithm 3.

Algorithm 3: Algorithm for training and testing prediction models

```
Let k be the number of folds
Vary k from 1 to 10
for each airport do
    for each k do
        randomly split the training-validation set into k groups
        for each group do
            set the selected group as the testing dataset
            set the remainder of the data as the training dataset
            train model and tune hyperparameters with the training dataset
            evaluate model with the testing dataset
            record the evaluation metrics
        end for
    end for
    return the best performing model and number of folds by comparing evaluation metrics
end for
return the best performing model and number of folds by comparing evaluation metrics
return ranking of predictor importance
return decision tree
return degree of support of predictions
```

5.2 Implementation and Testing of Methodology for Determining the Category That Daily Airport Operations Belong To

This section discusses the implementation and testing of Steps 1 through 5 of Experiment 2 with data from each of the 8 U.S. airports. As discussed in the previous section, the datasets were randomly divided into two sets, and the SMOTE algorithm was leveraged to reduce the imbalanced nature of the training-validation datasets of each airport to ensure optimal model performance. This was achieved by increasing the number of data points of each minority class to match that of the majority class. Combinations of hyperparameters needed for optimal performance of the algorithm were identified while training the models using k-fold cross-validation. Model performance was then evaluated using a set of metrics.

5.2.1 Boston Logan International Airport (BOS)

Table 5.4 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for BOS. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 500, 7, and 3, respectively.

Table 5.4: Summary of optimal set of hyperparameters and number of folds for BOS

Parameter	Optimal Value
Number of trees	500
Maximum depth	7
Number of folds	3

Tables 5.5 and 5.6 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the BOS test dataset, respectively. Table 5.5 shows that the algorithm accurately predicted all but two daily airport operations. Model performance was observed to be excellent as 99.4% of daily operations were predicted correctly.

Table 5.5: Confusion matrix for BOS on the test dataset

	Actual Cluster 1	Actual Cluster 2	Actual Cluster 3
Predicted Cluster 1	252	0	0
Predicted Cluster 2	2	80	0
Predicted Cluster 3	0	0	1

Table 5.6: Summary of evaluation metrics for BOS on the test dataset

Metric	Cluster 1	Cluster 2	Cluster 3
Specificity	1	0.992	1
Sensitivity	0.992	1	1
Balanced Accuracy	0.996	0.996	1

5.2.2 Baltimore/Washington International Thurgood Marshall International Airport (BWI)

Table 5.7 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for BWI. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 800, 8, and 3, respectively.

Table 5.7: Summary of optimal set of hyperparameters and number of folds for BWI

Parameter	Optimal Value
Number of trees	800
Maximum depth	8
Number of folds	3

Tables 5.8 and 5.9 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the BWI test dataset, respectively. Table 5.8 shows that the algorithm accurately predicted all of the daily airport operations, resulting in perfect values for all of the evaluation metrics in Table 5.9.

Table 5.8: Confusion matrix for BWI on the test dataset

	Actual Cluster 1	Actual Cluster 2	Actual Cluster 3
Predicted Cluster 1	318	0	0
Predicted Cluster 2	0	10	0
Predicted Cluster 3	0	0	7

Table 5.9: Summary of evaluation metrics for BWI on the test dataset

Metric	Cluster 1	Cluster 2	Cluster 3
Specificity	1	1	1
Sensitivity	1	1	1
Balanced Accuracy	1	1	1

5.2.3 Reagan National Airport (DCA)

Table 5.10 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for DCA. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 100, 13, and 3, respectively.

Table 5.10: Summary of optimal set of hyperparameters and number of folds for DCA

Parameter	Optimal Value
Number of trees	100
Maximum depth	13
Number of folds	3

Table 5.11 shows the confusion matrix for the Random Forests algorithm with the DCA test dataset. In particular, it shows that the algorithm accurately predicted all of the daily airport operations, resulting in perfect values for all of the evaluation metrics in Table 5.12.

Table 5.11: Confusion matrix for DCA on the test dataset

	Actual Cluster 1	Actual Cluster 2	Actual Cluster 3
Predicted Cluster 1	296	0	0
Predicted Cluster 2	0	18	0
Predicted Cluster 3	0	0	21

Table 5.12: Summary of evaluation metrics for DCA on the test dataset

Metric	**Cluster 1**	**Cluster 2**	**Cluster 3**
Specificity	1	1	1
Sensitivity	1	1	1
Balanced Accuracy	1	1	1

5.2.4 Newark Liberty International Airport (EWR)

Table 5.13 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for EWR. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 100, 6, and 3, respectively.

Table 5.13: Summary of optimal set of hyperparameters and number of folds for EWR

Parameter	**Optimal Value**
Number of trees	100
Maximum depth	6
Number of folds	3

Tables 5.14 and 5.15 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the EWR test dataset, respectively. Table 5.14 shows that the model accurately predicted all of the daily operations in the first and third clusters, but inaccurately placed 6 daily operations into the second cluster instead of the third cluster.

Table 5.14: Confusion matrix for EWR on the test dataset

	Actual Cluster 1	**Actual Cluster 2**	**Actual Cluster 3**
Predicted Cluster 1	141	0	0
Predicted Cluster 2	0	166	6
Predicted Cluster 3	0	0	22

Table 5.15: Summary of evaluation metrics for EWR on the test dataset

Metric	Cluster 1	Cluster 2	Cluster 3
Specificity	1	0.965	1
Sensitivity	1	1	0.786
Balanced Accuracy	1	0.982	0.893

5.2.5 Dulles International Airport (IAD)

Table 5.16 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for IAD. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 500, 8, and 3, respectively.

Table 5.16: Summary of optimal set of hyperparameters and number of folds for IAD

Parameter	Optimal Value
Number of trees	500
Maximum depth	8
Number of folds	3

Tables 5.17 and 5.18 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the IAD test dataset, respectively. Table 5.17 shows that the model predicted all of the daily operations in the second cluster, and a majority of daily operations in the first and third clusters correctly. However, the model inaccurately predicted 4 daily operations to be in the first cluster instead of the third cluster, and 1 daily operation in the third cluster instead of the first cluster. Overall, model performance was excellent as 98.5% of daily operations were predicted correctly.

Table 5.17: Confusion matrix for IAD on the test dataset

	Actual Cluster 1	**Actual Cluster 2**	**Actual Cluster 3**
Predicted Cluster 1	309	0	4
Predicted Cluster 2	0	8	0
Predicted Cluster 3	1	0	13

Table 5.18: Summary of evaluation metrics for IAD on the test dataset

Metric	**Cluster 1**	**Cluster 2**	**Cluster 3**
Specificity	0.840	1	0.997
Sensitivity	0.997	1	0.765
Balanced Accuracy	0.918	1	0.881

5.2.6 John F. Kennedy International Airport (JFK)

Table 5.19 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for JFK. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 700, 9, and 3, respectively.

Table 5.19: Summary of optimal set of hyperparameters and number of folds for JFK

Parameter	**Optimal Value**
Number of trees	700
Maximum depth	9
Number of folds	3

Tables 5.20 and 5.21 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the JFK test dataset, respectively. Table 5.20 shows that even though the model predicted a majority of daily operations correctly, it inaccurately predicted 5 daily operations. In particular, 3 daily operations were predicted to be in the third cluster instead of the second cluster, and one daily operation was predicted to be in the third cluster instead of the first one. One other

operation was also inaccurately predicted to be in the first cluster instead of the second cluster. Overall, model performance was excellent as 98.5% of daily operations were predicted correctly.

Table 5.20: Confusion matrix for JFK on the test dataset

	Actual Cluster 1	**Actual Cluster 2**	**Actual Cluster 3**
Predicted Cluster 1	245	1	0
Predicted Cluster 2	0	68	0
Predicted Cluster 3	1	3	17

Table 5.21: Summary of evaluation metrics for JFK on the test dataset

Metric	**Cluster 1**	**Cluster 2**	**Cluster 3**
Specificity	0.989	1	0.987
Sensitivity	0.996	0.944	1
Balanced Accuracy	0.992	0.972	0.994

5.2.7 LaGuardia Airport (LGA)

Table 5.22 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for LGA. In particular, it shows that the optimal number of trees, maximum depth, and number of folds was determined to be 600, 8, and 3, respectively.

Table 5.22: Summary of optimal set of hyperparameters and number of folds for LGA

Parameter	**Optimal Value**
Number of trees	600
Maximum depth	8
Number of folds	3

Tables 5.23 and 5.24 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the LGA test dataset, respectively. Table 5.23 shows that even though the model predicted a majority of daily operations correctly, it inaccurately predicted 2 daily operations to the second cluster instead of the first and third clusters. It also inaccurately predicted one daily operation to the first cluster instead of the third cluster. As with the other airports, model performance was excellent as 99.1% of daily operations were predicted accurately.

Table 5.23: Confusion matrix for LGA on the test dataset

	Actual Cluster 1	**Actual Cluster 2**	**Actual Cluster 3**
Predicted Cluster 1	210	0	1
Predicted Cluster 2	1	110	1
Predicted Cluster 3	0	0	12

Table 5.24: Summary of evaluation metrics for LGA on the test dataset

Metric	**Cluster 1**	**Cluster 2**	**Cluster 3**
Specificity	0.992	0.991	1
Sensitivity	0.995	1	0.857
Balanced Accuracy	0.994	0.9996	0.929

5.2.8 Philadelphia International Airport (PHL)

Table 5.25 provides a summary of the optimal set of hyperparameters and number of folds identified and used to develop the prediction model for PHL. In particular, it shows that the optimal number of trees, maximum depth, and number of folds were determined to be 300, 8, and 3, respectively.

Table 5.25: Summary of optimal set of hyperparameters and number of folds for PHL

Parameter	**Optimal Value**
Number of trees	300
Maximum depth	8
Number of folds	3

Tables 5.26 and 5.27 show the confusion matrix and evaluation metrics for the Random Forests algorithm with the PHL test dataset, respectively. Table 5.26 shows that even though the model predicted a majority of daily operations correctly, it inaccurately predicted 2 daily operations to the second cluster instead of the first cluster, and 1 daily operation to the third cluster instead of the second cluster. As with the other airports, model performance was excellent as 99.1% of daily operations were predicted accurately.

Table 5.26: Confusion matrix for PHL on the test dataset

	Actual Cluster 1	**Actual Cluster 2**	**Actual Cluster 3**
Predicted Cluster 1	276	0	0
Predicted Cluster 2	2	42	0
Predicted Cluster 3	0	1	14

Table 5.27: Summary of evaluation metrics for PHL on the test dataset

Metric	**Cluster 1**	**Cluster 2**	**Cluster 3**
Specificity	1	0.993	1
Sensitivity	0.993	0.977	0.993
Balanced Accuracy	0.996	0.985	0.996

5.3 Implementation and Testing of Methodology for Comparing Daily Operations in Similar and Different Airport Categories

The posterior probability or degree of support of predictions of each of the prediction models developed with the Random Forests algorithm were generated to assess the degree to which daily operations belong to airport categories. Each prediction was accompanied by how well the model thought a daily operation fit into each airport category based on historical data, as seen in Table 5.28. In particular, it shows five consecutive daily operations of Philadelphia International Airport (PHL) that were categorized differently. It also shows that even though days 3 through 5 were predicted to be in the third cluster, day 5 was observed to have very poor operational performance compared to days 3 and 4. This can be leveraged by stakeholders to compare daily operations in similar and different clusters so as to determine why consecutive daily airport operations were categorized differently, for example. This will allow stakeholders to identify any traffic management decisions that may have led to variations in airport operational performance, which may lead to the identification of trends and pattern for better decision making.

Table 5.28: Sample probability or degree of support of predictions

Day	Cluster	Cluster 1 probability	Cluster 2 probability	Cluster 3 probability
1	2	4.69	95.3	0
2	1	100	0	0
3	3	0	9.7	90.3
4	3	0	20.1	79.9
5	3	0	0	100

5.4 Implementation and Testing of Methodology for Analyzing and Assessing How Traffic Management decisions Impact Airport Operations

This section discusses the use of the ranking of predictor importance and Decision Trees of each prediction model to analyze and assess how traffic management decisions impact operations at the eight U.S. airports used for this work.

5.4.1 Ranking of Predictor Importance

Figure 5.3, and Figures D.1, D.3, D.4, D.6 and D.7 in Appendix D provide the ranking of predictor importance of the prediction models for PHL, BOS, BWI, DCA, EWR, IAD, JFK, and LGA, respectively. In particular, they show that Ground Delay Program Lead-In Time and Revisions, and TMI to AIrport Delays largely influenced the categorization of daily operations at BOS, DCA, EWR, JFK, LGA, and PHL, respectively. Indeed, these three metrics accounted for 52.3% to 82.9% of metric importance across these airports, which indicates, as expected, that the planning and implementation of Traffic Management Initiatives has significant impacts on operations of these airports.

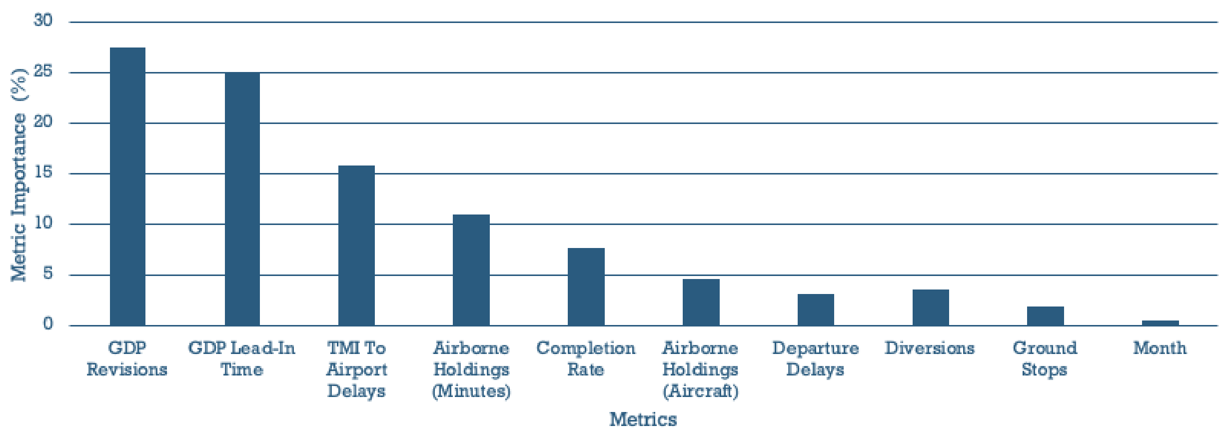

Figure 5.3: Ranking of predictor importance (PHL)

In particular, they show that Ground Delay Program Lead-In Time and Revisions, and TMI To Airport Delays largely influenced the categorization of daily operations at BOS, DCA, EWR, JFK, LGA, and PHL, respectively. Indeed, these three metrics accounted for 52.3% to 82.9% of metric importance across these airports, which indicates that the planning and implementation of Traffic Management Initiatives has significant impacts on operations of these airports.

Figure D.2 shows that Completion Rate, Airborne Holdings (Minutes and Aircraft), Ground Delay Program Revisions and Lead-In Times, and Diversions largely influenced the categorization of daily operations at Baltimore/Washington International Thurgood Marshall Airport (BWI). Indeed, these metrics combined accounted for over 88% of metric importance of the prediction model.

Figure D.5 shows that Airborne Holdings (Minutes and Aircraft) and Diversions accounted for 29.3%, 24.2%, and 11.7% of metric importance for the IAD prediction model, respectively. As such, it can be observed that these metrics significantly influenced the categorization of daily operations at Dulles International Airport (IAD).

As illustrated, the ranking of predictor importance provides some insight into how each metric influences the categorization of daily airport operations. This provides a means for assessing how traffic management decisions impact airport operations. The identification of key predictors also enables stakeholders to focus on improving decisions related to specific metrics so as to ensure efficient airport operations. Variations in the ranking of predictor importance of each model indicate that traffic management decisions and procedures affect operations at each airport differently. This observation thus invalidates OSPC's current assumption of using a broad set of predefined ranges of metrics across the eight airports. Finally, it is observed that the month of year consistently ranked as one of the least important metrics across the eight airports. As such, it can be concluded that its impact on airport operations is minimal.

5.4.2 Decision Trees

The Random Forests algorithm learns the relationships between predictors and targets so as to make accurate predictions. This is achieved by iterating over the data numerous times until the number of incorrect predictions is minimized. Decision Trees were thus leveraged to gain insight into how the optimal iteration of the data as well as the combinations of different predictors influenced model performance. Figure 5.4, and Figures E.1, E.2, E.3, E.4, E.5, E.6 and E.7 in Appendix E show how different traffic management decisions led to the prediction of daily airport categories for BOS, BWI, DCA, EWR, IAD, JFK, LGA, and PHL, respectively.

Figure 5.4 for example, shows that a daily operation at PHL will most likely be determined to have varying operational performance if it is characterized by a Completion Rate greater than 95.73%, GDP revisions greater than -0.5 ($\sim$ 0), and TMI to airport delays greater than 103.5. However, a daily operation at PHL will most likely be determined to have poor operational performance if it is characterized by a Completion Rate greater than 95.73%, GDP revisions greater than -0.5 ($\sim$ 0), TMI to airport delays less than or equal to 103.5, and Airborne Holdings (Minutes) greater than 229. Performing such analysis with the Decision Trees provides a means for stakeholders to assess how various traffic management decisions influence the categorization of daily airport operations, and thus impact airport operations. The Decision Trees can also be leveraged to create metric thresholds which would enable traffic management personnel to better plan and implement Traffic Management Initiatives and other operational decisions. Finally, Decision Trees and the posterior probability or degree of support of predictions can be leveraged to assess the effectiveness of traffic management decisions and procedures. In particular, once the degree to which a daily operation belongs to an airport category is determined using the posterior probability or degree of support of predictions, Decisions Trees can be leveraged to assess the effectiveness of traffic management decisions, and how they lead to “very good” operational performance or “barely good” operational performance, for example.

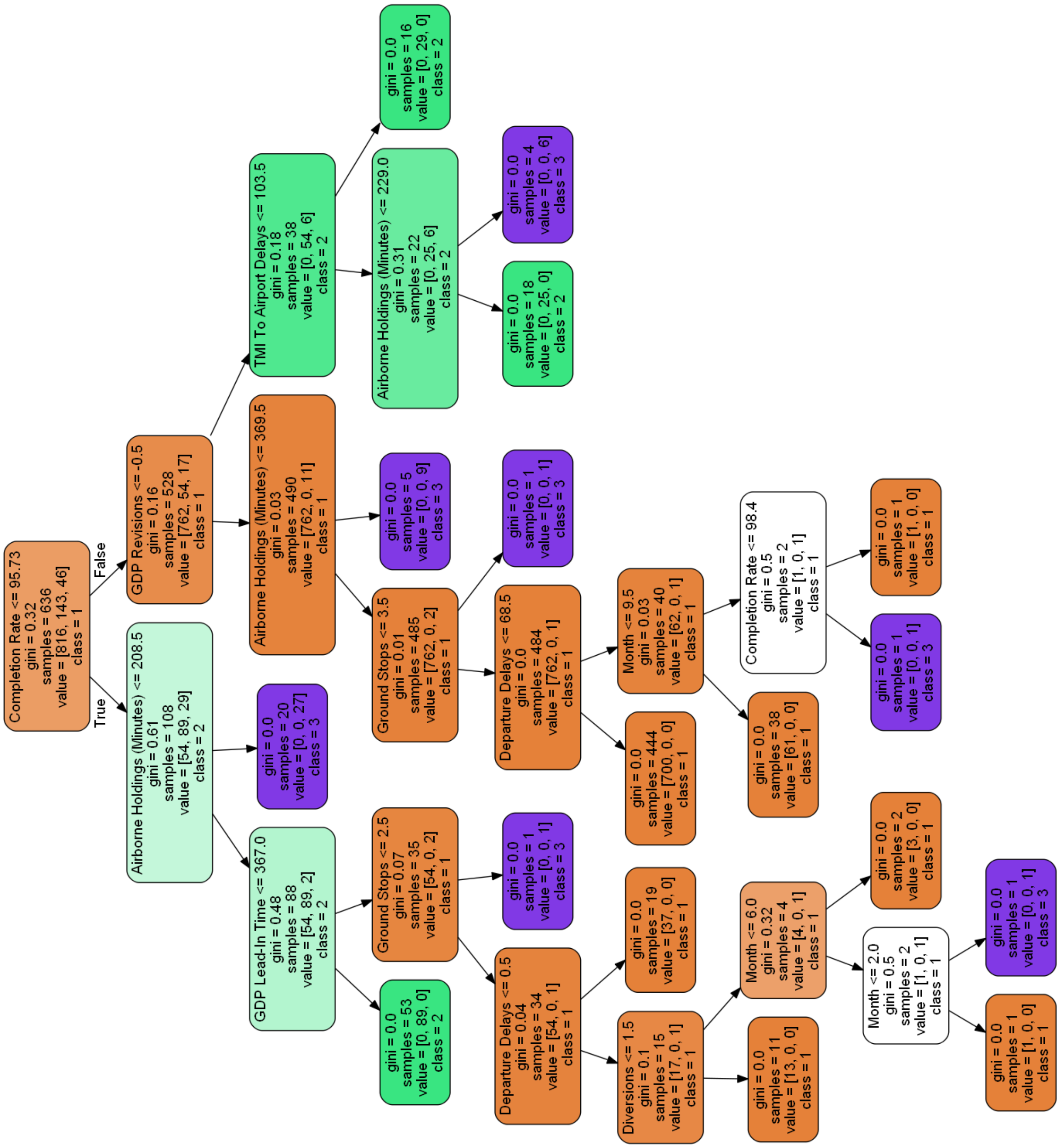

Figure 5.4: Decision Tree (PHL)

5.5 Summary of Findings from Experiment 2

Research Question 2.1 posed in Chapter 3 examines the capability of the methodology discussed herein to determine the category that daily airport operations belong to, instead of identifying the predominant class of parameters. As such, Experiment 2 was developed, implemented, and tested with metrics from 8 U.S. airports and the Random Forests Machine Learning algorithm. This experiment involved randomly splitting the data from each airport into training-validation and test datasets. The SMOTE algorithm was then leveraged to reduce the imbalanced nature of the datasets to ensure optimal model performance. Prediction models were then trained while tuning algorithm hyperparameters, and the performance of the models developed with the optimal combinations of parameters were evaluated with a set of metrics. Based upon the excellent performance of the prediction models, it is concluded that the conditions of Hypothesis 2.1 are satisfied, namely that developing prediction models with Machine Learning will provide a robust means for determining the category that daily airport operations belong to.

The remainder of Experiment 2 involved leveraging the ranking of predictor importance, Decision Trees, and the probability or degree of support of predictions to facilitate the analysis and comparison of daily airport operations. It is concluded that the conditions for Hypothesis 2.2 are satisfied, namely that daily operations in similar and different categories can be compared by leveraging the posterior probability or degree of support of predictions of the prediction models. It is also concluded that the conditions for Hypothesis 2.3 are satisfied, namely that the ranking of predictor importance and Decision Trees of prediction models provide a means for analyzing and assessing the impact of traffic management decisions on airport operations.

Therefore, the hypotheses for Research Questions 2.1, 2.2, and 2.3 are verified.

CHAPTER 6
DEVELOPMENT OF A FRAMEWORK FOR THE EXTRACTION, PROCESSING, AND STORAGE OF AIRPORT DATA

Chapter 4 outlined the implementation and testing of a methodology for categorizing daily airport operations. Chapter 5 outlined the implementation and testing of methodologies for determining the category that daily airport operations belong to, the comparison of daily airport operations in similar and different categories, and the analysis and assessment of the impact of traffic management decisions and procedures on airport operations. This chapter presents the development of a framework as a means to address the following research question:

Research Question 3

How can the efficient analysis and assessment of daily airport operations be automated from data extraction, through processing, analysis, and storage?

6.1 Methodology Overview

Figure 6.1 provides a broad overview of the methodology for Experiment 3, which consists of nine steps. The first step focuses on identifying technologies and software that are compatible with the FAA's Enterprise Information Management platform. Subsequent steps focus on the deployment of various software and scripts needed for the framework, the initiation of the framework, extraction of data needed, processing of data, temporary storage of data, execution of Machine Learning scripts, permanent storage of data, and analysis of data.

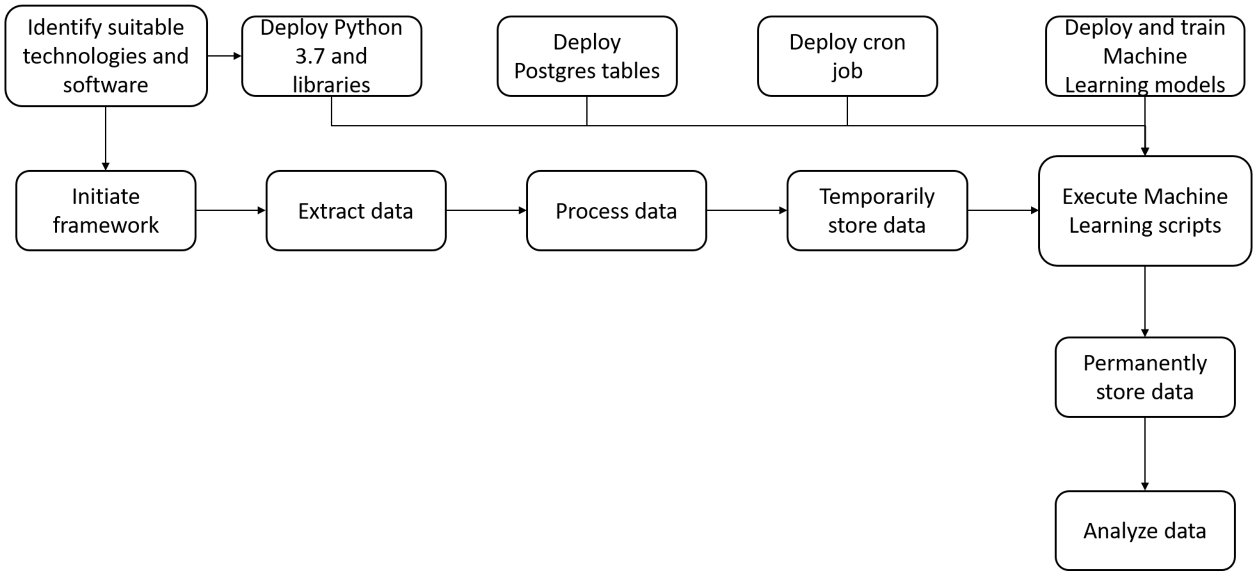

Figure 6.1: Overview of methodology for Experiment 3

6.1.1 Step #1 : Identify Suitable Technologies and Software

As discussed in Section 2.2, OSPC is a manual and time consuming process, as analysts spend the majority of their time extracting and processing data from ASPM. This research thus proposes the development of a framework to facilitate the efficient extraction, processing, and storage of data needed for the analysis and assessment of daily airport operations. The framework should offer a reliable and secure way to ingest, process, store, and analyze data from many sources. In addition, it should overcome constraints such as limited or expensive bandwidth, while ensuring data quality and reliability. The framework should also enable users to monitor data flow from ingestion to storage. This is particularly important for regulatory organizations such as the FAA, as they often need to retain and report chain of custody of data. Consequently, Apache NiFi, Amazon Simple Storage Service (S3), Elastic Search, Postgres, MuleSoft, and Tableau were identified as a suitable set of technologies and software for the development of the framework, as they satisfy the aforementioned requirements and are compatible with the FAA's Enterprise Information Management (EIM) platform. The remainder of this section provides an overview of each of these technologies.

Apache NiFi

Apache NiFi is a powerful and reliable system for the processing and distribution of Big Data [165–169]. It is a flow-based programming technology that is characterized by networks of processes, which exchange data across predefined connections. It is also data agnostic and reduces the risk of memory and performance related issues by facilitating data buffering through the application of back pressure [252, 253]. In addition, Apache NiFi automates the flow of data between systems by using data routing, transformation, and system mediation logic, and allows users to configure and regulate how data is distributed and consumed. It also enables users to track the flow of data in real time, and facilitates data provenance by logging the history of all processes. NiFi has a user interface that enables users to edit, design, visualize, and monitor the flow of data. It also provides users with the opportunity to monitor all data received, forked, joined, cloned, modified, sent, and ultimately dropped upon reaching its configured end-state. The user interface, as seen in Figure 6.2 can be assessed via a web browser and is composed of several segments with specific functions [165–168].

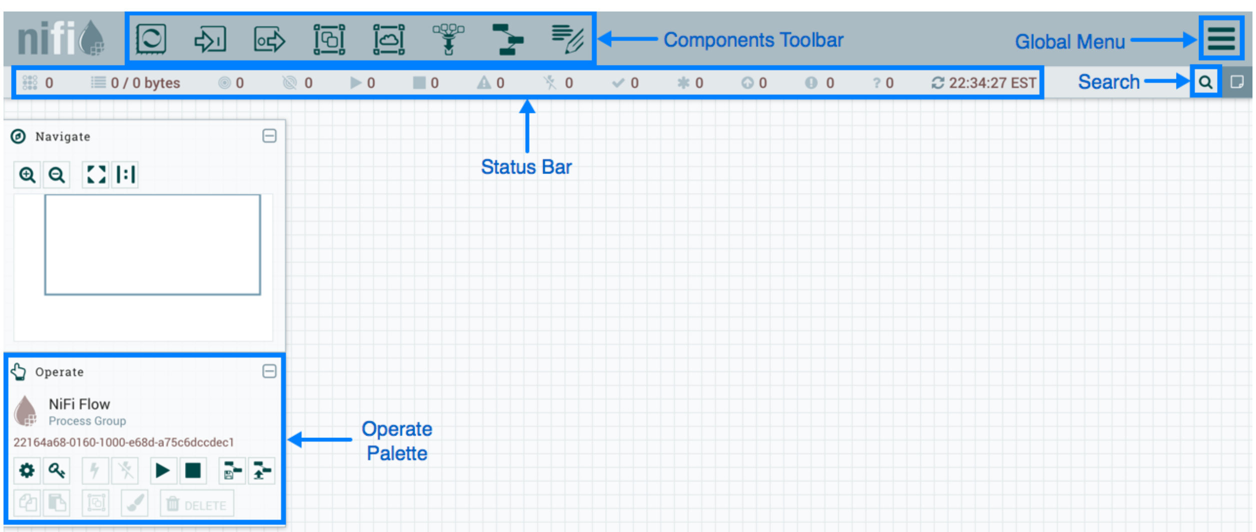

Figure 6.2: Web based user interface for Nifi [169]

The Components Toolbar, located in the top left section of the canvas is comprised of components, each with specific functions, that can be dragged onto the canvas to develop a dataflow. The Status bar, located below the Components Toolbar provides information such as the number of active threads in the dataflow, the amount of data being handled, the state of the different processes (transmitting or not transmitting), the timestamp at which the information was last refreshed, etc. The Operate Palette on the left side of the screen consists of buttons that can be used to start and end parts or all of the data flow, and manage the flow. The Search bar, located at the top right of the canvas can be used to search components by their names, types, identifiers, configuration properties, and values. The Global Menu component contains options that allow users to manipulate existing components on the canvas.

The flow of data in NiFi is facilitated by *Processors*, represented by the boxes in Figure 6.3. The processors are developed in JAVA to perform specific functions such as ingesting, converting, processing or merging data. They are developed as NiFi Archive (NAR) files that mirror Java Web application Archive (WAR) or Java Application Archive (JAR) files. NAR files are used in NiFi because they provide isolation from the potential issue of “NoClassDefFoundError” exceptions that can be generated when wrong versions of dependencies are loaded in the ClassLoader from a different processor, which often occurs with JAR and WAR files. Each developed NAR file is deployed into NiFi with Ansible scripts which automate the deployment of system configurations, software, codes, files, etc. into Big Data platforms [178–180]. This replaces the manual and time consuming effort required to transfer the NAR files to NiFi.

Each arrow in Figure 6.3, referred to as *Connections*, connects processors and enables various processes to interact at different rates. The connections enable users to keep track of how many files are transferred between processors. NiFi also enables users to identify any errors in the data flow. This is done using **LogMessage** processors which log the location and cause of errors in the data flow. This is particularly useful as programmers do not have to review multiple scripts or codes to identify the locations and causes of errors. Each processor also provides users with a summary of files that they ingested and processed, as well as the time taken to execute tasks. This enables users to efficiently monitor data flow in real time.

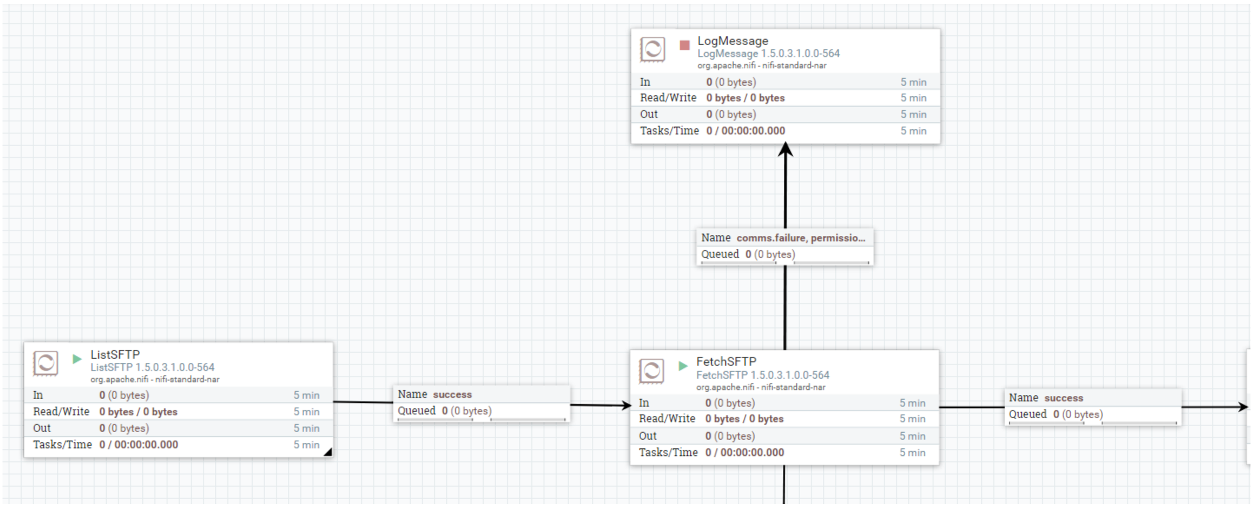

Figure 6.3: NiFi flow showing NiFi processors

Apache NiFi was thus leveraged to automate the efficient extraction, processing, and storage of data as it replaces the need for analysts to manually extract and process data. NiFi also allows analysts to track the flow of data in real time and to easily identify errors and their causes. Analysts will also be able to expand the scope of this work to include additional airports in the National Airspace System, while avoiding any computer processing and memory related issues that they may face with a local machine.

Amazon Simple Storage Service (S3)

Amazon's Simple Storage Service (S3) provides a scalable and secure platform for storing and protecting data [159, 254]. Consequently, S3 was identified and used to store the data used in this work.

Elastic Search

Elastic Search [162–164] provides a secure platform for storing, searching, and analyzing Big Data. As such, it was leveraged to store the indices or locations of data stored in S3.

MuleSoft Application Programming Interfaces (API)

Application Programming Interface (APIs) are software intermediaries that allows applications to communicate with each other [255]. As such, APIs developed with Mulesoft were leveraged to query data stored in S3. The MuleSoft APIs initially extract the indices or locations of desired data files from Elastic Search, and use this information to return the requested data.

PostgreSQL

PostgreSQL is a relational database that stores Big Data in a safe and secure manner [170–172]. It was leveraged as an additional store of data as analysts can use SQL-like queries to query data needed. It was also leveraged to provide data to Tableau dashboards developed to facilitate the analysis and assessment of daily airport operations.

Tableau

Tableau is a visualization tool that enables users to analyze, understand, interpret, and manipulate data [256, 257]. As such, it was leveraged to provide a means for stakeholders to efficiently analyze and assess daily airport operations.

6.1.2 Step #2 : Deployment of software and scripts

The overaching objective of this work is to develop a framework to facilitate the analysis and assessment of daily airport operations. It was observed from Experiments 1 and 2 discussed in Chapters 4 and 5, respectively, that unsupervised and supervised Machine Learning techniques can be leveraged to analyze and assess daily airport operations. As such, there is a need to incorporate the classification models developed in Chapter 5 into the framework.

As previously discussed, Ansible scripts automate the deployment of system configurations, software, codes, files, etc. into Big Data platforms, and replaces the manual and time consuming effort required to transfer and/or install software, codes, etc. Consequently, Ansible scripts were written to deploy Python 3.7 and libraries needed for the Machine Learning models, the Machine Learning scripts, and the data from Experiment 2 which is needed to train the models. Two Postgres tables were also created and deployed - one to store the data needed for the prediction models and the other to store the outcomes of the prediction models. Cron jobs were then created and deployed to schedule the execution of the Machine Learning scripts. Each of these were deployed with Ansible Tower [258] into the FAA's Computing Analytics and Shared Services Integrated Environment (CASSIE). NiFi is currently leveraged by CASSIE. As such, there was no need to deploy it. The next four steps of this experiment were implemented using NiFi.

6.1.3 Step #3: Initiate Framework

FAA Analysts and researchers manually extract and process the previous day's metrics each weekday morning. The next step of Experiment 3 thus involves enabling the framework to be automatically initiated or triggered each weekday morning to extract and process the data needed from ASPM's database. This was achieved by leveraging the seven processors shown in Figure 6.4.

Generatefiles processors create files with random data or custom content at a predefined time [259]. The **Generatefiles** processor labelled "Test Connection" serves as an initial trigger of the framework at 7:30 AM EST each weekday, as ASPM is updated with the previous day's metrics by 7AM EST each weekday morning. The initial trigger is used to test the framework's connection to ASPM each weekday morning, prior to the extraction and processing of data. The **Generatefiles** processor labelled "24 hours (Mon-Fri)" is used to generate a file comprised of the previous day's date, at 8AM EST each weekday. Two additional **Generatefiles** processors, labelled "Previous 48

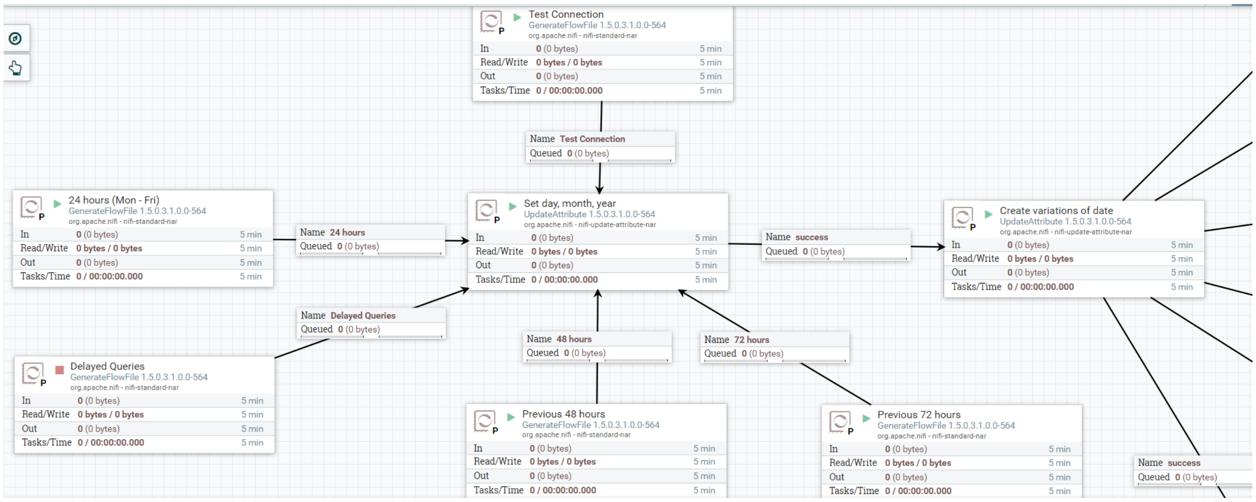

Figure 6.4: Generatefiles and UpdateAttribute Processors

hours" and "Previous 72 hours" are scheduled to generate files containing the dates of the previous Friday and Saturday at 8:00AM EST every Monday, as ASPM is updated with the prior weekend's data each Monday morning. The **Generatefiles** processor labelled "Delayed Queries" is used to trigger the initiation of the framework whenever there is a need to extract and process data from a specific date.

The generated files containing dates serve as inputs to SQL queries needed to extract metrics from ASPM, and replaces the need for analysts to manually enter the desired day's date into the SQL query of each table in ASPM's database. The desired day's date generated by the **Generatefiles** processor is of the format "YYYY-MM-DD", where YYYY, MM, and DD correspond to the year, month, and day, respectively. The various tables in the database require the date to be queried in varying formats. As such, the two **UpdateAttribute** processors [260] labelled, "Set day, month, years" and "Create variations of date", as seen in Figure 6.4, are used to transform the predefined date into parameters needed for the various queries. A summary of these parameters, as well as their corresponding time-related variables are provided in Table 6.1.

Table 6.1: Time-related parameters for querying of data

Parameter Formats	Context
DD	Day
MM	Month
YYYY	Year
YYYYMM	YearMonth
YYYYMMDD	YearMonthDay
MM/DD/YYYY	Month/Day/Year

6.1.4 Step #4: Extract data

The next step involves querying the metrics needed for the categorization of daily airport operations. The metrics are stored in different tables in ASPM's database. The files comprised of the time-related parameters in Table 6.1 are passed into specific **ExecuteSQL** processors [261], one for each table in the database. These parameters are dynamically inserted into SQL queries needed to extract data from the tables. A majority of the metrics needed for the categorization of daily airport operations are directly extracted from the database. However, completion rate, number of Ground Stops, and Ground Delay Program lead-in time and number of revisions are computed using data extracted from ASPM. Completion rate is computed using the number of arrivals and cancellations at an airport, while the number of Ground Stops, and Ground Delay Program lead-in time and number of revisions are computed using Ground Stop and Ground Delay Program data, respectively. Table 6.2 provides a summary of metrics to be extracted from ASPM. It also shows that certain metrics are extracted from the same tables.

Table 6.2: Summary of metrics and their SQL Tables

Metric	**Table**
Arrival count	1
Diversions	2
TMI to airport delays	3
Departure delays	3
Airborne Holding events	3
Airborne Holding minutes	3
Cancellation count	4
Ground Stops	5
Ground Delay Programs	6

This step is automated as the **ExecuteSQL** processors are triggered by the incoming files comprised of the predefined date. The **ExecuteSQL** processors are connected to ASPM's database by using a Database Connection Pooling Service [262] such as DBCPConnectionPool [263, 264]. The DBCPConnectionPool is configured with the database's connection url, driver class name and location(s), username, and password. The queried data is extracted in avro format [265], which uses json [266] objects to define data types and protocols. Avro formats also serialize data in a compact binary format, which ensures the rapid ingestion of large amounts of data for processing. Each **ExecuteSQL** processor is also configured with a feedback loop, as seen in Figure 6.5 to ensure that failed queries are re-executed. This will enable analysts to keep track of failed queries and to identify any issues pertaining to the database such as missing or inconsistent data, wrong query statements, etc. It is also important to note that the duration of each query differs, which may impact the order that metrics are compiled by airport for analysis. Consequently, an **UpdateAttribute** processors is used to tag the output of each query with a priority number corresponding to their table number in Table 6.2, to ensure a consistent compilation of data by airport, later on in the data flow.

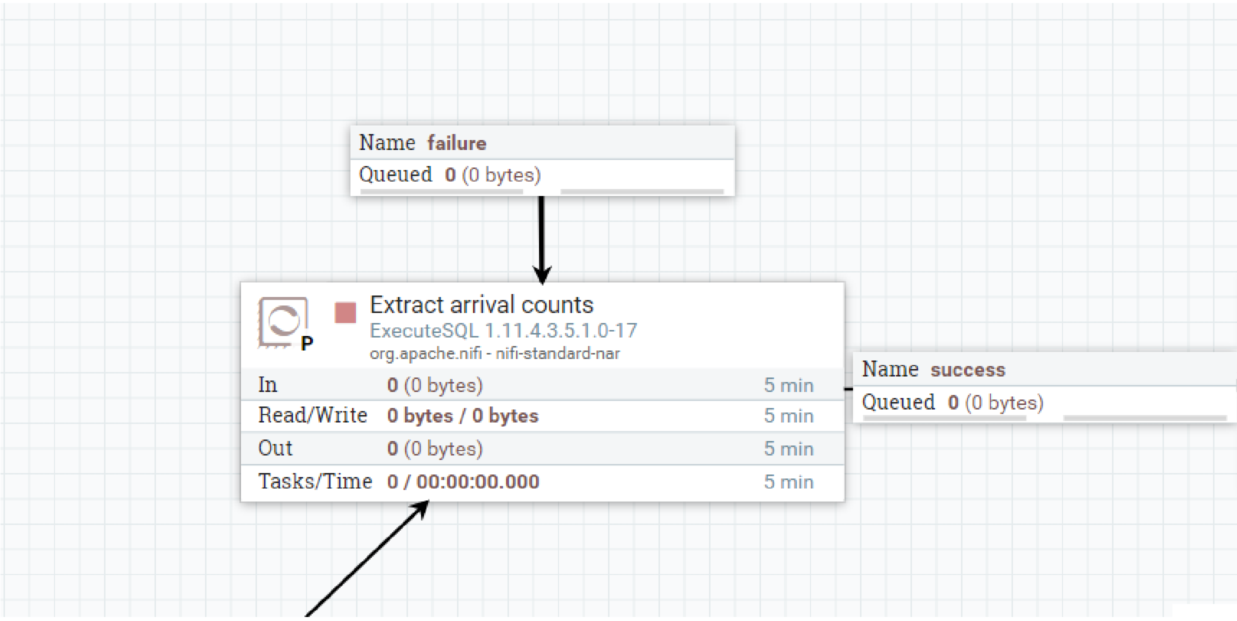

Figure 6.5: ExecuteSQL Processor

6.1.5 Step #5: Process data

The next step in the methodology focuses on processing the extracted data using the following process:

Split files

Each query outputs a single file composed of specific metrics for the eight airports. However, there is a need to generate individual records of each metric per airport to facilitate their categorization. The **SplitAvro** processor [267], shown in Figure 6.6 is thus used to split up the binary encoded files into individual records. As seen in Figure 6.6, **LogMessage** processors are connected to the **SplitAvro** processors to allow analysts to identify files that are not split due to corrupted content, inconsistent formats, etc. The processors also provides details of the cause(s) of failure.

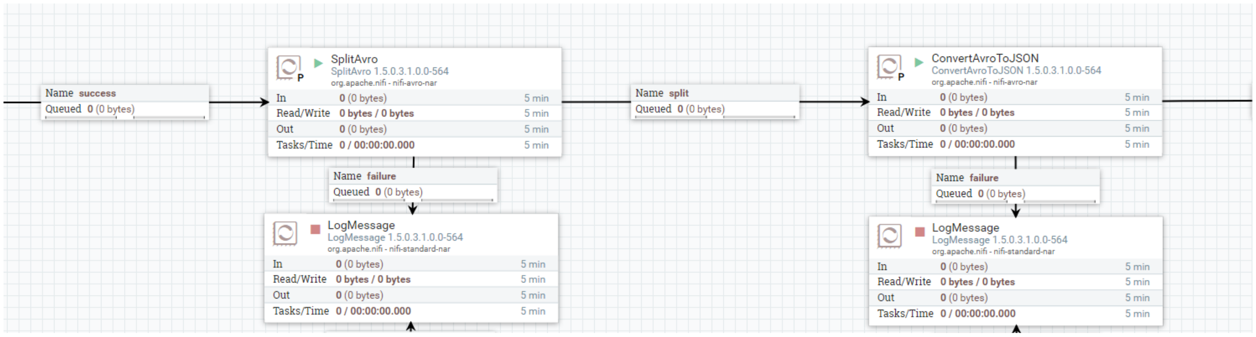

Figure 6.6: SplitAvro and ConvertAvroToJSON Processors

Convert files

As previously mentioned, the metrics are ingested from ASPM in avro format. Avro is appropriate for data transfer and storage purposes as data is compressed into binary records. There is thus a need for the ingested data to be converted into a format suitable for data analytics purposes. The **ConvertAvroToJSON** processor, shown in Figure 6.6 is thus used to convert the files from avro to json. This is achieved by using the schema of the avro files to map each avro field to a json field. This ensures that the resulting json files have the same hierarchical structure as the Avro file. **LogMessage** processors are used to identify files that are not successfully converted, as well as the cause(s) of failure, as seen in Figure 6.6.

Fuse metrics by airport

The next step focuses on fusing the json files comprised of metrics by airport. This is achieved by using the **EvaluateJsonPath**, **EnforceOrder**, and **MergeContent** processors. Six **EvaluateJsonPath** Processors [268], one for each APSM table and labelled “Tag file with airport name”, as seen in Figure 6.7, are used to tag each json file with the name of the airport that the data belongs to. This is achieved by evaluating the contents of incoming json files and extracting the json key corresponding to the name of the airport. Incoming files that do not contain the specified json key are routed to **LogMessage** processors as *unmatched*, as seen in Figure 6.7. This enables analysts to identify any files that are not tagged correctly, as they will hinder the eventual compilation of metrics by airport. **LogMessage** processors are also used to identify any invalid json files, which are routed as *failures*.

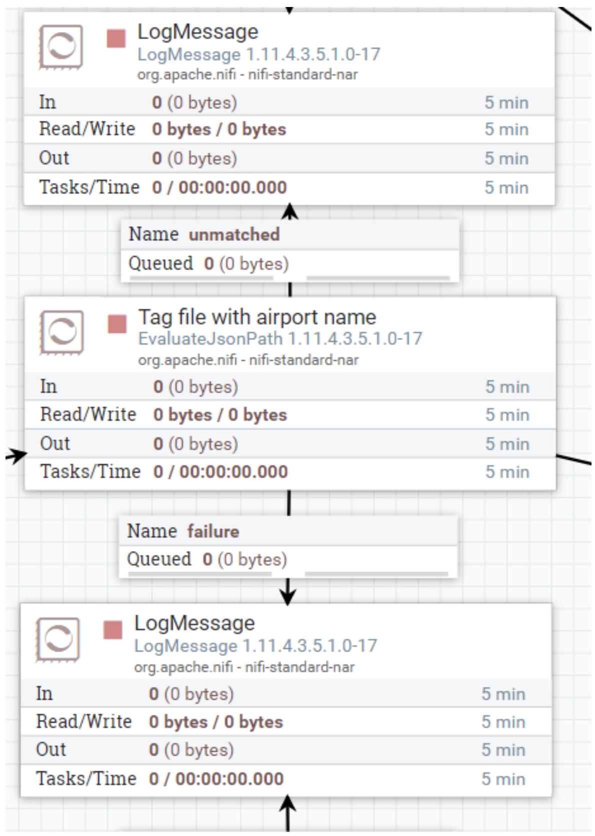

Figure 6.7: EvaluateJsonPath Processors

Successfully tagged files are then routed to the **EnforceOrder** processor which is used to ensure that incoming files containing various metrics are merged/compiled by airport in a consistent order. This processor facilitates the consistent merging of metrics by ensuring that metrics for each airport are ordered according to their previously assigned priority number. This is particularly important as the compilation of metrics by airport may be impacted by errors, branching, and other flow designs. The **EnforceOrder** processor keeps track of each airport's current priority number as it processes the files, starting from one to six, which is the number of ASPM tables used for this work and thus, the predefined maximum priority number. As such, any files containing metrics that arrive late and have a priority number lower than an airport's current number are routed as *skipped*, as seen in Figure 6.8. These files are routed back into the flow and placed in order once the maximum priority number is reached. Any files that arrive early with a priority number higher than current, are routed to *wait*. These files are routed back into the flow once their priority number is reached. Files with higher priority numbers may be routed to wait indefinitely when files with lower priority numbers are not processed due to prior failures. A wait timeout is thus set to route waiting files to *overtook*, which are then routed back to the flow, as seen in Figure 6.8. Any files without an airport tag or priority number are routed to **failure**, while successfully ordered files are routed to the **MergeContent** processor, as seen in Figure 6.8. The **EnforceOrder** processor is also configured to reset the priority number to one after 40 minutes to ensure the consistent merging of files the next time the framework is initiated. The **MergeContent** processor then merges the metrics and packages them into individual files for each airport.

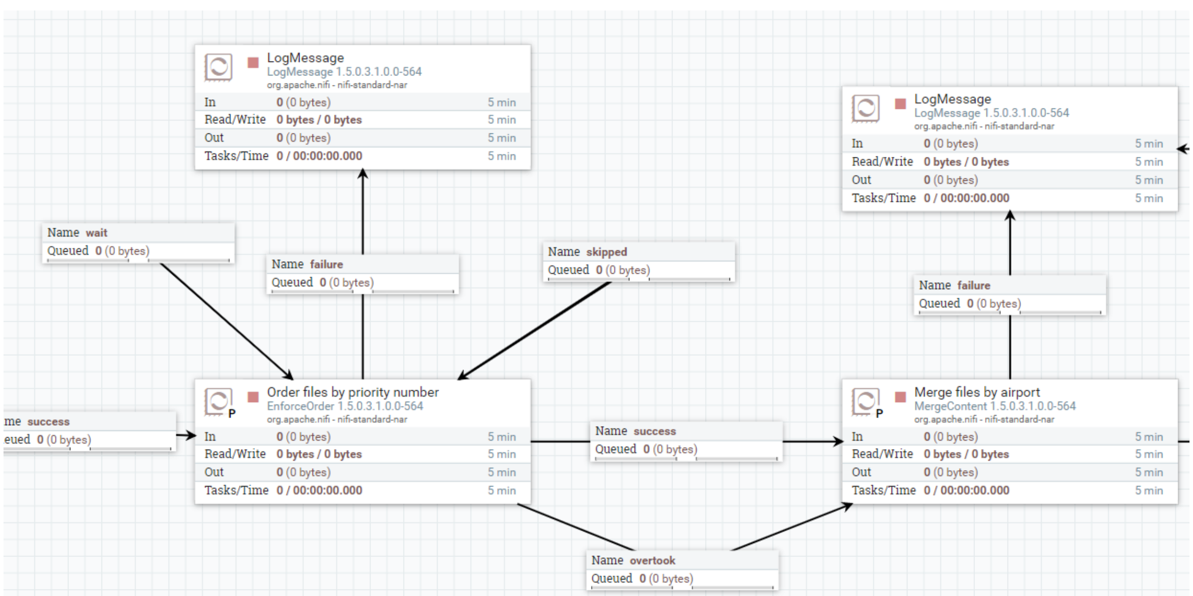

Figure 6.8: EnforceOrder and MergeContent files Processors

Ensure consistency of data

The next step of the proposed methodology focuses on ensuring the consistency of data. Diversions, Cancellations, Ground Stops, and Ground Delay Programs are extracted as null values whenever an airport is not characterized by any of these events, unlike the other metrics which output zero in the case of no event. There is thus a need to convert the null values to zero to facilitate the categorization of daily airport operations. This is achieved by using a **RouteOnAttribute** processor [269] to identify the number of metrics with null values for each airport in order for them to be updated to zero. The processor is used to route each file to different processors based on how many null values are present, as seen in Figure 6.9, where the name of the connection corresponds to the number of metrics with null values. **EvaluateJsonPath** processors are then used to identify metrics with null values which are replaced with zero using **UpdateAttribute** processors.

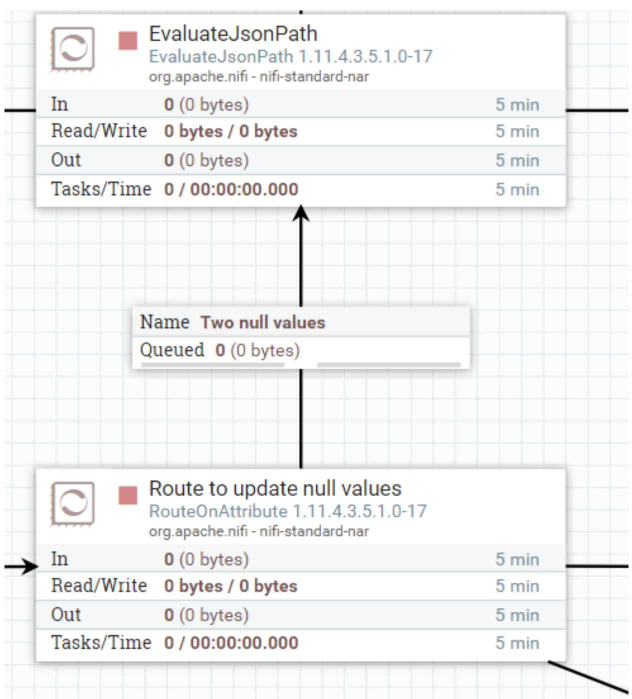

Figure 6.9: RouteOnAttribute Processor

6.1.6 Step #6: Temporarily store data

The next step involves temporarily storing the extracted metrics into Postgres tables. The data in these tables serve as inputs to the Machine Learning models used to categorize daily airport operations. Json files containing the metrics are stored in the Postgres tables using **PostgreSQL** processors, as seen in Figure 6.10. It also shows that a feedback loop is used to ensure that additional attempts are made to insert the data into the table, if needed. The framework is also configured to notify analysts via email after retry attempts have failed due to invalid queries or integrity constraint violations. This is achieved by leveraging the **PutEmail** processor, as seen in Figure 6.10, which notifies analysts of the causes and locations of errors in the data flow that hinder the storage of airport data in the Postgres tables.

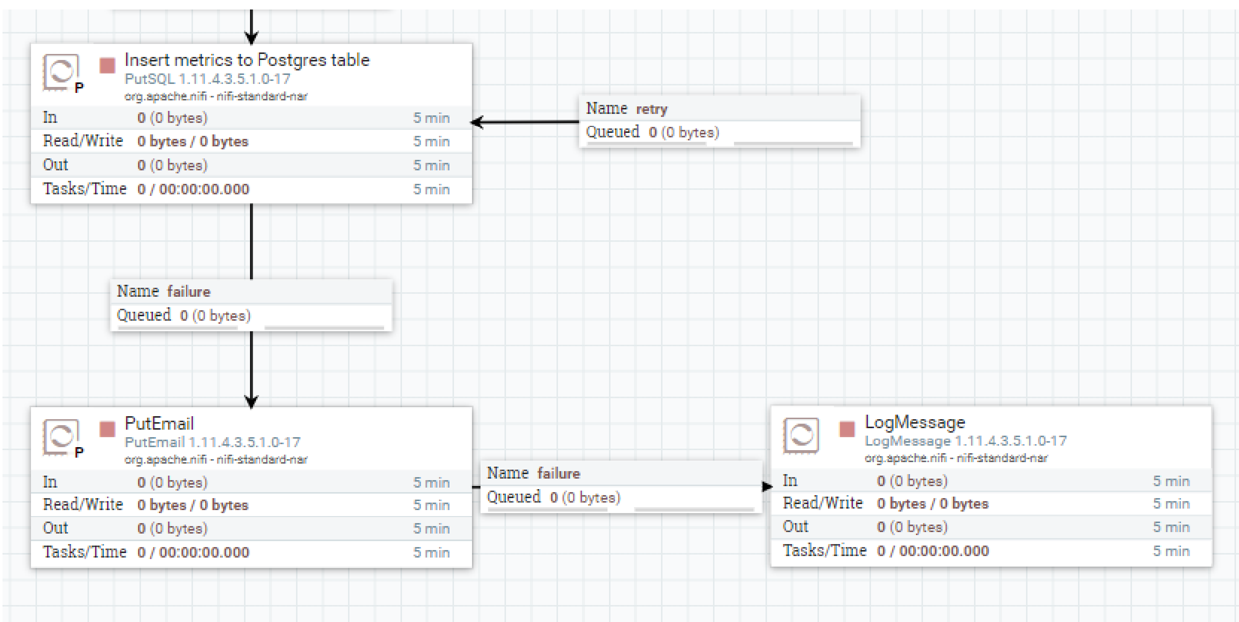

Figure 6.10: PostgreSQL and PutEmail Processors

6.1.7 Step #7: Execute Scripts

The deployed cron jobs are used to schedule the execution of the Machine Learning scripts at 8:45AM, each workday to categorize airport operations from the previous day. Cron jobs are also scheduled to execute the Machine Learning scripts at 9AM and 10AM each Monday to categorize airport operations from the previous Friday and Saturday, respectively. The Machine Learning scripts extract data from the temporary Postgres tables, compute the Completion Rate, number of Ground Stops, and Ground Delay Program revisions and lead-in time for each airport. The computed metrics as well as the other extracted metrics are then used as inputs of the prediction models. The contents of the temporary Postgres tables are deleted after the successful execution of the scripts. The category that each daily airport operation belongs to, posterior probability or degree of support of predictions, and the computed and extracted metrics are then stored in another Postgres table, permanently. Finally, cron jobs are also scheduled to retrain the models on the first day of each month so as to learn from the additional data obtained from the previous month.

6.1.8 Step #8: Permanently store data

The next step involves using Apache Nifi to extract the previous day's data from the Postgres table for storage in S3. This is done to facilitate the use of APIs to query data. This is achieved by configuring and scheduling an **ExecuteSQL** processor to query the previous day's data from the permanent Postgres table an hour after the framework is triggered. The queried data is then stored in S3 using the process shown in Figure 6.11.

The extracted files are stored in S3 in json format using the **PutS3Object** processor [270]. The **GenerateMetadataJson** processor is then used to create metadata for each file comprised of the date, S3 location, and other pertinent information which is stored in Elastic Search with the **PutElasticsearchHttp** processor [271]. The metadata facilitates the querying of data with an API.

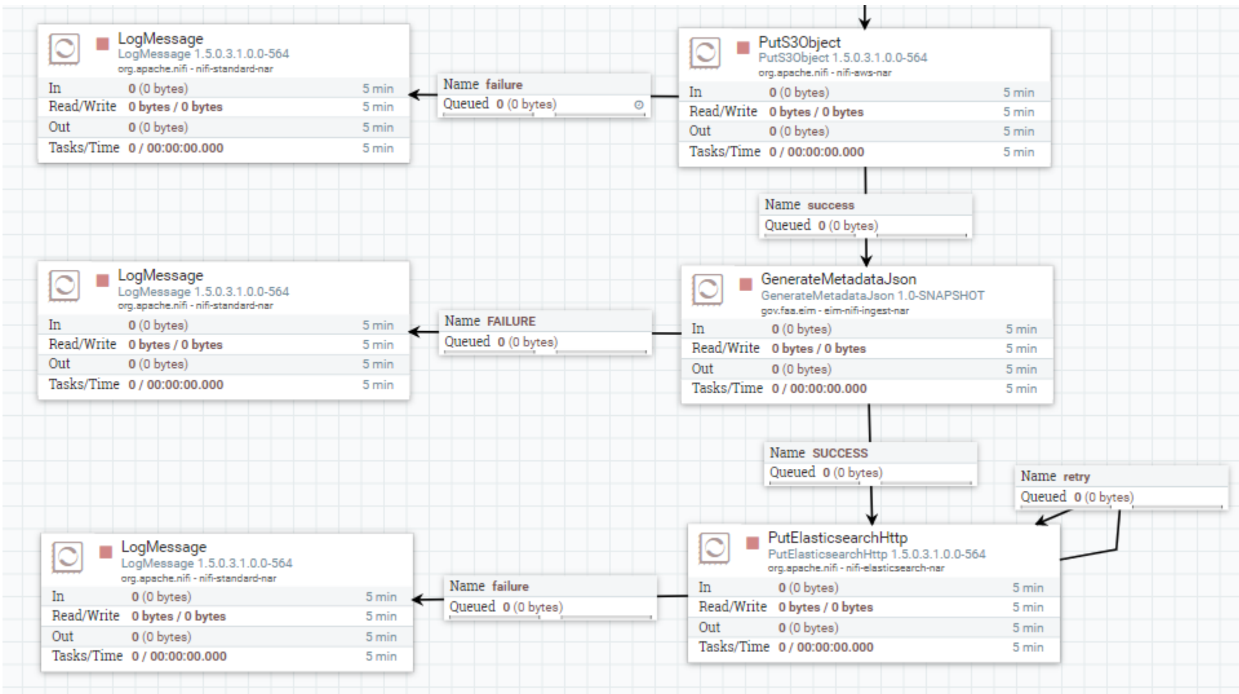

Figure 6.11: PutS3Object, GenerateMetadataJson and PutElasticsearchHttp Processors

6.1.9 Step #9: Analyze data

The final step of this methodology involves providing a means for analysts to analyze the data. This is achieved by leveraging an API developed with MuleSoft [255] to facilitate the querying of data. Credentials and instructions for querying data across different time periods and airports will be provided to analysts to enable them access to the data needed for their work. Interactive dashboards with Tableau [256, 257] were also developed to enable analysts to efficiently analyze daily airport operations. These dashboards are updated by 9:15AM every weekday morning with airport operations data to facilitate their analysis.

6.2 Implementation and Testing of Methodology for Developing a Framework for the Extraction, Processing, and Storage of Aviation Data

The methodology outlined in the previous section was implemented to develop a framework to facilitate the efficient extraction, processing, storage, and analysis of daily airport operations, as seen in Figure 6.12. In particular, it shows the different technologies and software that were leveraged for each component of the framework.

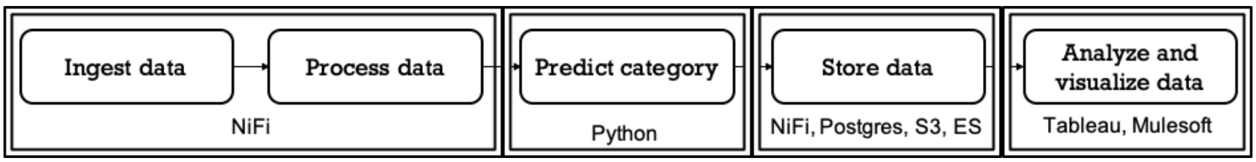

Figure 6.12: High level overview of framework

It was observed from a four-month long monitoring of the framework that its performance was excellent as data was extracted, processed, and stored in a reliable and secure manner, as expected. In addition, the accuracy of the prediction models remained excellent as the models were automatically retrained on the first day of each month using the original training data as well as the data processed from the previous month. Figures 6.13, 6.14, 6.15, and 6.16 of Appendix **??** show dashboards that were developed to facilitate the analysis of daily airport operations.

The dashboard shown in Figure 6.13 enables analysts to review the category that airport operations belong to, the degree to which daily operations belong to each airport category, and airport metrics for the selected day. This is done by selecting a date and one or more airports from the dropdown menus, and hovering over the airport. As observed in Chapter 4, daily airport operations in the first, second, and third clusters are characterized by good, varying, and poor operational performance, respectively. As such, green, yellow, and red were used to indicate good, varying, and poor operational performance of the airports, as seen in Figure 6.13.

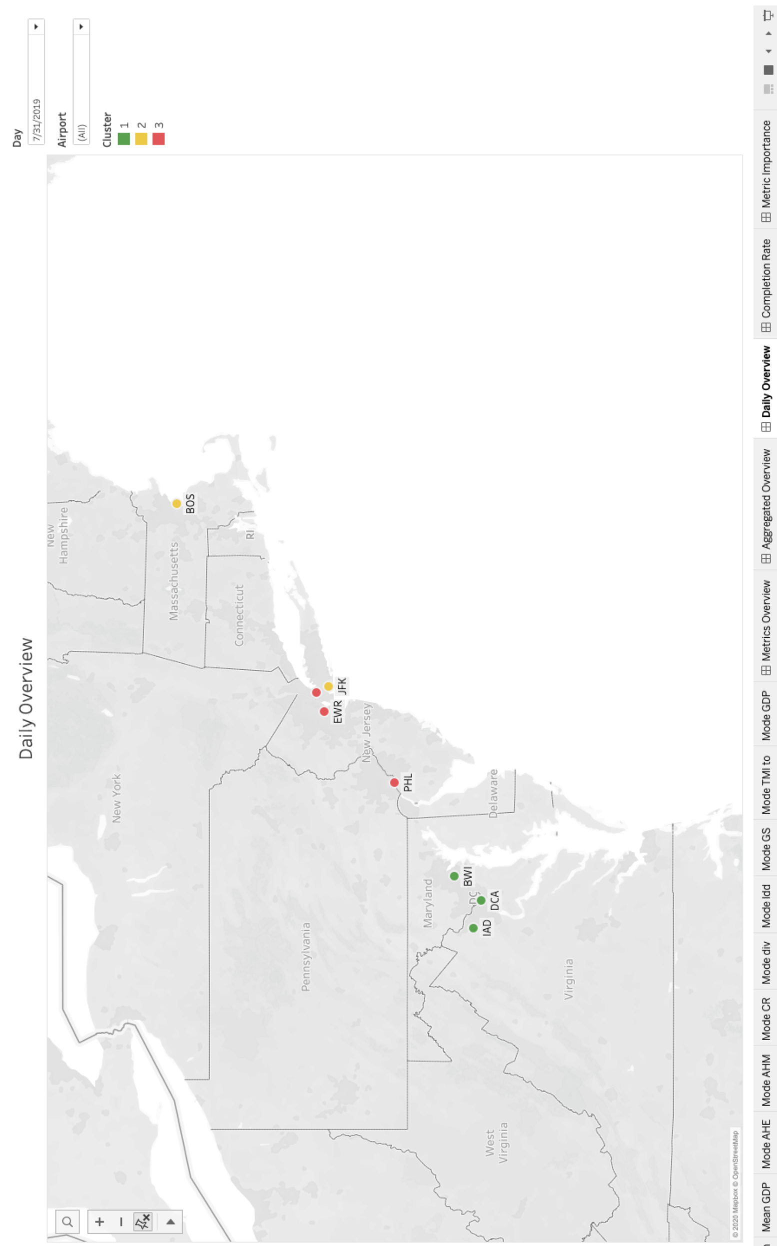

Figure 6.13: Dashboard for reviewing daily airport operations on a specific day

Figure 6.14 shows a dashboard that provides an aggregated overview of daily operations across airport categories. This enables analysts to compare the breakdown of airport operations across clusters and airports. Hovering over an airport also provides a summary of the various metrics over the time period selected in any of the drop down menus.

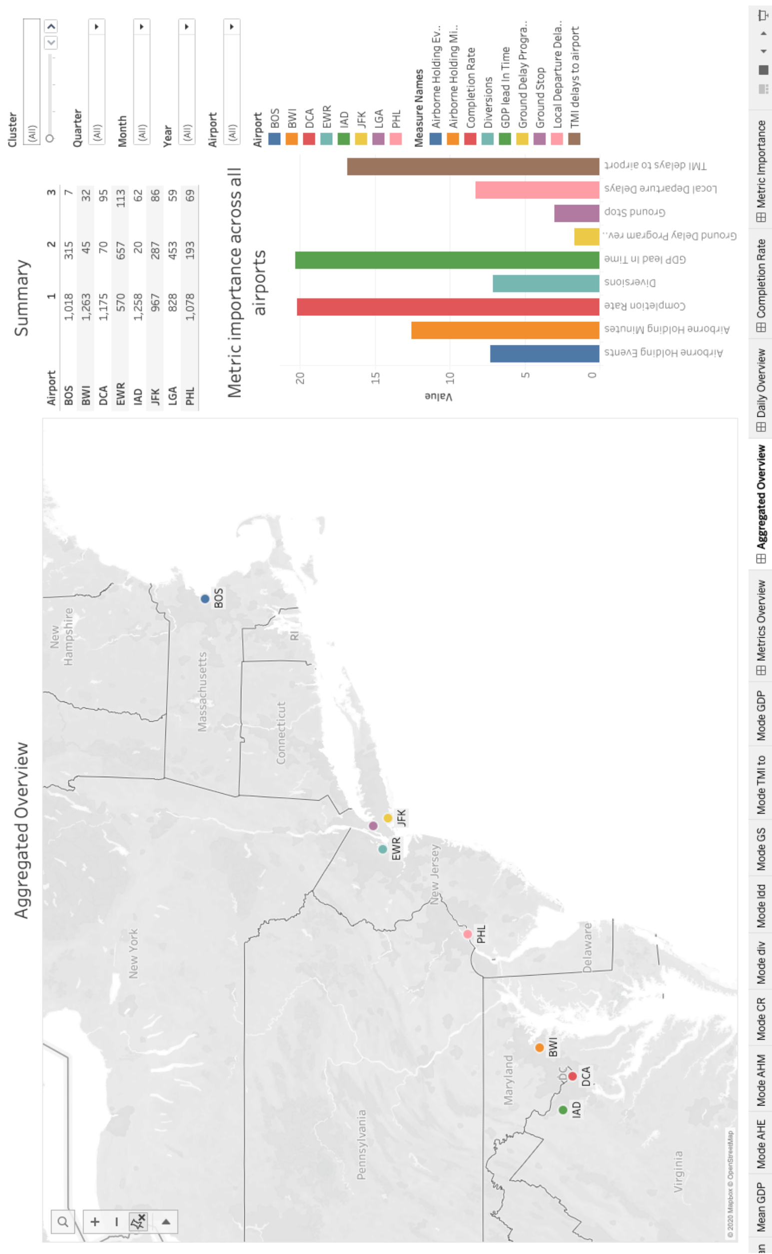

Figure 6.14: Dashboard for reviewing daily airport operations across different time periods

Figure 6.15 enables analysts to analyze the distribution of metrics across the different airport clusters. The distributions can be analyzed across multiple days, seasons, months or years by using their drop down menus. This facilitates the analysis of airport operations for the identification of trends and patterns.

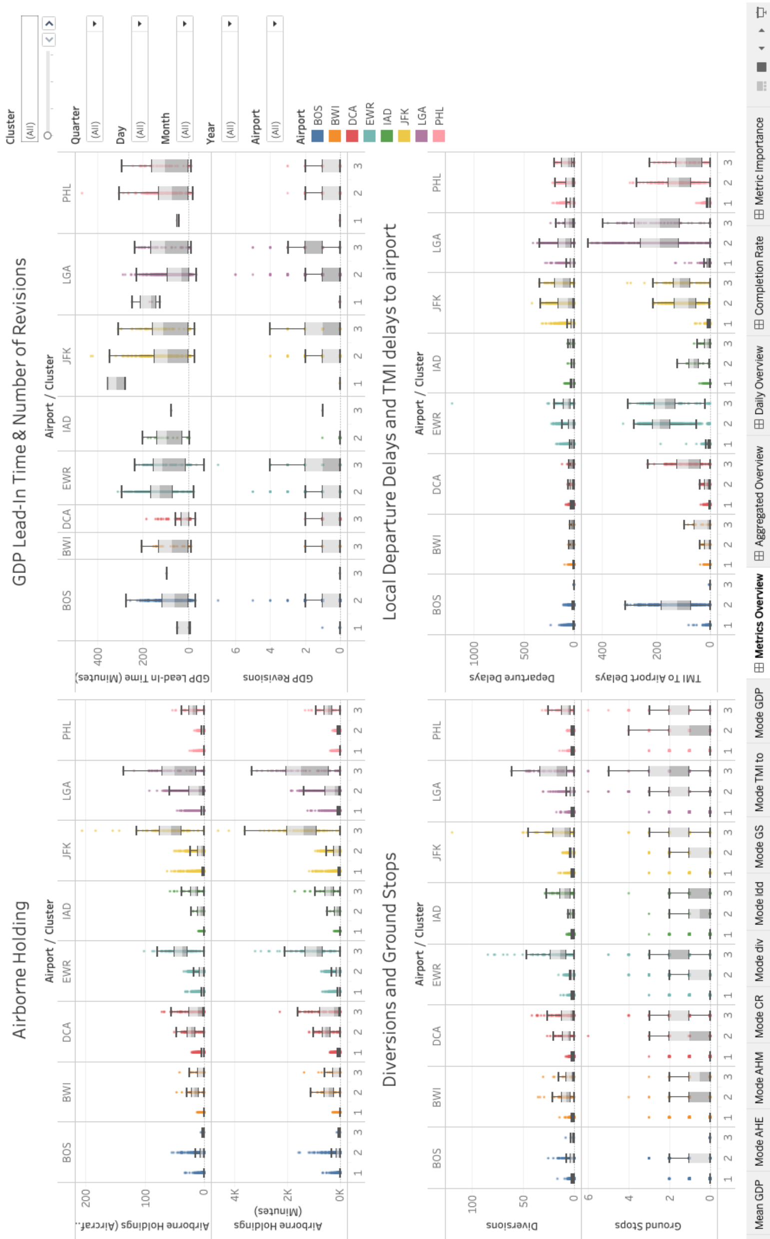

Figure 6.15: Dashboard for analyzing metrics across airport categories

Figure 6.16 shows the ranking of metric airport of each airport obtained in Experiment 4. This interactive dashboard enables analysts to compare how metrics impact daily airport operations.

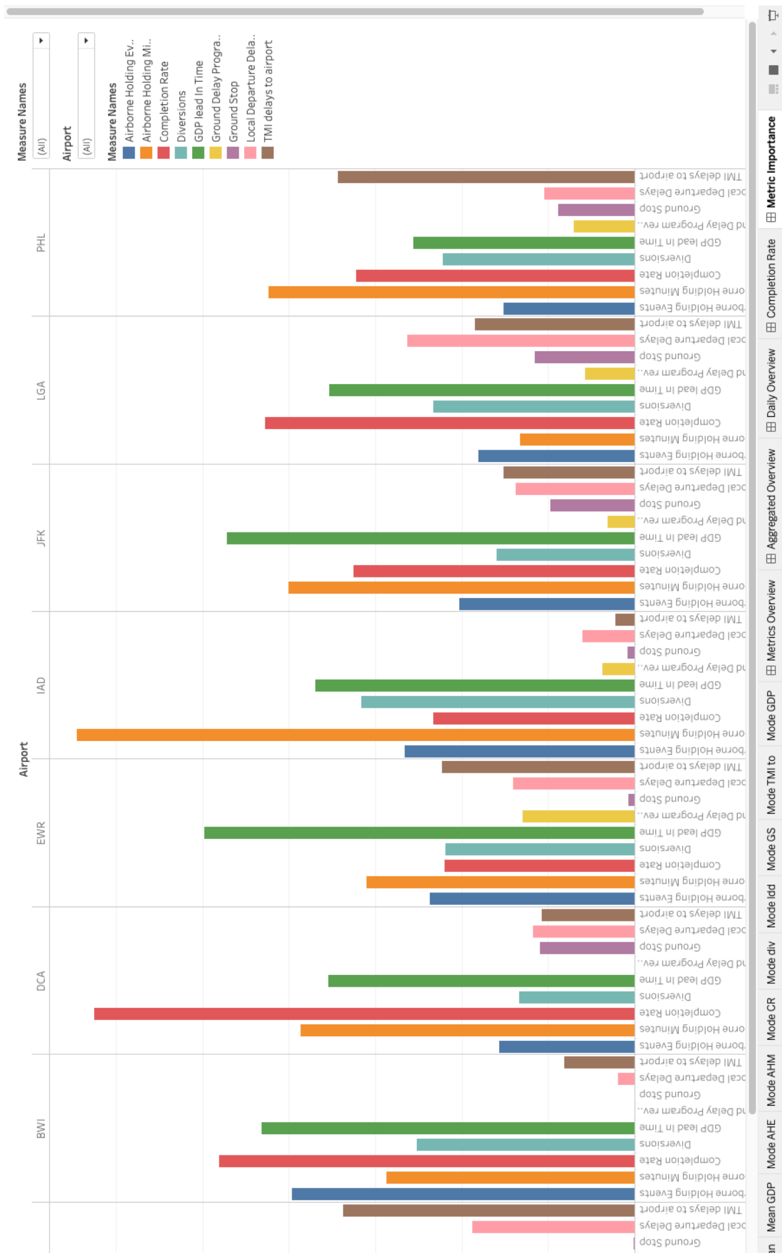

Figure 6.16: Dashboard for analyzing the ranking of metric importance across airports

6.3 Summary of Findings from Experiment 3

Research Question 3 posed in Chapter 3 examines the capability of the methodology discussed herein to develop a framework for the efficient extraction, processing, storage, and analysis of data needed for the analysis and assessment of daily airport operations. As such, Experiment 3 was developed, implemented, and tested by incorporating the outcomes of Experiments 1 and 2, and identifying and leveraging a set of technologies and software that offer a reliable and secure way for handling aviation data. The developed framework automates the flow of data from extraction through storage and enables users to track the flow of data in real time. It also facilitates data provenance by logging the history of all processes, and is equipped with the capability to log errors and their causes, and to notify analysts via email whenever they occur. In addition, it scales well and has the capacity to facilitate the analysis and assessment of the daily operations of all airports in the NAS, if needed. The framework is also equipped with interactive dashboards that are updated by 9:15AM every weekday to enable analysts to efficiently analyze and compare daily airport operations in real time.

The eventual deployment of the framework into the FAA's Enterprise Information Management (EIM) platform will enable FAA analysts and researchers to perform a comprehensive analysis of daily airport operations in real time in a cloud-based environment. Indeed, this framework will be one of the first of its kind to be deployed in the FAA's EIM platform, and will serve as a template for leveraging cloud-based services and Big Data technologies to maintain the safety and improve operations in the National Airspace System. Based upon these findings, it is concluded that the conditions of Hypothesis 3 are satisfied, namely that developing an automated framework will facilitate the efficient analysis and assessment of daily airport operations, from data extraction, through processing, analysis, and storage.

Therefore, the hypothesis for Research Question 3 are verified.

CHAPTER 7

CONCLUSION

Even though tremendous progress has been made to modernize the NAS by way of technological advancements and the introduction of procedures and policies that have maintained the safety of the United States airspace, much more needs to be done to ensure that operations in the NAS, and particularly at airports, are as efficient as possible. Traffic management personnel regularly analyze projected airport demand, forecasted weather conditions, and the statuses of airport systems, equipment, and infrastructure in order to plan daily airport operations. Ideally, the impact and effectiveness of traffic management decisions and procedures on daily airport operations should be analyzed and assessed in an efficient manner, so as to identify trends and patterns, which will inform better decision making and thus, improve airport operational performance. However, it was observed from a survey of the literature that a robust approach to analyze and assess daily airport operations is lacking. This gap prompted the overaching objective of this research:

Develop a framework to facilitate the analysis and assessment of daily airport operations to improve airport operational performance

To meet this objective, it was hypothesized that:

A framework that automates the extraction and processing of airport data, and facilitates the analysis and assessment of daily airport operations in a comprehensive, robust, and repeatable manner will enable stakeholders to identify trends and patterns for better decision making and as a consequence lead to improved airport operational performance

This overarching hypothesis is associated with methodologies that categorize daily airport operations, determine the category that daily airport operations belong to, provide a means for analyzing and assessing daily airport operations, and develop a framework to automate the ingestion, processing, analysis, and storage of airport data. The methodologies were developed by building upon methods and techniques within the literature in order to formulate a set of tools which are tailored for use within the given context. Following their development, each methodology was further examined according to the set of research questions formulated within Chapter 3. A summary of these research questions and their relative relationship to the major components of the overall methodology is provided in Figure 7.1.

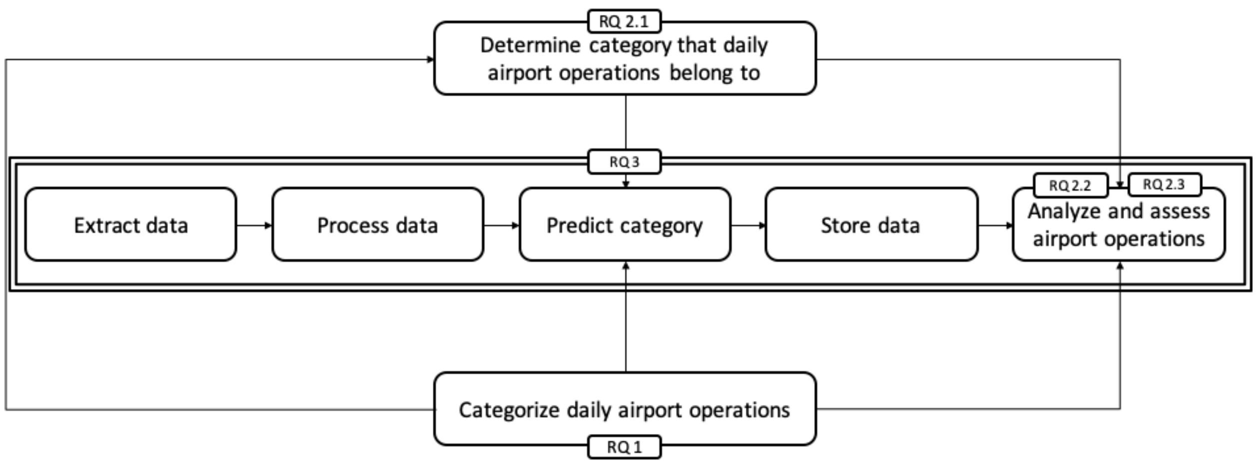

Figure 7.1: Summary of overall methodology with examined research questions

7.0.1 Research Question 1

Research Question 1 was tested through Experiment 1 and is stated as:

Research Question 1

How can daily airport operations be categorized in a systematic, robust and repeatable manner?

This research question examined the capability of the methodology herein discussed to categorize daily airport operations to facilitate their analysis and assessment. While the existing literature outlined various efforts pursued to categorize and analyze airports and their operations, it was observed that a systematic, robust, and repeatable approach is lacking. As such, Experiment 1 was developed, implemented, and tested with data from 8 U.S. airports. This involved extracting and computing the necessary metrics, normalizing the data, reducing the dimensionality and assessing the clustering tendency of the datasets, benchmarking and evaluating the performance of clustering algorithms, identifying the best combination of algorithm and number of clusters, and determining the best suited approach for categorizing daily operations of each airport. The dimensionality of all but two of the datasets was reduced from 9 to 3 using Principal Component Analysis. The dimensionality of the Boston Logan and Dulles International Airport datasets was reduced from 9 to 4. The clustering tendency of the datasets was observed to be very high, as indicated by the high Hopkins Statistic values and presence of multiple boxes along the diagonal of the Visual Assessment of clustering Tendency (VAT) plots of each airport. The best combination of clustering algorithm and number of clusters was also determined to be the Single Linkage Hierarchical algorithm and 3 clusters for each airport by a majority of metrics. The three clusters of each airport were analyzed and observed to either exhibit good, varying, or poor operational performance, as seen in Table 4.35. The number of clusters and their characteristics were consistent with the approach currently employed by the Operational Service Performance Criteria (OSPC), which classifies daily airport operations into “Good days”, “Average days”, and “Bad days”. The outcomes of clustering and OSPC were then compared and reviewed by Subject Matter Experts to determine the best approach for categorizing daily operations at each airport. The comparison revealed that OSPC classified a majority of daily airport operations as "Good days", even though several of them exhibited sub-optimal to poor operational performance due to very low completion rates and Ground Delay Program lead-in times, and high airborne holdings (minutes and number of aircraft). As such, FAA analysts may have to manually validate these classifications each day and/or regularly update the predefined ranges of metrics. The clustering algorithm on the other hand, correctly classified a majority of these as either having varying or poor operational performance. Variations of the means, medians, and modes of the airport clusters also indicates that the same predefined set of ranges of metrics should not be applied across the eight airports, as is currently done in OSPC. This observation in addition to the distributions of metrics across airport clusters shows that the metrics cannot be weighted equally, as is currently assumed with OSPC. Based upon this comparison and the review of the developed clusters by Subject Matter Experts, it was concluded that the conditions of Hypothesis 1 were satisfied, namely that benchmarking clustering algorithms while varying the number of clusters will facilitate the categorization of daily airport operations in a systematic, robust, and repeatable manner. Therefore, the hypothesis for Research Question 1 was verified.

7.0.2 Research Questions 2

The second set of research questions focused on the development of a robust and repeatable methodology for determining the category that daily airport operations belong to. They also focus on developing methodologies to facilitate the comparison of daily airport operations in similar and different categories, and the analysis and assessment of the impact of traffic management decisions on daily airport operations, both of which are currently lacking. Consequently, the second research question is threefold:

Research Question 2.1

How can the category that a daily airport operation belongs to be better determined?

Research Question 2.2

How can daily airport operations in similar and different categories be compared for the identification of trends and patterns?

Research Question 2.3

How can the impact of traffic management personnel actions on airport operations be analyzed and assessed?

Research Question 2.1 posed in Chapter 3 examined the capability of the methodology herein discussed to determining the category that a daily airport operation will belong to, instead of identifying the predominant class of parameters. As such, Experiment 2 was developed, implemented, and tested with metrics from 8 U.S. airports and the Random Forests Machine Learning Algorithm. This experiment involved randomly splitting the data from each airport into training-validation and test datasets. The SMOTE algorithm was then leveraged to reduce the imbalanced nature of the datasets to ensure optimal model performance. Prediction models were then trained while tuning algorithm hyperparameters, and the performance of the models developed with the optimal combinations of hyperparameters were then evaluated with a set of metrics. Based upon the excellent performance of the prediction models, it is observed that the conditions of Hypothesis 2.1 are satisfied, namely that developing prediction models with Machine Learning will provide a robust means for determining the category that daily airport operations will belong to.

The remainder of Experiment 2 involved leveraging the ranking of predictor importance, Decision Trees, and the probability or degree of support of predictions to facilitate the analysis and comparison of daily airport operations. It was observed in Section 5.3 that the conditions for Hypothesis 2.2 are satisfied, namely that daily operations in similar and different categories can be compared by leveraging the posterior probability or degree of support of predictions of the prediction models. It was also observed in Section 5.4 that the conditions for Hypothesis 2.3 are satisfied, namely that the ranking of predictor importance and Decision Trees of prediction models provide a means for analyzing and assessing the impact of traffic management decisions on airport operations. Therefore, the hypotheses for Research Questions 2.1, 2.2, and 2.3 are verified.

7.0.3 Research Questions 3

The final Research Question tested through Experiment 3 is:

Research Question 3

How can the efficient analysis and assessment of daily airport operations be automated from data extraction, through processing, analysis, and storage?

Research Question 3 posed in Chapter 3 examined the capability of the methodology herein discussed to develop a framework for the efficient extraction, processing, storage, and analysis of data needed for the analysis and assessment of daily airport operations. As such, Experiment 3 was developed, implemented, and tested by incorporating the outcomes of Experiments 1 and 2, and identifying and leveraging a set of technologies and software that offer a reliable and secure way for handling Big Data. The developed framework automates the flow of data from extraction through storage, and enables users to track the flow of data in real time, and facilitates data provenance by logging the history of all processes. The framework is also equipped with the capability to log errors and their causes, and to notify analysts via email whenever they occur. In addition, it scales well and has the capacity to facilitate the analysis and assessment of the daily operations of all airports in the NAS, if needed. The framework is also equipped with interactive dashboards that allow analysts to efficiently analyze and compare daily airport operations in real time. The dashboards are updated by 9:15AM every weekday, which prevents analysts from spending a lot of their their time manually extracting and processing the data needed. It was observed from a two-month long monitoring of the framework that its performance was excellent as data was extracted, processed, and stored in a reliable and secure manner, as expected. Based upon these findings, it is observed that the conditions of Hypothesis 3 are satisfied, namely that developing a framework will automate the analysis and assessment of daily airport operations, from data extraction, through processing, analysis, and storage.

Based upon the verification of the hypotheses of the Research Questions, it is concluded that the conditions of the overaching Hypothesis are satisfied, namely that developing a framework that automates the extraction and processing of airport data, and facilitates the analysis and assessment of daily airport operations in a comprehensive, robust, and repeatable manner enables stakeholders to identify trends and patterns for better decision making and consequently leads to improved airport operational performance.

7.1 Contributions

During the course of this work, several contributions have been made which ware discussed below. The first contribution of this work is the development and testing of a methodology for categorizing daily airport operations airport as a means to facilitate their analysis and assessment. The developed methodology leveraged clustering algorithms to categorize the daily operations of eight U.S. airports, instead of using predefined ranges of metrics as is currently done by FAA analysts and researchers. The outcomes of this methodology also highlighted the need to benchmark the performance of clustering algorithms while varying the number of clusters in order to identify the optimal combination of algorithm(s) and number of clusters.

The second contribution of this work is the development of a methodology for determining the category that subsequent daily airport operations belong to, instead of identifying the predominant class of metrics, as is currently done by FAA analysts. Indeed, the excellent performance of the prediction models for each airport indicates that this methodology serves as a robust approach for determining the category that daily airport operations belong to. The repeatable nature of this methodology will enable FAA analysts to expand the scope of this work to include additional airports in the NAS.

Another contribution of this work is the development of methodologies to facilitate the comparison of daily airport operations, and the analysis and assessment of the impact of traffic management decisions on daily airport operations which are currently lacking. These will enable FAA analysts and traffic management personnel to make better decisions to ensure safe and efficient airport operations.

The final contribution of this work is the development of a framework for the efficient extraction, processing, storage, and analysis of data needed for the analysis and assessment of daily airport operations. The developed framework automates the flow of data from extraction through storage and enables users to track the flow of data in real time. It also facilitates data provenance by logging the history of all processes and is equipped with the capability to log errors and their causes, and to notify analysts via email whenever they occur. In addition, it scales well and has the capacity to facilitate the analysis and assessment of the daily operations of all airports in the NAS, if needed. The successful deployment of the framework into the FAA's EIM platform will enable FAA analysts and researchers to perform a comprehensive analysis of daily airport operations in real time in a cloud-based environment. Indeed, this framework will be one of the first of its kind to be deployed in the FAA's EIM platform, and will serve as a template for leveraging cloud-based services and Big Data technologies to maintain the safety and improve operations in the National Airspace System.

7.2 Recommendations for Future Work

This dissertation has outlined how the development of a framework will lead to the efficient extraction, processing, and storage of data needed for the analysis of daily airport operations. However, this present work was carried out using data from 8 U.S. airports. As such, the first recommendation for future work involves expanding the scope of the developed framework to include additional airports of interest to FAA analysts and researchers. Indeed, the scope of this work can be expanded to include all airports in the NAS due to the scalable nature of the framework.

Traffic management personnel use forecasted weather data (Terminal Aerodrome Forecasts), projected airport demand, etc. to plan daily airport operations. The next recommendation for future work thus focuses on fusing airport operations and weather data, and leveraging supervised Machine Learning algorithms to predict the category that future daily airport operations will belong to. Doing so will enable stakeholders to analyze and assess how various possible traffic management decisions may impact airport operations, which will then lead to better decision making.

It was observed from Chapter 5 that the planning and implementation of Traffic Management Initiatives has significant impacts on operations at BOS, DCA, EWR, JFK, LGA, and PHL. As such, the next recommendation for future work focuses on developing a methodology to determine how the time between the announcement and implementation of Traffic Management Initiatives impacts daily airport operations. This can be achieved by leveraging Data Fusion, supervised Machine Learning algorithms, and Partial Dependence Plots (PDP) to assess how Traffic Management Initiative lead-in times impact daily airport operations. The implementation of this methodology will allow traffic management personnel to better plan and implement Traffic Management Initiatives. It will also provide insights into how early Traffic Management Initiatives should be announced prior to their implementation to ensure the efficient operation of airports.

As previously discussed, the framework developed for this research was implemented and tested in the FAA's Computing Analytics and Shared Services Integrated Environment (CASSIE). The final recommendation for future work thus involves successfully deploying and testing the framework in the FAA's Enterprise Information Management platform. Doing so will allow FAA analysts to readily analyze and assess daily airport operations. Furthermore, this platform contains different NAS-related datasets that can be fused and leveraged by FAA analysts to further analyze daily airport operations.

Appendices

APPENDIX A

TABLES

Table A.1: Air Route Traffic Control Centers (ARTCCs) and their locations in the United States [272]

ARTCC	Location
ZAB	Albuquerque
ZAN	Anchorage
ZAU	Chicago
ZBW	Boston
ZDC	Washington
ZDV	Denver
ZFW	Dallas-Fort Worth
ZHN	Honolulu
ZHU	Houston
ZID	Indianapolis
ZJX	Jacksonville
ZKC	Kansas City
ZLA	Los Angeles
ZLC	Salt lake City
ZMA	Miami
ZME	Memphis
ZMP	Minneapolis
ZNY	New York
ZOA	Oakland
ZOB	Cleveland
ZSE	Seattle
ZTL	Atlanta

Table A.2: Machine Learning Algorithms And Their Learning Tasks

Algorithm	Learning Task
Nearest neighbor	Classification
Naive Bayes	Classification
Decision Trees	Classification
Classification Rule Learners	Classification
Linear Regression	Numeric Prediction
Regression Trees	Numeric Prediction
Model Trees	Numeric Prediction
Neural Networks	Dual use
Support Vector Machines	Dual use
Association Rules	Pattern detection
K-means	Clustering
Random forests	Dual use
Bagging Ensemble	Dual use
Boosting Ensemble	Dual use
Divisive Analysis (DIANA)	Clustering
Self Organizing Tree Algorithm (SOTA)	Clustering
Complete Linkage	Clustering
Average Linkage	Clustering
Single Linkage	Clustering
Centroid Linkage	Clustering
Ward	Clustering
Partitioning Around Medoids (PAM)	Clustering
Clustering for Large Applications (CLARA)	Clustering
Model-based	Clustering

APPENDIX B

BENCHMARKING AND EVALUATION OF CLUSTERING ALGORITHMS

B.1 Baltimore/Washington International Thurgood Marshall Airport (BWI)

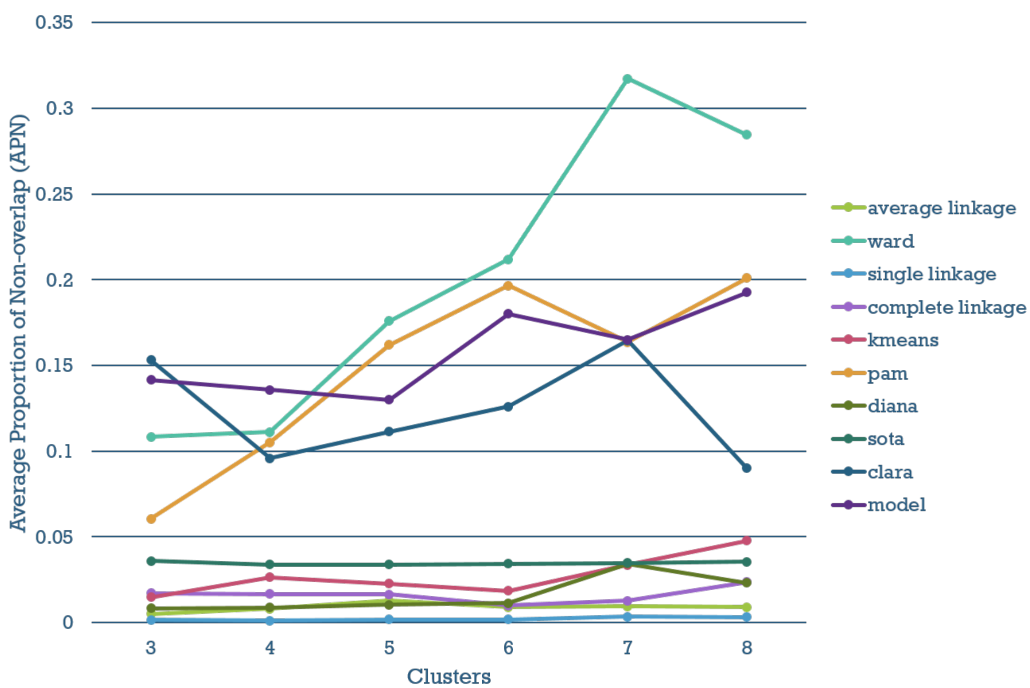

Figure B.1: Average Proportion of Non-overlap (APN) for Baltimore/Washington International Thurgood Marshall Airport (BWI)

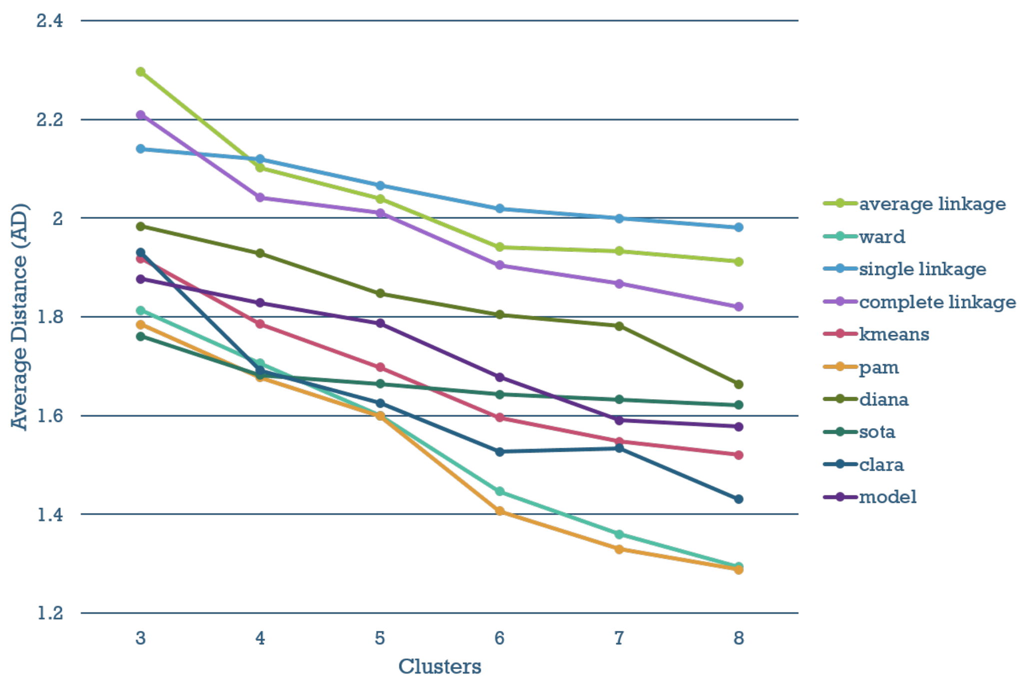

Figure B.2: Average Distance (AD) for Baltimore/Washington International Thurgood Marshall Airport (BWI)

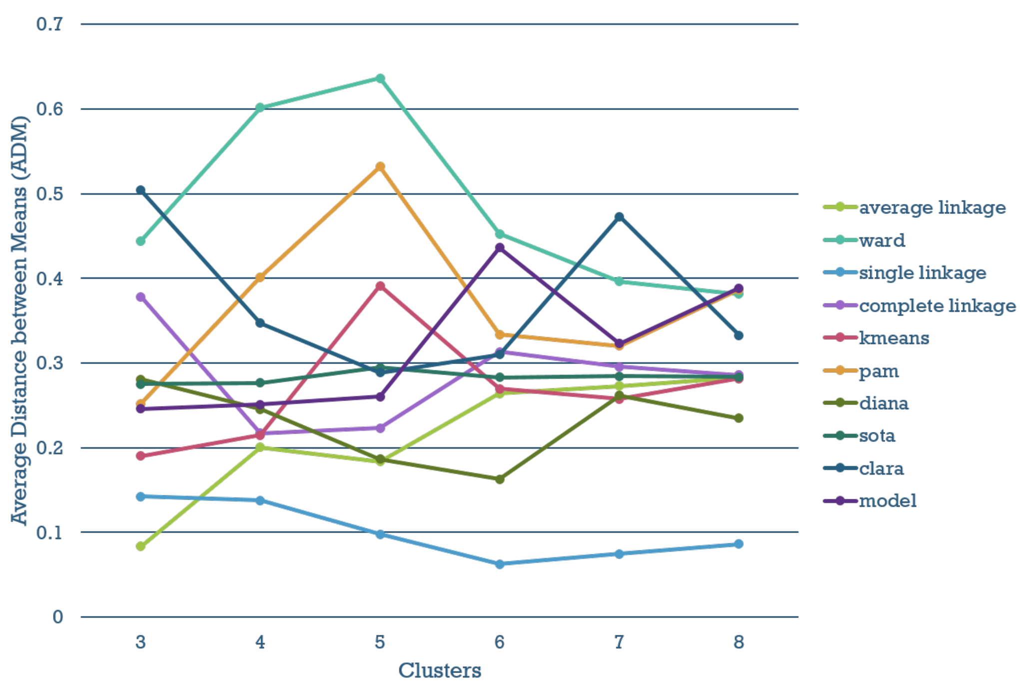

Figure B.3: Average Distance between Means (ADM) for Baltimore/Washington International Thurgood Marshall Airport (BWI)

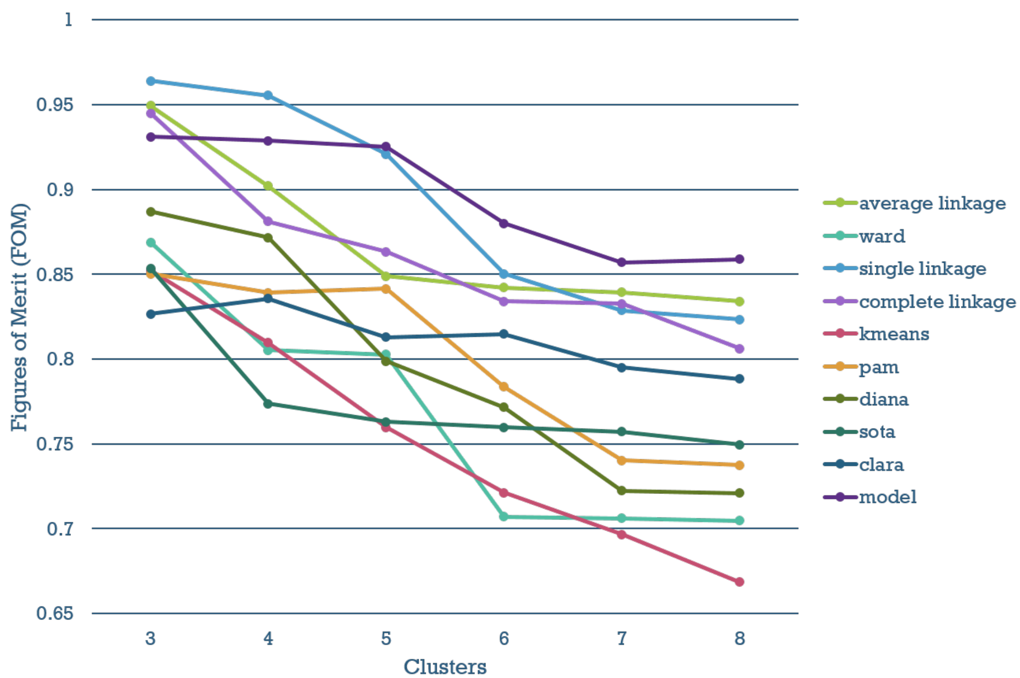

Figure B.4: Figures of Merit (FOM) for Baltimore/Washington International Thurgood Marshall Airport (BWI)

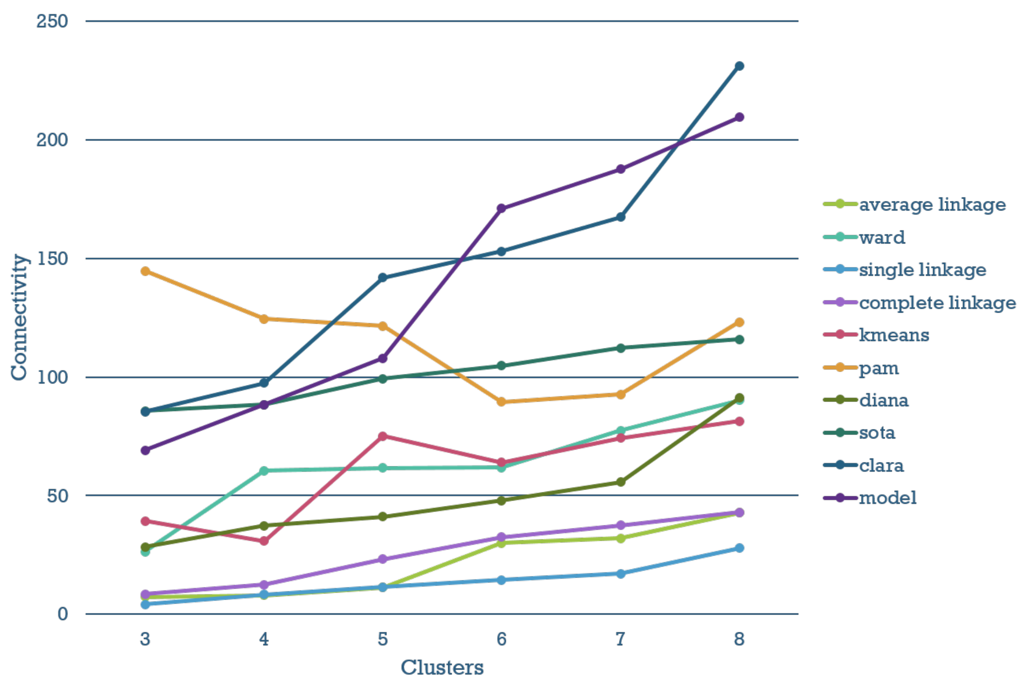

Figure B.5: Connectivity for Baltimore/Washington International Thurgood Marshall Airport (BWI)

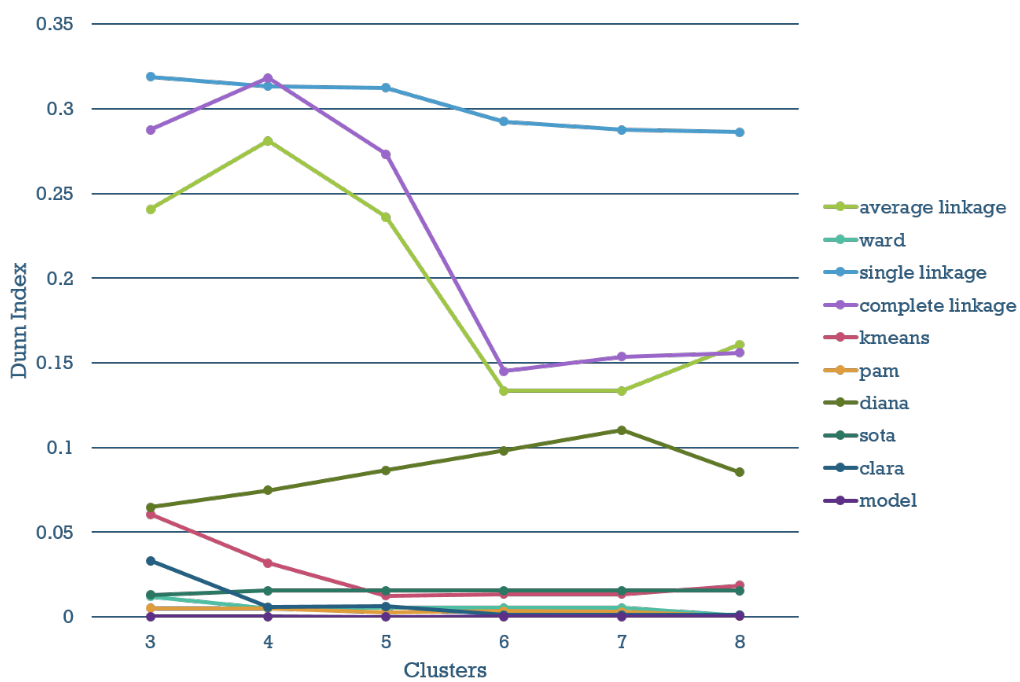

Figure B.6: Dunn Index for Baltimore/Washington International Thurgood Marshall Airport (BWI)

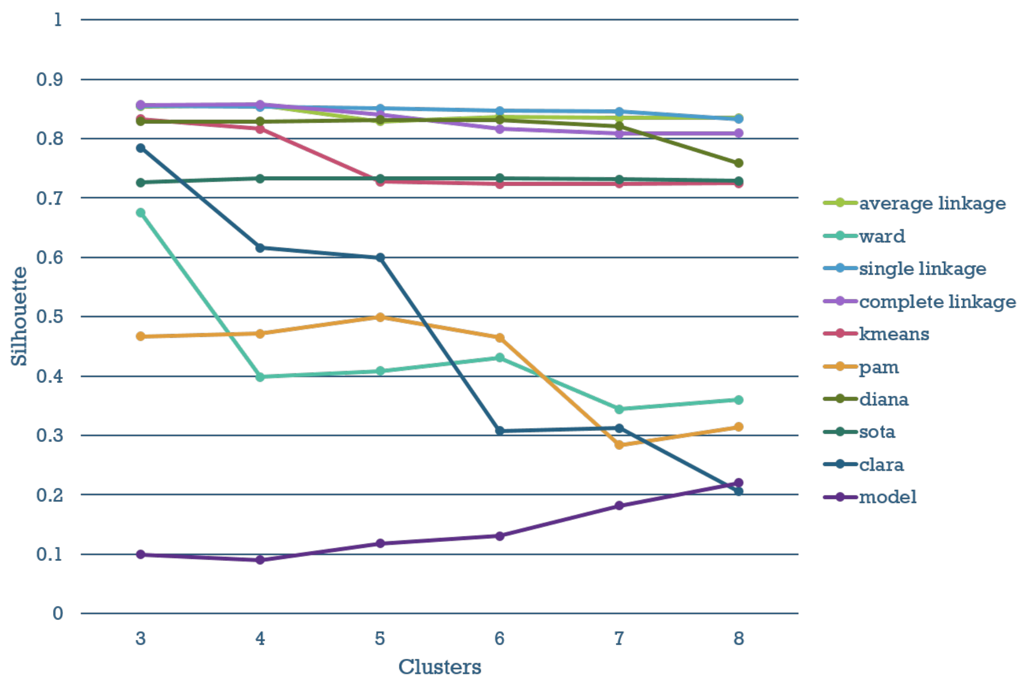

Figure B.7: Silhouette for Baltimore/Washington International Thurgood Marshall Airport (BWI)

B.2 Reagan National Airport (DCA)

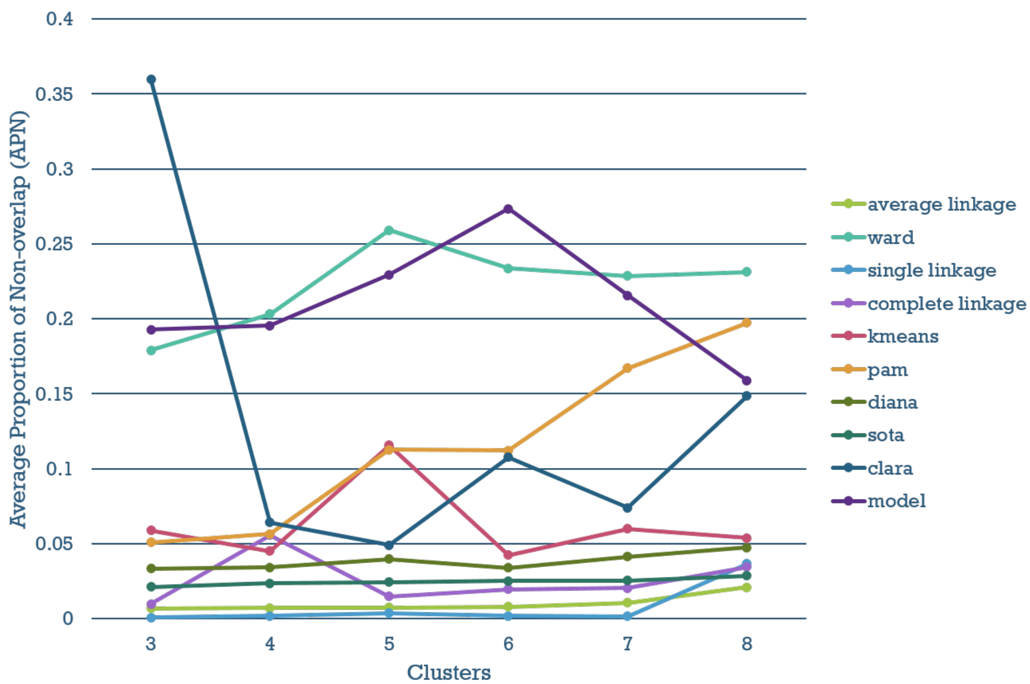

Figure B.8: Average Proportion of Non-overlap (APN) for Reagan National Airport (DCA)

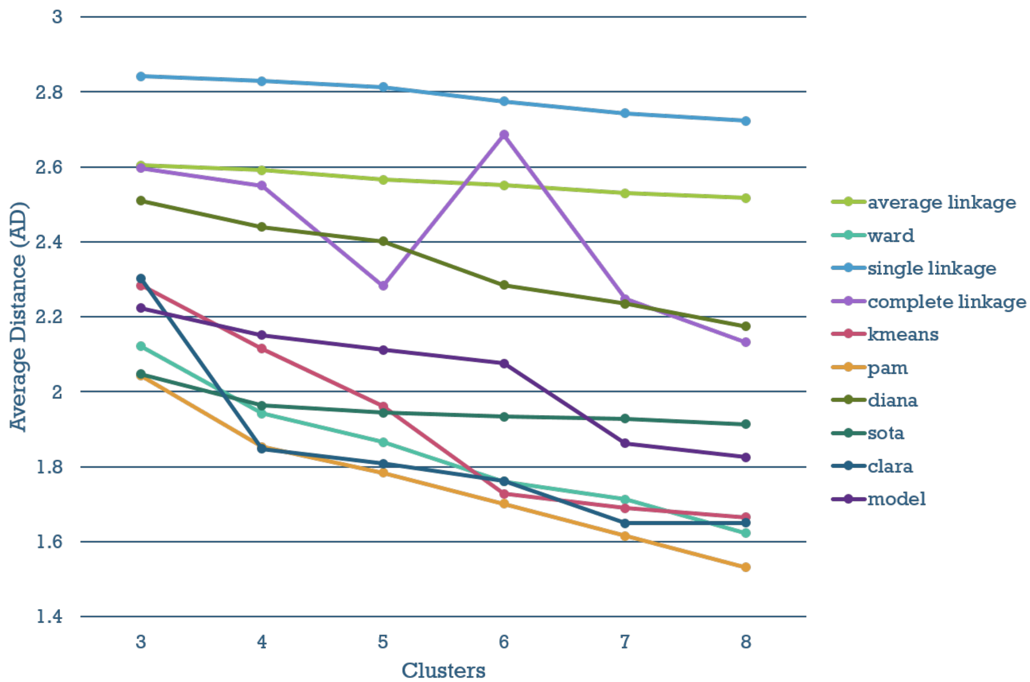

Figure B.9: Average Distance (AD) for Reagan National Airport (DCA)

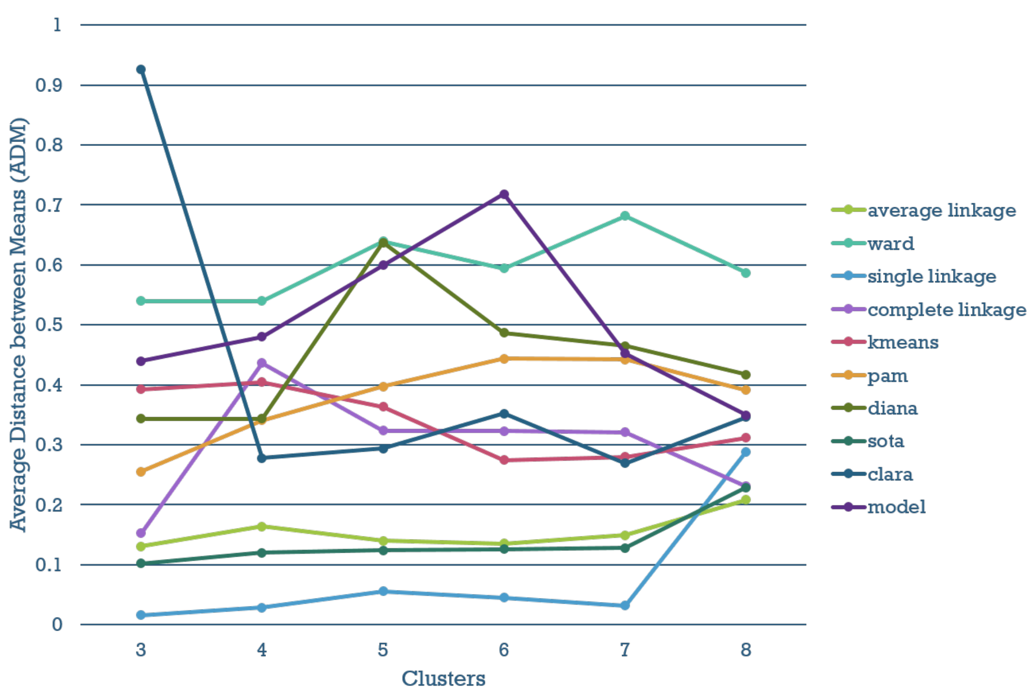

Figure B.10: Average Distance between Means (ADM) for Reagan National Airport (DCA)

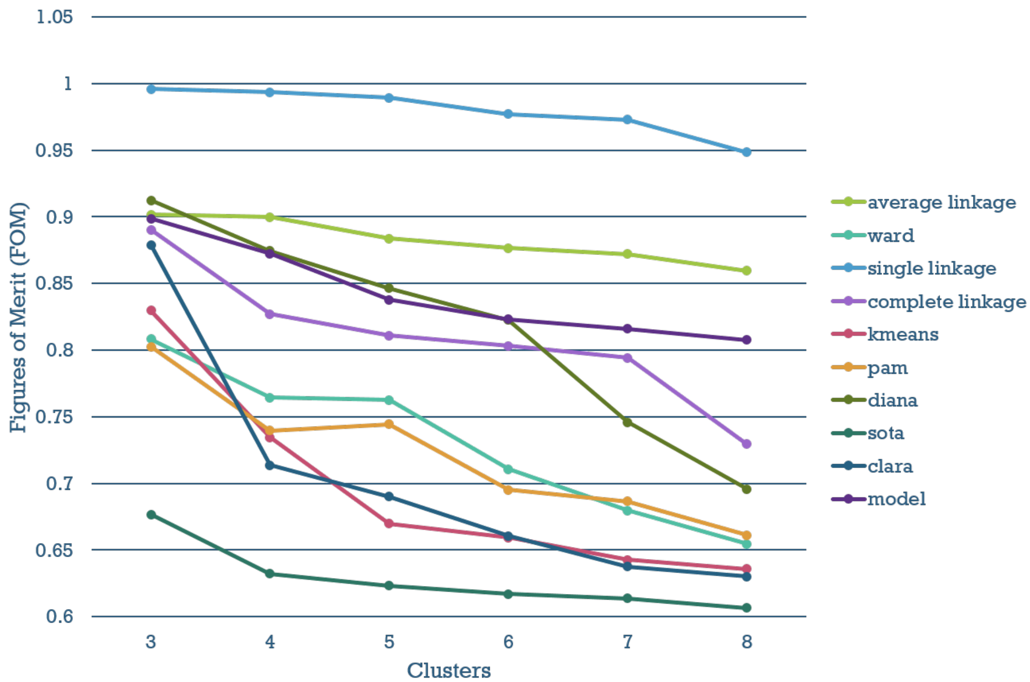

Figure B.11: Figures of Merit (FOM) for Reagan National Airport (DCA)

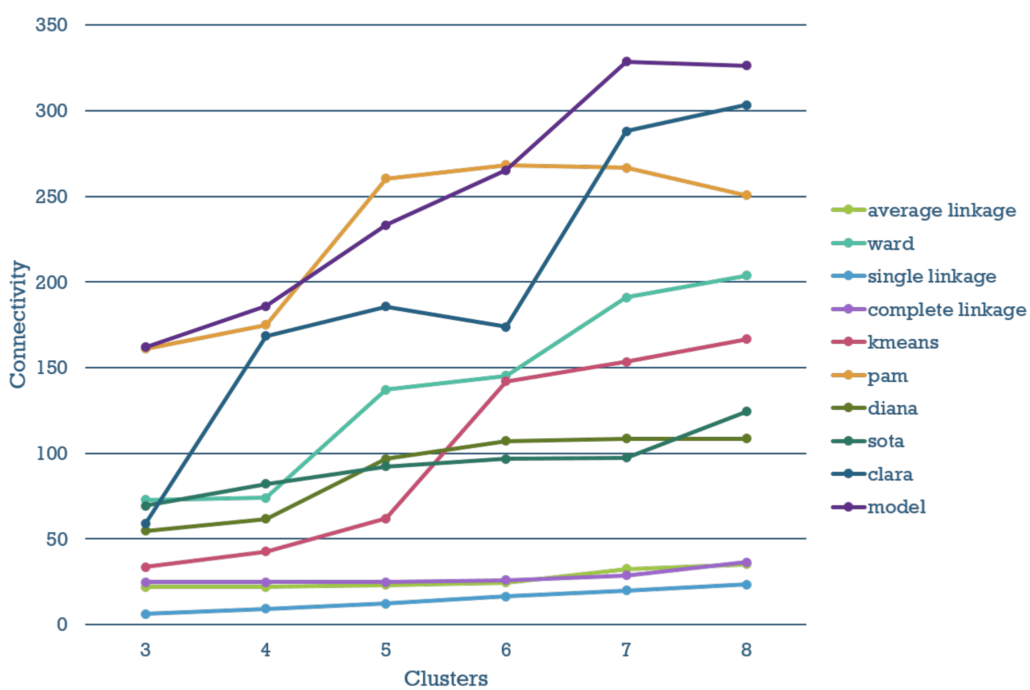

Figure B.12: Connectivity for Reagan National Airport (DCA)

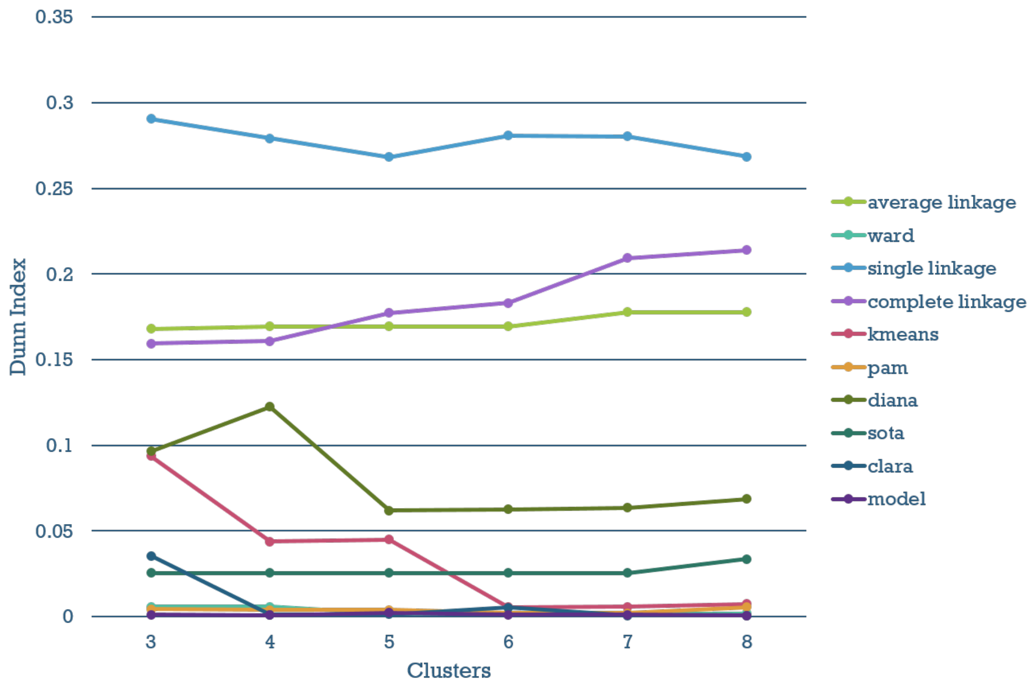

Figure B.13: Dunn Index for Reagan National Airport (DCA)

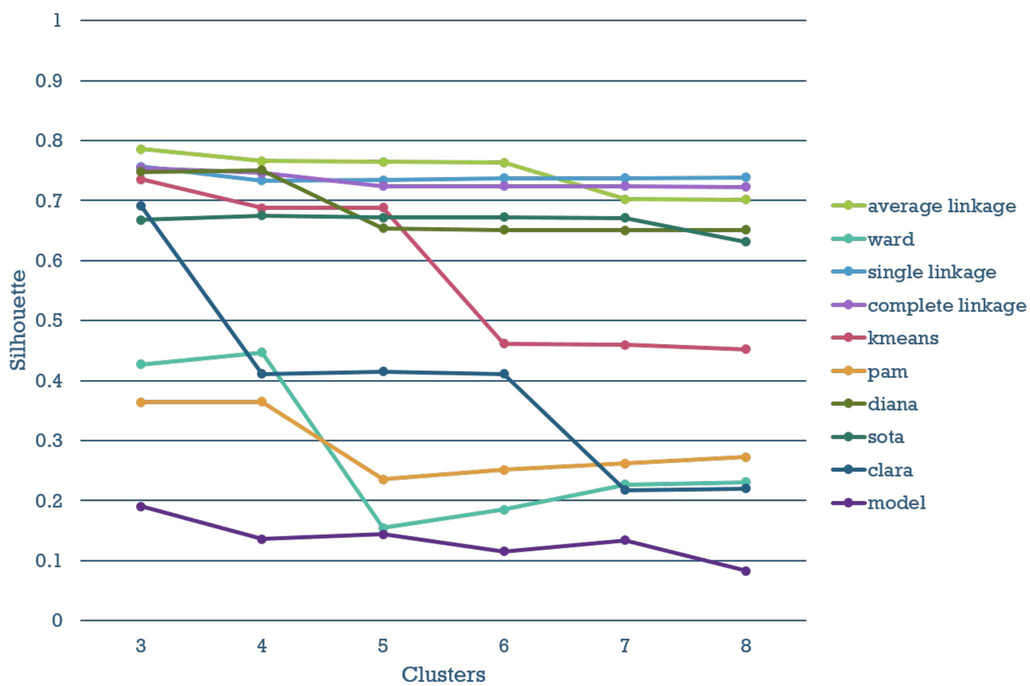

Figure B.14: Silhouette for Reagan National Airport (DCA)

B.3 Newark Liberty International Airport (EWR)

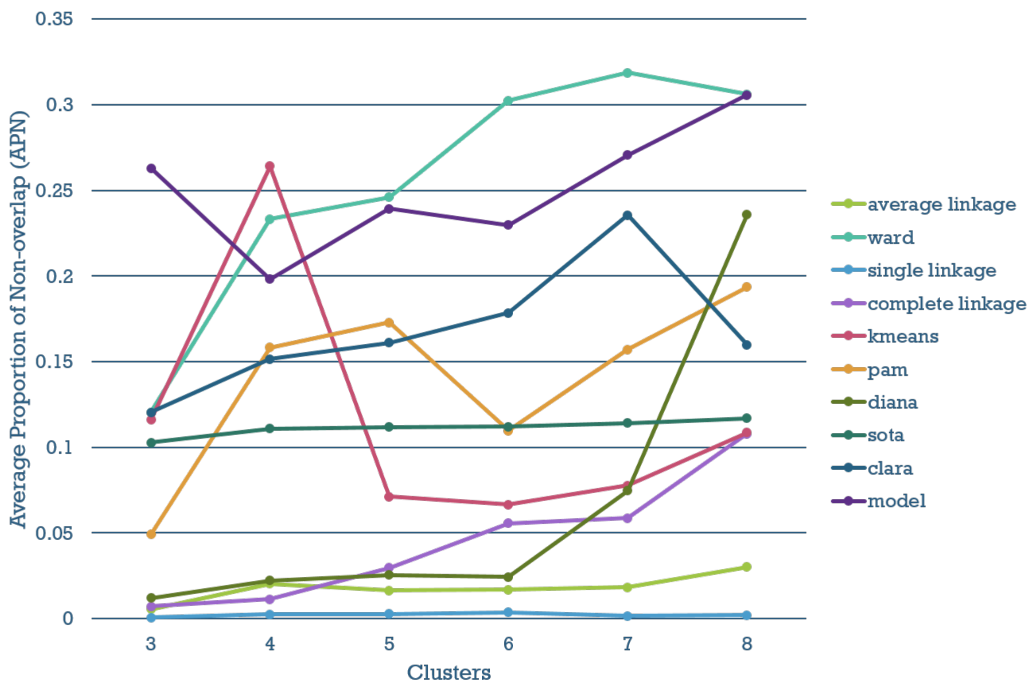

Figure B.15: Average Proportion of Non-overlap (APN) for Newark Liberty International Airport (EWR)

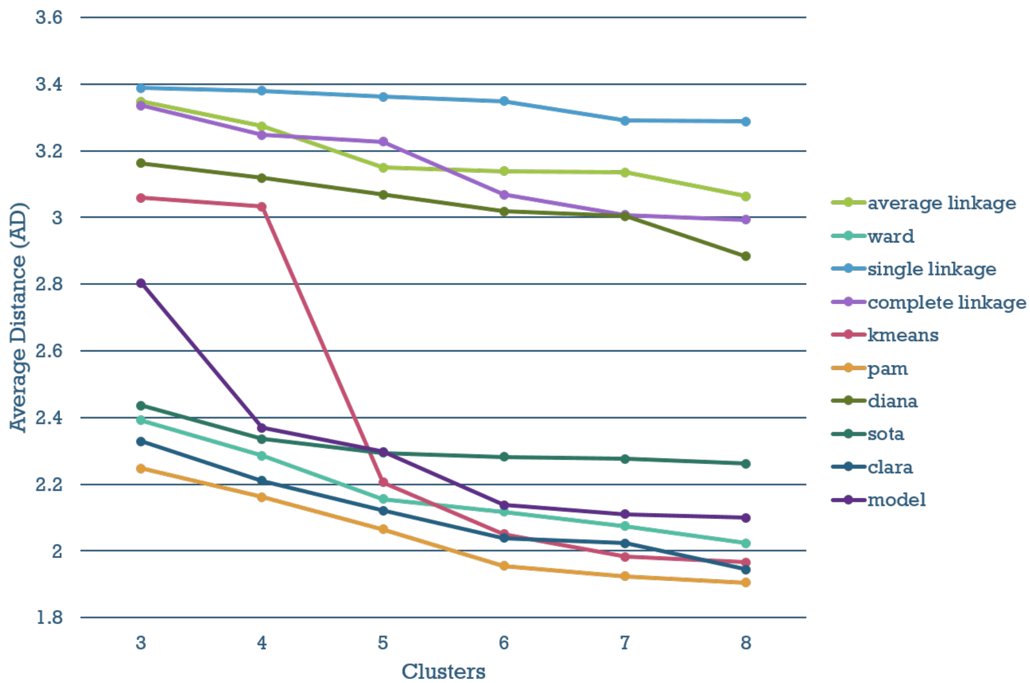

Figure B.16: Average Distance (AD) for Newark Liberty International Airport (EWR)

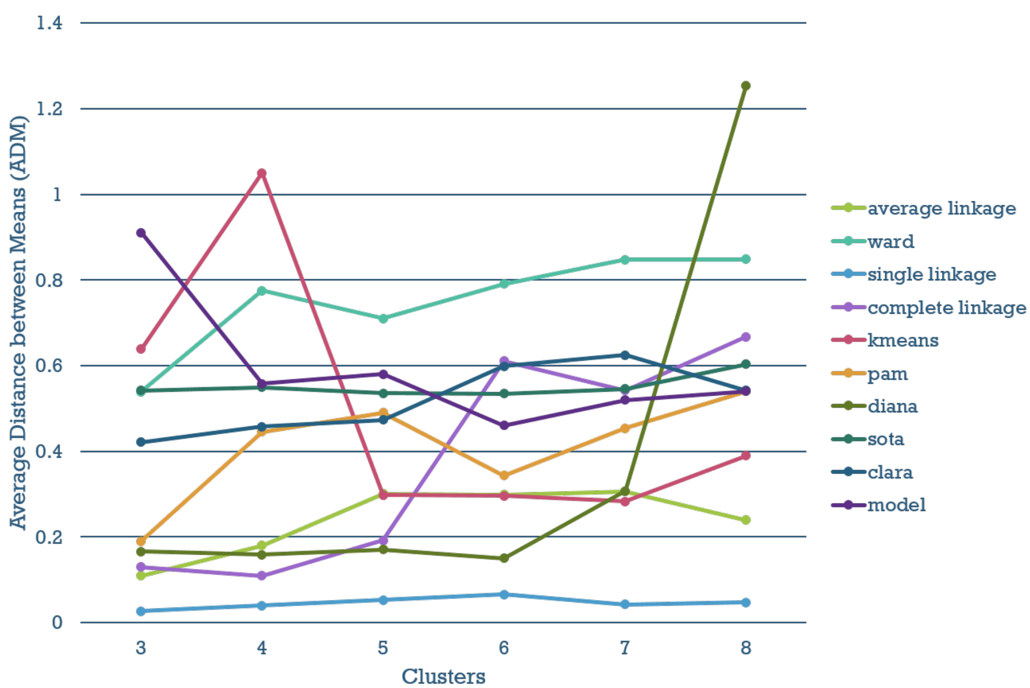

Figure B.17: Average Distance between Means (ADM) for Newark Liberty International Airport (EWR)

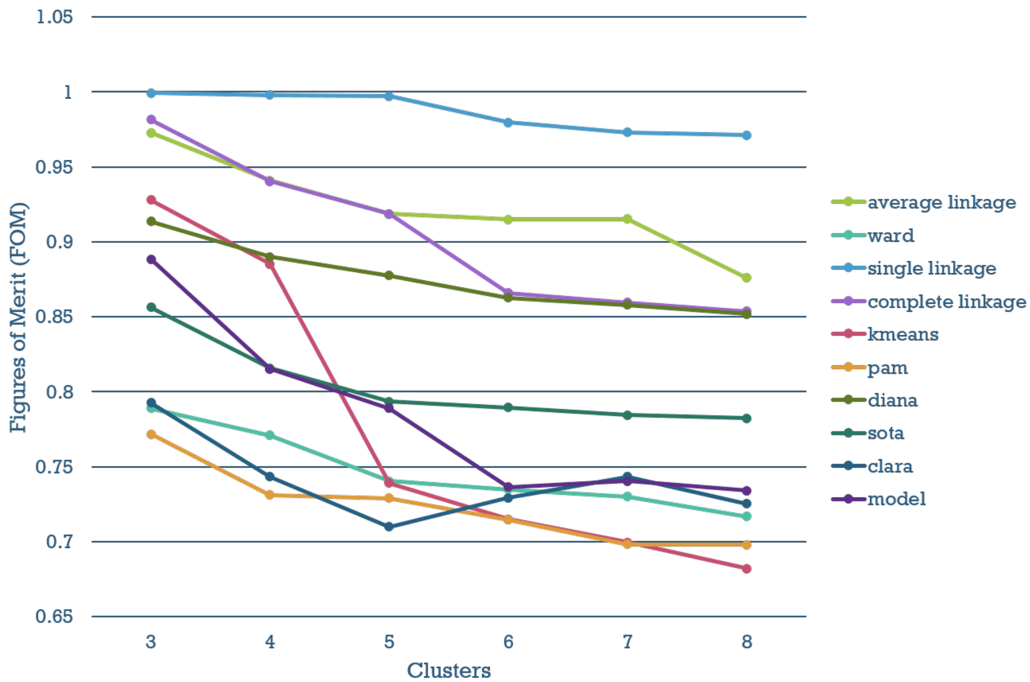

Figure B.18: Figures of Merit (FOM) for Newark Liberty International Airport (EWR)

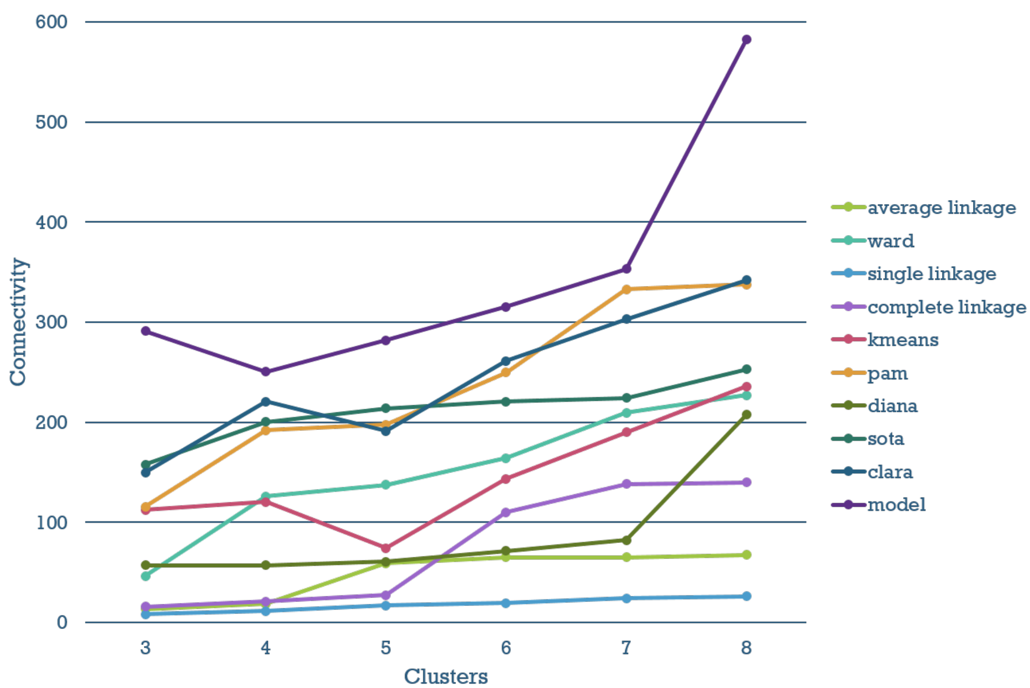

Figure B.19: Connectivity for Newark Liberty International Airport (EWR)

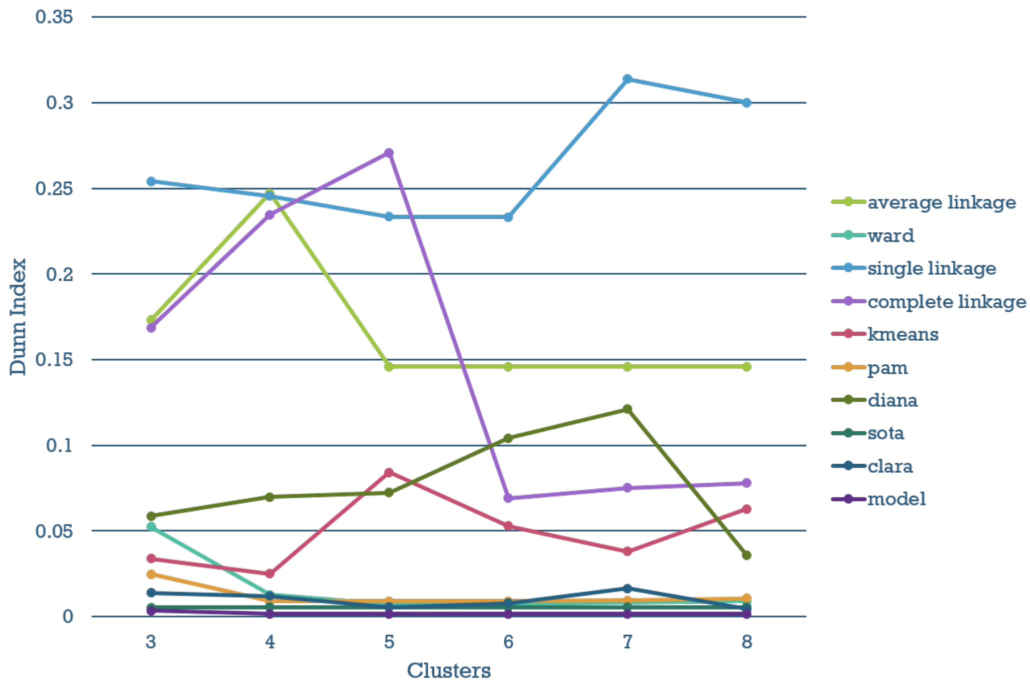

Figure B.20: Dunn Index for Newark Liberty International Airport (EWR)

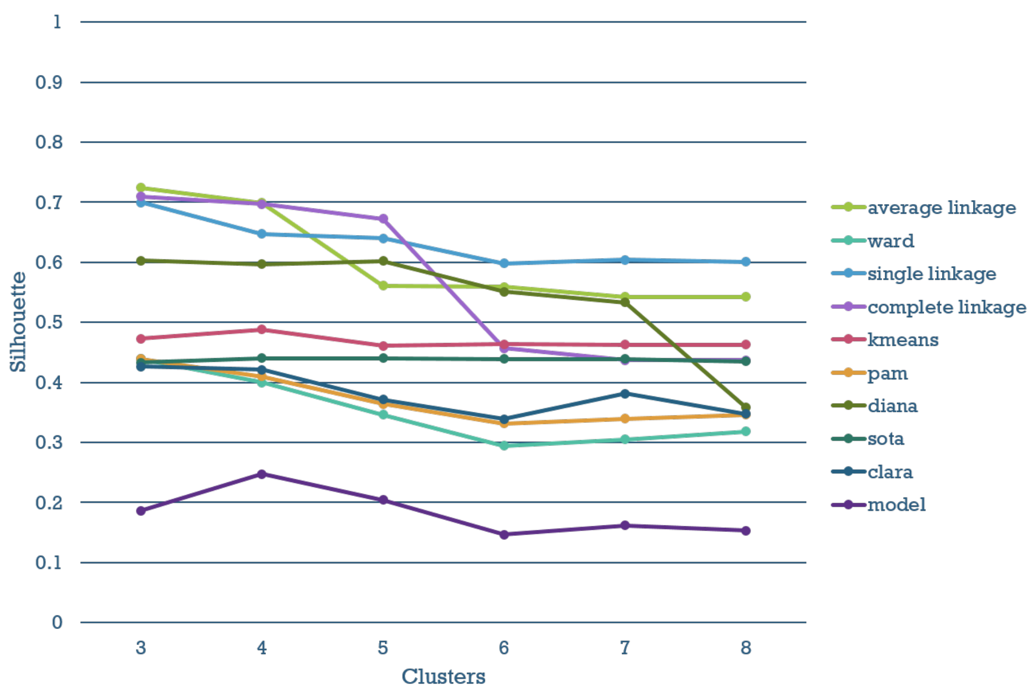

Figure B.21: Silhouette for Newark Liberty International Airport (EWR)

B.4 Dulles International Airport (IAD)

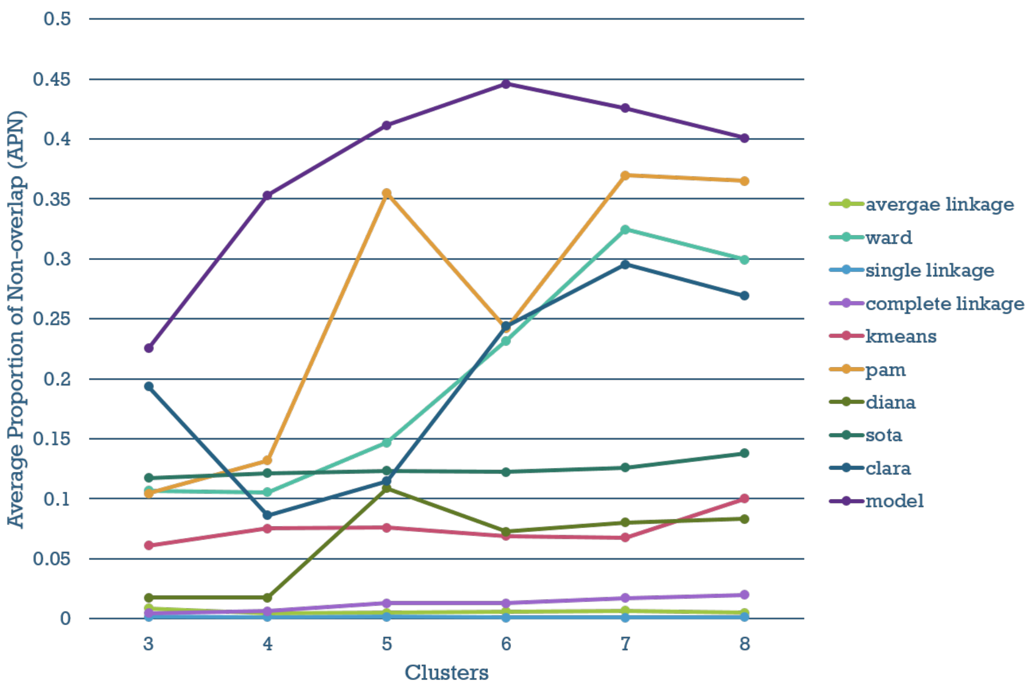

Figure B.22: Average Proportion of Non-overlap (APN) for Dulles International Airport (IAD)

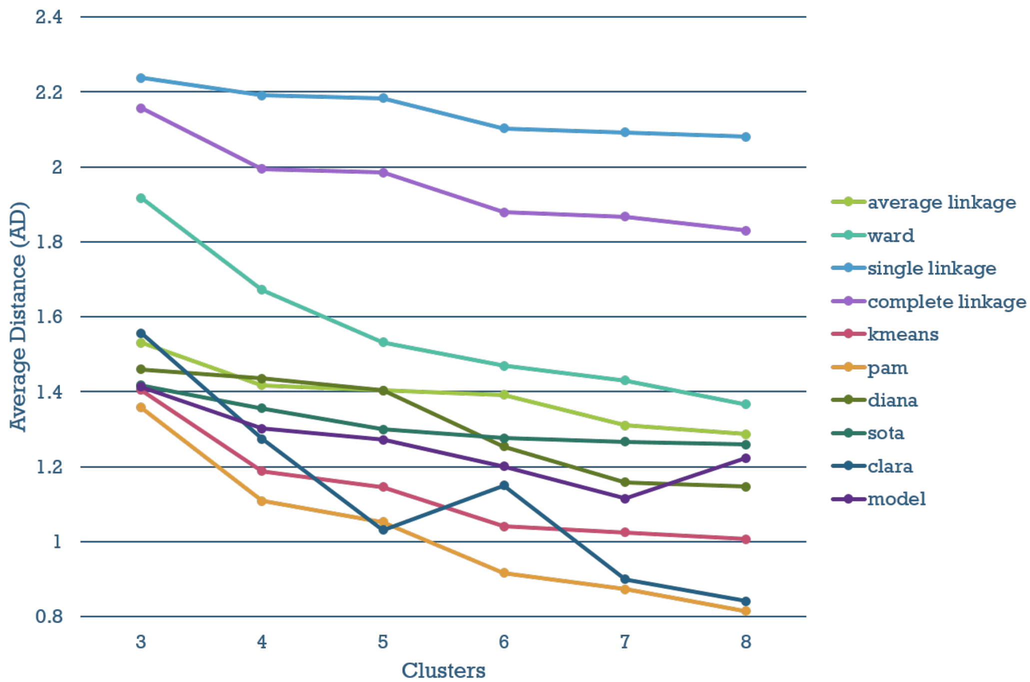

Figure B.23: Average Distance (AD) for Dulles International Airport (IAD)

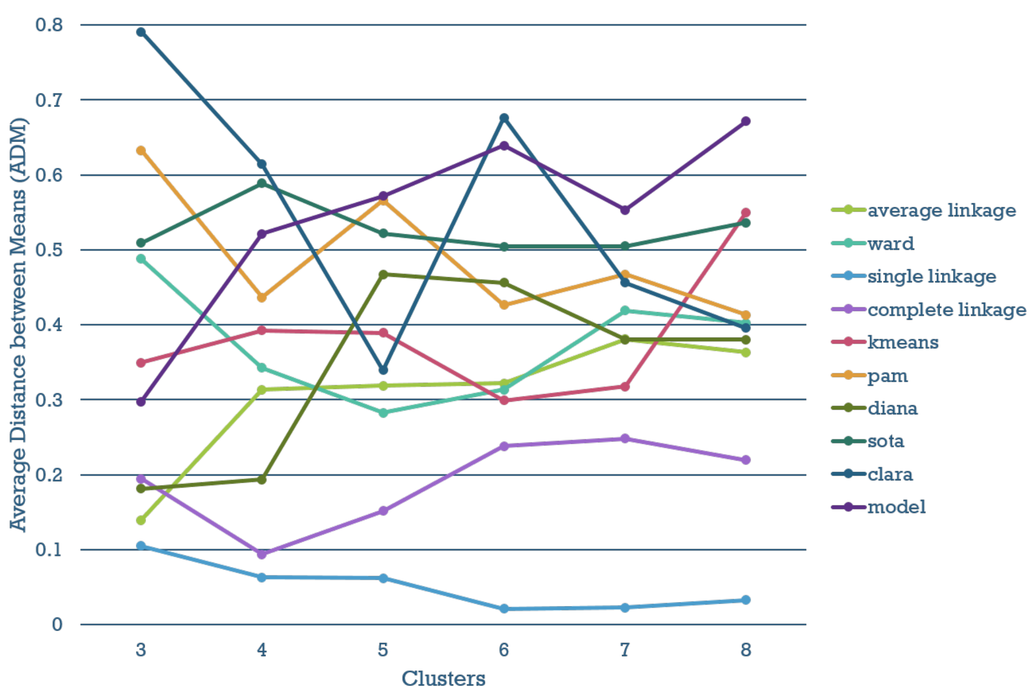

Figure B.24: Average Distance between Means (ADM) for Dulles International Airport (IAD)

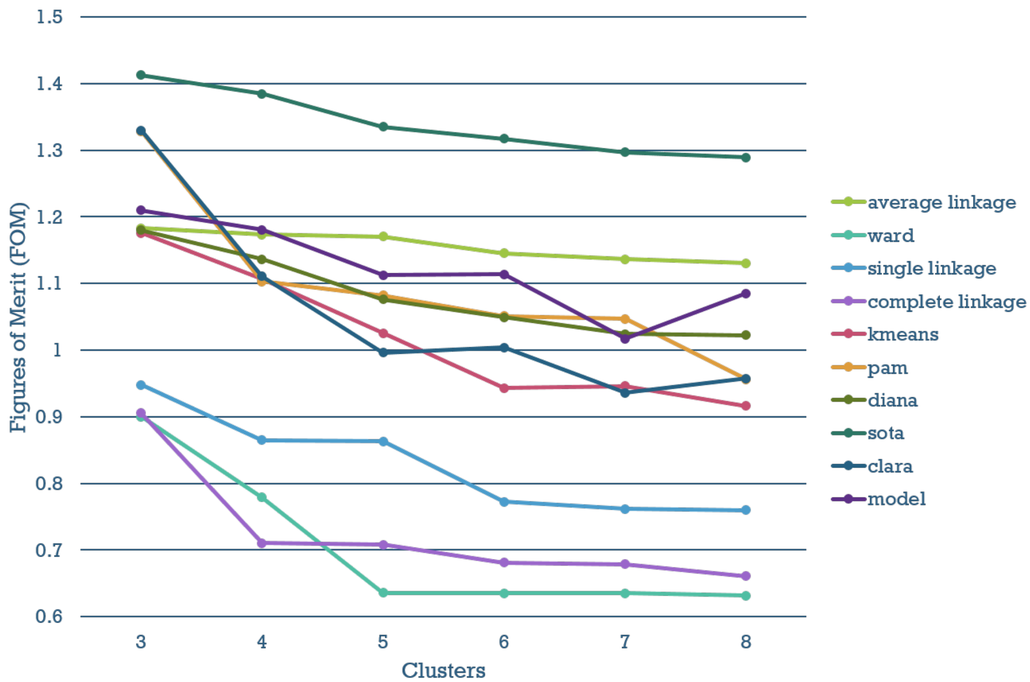

Figure B.25: Figures of Merit (FOM) for Dulles International Airport (IAD)

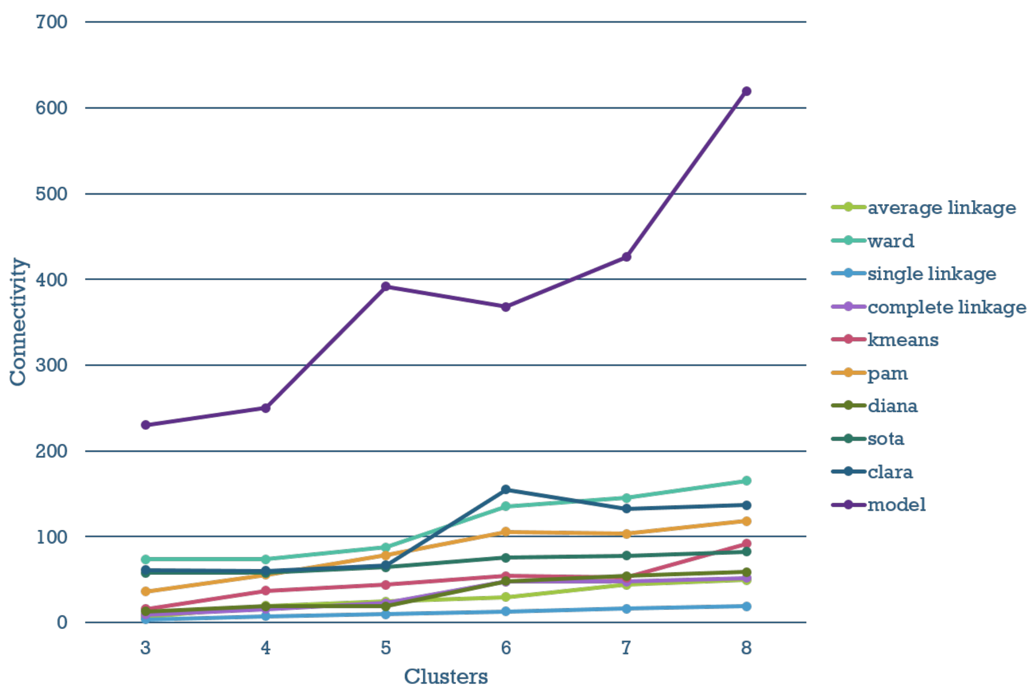

Figure B.26: Connectivity for Dulles International Airport (IAD)

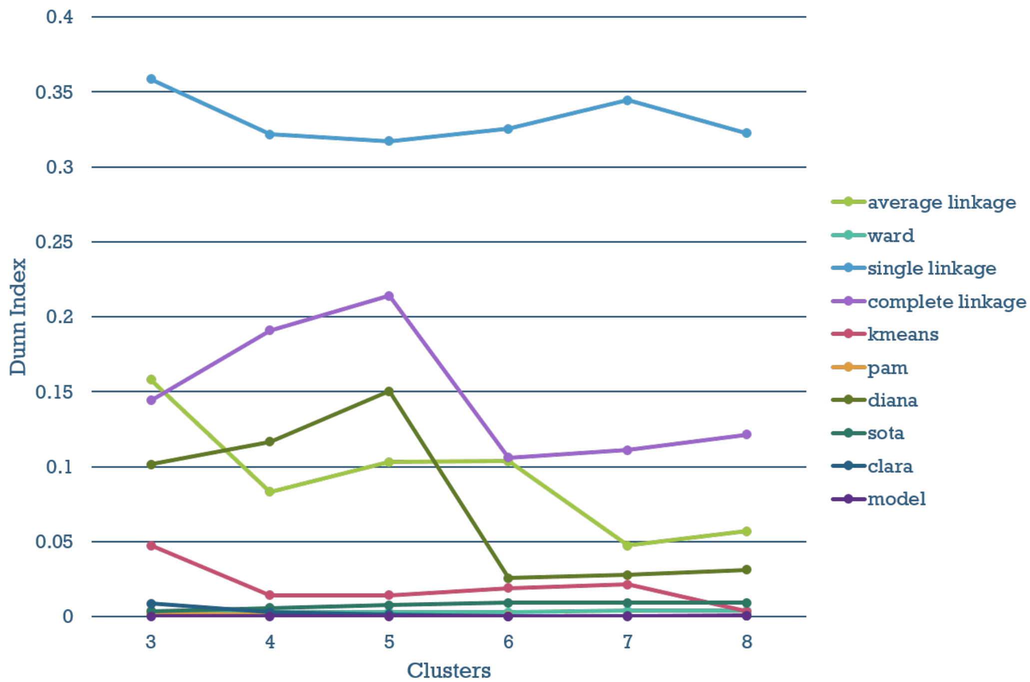

Figure B.27: Dunn Index for Dulles International Airport (IAD)

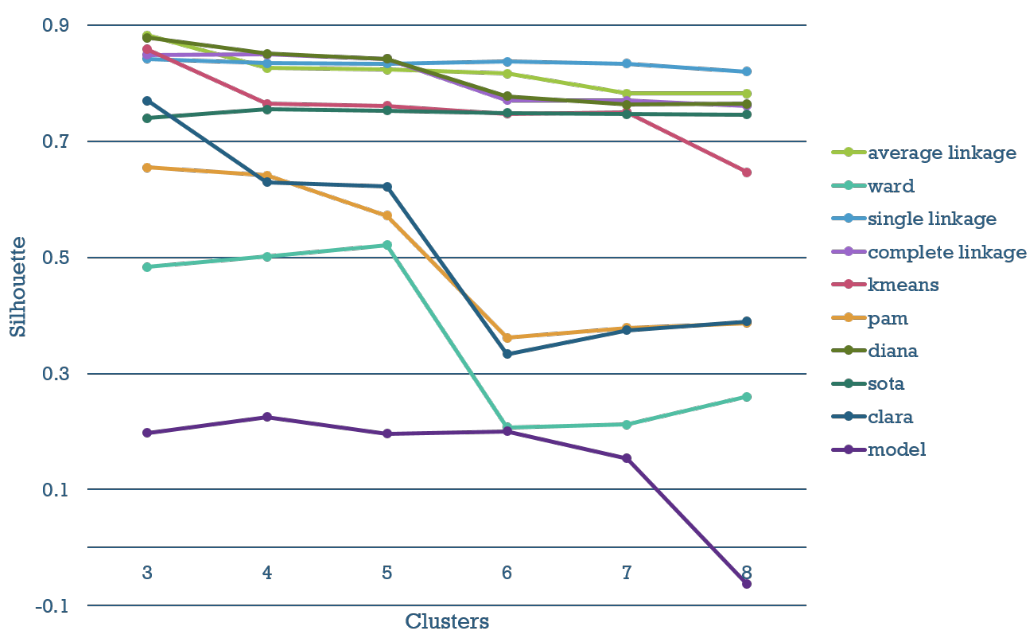

Figure B.28: Silhouette for Dulles International Airport (IAD)

B.5 John F. Kennedy International Airport (JFK)

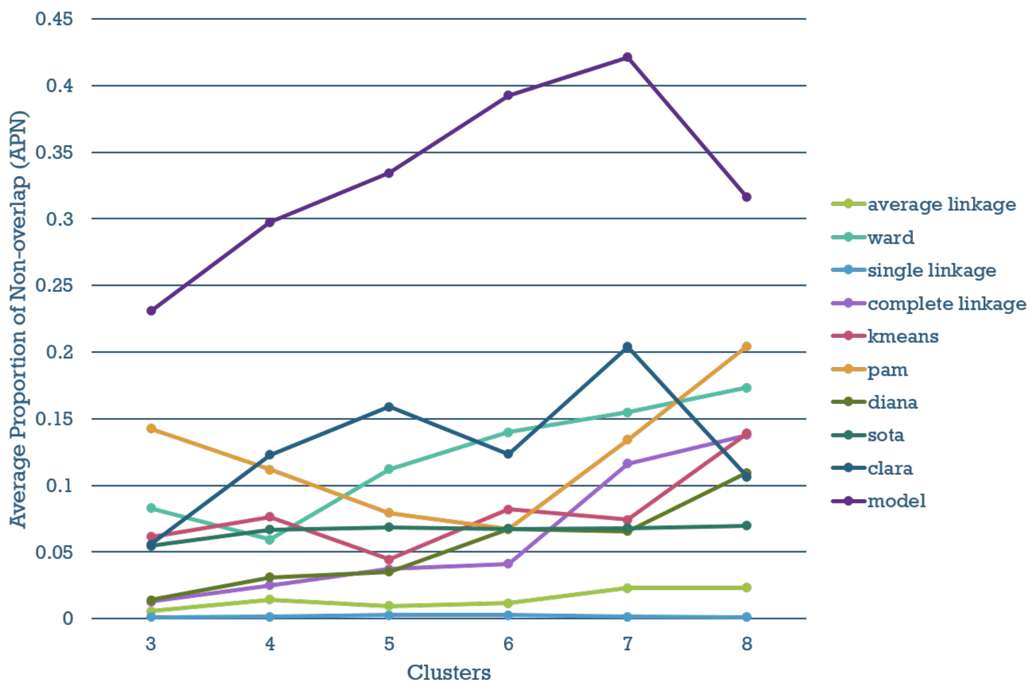

Figure B.29: Average Proportion of Non-overlap (APN) for John F. Kennedy International Airport (JFK)

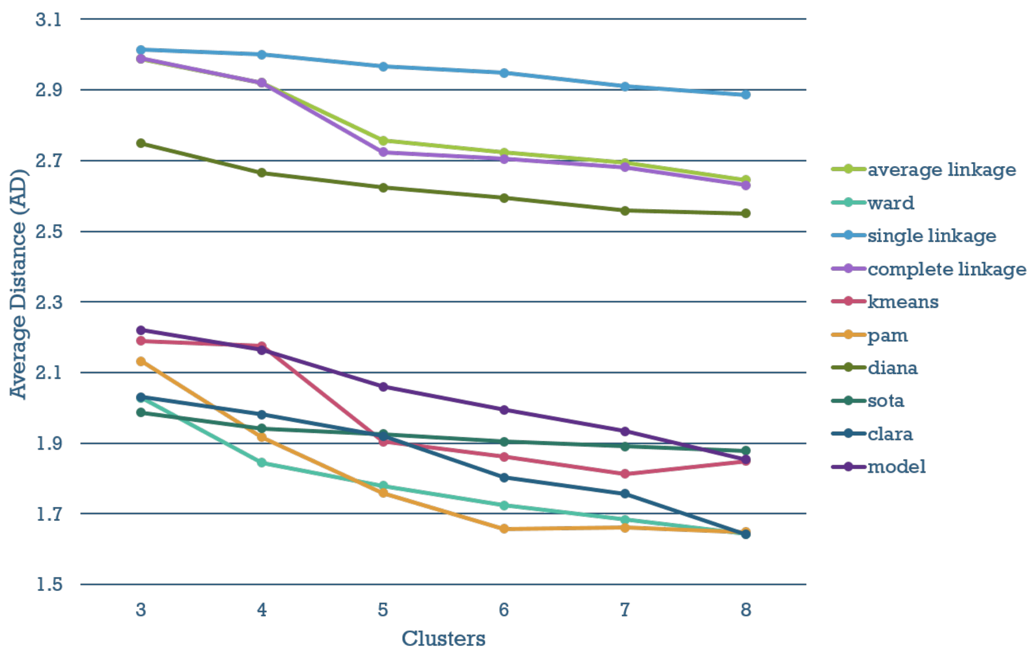

Figure B.30: Average Distance (AD) for John F. Kennedy International Airport (JFK)

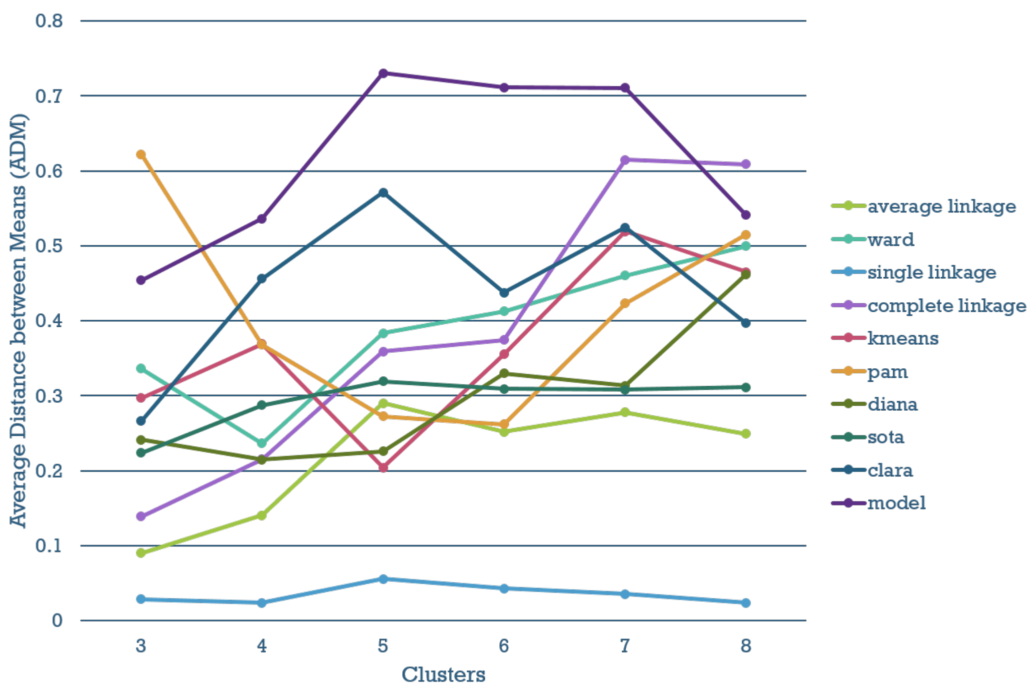

Figure B.31: Average Distance between Means (ADM) for John F. Kennedy International Airport (JFK)

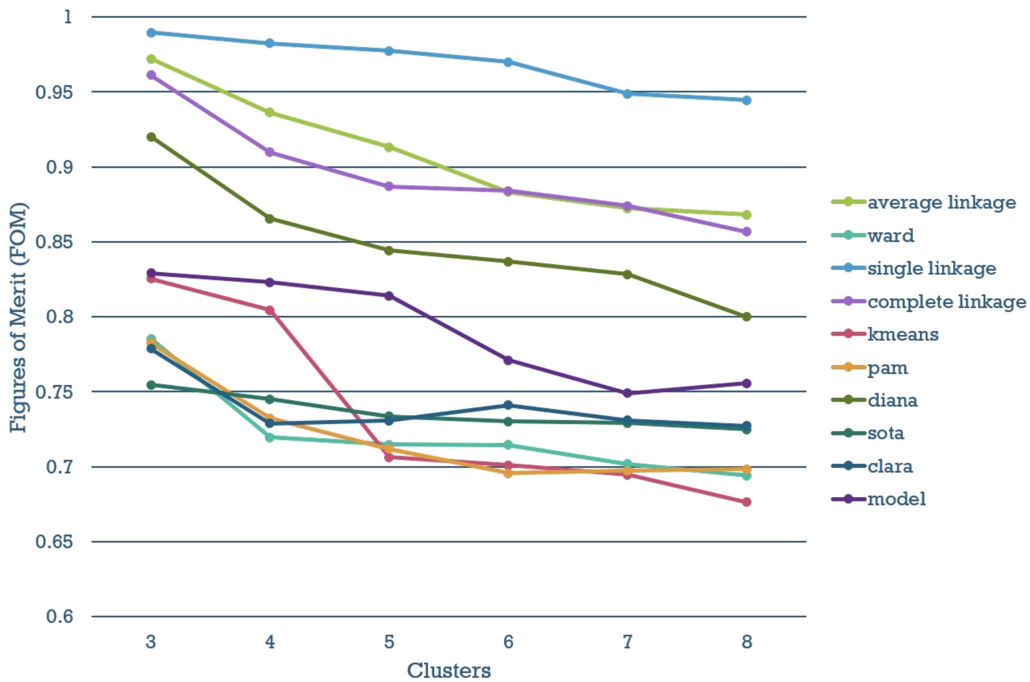

Figure B.32: Figures of Merit (FOM) for John F. Kennedy International Airport (JFK)

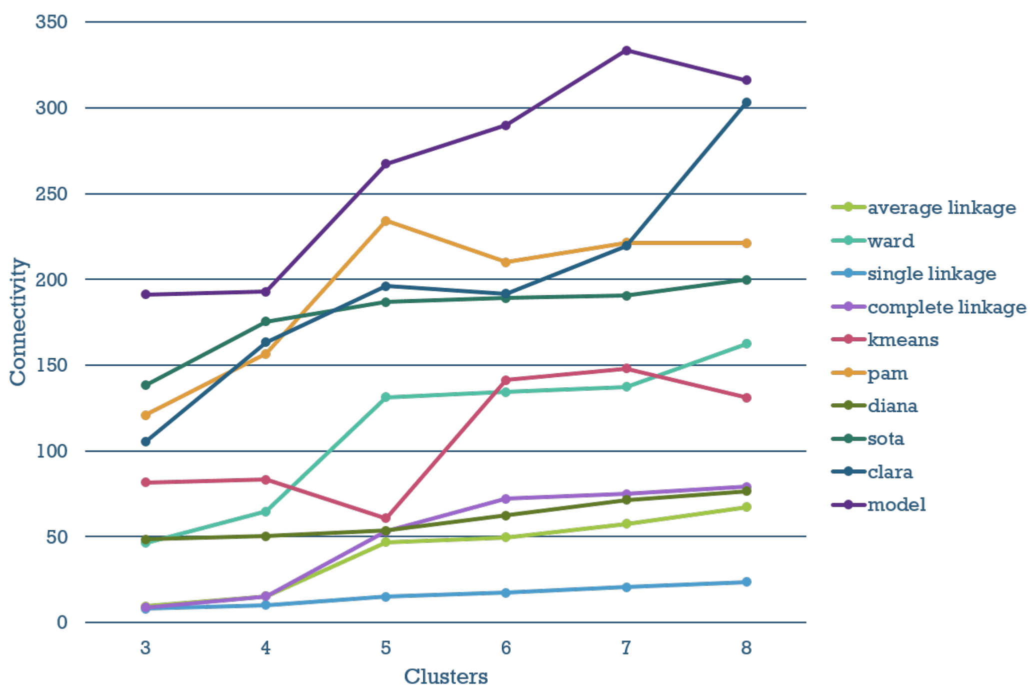

Figure B.33: Connectivity for John F. Kennedy International Airport (JFK)

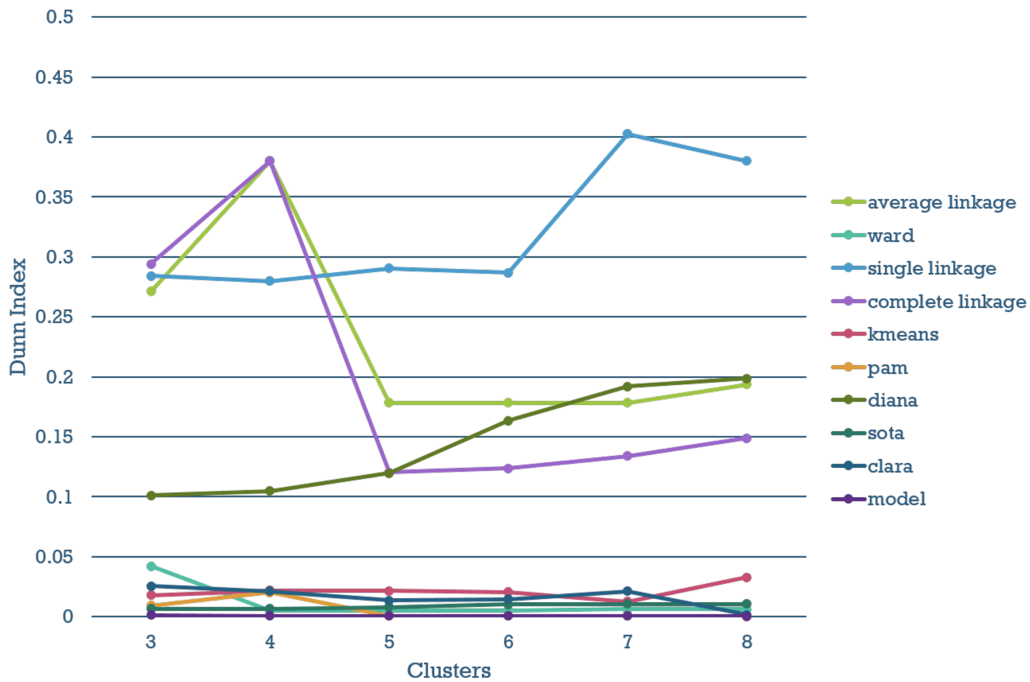

Figure B.34: Dunn Index for John F. Kennedy International Airport (JFK)

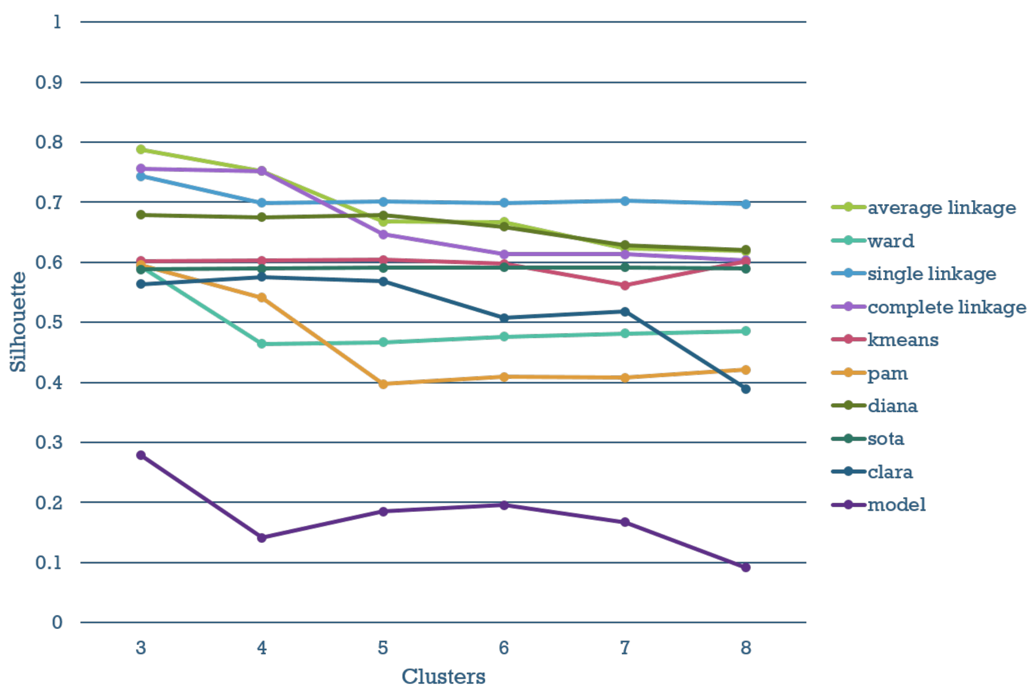

Figure B.35: Silhouette for John F. Kennedy International Airport (JFK)

B.6 LaGuardia Airport (LGA)

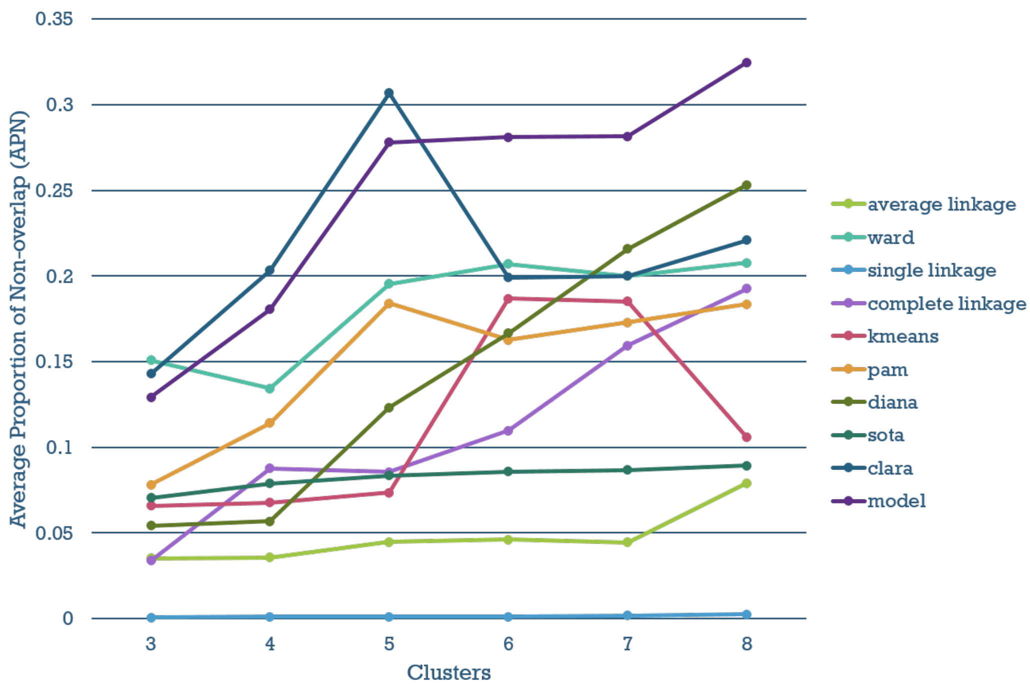

Figure B.36: Average Proportion of Non-overlap (APN) for LaGuardia Airport (LGA)

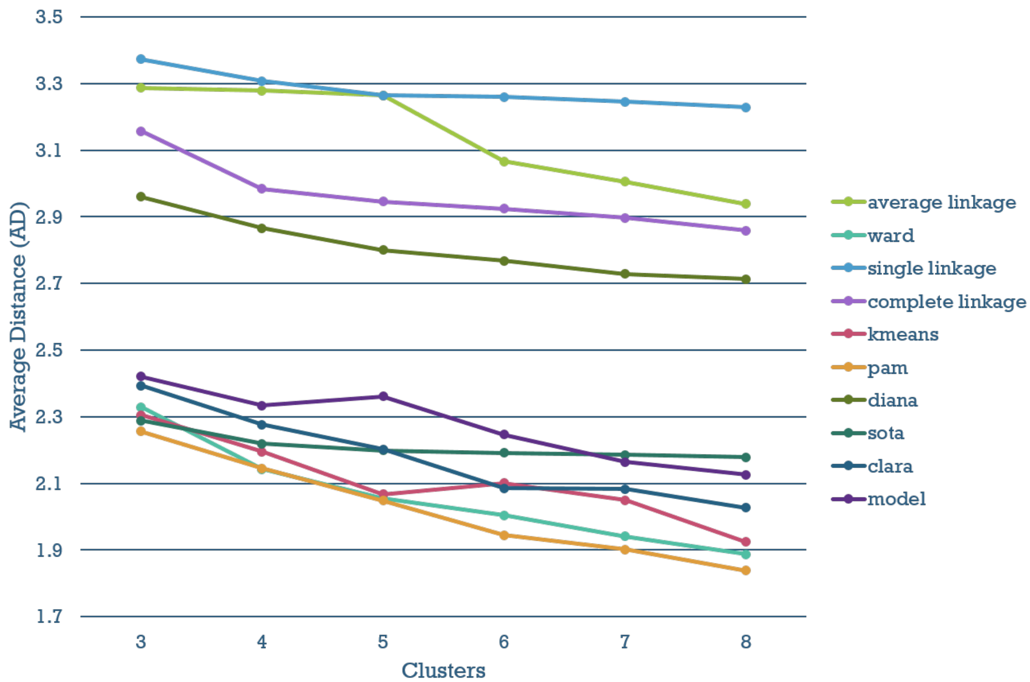

Figure B.37: Average Distance (AD) for LaGuardia Airport (LGA)

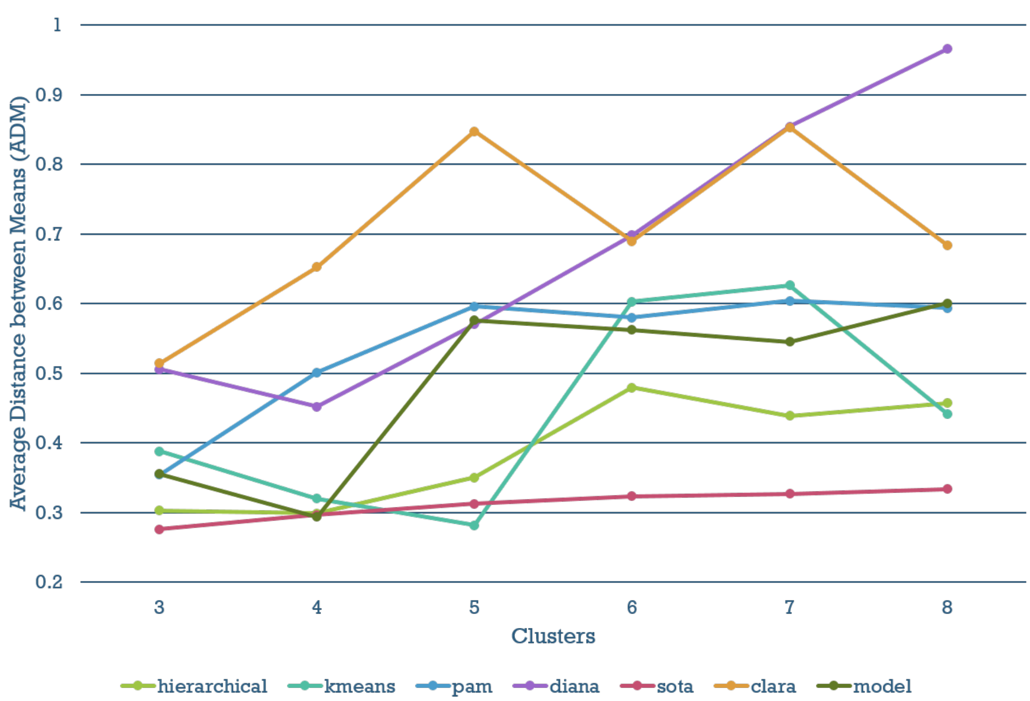

Figure B.38: Average Distance between Means (ADM) for LaGuardia Airport (LGA)

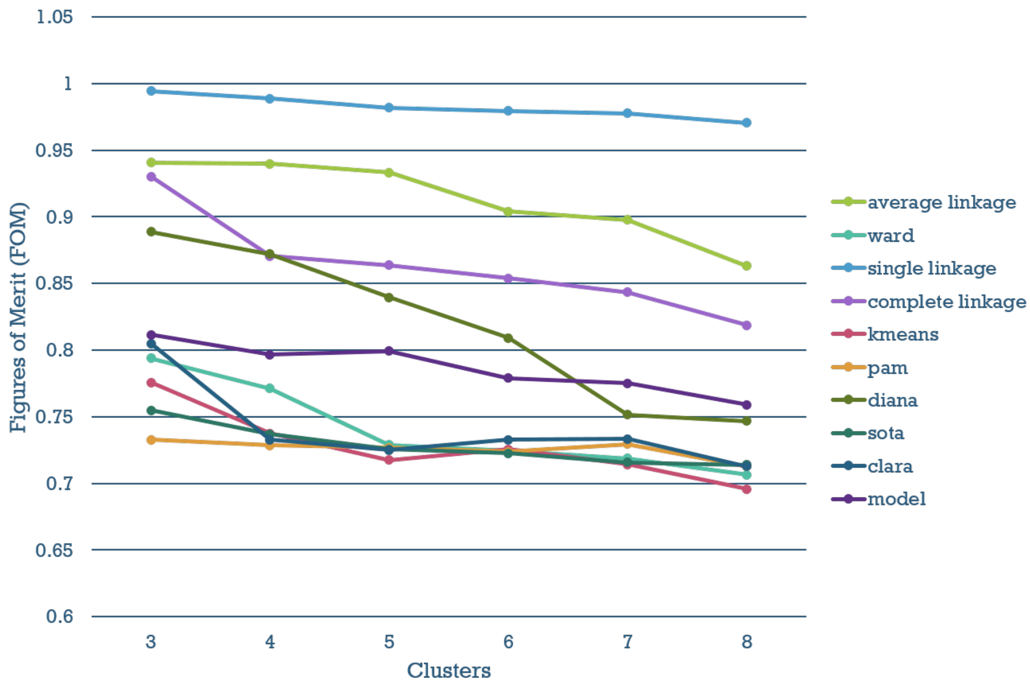

Figure B.39: Figures of Merit (FOM) for LaGuardia Airport (LGA)

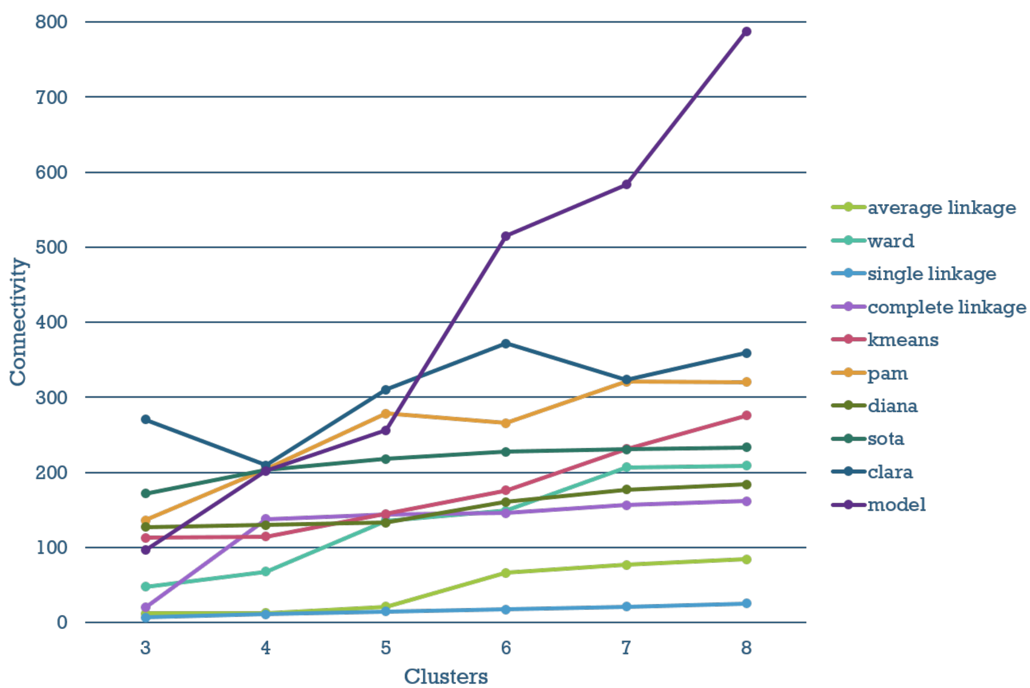

Figure B.40: Connectivity for LaGuardia Airport (LGA)

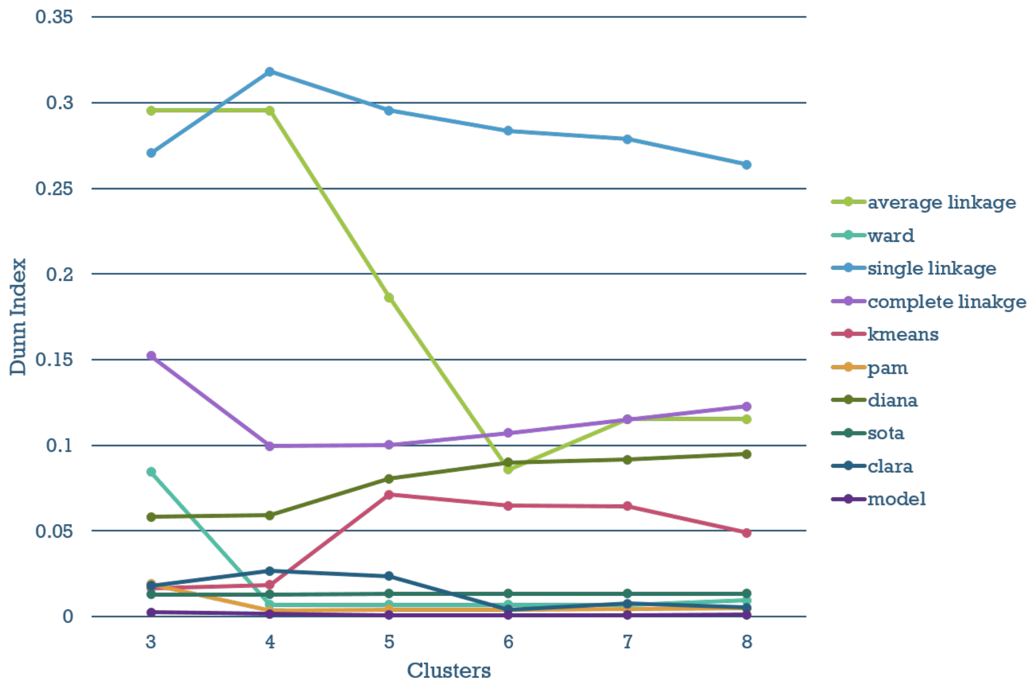

Figure B.41: Dunn Index for LaGuardia Airport (LGA)

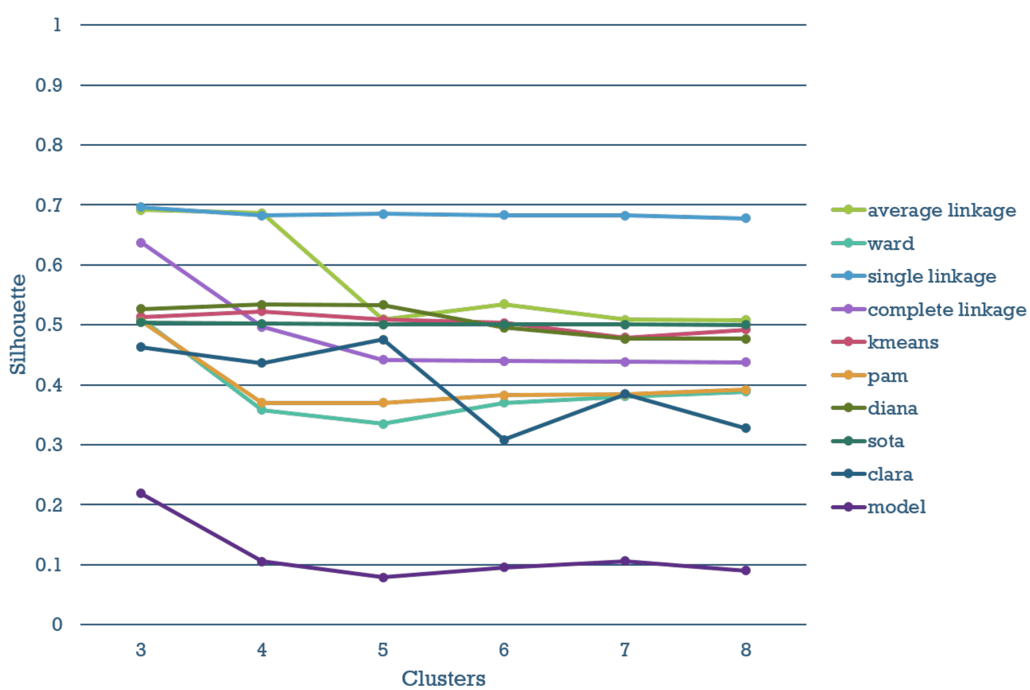

Figure B.42: Silhouette for LaGuardia Airport (LGA)

B.7 Philadelphia International Airport (PHL)

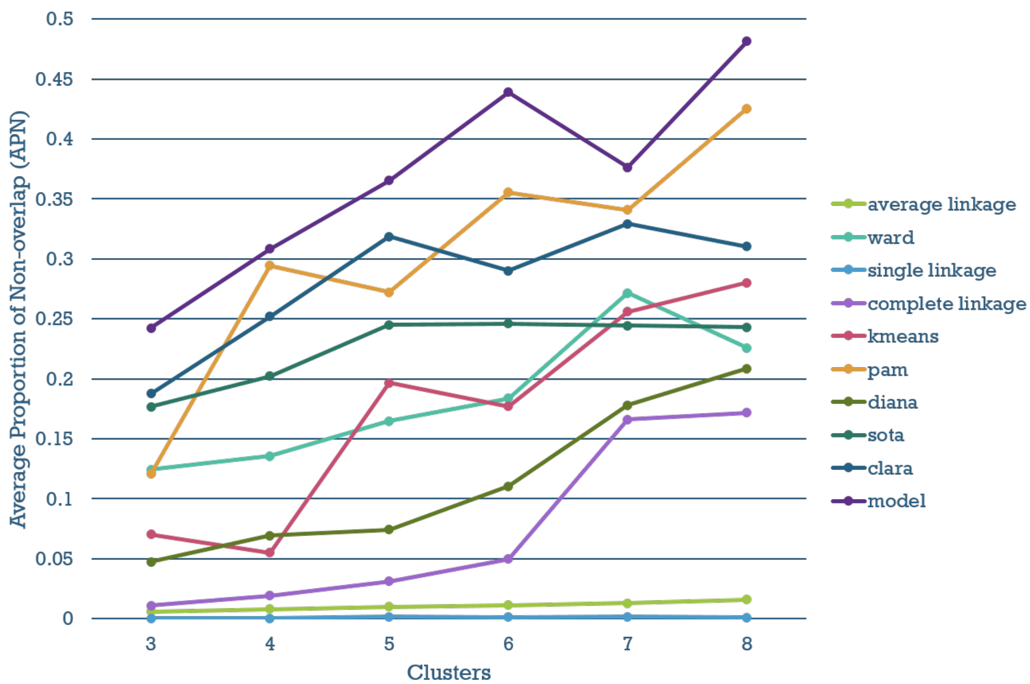

Figure B.43: Average Proportion of Non-overlap (APN) for Philadelphia International Airport (PHL)

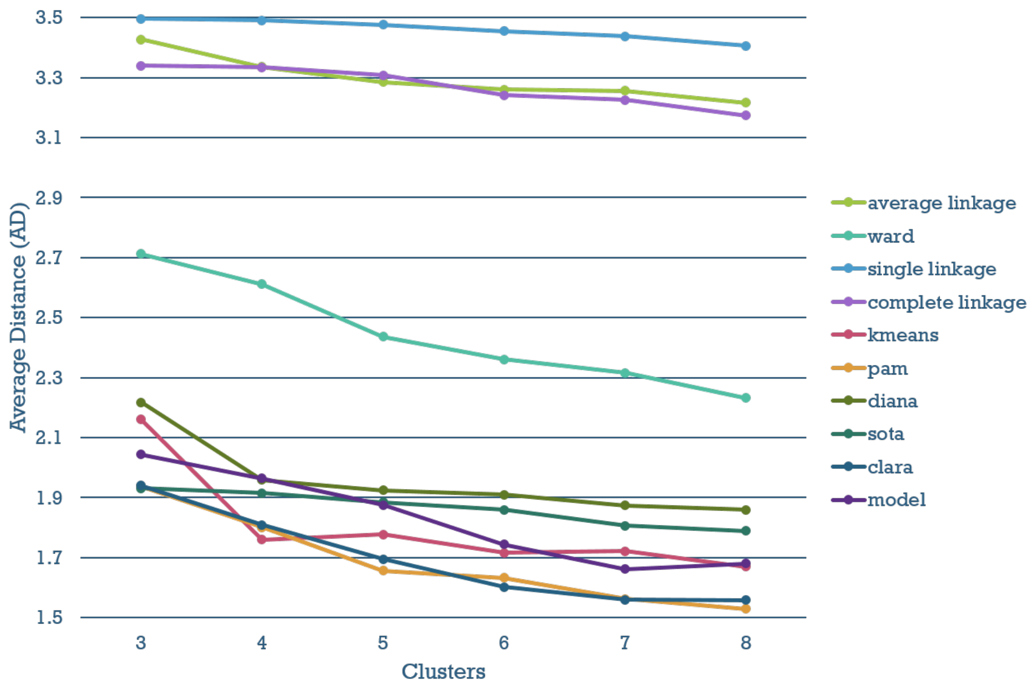

Figure B.44: Average Distance (AD) for Philadelphia International Airport (PHL)

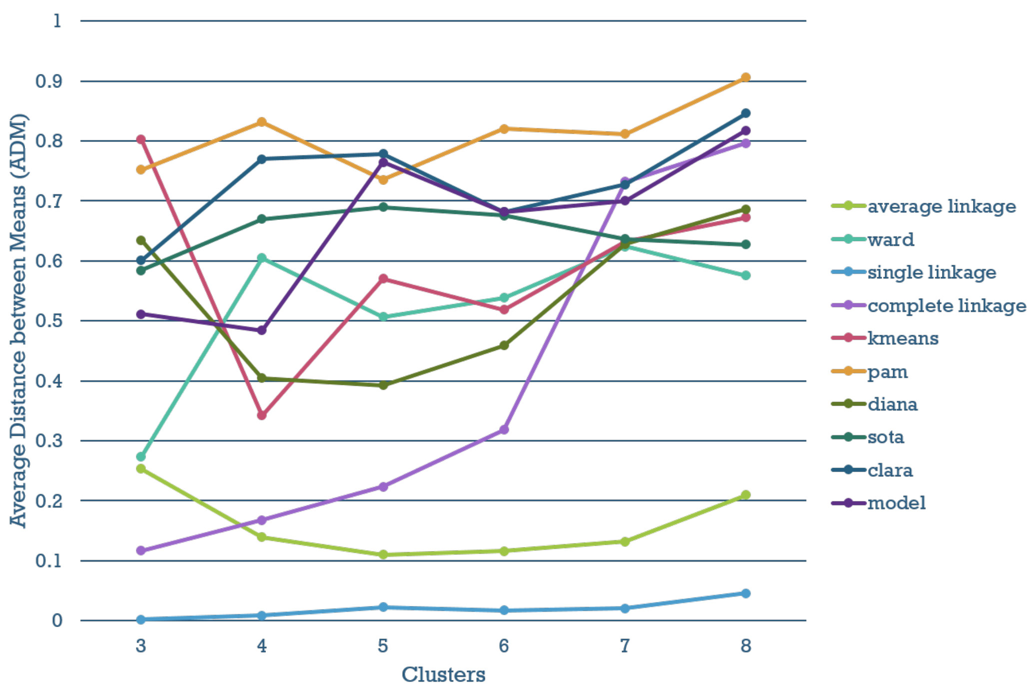

Figure B.45: Average Distance between Means (ADM) for Philadelphia International Airport (PHL)

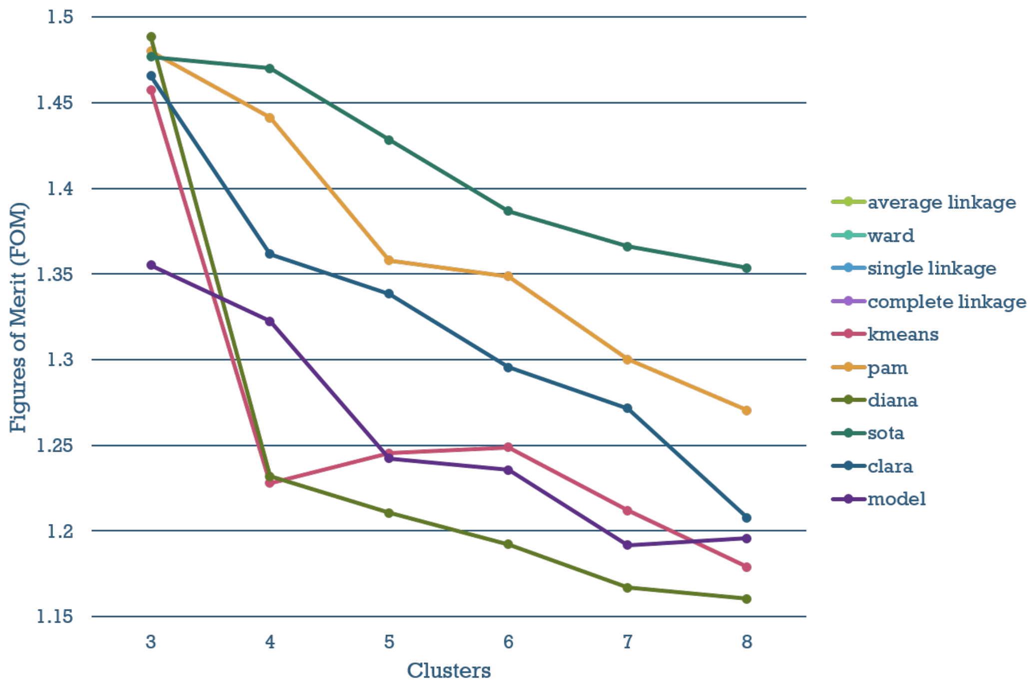

Figure B.46: Figures of Merit (FOM) for Philadelphia International Airport (PHL)

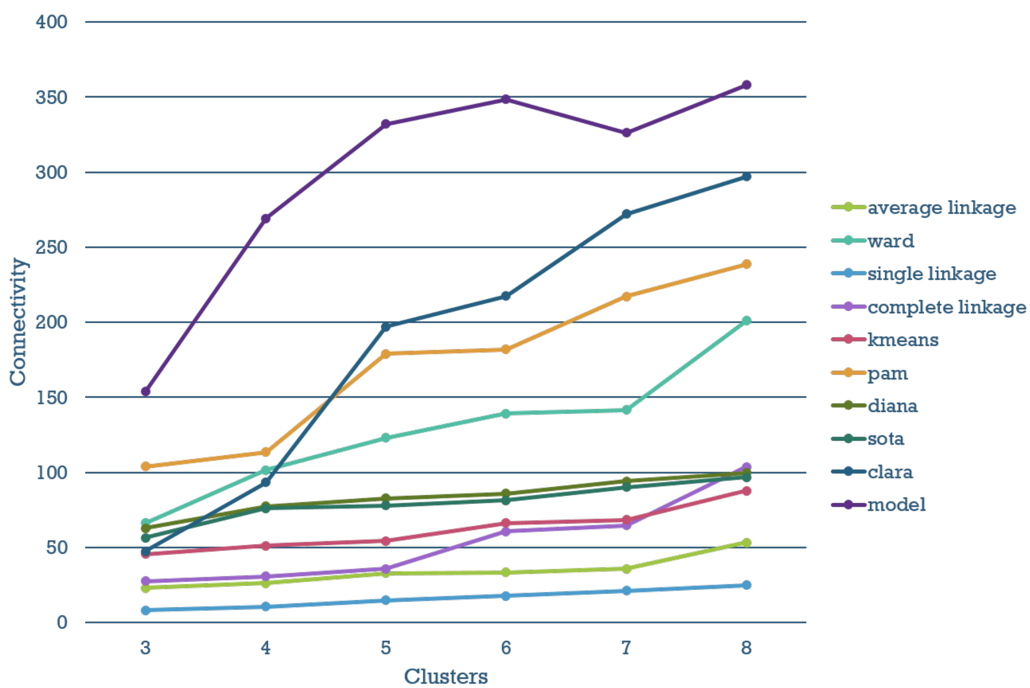

Figure B.47: Connectivity for Philadelphia International Airport (PHL)

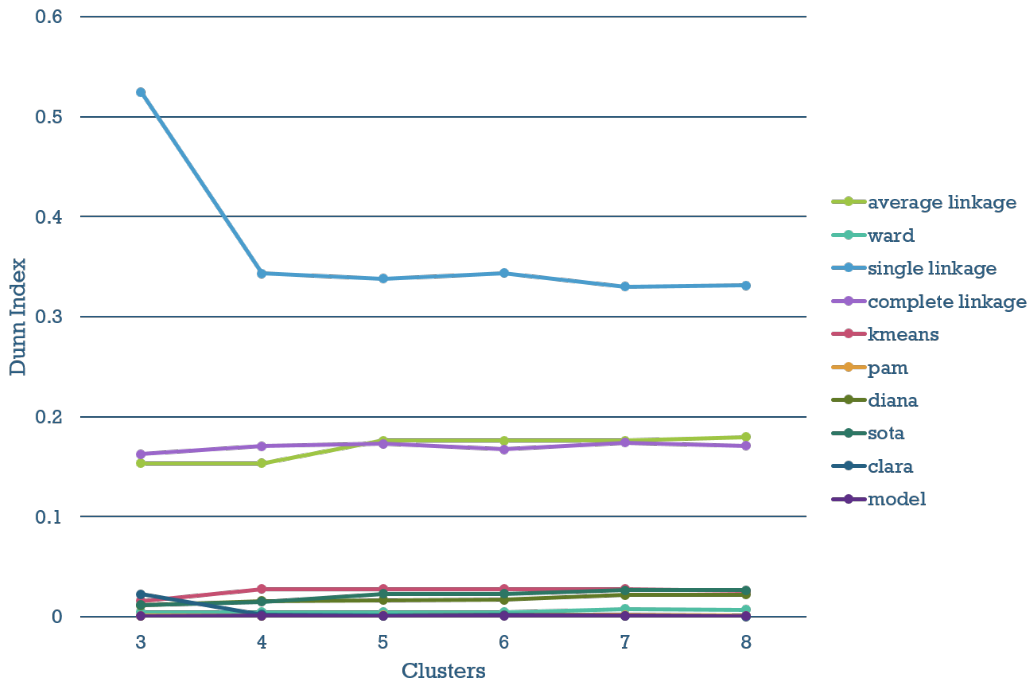

Figure B.48: Dunn Index for Philadelphia International Airport (PHL)

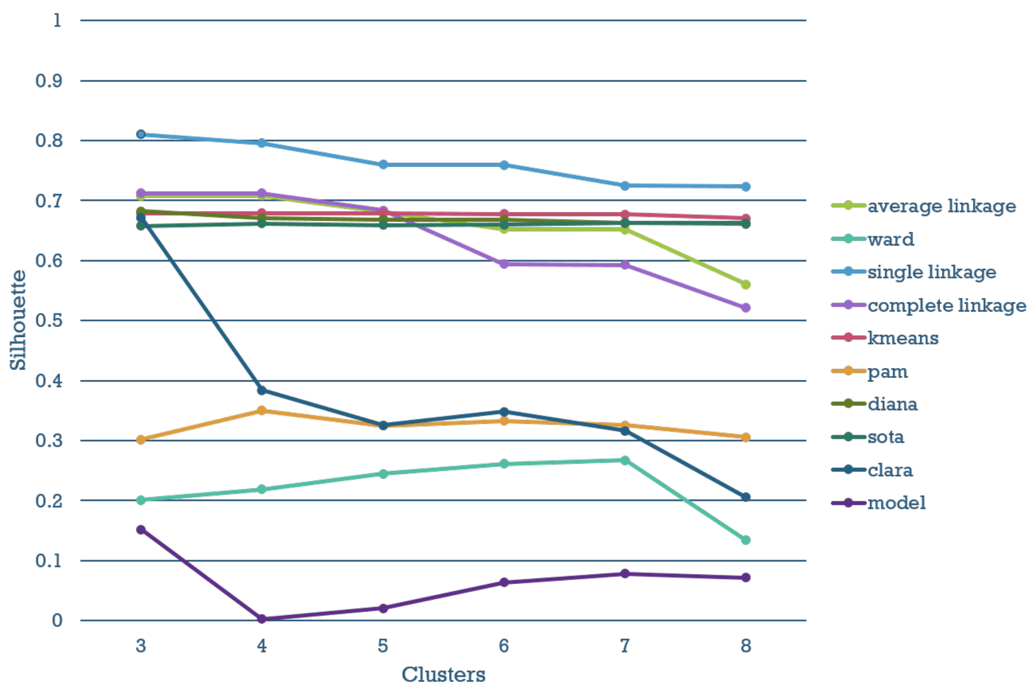

Figure B.49: Silhouette for Philadelphia International Airport (PHL)

APPENDIX C

COMPARISON OF RESULTS FROM CLUSTERING AND PREDEFINED RANGES OF METRICS

Table C.1: Comparison of the categorization of daily airport operations at BOS with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	4	66	1	500	-1	0	0	0	63.7	2	Good
2	7	149	1	184	0	0	0	50	62.3	2	Good
3	51	1506	15	500	-1	1	15	4	99.7	2	Good
4	45	1127	1	58	2	2	0	266	90.1	2	Good
5	0	0	1	-31	0	1	0	87	99.5	2	Good
6	17	382	15	-27	0	1	21	104	98.8	2	Good
7	0	0	1	500	-1	0	0	0	11.9	3	Good
8	0	0	1	500	-1	0	0	0	7.7	3	Good
9	3	65	3	500	-1	0	0	0	19.9	3	Good
10	0	0	0	500	-1	0	0	0	44.5	3	Good
11	0	0	0	500	-1	0	0	0	11.9	3	Good
12	0	0	2	500	-1	0	0	0	57.2	3	Good
13	5	115	8	96	0	0	0	4	40.1	3	Good
14	24	528	12	30	0	1	97	229	77.5	2	Bad
15	36	665	3	44	0	1	0	218	98.8	2	Bad
16	37	968	11	12	1	2	0	166	98.3	2	Bad
17	10	229	15	174	1	1	57	152	85.1	2	Bad
18	5	92	6	46	1	1	97	79	98.2	1	Average

Table C.2: Comparison of the categorization of daily airport operations at BWI with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	46	1067	14	500	-1	1	27	12	96.6	2	Good
2	37	1096	33	500	-1	1	18	9	97.9	2	Good
3	37	1096	33	500	-1	1	18	9	97.9	2	Good
4	8	165	35	500	-1	1	0	18	98.7	2	Good
5	22	700	29	500	-1	4	20	37	88.1	2	Good
6	7	212	10	500	-1	0	0	0	73.2	3	Good
7	0	0	1	500	-1	0	0	0	23.0	3	Good
8	0	0	1	500	-1	0	8	0	76.7	3	Good
9	3	72	3	500	-1	1	0	2	42.9	3	Good
10	1	20	1	500	-1	0	0	0	65.4	3	Good
11	6	119	1	0	0	0	4	0	98.8	3	Good
12	0	0	0	-2	0	2	4	0	96.3	3	Good
13	41	1350	30	-7	0	1	10	55	89.8	3	Good
14	11	392	7	0	1	1	14	5	88.2	3	Good
15	24	806	20	-8	0	2	15	61	94.2	3	Good
16	5	150	2	-12	0	1	9	12	90.9	3	Good

Table C.3: Comparison of the categorization of daily airport operations at DCA with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	9	158	3	0	1	0	21	149	98.5	3	Good
2	0	0	1	0	0	2	49	207	86.8	3	Good
3	46	1167	19	500	-1	2	33	38	89.3	2	Good
4	50	1149	13	500	-1	1	41	2	90.2	2	Good
5	33	1000	26	500	-1	0	7	15	98.9	2	Good
6	22	626	5	-7	0	1	21	84	95.7	3	Good
7	19	628	10	-30	0	1	19	34	93.3	3	Good
8	67	1565	17	500	-1	4	26	29	92.1	3	Good
9	2	33	5	500	-1	0	1	0	38.9	3	Good
10	0	0	8	500	-1	0	2	0	28.3	3	Good
11	11	328	6	500	-1	1	0	3	35.4	3	Good
12	25	582	1	500	-1	0	0	0	60.9	3	Good
13	35	1031	20	119	1	2	27	115	84.5	3	Average
14	7	176	4	58	1	3	43	106	81.6	3	Average

Table C.4: Comparison of the categorization of daily airport operations at EWR with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	0	0	0	-24	0	1	11	126	98.7	2	Good
2	0	0	0	-18	0	0	43	169	98.5	2	Good
3	5	94	1	-15	0	1	0	147	99.7	2	Good
4	6	164	1	0	5	4	77	264	84.7	2	Bad
5	1	17	1	21	5	2	152	260	85.2	2	Bad
6	8	215	4	500	-1	0	1	0	9.9	3	Good
7	13	341	21	500	-1	0	0	0	13.2	3	Good
8	0	0	1	500	-1	0	1210	12	96.4	3	Good
9	41	1414	19	-35	2	2	55	77	80.9	3	Average

Table C.5: Comparison of the categorization of daily airport operations at IAD with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	1	16	1	500	-1	0	2	0	56.0	3	Good
2	0	0	1	500	-1	0	0	0	52.4	3	Good
3	17	473	21	500	-1	1	0	3	58.8	3	Good
4	1	15	0	500	-1	0	0	0	62.5	3	Good
5	50	1127	13	500	-1	1	33	27	95.3	3	Good
6	45	1313	28	500	-1	0	16	7	99.3	3	Good
7	50	1700	24	500	-1	1	51	34	97.0	3	Good
8	57	1333	21	500	-1	2	4	33	98.8	3	Good
9	38	1113	19	500	-1	1	31	37	96.1	3	Good
10	21	495	1	147	1	1	11	85	99.5	2	Good
11	18	340	4	-3	0	2	2	53	97.7	2	Good
12	24	745	10	500	-1	2	30	47	89.3	3	Average
13	16	284	12	75	1	1	55	59	84.9	3	Average

Table C.6: Comparison of the categorization of daily airport operations at JFK with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	2	67	1	500	-1	0	0	0	21.6	3	Good
2	17	379	2	500	-1	0	18	0	55.3	3	Good
3	0	0	0	500	-1	0	0	2	39.2	3	Good
4	143	3007	9	-6	3	3	0	95	96.4	3	Good
5	50	2010	28	500	-1	1	49	30	98.4	3	Good
6	73	2031	7	-1	0	2	87	50	95.1	3	Good
7	80	1773	3	-7	0	2	0	40	99.5	3	Good
8	88	2120	6	63	3	2	117	168	84.5	3	Average
9	77	1409	12	114	2	2	96	100	99.0	3	Average
10	66	1476	8	86	2	1	18	124	85.3	3	Average
11	183	4192	8	107	2	3	269	111	86.7	3	Average

Table C.7: Comparison of the categorization of daily airport operations at LGA with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	4	168	2	500	-1	0	0	0	12.6	3	Good
2	0	0	0	500	-1	0	0	0	7.8	3	Good
3	0	0	0	500	-1	0	0	0	2.1	3	Good
4	2	102	1	500	-1	0	0	0	3.7	3	Good
5	88	2598	44	500	-1	1	25	32	96.6	3	Good
6	62	1840	47	500	-1	2	6	21	40.5	3	Good
7	38	1273	16	-5	0	1	61	59	92.7	2	Good
8	49	1790	4	-10	2	2	52	3	93.2	2	Good
9	79	1922	11	118	2	2	42	237	90.9	3	Average
10	9	181	11	215	2	1	84	125	69.4	3	Average
11	39	1282	32	200	3	3	51	115	60.2	3	Average

Table C.8: Comparison of the categorization of daily airport operations at PHL with Experiment 1 and Operational Service Performance Criteria

Day	Airborne Holdings (Aircraft)	Airborne Holdings (Minutes)	Diversions	GDP Lead-in Time (minutes)	GDP Revisions	Ground Stops	Dep. Delays	TMI to Airport Delays	Completion Rate (%)	Cluster	OSPC Category
1	0	0	4	500	-1	0	0	0	18.9	3	Good
2	18	605	22	500	-1	1	0	6	36.2	3	Good
3	4	158	27	500	-1	1	0	13	39.1	3	Good
4	0	0	2	500	-1	0	0	1	26.5	3	Good
5	48	1075	25	500	-1	1	0	14	97.2	3	Good
6	47	1062	9	500	-1	2	14	19	97.4	3	Good
7	53	1312	18	-7	0	1	31	58	93.2	3	Good
8	2	45	2	-18	1	2	8	100	91.7	2	Good
9	7	157	14	-13	0	2	41	61	95.9	3	Good
10	0	0	2	-19	1	2	68	77	84.5	2	Good
11	0	0	0	161	2	2	201	220	81.8	2	Good
12	0	0	1	-5	1	2	213	270	83.5	2	Good

APPENDIX D

RANKING OF PREDICTOR IMPORTANCE

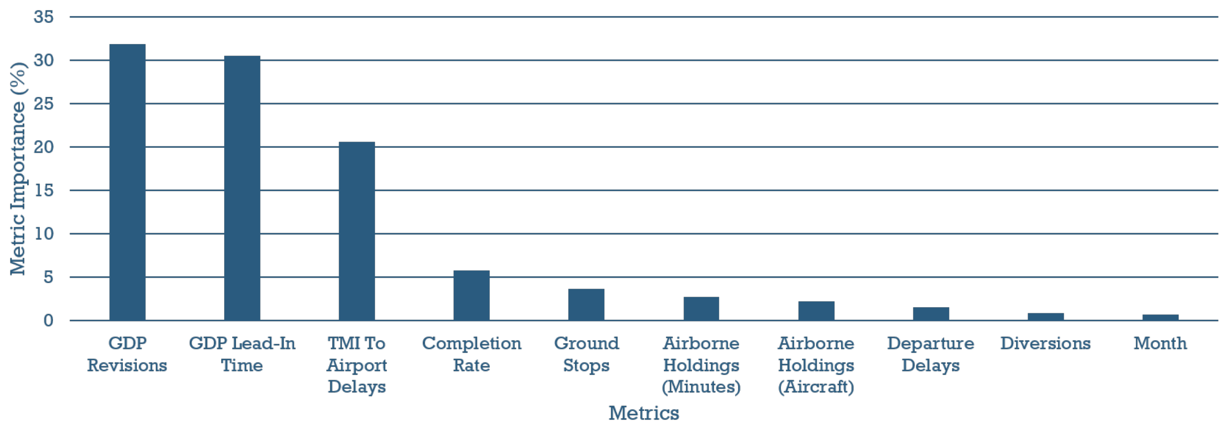

Figure D.1: Ranking of predictor importance (BOS)

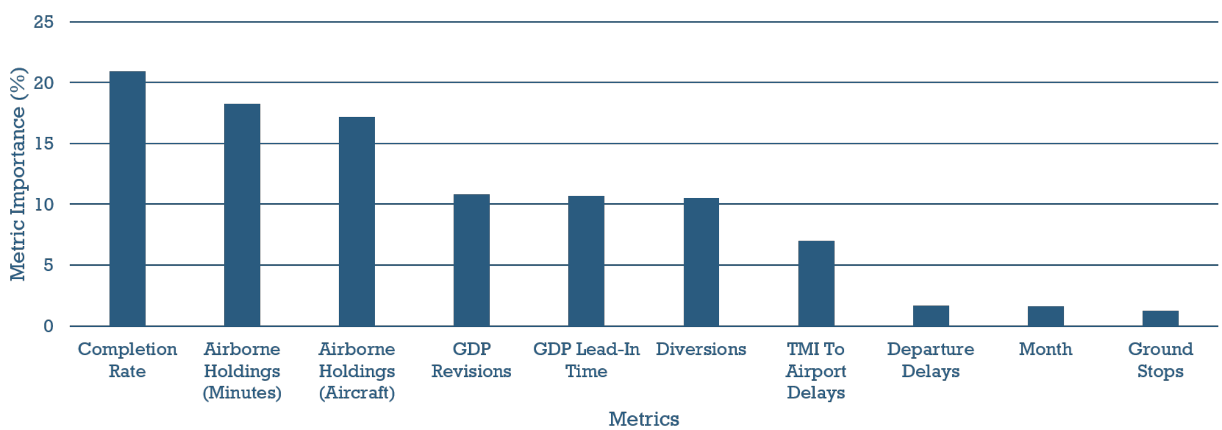

Figure D.2: Ranking of predictor importance (BWI)

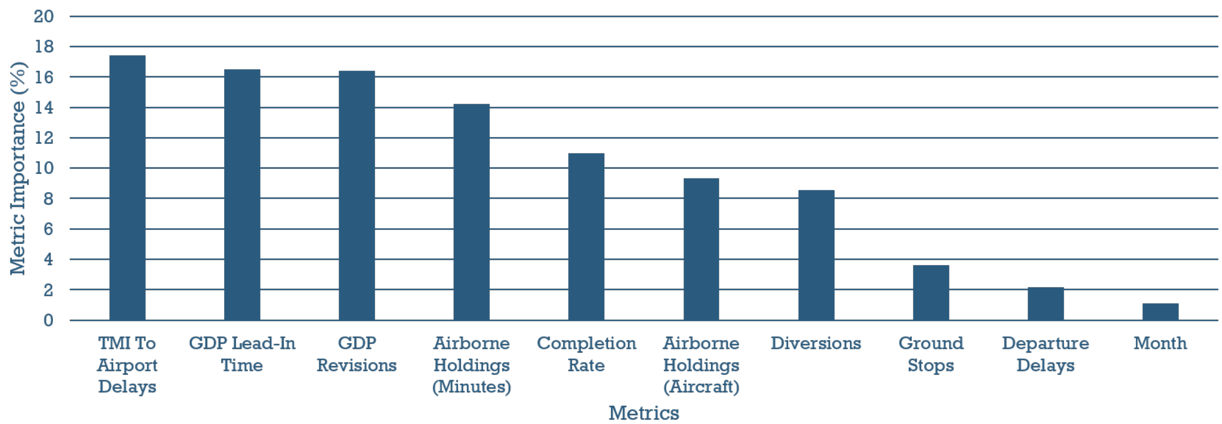

Figure D.3: Ranking of predictor importance (DCA)

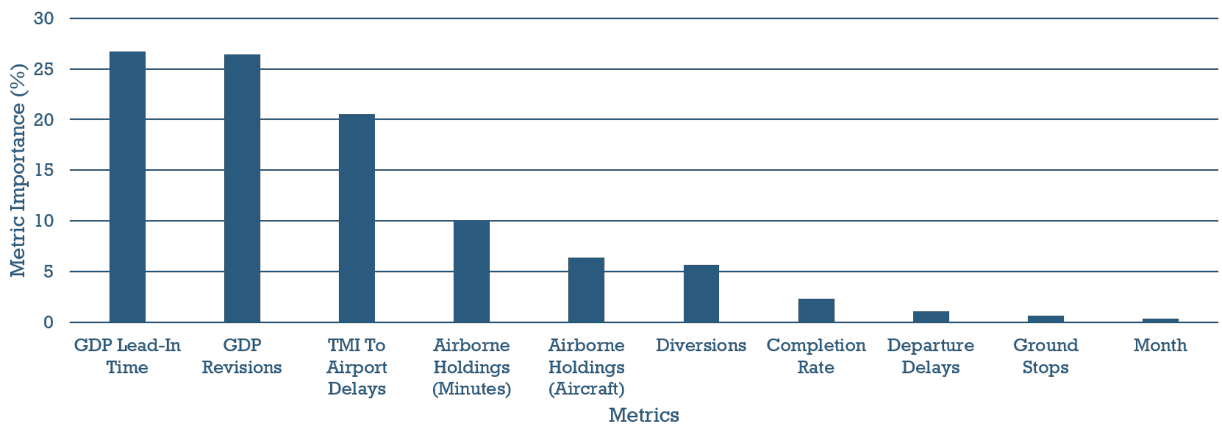

Figure D.4: Ranking of predictor importance (EWR)

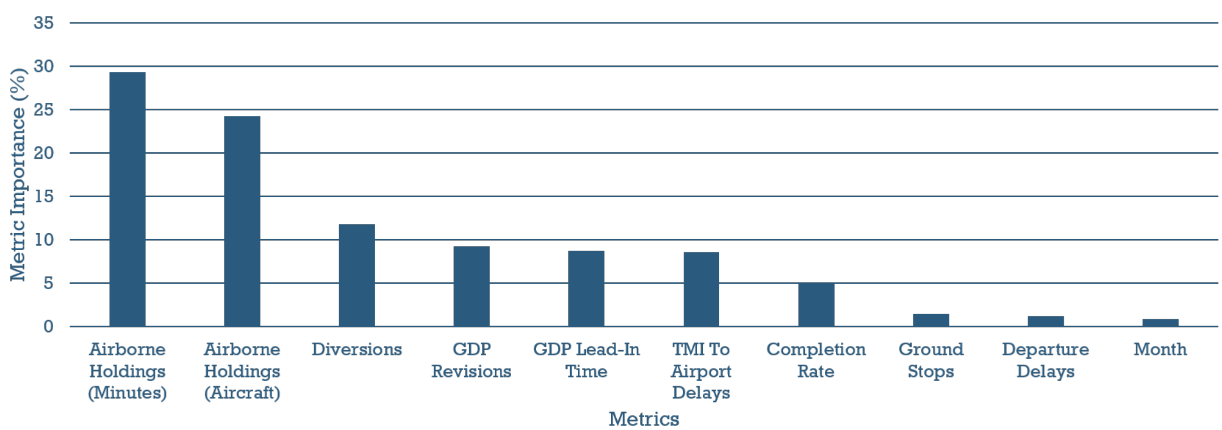

Figure D.5: Ranking of predictor importance (IAD)

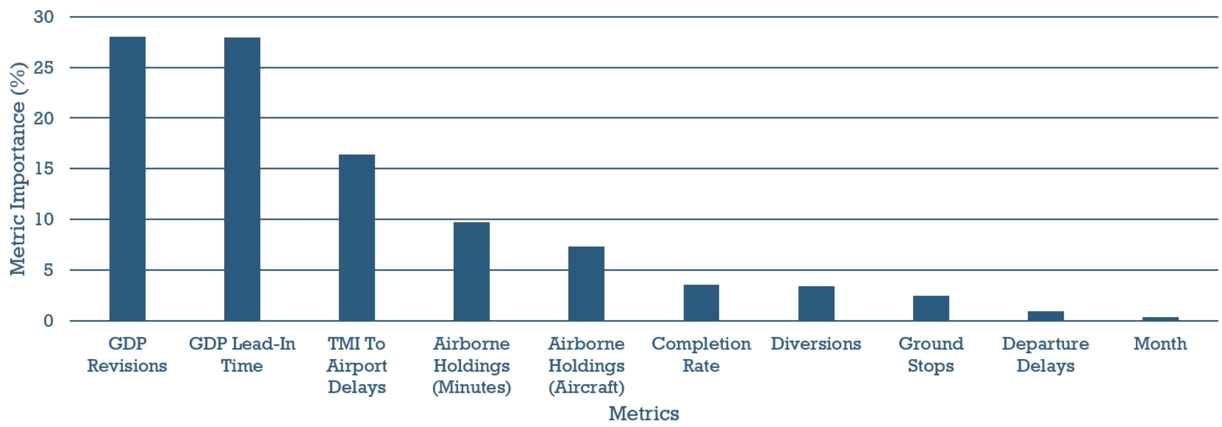

Figure D.6: Ranking of predictor importance (JFK)

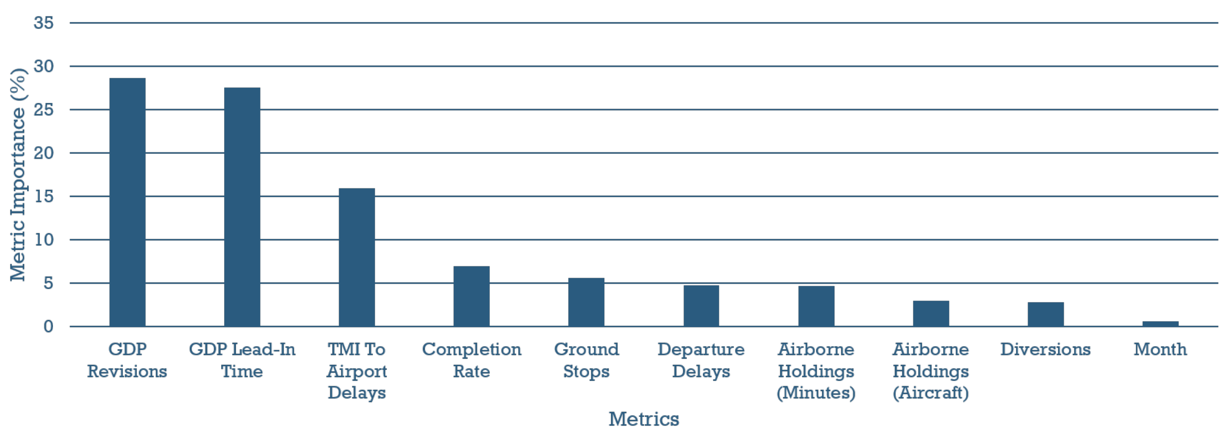

Figure D.7: Ranking of predictor importance (LGA)

APPENDIX E

DECISION TREES

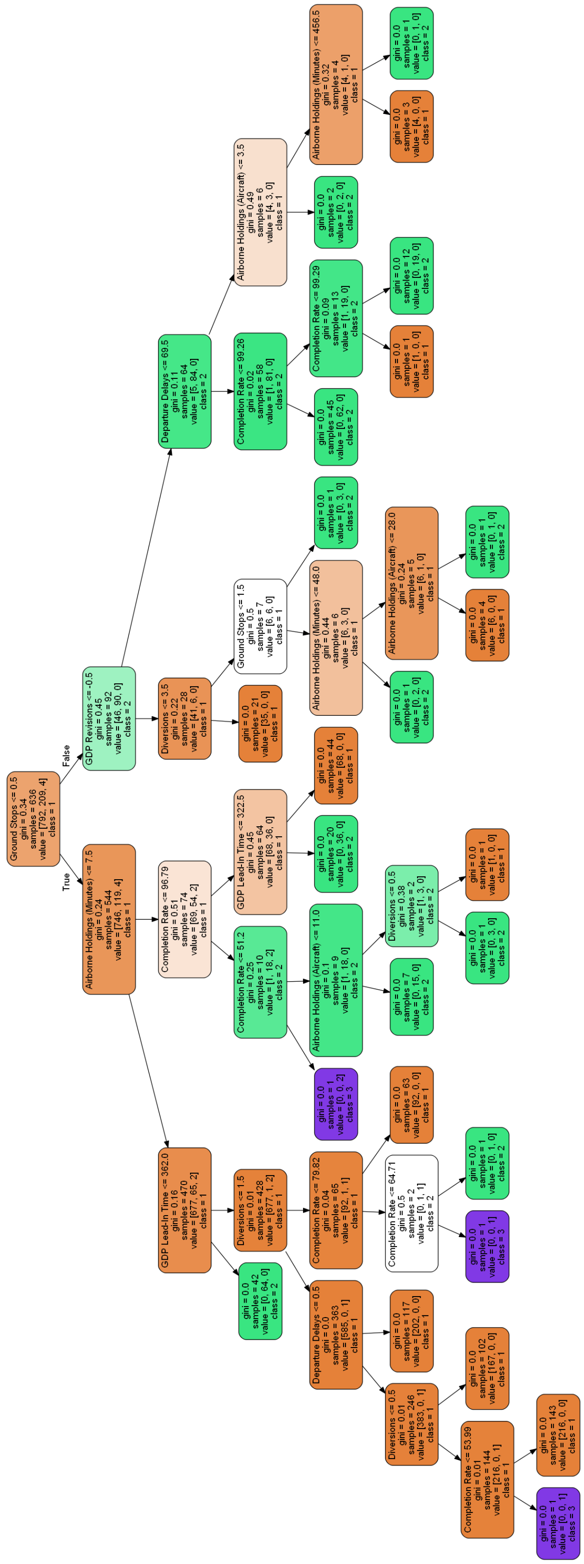

Figure E.1: Decision Tree (BOS)

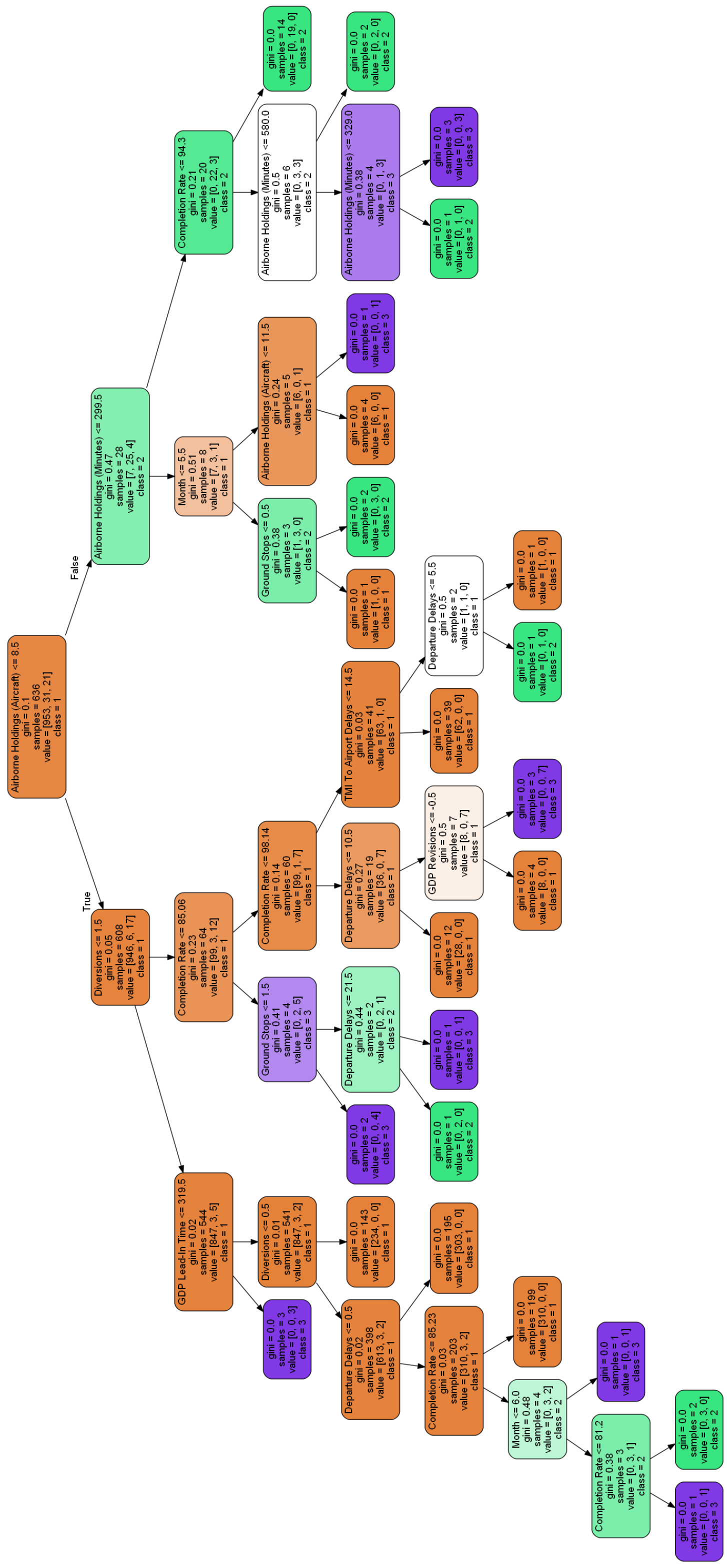

Figure E.2: Decision Tree (BWI)

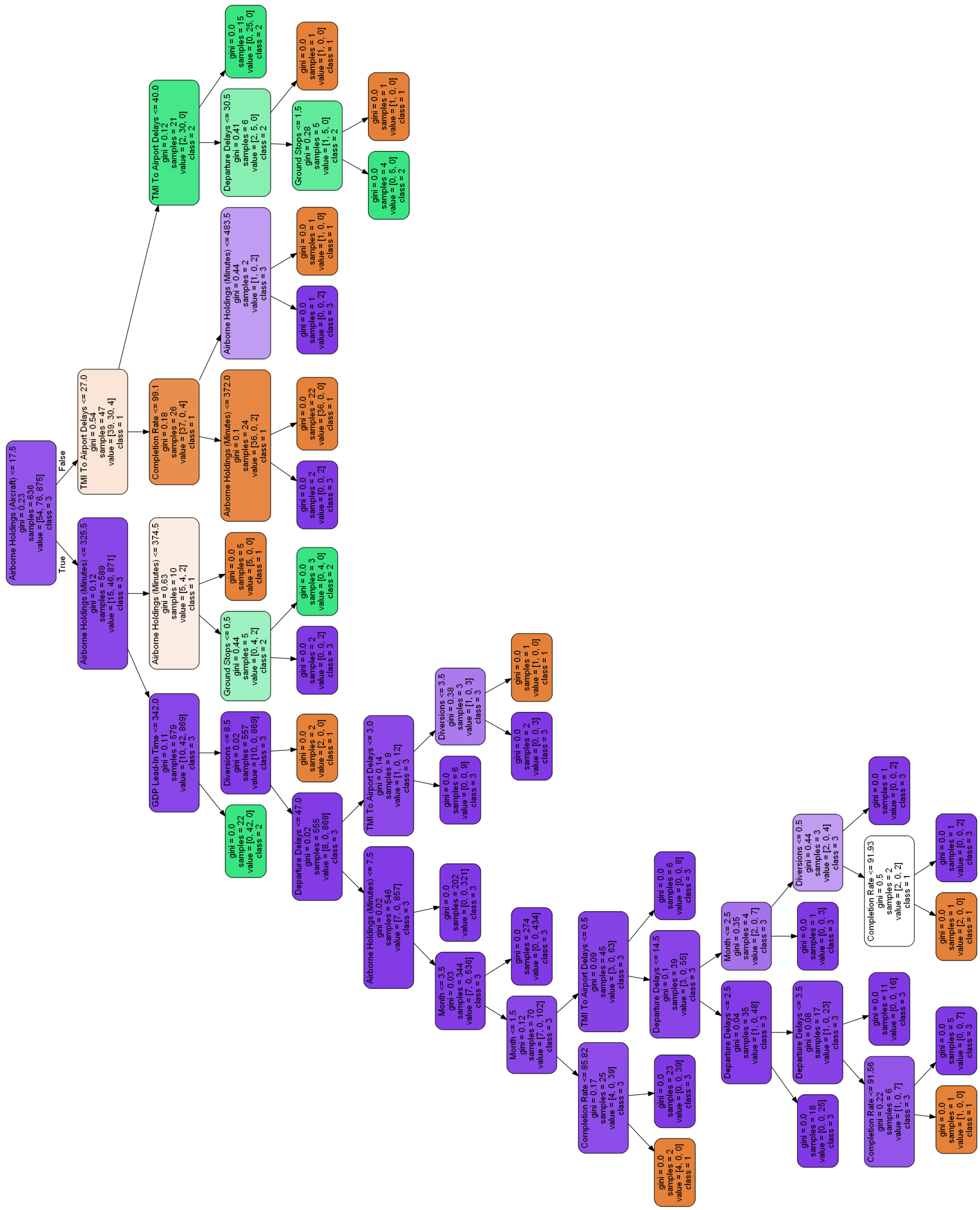

Figure E.3: Decision Tree (DCA)

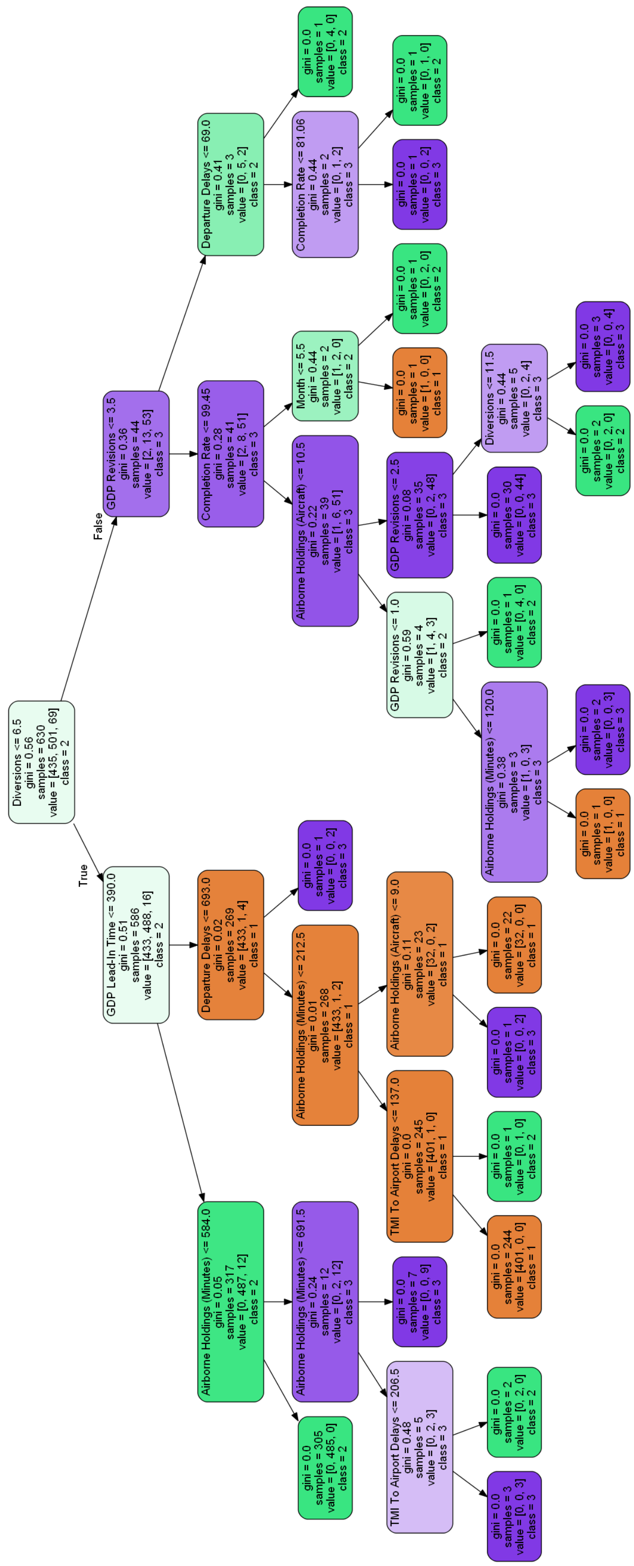

Figure E.4: Decision Tree (EWR)

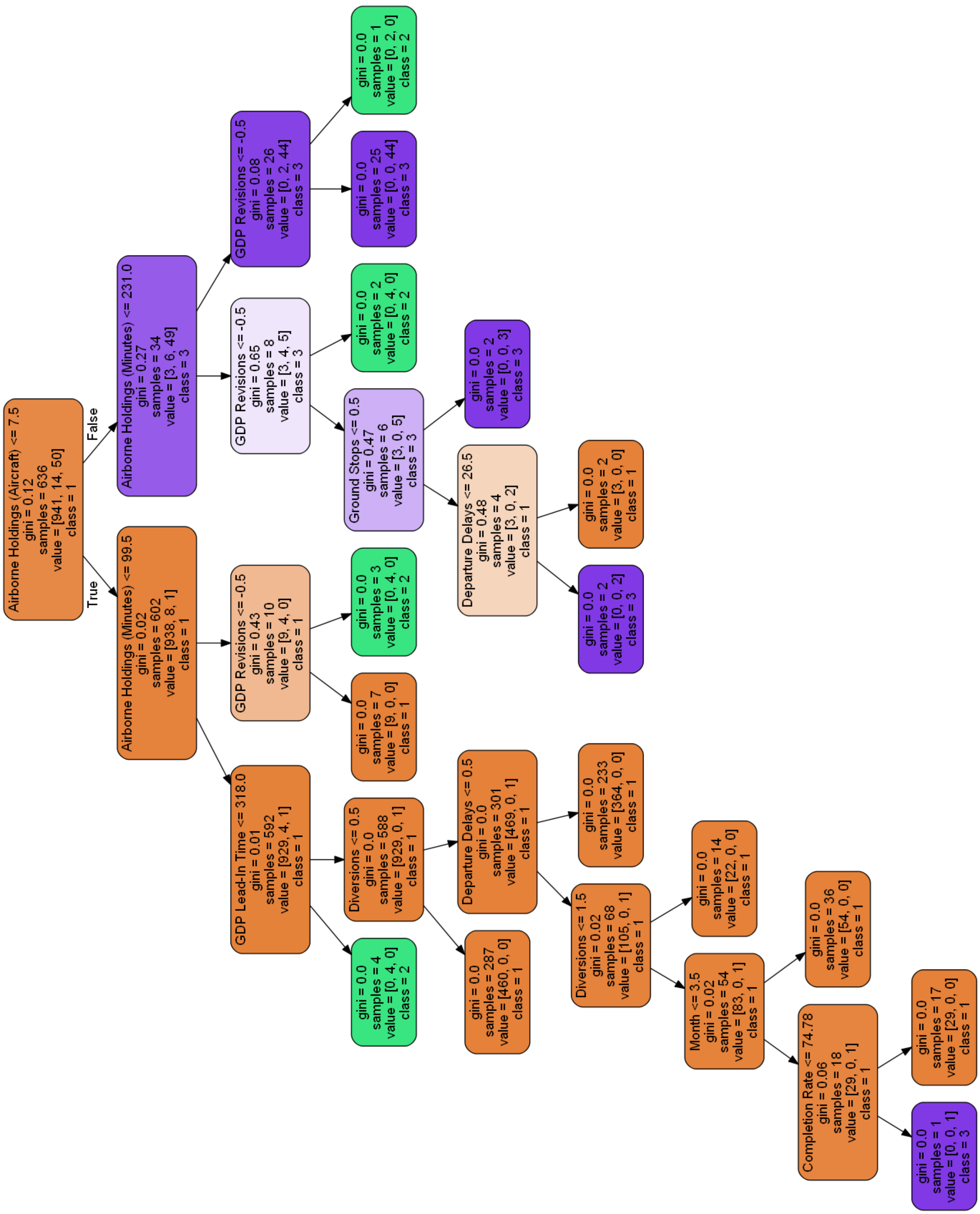

Figure E.5: Decision Tree (IAD)

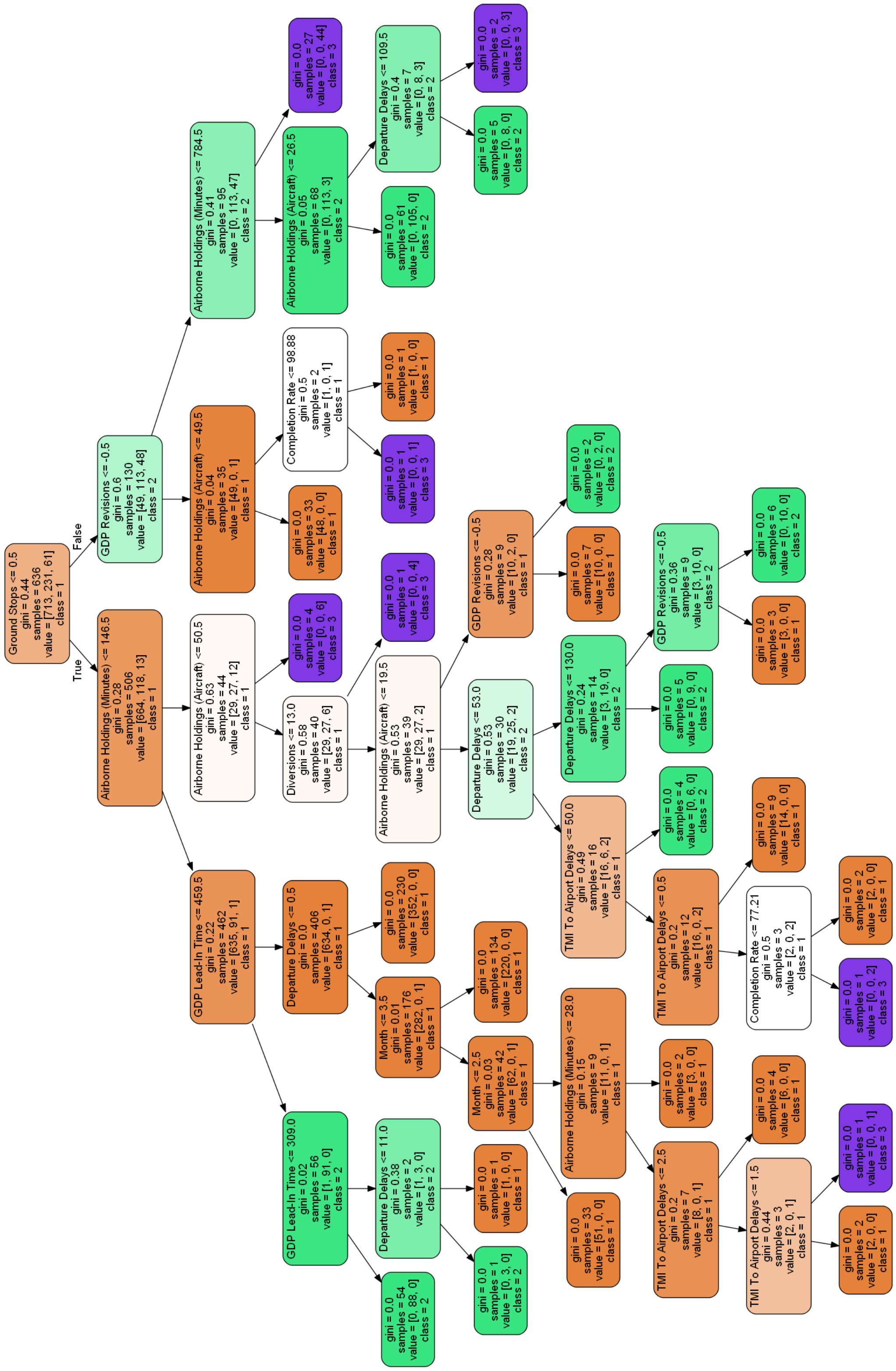

Figure E.6: Decision Tree (JFK)

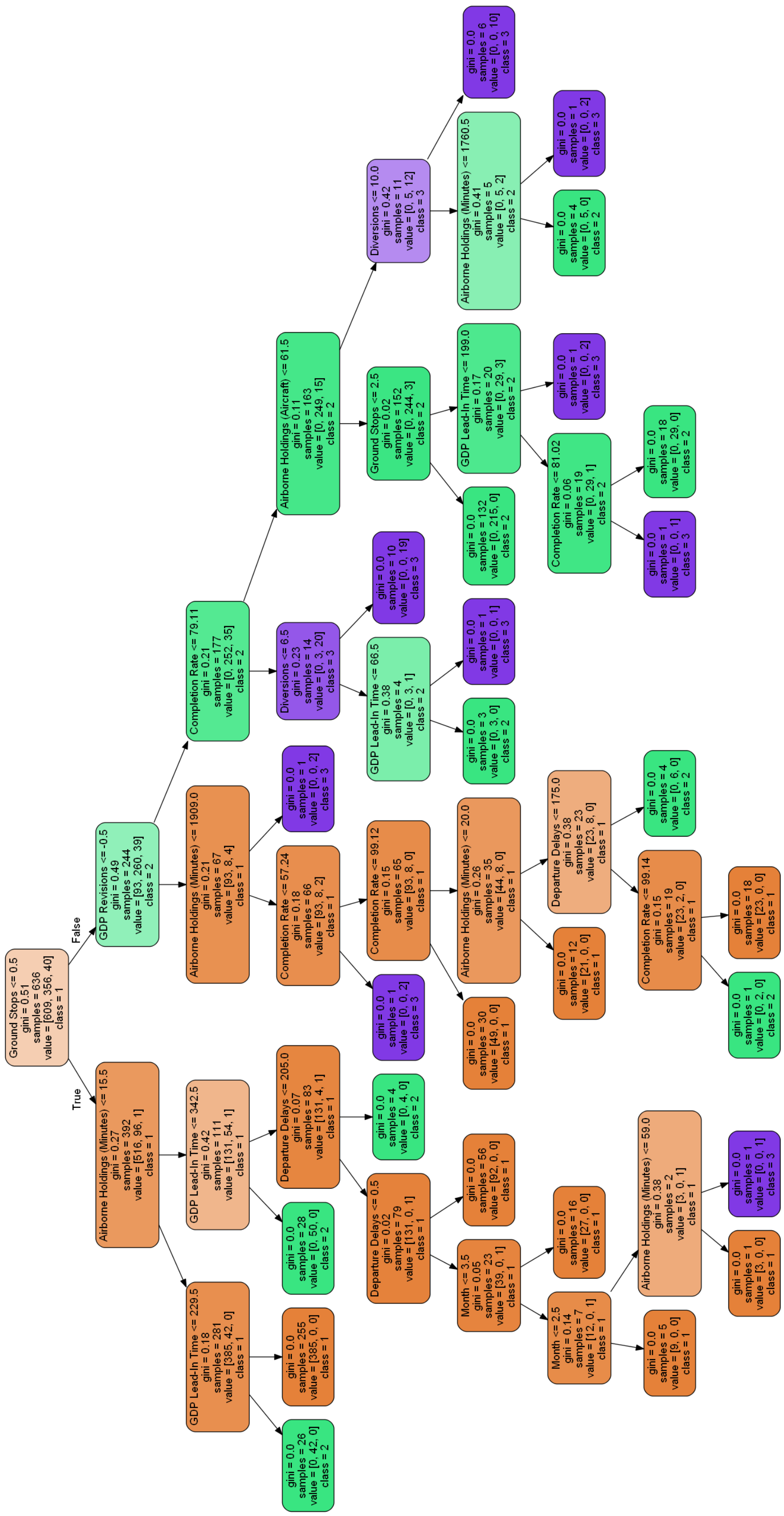

Figure E.7: Decision Tree (LGA)

REFERENCES

[1] International Air Transport Association and others. "Air passenger market analysis". In: *Montreal: International Air Transport Association* (2014).

[2] Tiago Mendes Dantas, Fernando Luiz Cyrino Oliveira, and Hugo Miguel Varela Repolho. "Air transportation demand forecast through Bagging Holt Winters methods". In: *Journal of Air Transport Management* 59 (2017), pp. 116 –123.

[3] DP Singh, Narendra N Dalei, and T Bangar Raju. "Forecasting investment and capacity addition in Indian airport infrastructure: Analysis from post-privatization and post-economic regulation era". In: *Journal of Air Transport Management* 53 (2016), pp. 218–225.

[4] Junwook Chi and Jungho Baek. "Dynamic relationship between air transport demand and economic growth in the United States: A new look". In: *Transport Policy* 29 (2013), pp. 257 –260.

[5] U.S. Department of Labor, Bureau of Labor Statistics. *Labor Force Statistics from the Current Population Survey*. `http://www.bls.gov/data/`. 2016.

[6] Federal Aviation Administration. *The Economic Impact of Civil Aviation on the U.S. Economy*. `https://www.faa.gov/air_traffic/publications/media/2016-economic-impact-report_FINAL.pdf`. 2016.

[7] Matthew Russell. *Economic Productivity and Air Transportation: Multifactor and Labor Productivity Trends, 1990-2014*. June 2016.

[8] Federal Aviation Administration. *National Airspace System Overview*. `https://www.faa.gov/air_traffic/nas/nynjphl_redesign/documentation/feis/media/Appendix_A-National_Airspace_System_Overview.pdf`. 2020.

[9] K. L. Clark et al. "Resilience of the U.S. National Airspace System Airport Network". In: *IEEE Transactions on Intelligent Transportation Systems* 19.12 (Dec. 2018), pp. 3785–3794.

[10] D. Miller. "An ontology for future airspace system architectures". In: *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. Sept. 2017, pp. 1–8.

[11] G. Burke. “Shaping the National Airspace System for the 21st century”. In: *16th DASC. AIAA/IEEE Digital Avionics Systems Conference. Reflections to the Future. Proceedings*. Vol. 1. Oct. 1997, pp. 0.4–1.

[12] S. L. Waslander, R. L. Raffard, and C. J. Tomlin. “Toward efficient and equitable distributed air traffic flow control”. In: *2006 American Control Conference*. June 2006, p. 6.

[13] Federal Aviation Administration. *Air Traffic By The Numbers*. `https://www.faa.gov/air_traffic/by_the_numbers/`. 2019.

[14] X. Liu et al. “Web-Based Intelligent Computational Argumentation Based Conflict Resolution in Air Traffic Management”. In: *2010 10th IEEE/IPSJ International Symposium on Applications and the Internet*. July 2010, pp. 117–120.

[15] Bojia Ye et al. “Analysis of alternative collaborative route selection strategies based on cost and throughput”. In: *2012 Integrated Communications, Navigation and Surveillance Conference*. Apr. 2012, pp. I2–1–I2–12.

[16] Federal Aviation Administration. *NAS Operations*. `https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/systemops/nas_ops/`. 2020.

[17] Ghislain Dard. “APPLICATION OF DATA FUSION AND MACHINE LEARNING TO THE ANALYSIS OF THE RELEVANCY OF RECOMMENDED FLIGHT REROUTES”. MA thesis. Georgia Institute of Technology, 2019.

[18] C. Yao et al. “Aircraft conflict resolution cataloguer”. In: *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. Sept. 2017, pp. 1–10.

[19] G. Tokadli, A. Marzuoli, and E. Boidot. “Resilience of the national airspace system: A case study of the fire at the Chicago ARTCC”. In: *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. Sept. 2016, pp. 1–10.

[20] Richard A Lamb and Jack J Gilbert. *Instrument flight rules (IFR) training device*. US Patent 5,252,069. Oct. 1993.

[21] J. R. Clymer and V. Harrison. “Simulation of air traffic control at a VFR airport using OpEMCSS”. In: *Proceedings. The 21st Digital Avionics Systems Conference*. Vol. 1. Oct. 2002, 3B1–3B1.

[22] R. J. Marske, D. M. Hechtman, and M. J. Yablonski. “Developing an En Route Facility Traffic Forecasting Capability Using Origin-Destination Segment Passenger Demand”. In:

2019 Integrated Communications, Navigation and Surveillance Conference (ICNS). Apr. 2019, pp. 1–6.

[23] avweb.com. *artcc-map*. https://www.avweb.com/flight-safety/faa-regs/air-route-traffic-control/. 2020.

[24] Bureau of Transportation Statistics. *Airline On-Time Statistics and Delay Causes*. https://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp?pn=1. 2020.

[25] Eugene Mangortey. "PREDICTING THE OCCURRENCE OF GROUND DELAY PROGRAMS AND THEIR IMPACT ON AIRPORT AND FLIGHT OPERATIONS". MA thesis. Georgia Institute of Technology, 2019.

[26] Peterson, Everett B. "The Economic Cost of Airline Flight Delay". In: *Journal of Transport Economics and Policy* 47 (1 2013), pp. 107–121.

[27] Ball, Michael and Barnhart, Cynthia. *Total Delay Impact Study: A Comprehensive Assessment of the Costs and Impacts of Flight Delay in the United States*. NEXTOR, 2010.

[28] Coronavirus Pneumonia Emergency Response Epidemiology Novel et al. "The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China". In: *Zhonghua liu xing bing xue za zhi= Zhonghua liuxingbingxue zazhi* 41.2 (2020), p. 145.

[29] Majdouline Mhalla. "The Impact of Novel Coronavirus (COVID-19) on the Global Oil and Aviation Markets". In: *Journal of Asian Scientific Research* 10.2 (2020), p. 96.

[30] Maria Nicola et al. "The socio-economic implications of the coronavirus and COVID-19 pandemic: a review". In: *International Journal of Surgery* (2020).

[31] Stefano Maria Iacus et al. "Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socio-economic impact". In: *Safety Science* (2020), p. 104791.

[32] Cynthia Barnhart et al. "Demand and capacity management in air transportation". In: *EURO Journal on Transportation and Logistics* 1.1-2 (2012), pp. 135–155.

[33] Federal Aviation Administration. *FACT3: Airport Capacity Needs in the National Airspace System*. https://www.faa.gov/airports/planning_capacity/media/FACT3-Airport-Capacity-Needs-in-the-NAS.pdf. 2015.

[34] Marcial Marazzo, Rafael Scherre, and Elton Fernandes. “Air transport demand and economic growth in Brazil: A time series analysis”. In: *Transportation Research Part E: Logistics and Transportation Review* 46.2 (2010), pp. 261–269.

[35] Abdullah O Ba-Fail et al. “The determinants of domestic air travel demand in the Kingdom of Saudi Arabia”. In: *Journal of Air Transportation World Wide* 5.2 (2000), pp. 72–86.

[36] Matthew G Karlaftis. “Demand forecasting in regional airports: dynamic Tobit models with GARCH errors”. In: *Sitraer* 7 (2008), pp. 100–111.

[37] Dipasis Bhadra. “Demand for air travel in the United States: Bottom-up econometric estimation and implications for forecasts by origin-destination pairs”. In: *AIAA’s Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*. 2002, p. 5861.

[38] Mott McDonald. “Annual Analyses of the EU Air Transport Market 2016”. In: (2017), p. 7.

[39] Bureau of Transportation Statistics. *Transportation Accidents by Mode*. `https://www.bts.gov/content/transportation-accidents-mode`. 2020.

[40] Mariya A Ishutkina and R John Hansman. *Analysis of the interaction between air transportation and economic activity: a worldwide perspective*. Tech. rep. 2011.

[41] James M Lyneis. “System dynamics for market forecasting and structural analysis”. In: *System Dynamics Review: The Journal of the System Dynamics Society* 16.1 (2000), pp. 3–25.

[42] Jing Xiong. “Revealed Preference of Airlines’ Behavior under Air Traffic Management Initiative”. PhD thesis. University of California, Berkeley, 2010.

[43] Eugene Mangortey et al. “Development of a Data Fusion Framework to support the Analysis of Aviation Big Data”. In: *AIAA Scitech 2019 Forum, AIAA SciTech Forum, (AIAA 2019-1538)* (2019).

[44] Mangortey, E., Pinon, O., Puranik, T., and Mavris, D. “Predicting The Occurrence of Weather And Volume Related Ground Delay Programs”. In: *AIAA AVIATION Forum* (2019).

[45] Mangortey, E., Puranik, T., Pinon, O., and Mavris, D. “Prediction and Analysis of Ground Stops with Machine Learning”. In: *AIAA Science and Technology Forum (AIAA Scitech)* (2020).

[46] Dard, G., Mangortey, E., Pinon, O., and Mavris, D. “Application Of Data Fusion And Machine Learning To The Analysis Of The Relevance Of Recommended Flight Reroutes”. In: *AIAA AVIATION Forum* (2019).

[47] Hanif D Sherali et al. “Integrating slot exchange, safety, capacity, and equity mechanisms within an airspace flow program”. In: *Transportation Science* 45.2 (2011), pp. 271–284.

[48] Jimmy Krozel, Ray Jakobovits, and Steve Penny. “An algorithmic approach for airspace flow programs”. In: *Air Traffic Control Quarterly* 14.3 (2006), pp. 203–229.

[49] Khaled Abdelghany, Ahmed Abdelghany, and Tim Niznik. “Managing severe airspace flow programs: The Airlines’ side of the problem”. In: *Journal of Air Transport Management* 13.6 (2007), pp. 329–337.

[50] Amy Spencer et al. “Future concepts for collaborative traffic flow management in the national airspace system”. In: *2006 ieee/aiaa 25TH Digital Avionics Systems Conference*. IEEE. 2006, pp. 1–8.

[51] K. Sheth and S. Gutierrez-Nolasco. “Development of miles-in-trail passback restrictions for air traffic management”. In: *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*. Oct. 2014, pp. 1D3–1–1D3–11.

[52] P. T. R. Wang, C. R. Wanke, and F. P. Wieland. “Modeling time and space metering of flights in the National Airspace System”. In: *Proceedings of the 2004 Winter Simulation Conference, 2004.* Vol. 2. Dec. 2004, 1299–1304 vol.2.

[53] K. . White. “A study of Chicago O’Hare’s March 2005 restrictions using an MIT analysis tool”. In: *24th Digital Avionics Systems Conference*. Vol. 1. Oct. 2005, 3.E.5–31.

[54] Michael Ball and Lulli Guglielmo. “Ground Delay Programs: Optimizing over the Included Flight Set Based on Distance”. In: *14th AIAA Aviation Technology, Integration, and Operations Conference* (2014).

[55] Avijit Mukherjee, Shon Grabbe, and Banavar Sridhar. “Predicting Ground Delay Program At An Airport Based On Meteorological Conditions”. In: *Air Traffic Control Quarterly, vol. 12, no. 1* (2004).

[56] Mark Hansen, Avijit Mukherjee, and Shon Grabbe. “Ground Delay Program Planning under Uncertainty in Airport Capacity”. In: *Transportation Planning and Technology, vol. 35, no. 6* (2012).

[57] David Smith and Sherry Lance. “Decision Support Tool for Predicting Aircraft Arrival Rates from Weather Forecasts”. In: *2008 Integrated Communications, Navigation and Surveillance Conference* (2008).

[58] Y. Wang. “Analysis and prediction of weather impacted ground stop operations”. In: *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*. Oct. 2014, 7A2–1–7A2–14.

[59] Mangortey, E., Bleu Laine, M., Puranik, T., Pinon, O., and Mavris, D. “Application of Machine Learning to the Analysis and Prediction of the Coincidence of Ground Delay Programs and Ground Stops”. In: *AIAA Science and Technology Forum (AIAA Scitech)* (2020).

[60] J. R. Armstrong and D. Henry. “Competencies required for successful acquisition of the Next Generation Air Transportation System”. In: *2009 3rd Annual IEEE Systems Conference*. Mar. 2009, pp. 419–424.

[61] P. Park et al. “High Confidence Networked Control for Next Generation Air Transportation Systems”. In: *IEEE Transactions on Automatic Control* 59.12 (Dec. 2014), pp. 3357–3372.

[62] K. Sampigethaya et al. “Future E-Enabled Aircraft Communications and Security: The Next 20 Years and Beyond”. In: *Proceedings of the IEEE* 99.11 (Nov. 2011), pp. 2040–2055.

[63] Radio Technical Commission for Aeronautics. *Minimum Aviation System Performance Standards for Automatic Dependent Surveillance Broadcast (ADS-S)*. RTCA, Incorporated, 2002.

[64] Joint Planning et al. “Concept of operations for the next generation air transportation system”. In: *Technical Report*. Citeseer, 2007.

[65] Federal Aviation Administration. *Modernization of U.S. Airspace*. `https://www.faa.gov/nextgen/`. 2019.

[66] J. Song et al. “Application of maritime AIS (Automatic Identification System) to ADS-B (Automatic Dependent Surveillance — Broadcast) transceiver”. In: *ICCAS 2010*. Oct. 2010, pp. 2233–2237.

[67] P. Roy and D. Dawn. “A high-power and high-efficiency CMOS VCO for Automatic Dependent Surveillance-Broadcast (ADS-B) system”. In: *2015 IEEE International Conference on Electro/Information Technology (EIT)*. May 2015, pp. 511–514.

[68] S. S. Bharadwaj, D. Yadav, and M. Tiwari. “Design of an Automatic Dependent Surveillance - Broadcast (ADS-B) Monopole Antenna with Modified Ground Plane”. In: *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*. Feb. 2019, pp. 1–3.

[69] T. Hnilicka, T. Zalabsky, and V. Krcmar. “Design of broadband antenna system for automatic dependent surveillance - broadcast”. In: *2017 Conference on Microwave Techniques (COMITE)*. Apr. 2017, pp. 1–4.

[70] Federal Aviation Administration. *Airport Surface Detection Equipment, Model X (ASDE-X)*. `https://www.faa.gov/air_traffic/technology/asde-x/`. 2014.

[71] A. Srivastava. “Improving departure taxi time predictions using ASDE-X surveillance data”. In: *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. Oct. 2011, 2B5–1–2B5–14.

[72] M. Farneth. “ASSC AeroMACS system”. In: *2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC)*. Oct. 2013, pp. 1–17.

[73] Federal Aviation Administration. *ADS-B Airport Surface Surveillance Capability (ASSC)*. `https://www.faa.gov/nextgen/programs/adsb/atc/assc/`. 2019.

[74] S. Mondoloni. “Trajectory-based operations — Robust planning under trajectory uncertainty”. In: *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. Sept. 2016, pp. 1–10.

[75] T. Prevot and P. U. Lee. “Trajectory-Based Complexity (TBX): A modified aircraft count to predict sector complexity during trajectory-based operations”. In: *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. Oct. 2011, 3A3–1–3A3–15.

[76] S. Torres. “Accuracy impact of trajectory sampling and representation for Trajectory-Based Operations”. In: *2013 IEEE/AIAA 32nd Digital Avionics Systems Conference (DASC)*. Oct. 2013, pp. 5D2–1–5D2–16.

[77] S. Torres et al. “En-Route Automation Modernization (ERAM) trajectory model evolution to support trajectory-based operations (TBO)”. In: *2015 IEEE/AIAA 34th Digital Avionics Systems Conference (DASC)*. Sept. 2015, 1A6–1–1A6–15.

[78] J. Klein and S. Morey. “Use of eram SWIM for NAS system enhancements”. In: *2011 Integrated Communications, Navigation, and Surveillance Conference Proceedings*. May 2011, E1–1–E1–8.

[79] M. Paglione et al. "Operational impact of trajectory prediction accuracy on air traffic automation tools". In: *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. Sept. 2017, pp. 1–10.

[80] M. Paglione et al. "Trajectory predictor performance experiment using Required Time of Arrival during descent". In: *2011 IEEE/AIAA 30th Digital Avionics Systems Conference*. Oct. 2011, pp. 2D3–1–2D3–10.

[81] Federal Aviation Administration. *NextGen Implementation Plan 2018-19*. `https://www.faa.gov/nextgen/media/NextGen_Implementation_Plan-2018-19.pdf`. 2019.

[82] R. Klein et al. "Aircraft access to SWIM — A collaborative ATM pathfinder to support a TBO environment". In: *2015 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. Apr. 2015, Q4–1–Q4–10.

[83] R. J. Stamm et al. "Advances in the Use of NAS Infrastructure and GBDAA for UAS Operations". In: *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. Sept. 2018, pp. 1–9.

[84] A. E. Leu, R. E. Snow, and L. C. Monticone. "SWIM optimization methodolgy". In: *2015 Integrated Communication, Navigation and Surveillance Conference (ICNS)*. Apr. 2015, G1–1–G1–14.

[85] K. Witzberger et al. "NextGen technologies on the FAA's standard terminal automation replacement system". In: *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*. Oct. 2014, pp. 1C1–1–1C1–15.

[86] J. Mercer et al. "An evaluation of airborne spacing in the terminal area". In: *24th Digital Avionics Systems Conference*. Vol. 1. Oct. 2005, 2.A.3–21.

[87] K. A. Weiss, E. C. Ong, and N. G. Leveson. "Supporting the development of new air traffic management software". In: *Proceedings. The 21st Digital Avionics Systems Conference*. Vol. 1. Oct. 2002, 2E3–2E3.

[88] X. Lu et al. "Air-Ground System Wide Information Management to Achieve Safe Flight Operation". In: *2019 IEEE 19th International Symposium on High Assurance Systems Engineering (HASE)*. Jan. 2019, pp. 42–49.

[89] Q. Ming and L. Songtao. "Overview of System Wide Information Management and Security Analysis". In: *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*. Mar. 2017, pp. 191–194.

[90] X. Lu and T. Koga. “System Wide Information Management for Heterogeneous Information Sharing and Interoperability”. In: *2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS)*. Mar. 2017, pp. 199–204.

[91] R. Raju, R. Mital, and D. Finkelsztein. “Data Lake Architecture for Air Traffic Management”. In: *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. Sept. 2018, pp. 1–6.

[92] Federal Aviation Administration. *Collaborative Air Traffic Management*. `https://www.faa.gov/nextgen/cip/catm/`. 2017.

[93] Federal Aviation Administration. *Environment and Energy*. `https://www.faa.gov/nextgen/cip/eande/`. 2018.

[94] Federal Aviation Administration. *Improved Approaches and Low-Visibility Operations*. `https://www.faa.gov/nextgen/cip/ialvo/`. 2017.

[95] Federal Aviation Administration. *Improved Multiple Runway Operations*. `https://www.faa.gov/nextgen/cip/imro/`. 2017.

[96] Federal Aviation Administration. *Improved Surface Operations*. `https://www.faa.gov/nextgen/cip/iso/`. 2017.

[97] Federal Aviation Administration. *NAS Infrastructure*. `https://www.faa.gov/nextgen/cip/ni/`. 2017.

[98] Federal Aviation Administration. *On-Demand NAS Information*. `https://www.faa.gov/nextgen/cip/odnasi/`. 2017.

[99] Federal Aviation Administration. *Performance Based Navigation*. `https://www.faa.gov/nextgen/cip/pbn/`. 2017.

[100] Federal Aviation Administration. *Separation Management*. `https://www.faa.gov/nextgen/cip/sm/`. 2017.

[101] Federal Aviation Administration. *System Safety Management*. `https://www.faa.gov/nextgen/cip/ssm/`. 2017.

[102] Federal Aviation Administration. *Time Based Flow Management*. `https://www.faa.gov/nextgen/cip/tbfm/`. 2017.

[103] J. DeArmon, A. Mahashabde, and B. Hogan. "Departure flow efficiency and the identification of causes for inefficiencies". In: *2014 IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC)*. Oct. 2014, 7A4–1–7A4–8.

[104] Federal Aviation Administration. *TAER*. `https://aspmhelp.faa.gov/index.php/SAER`. 2019.

[105] Federal Aviation Administration. *TAER*. `https://aspmhelp.faa.gov/index.php/TAER`. 2019.

[106] Federal Aviation Administration. *OPSNET Reports: Definitions of Variables*. `https://aspmhelp.faa.gov/index.php/OPSNET_Reports:_Definitions_of_Variables`. 2018.

[107] Bengi Manley and Lance Sherry. "Analysis of performance and equity in ground delay programs". In: *Transportation research part c: Emerging technologies* 18.6 (2010), pp. 910–920.

[108] Kan Chang et al. "Enhancements to the FAA ground-delay program under collaborative decision making". In: *Interfaces* 31.1 (2001), pp. 57–76.

[109] Thomas Vossen and Michael Ball. "Optimization and mediated bartering models for ground delay programs". In: *Naval research logistics (NRL)* 53.1 (2006), pp. 75–90.

[110] Sridhar Banavar et al. "Integration of Traffic Flow Management Decisions". In: *AIAA Guidance, Navigation, and Control Conference and Exhibit, Guidance, Navigation, and Control and Co-located Conferences* (2002).

[111] Federal Aviation Administration. *ASPM Efficiency: Definitions of Variables*. `https://aspmhelp.faa.gov/index.php/ASPM_Efficiency:_Definitions_of_Variables`. 2018.

[112] Federal Aviation Administration. *FAA Operations & Performance Data*. `https://aspm.faa.gov/`. 2020.

[113] Peter Bießlich et al. "Developing Generic Flight Schedules for Airport Clusters". In: *5th CEAS Air and Space Conference* (2015). `https://elib.dlr.de/95309/`.

[114] Mark Azzam. "Evolution of airports from a network perspective – An analytical concept". In: *Chinese Journal of Aeronautics* 30.2 (2017), pp. 513 –522.

[115] OAG DESKTOP FLIGHT GUIDE. “Official Airline Guides”. In: *Inc., Oak Brook, Illinois* (1991).

[116] John Bowen. “Network change, deregulation, and access in the global airline industry”. In: *Economic Geography* 78.4 (2002), pp. 425–439.

[117] Theodore W Anderson and Leo A Goodman. “Statistical inference about Markov chains”. In: *The Annals of Mathematical Statistics* (1957), pp. 89–110.

[118] Gerald Öttl and Philipp Böck. “An approach to airport categorization–category definition and quantification for air traffic related simulation purposes”. In: *Journal of Aerospace Operations* 1.1-2 (2011), pp. 155–170.

[119] Z Marta. “Cluster analysis of world’s airports on the basis of number of passengers handled (case study examining the impact of significant events)”. In: *Statistika* 97 (Jan. 2017), pp. 74–88.

[120] Darda P. “The economic importance of passengers for airports”. MA thesis. Univerzita J. E. Purkyně, ESF: Ústí nad Labem, 2014.

[121] Sreeta Gorripaty et al. “Identifying similar days for air traffic management”. In: *Journal of Air Transport Management* 65 (2017), pp. 144 –155.

[122] Shon R. Grabbe, Banavar Sridhar, and Avijit Mukherjee. “Clustering Days with Similar Airport Weather Conditions”. In: *14th AIAA Aviation Technology, Integration, and Operations Conference*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2014-2712`.

[123] Shon R. Grabbe, Banavar Sridhar, and Avijit Mukherjee. “Similar Days in the NAS: an Airport Perspective”. In: *2013 Aviation Technology, Integration, and Operations Conference*. eprint: `https://arc.aiaa.org/doi/pdf/10.2514/6.2013-4222`.

[124] Shon Grabbe, Banavar Sridhar, and Avijit Mukherjee. “Clustering Days and Hours with Similar Airport Traffic and Weather Conditions”. In: *Journal of Aerospace Information Systems* 11.11 (2014), pp. 751–763. eprint: `https://doi.org/10.2514/1.I010212`.

[125] Bob Hoffman et al. “A cluster analysis to classify days in the National Airspace System”. In: *AIAA Guidance, Navigation, and Control Conference and Exhibit*. 2003, p. 5711.

[126] Gano Chatterji and Bassam Musaffar. “Characterization of Days Based on Analysis of National Airspace System Performance Metrics”. In: *AIAA Guidance, Navigation and Control*

Conference and Exhibit. eprint: https://arc.aiaa.org/doi/pdf/10.2514/6.2007-6449.

[127] Lantz, Brett. *Machine Learning with R: Discover How to Build Machine Learning Algorithms, Prepare Data, and Dig Deep into Data Prediction Techniques with R*. Packt Publishing, 2015.

[128] Cosma Rohilla Shalizi, Kristina Lisa Shalizi, and James P Crutchfield. "Pattern discovery in time series, Part I: Theory, algorithm, analysis, and convergence". In: *Journal of Machine Learning Research* (2002), pp. 02–10.

[129] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. "Quantum algorithms for supervised and unsupervised machine learning". In: *arXiv preprint arXiv:1307.0411* (2013).

[130] Alboukadel Kassambara. *Practical guide to cluster analysis in R: Unsupervised machine learning*. Vol. 1. Sthda, 2017.

[131] Robert Gentleman and Vincent J Carey. "Unsupervised machine learning". In: *Bioconductor case studies*. Springer, 2008, pp. 137–157.

[132] Michael Steinbach, George Karypis, Vipin Kumar, et al. "A comparison of document clustering techniques". In: *KDD workshop on text mining*. Vol. 400. Boston. 2000, pp. 525–526.

[133] Stephen C Johnson. "Hierarchical clustering schemes". In: *Psychometrika* 32.3 (1967), pp. 241–254.

[134] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim. "CURE: an efficient clustering algorithm for large databases". In: *ACM Sigmod record* 27.2 (1998), pp. 73–84.

[135] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim. "ROCK: A robust clustering algorithm for categorical attributes". In: *Information systems* 25.5 (2000), pp. 345–366.

[136] Sudipto Guha, Rajeev Rastogi, and Kyuseok Shim. "Cure: an efficient clustering algorithm for large databases". In: *Information systems* 26.1 (2001), pp. 35–58.

[137] Mangortey, E., Puranik, T., Pinon, O., and Mavris, D. "Classification, Analysis, and Prediction of the Daily Operations of Airports Using Machine Learning". In: *AIAA Science and Technology Forum (AIAA Scitech)* (2020).

[138] Ross Ihaka and Robert Gentleman. "R: A Language for Data Analysis and Graphics". In: *Journal of Computational and Graphical Statistics, 5:3, 299-314* (1996).

[139] Maglogiannis, Ilias and and Karpouzis, Kostas and Wallace, Manolis. *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in eHealth, HCI, Information Retrieval and Pervasive Technologies - Volume 160 Frontiers in Artificial Intelligence and Applications*. I O S Press, Incorporated, 2007.

[140] Mangortey, E., Monteiro, D., Ackley, J., Gao, Z., Puranik, T., Kirby, M., Pinon, O., and Mavris, D. "Application of Machine Learning Techniques to Parameter Selection for Flight Risk Identification". In: *AIAA Science and Technology Forum (AIAA Scitech)* (2020).

[141] Sheridan, K., Puranik, T., Mangortey, E., Kirby, M., Pinon, O., and Mavris, D. "An Application of DBSCAN Clustering For Flight Anomaly Detection During The Approach Phase". In: *AIAA Science and Technology Forum (AIAA Scitech)* (2020).

[142] Feyza Gürbüz, Lale Özbakir, and Hüseyin Yapici. "Classification rule discovery for the aviation incidents resulted in fatality". In: *Knowledge-Based Systems* 22.8 (2009), pp. 622–632.

[143] Ali Rahim Taleqani et al. "Machine Learning Approach to Cyber Security in Aviation". In: *2018 IEEE International Conference on Electro/Information Technology (EIT)*. IEEE. 2018, pp. 0147–0152.

[144] Varsha Venkatesh et al. "Iterative machine and deep learning approach for aviation delay prediction". In: *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*. IEEE. 2017, pp. 562–567.

[145] John K Williams. "Using random forests to diagnose aviation turbulence". In: *Machine learning* 95.1 (2014), pp. 51–70.

[146] Christopher Molnar. *A Guide for Making Black Box Models Explainable*. 2019.

[147] Travis E Oliphant. "Python for scientific computing". In: *Computing in Science & Engineering* 9.3 (2007), pp. 10–20.

[148] Mark Lutz. *Programming python*. " O'Reilly Media, Inc.", 2001.

[149] Marr, Bernard. *What Is Big Data? A Super Simple Explanation for Everyone.* `http://www.space.com/2995-nasa-jet-bears-nose-grows-sonic-boom-tests.html`. 2012.

[150] Katharina Ebner, Thilo Bühnen, and Nils Urbach. "Think big with Big Data: Identifying suitable Big Data strategies in corporate environments". In: *2014 47th Hawaii International Conference on System Sciences*. IEEE. 2014, pp. 3748–3757.

[151] Fadi H Gebara, H Peter Hofstee, and Kevin J Nowka. "Second-generation big data systems". In: *Computer* 48.1 (2015), pp. 36–41.

[152] Rahul Beakta. "Big Data And Hadoop: A Review Paper". In: *International Journal of Computer Science & Information Technology* 2.2 (2015), pp. 13–15.

[153] Bhoopathi Rapolu and AviationWeek.com. *Internet Of Aircraft Things: An Industry Set To Be Transformed*. http://aviationweek.com/connected-aerospace/internet-aircraft-things-industry-set-be-transformed. 2016.

[154] Bradbury, Danny. *How Big Data in Aviation Is Transforming the Industry*. https://hortonworks.com/article/how-big-data-in-aviation-is-transforming-the-industry/. 2018.

[155] Airlines For America. *A4A Passenger Airline Cost Index (PACI)*. http://airlines.org/dataset/a4a-quarterly-passenger-airline-cost-index-u-s-passenger-airlines/. 2018.

[156] Noyes, Katherine. *For the airline industry, big data is cleared for take-off*. http://fortune.com/2014/06/19/big-data-airline-industry/. 2014.

[157] Aviation Maintenance. *Big Data Takes Off But Flight Is Just Beginning*. https://www.avm-mag.com/big-data-takes-off-flight-just-beginning/. 2017.

[158] Prashant Tyagi and Haluk Demirkan. *Data Lakes: The biggest big data challenges - Why data lakes are an important piece of the overall big data strategy*. http://analytics-magazine.org/data-lakes-biggest-big-data-challenges/. 2016.

[159] Ignacio Bermudez et al. "Exploring the cloud from passive measurements: The Amazon AWS case". In: *2013 Proceedings IEEE INFOCOM*. IEEE. 2013, pp. 230–234.

[160] Keith R Jackson et al. "Performance analysis of high performance computing applications on the amazon web services cloud". In: *2010 IEEE second international conference on cloud computing technology and science*. IEEE. 2010, pp. 159–168.

[161] James Murty. *Programming amazon web services: S3, EC2, SQS, FPS, and SimpleDB*. " O'Reilly Media, Inc.", 2008.

[162] Clinton Gormley and Zachary Tong. *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine*. " O'Reilly Media, Inc.", 2015.

[163] Manda Sai Divya and Shiv Kumar Goyal. “ElasticSearch: An advanced and quick search technique to handle voluminous data”. In: *Compusoft* 2.6 (2013), p. 171.

[164] Rafał Kuć and Marek Rogoziński. *Mastering elasticsearch*. Packt Publishing Ltd, 2013.

[165] BiswaRanjan Samal and Mrutyunjaya Panda. “Real Time Product Feedback Review and Analysis Using Apache Technologies and NOSQL Database”. In: *International Journal of Engineering and Computer Science* 6.10 (2017), pp. 22551–22558.

[166] Lauren Ajamie. “Flood Control: Using Apache NiFi to Streamline Data Integration”. In: (2018).

[167] Haruna Isah and Farhana Zulkernine. “A scalable and robust framework for data stream ingestion”. In: *2018 IEEE International Conference on Big Data (Big Data)*. IEEE. 2018, pp. 2900–2905.

[168] M Sarnovsky, P Bednar, and M Smatana. “Data integration in scalable data analytics platform for process industries”. In: *2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES)*. IEEE. 2017, pp. 000187–000192.

[169] Apache NiFi Team. *Apache NiFi User Guide*. `https://nifi.apache.org/docs/nifi-docs/html/user-guide.html#introducti`. 2019.

[170] Michael Stonebraker and Lawrence A Rowe. “The design of POSTGRES”. In: *ACM Sigmod Record* 15.2 (1986), pp. 340–355.

[171] Michael Stonebraker, Lawrence A Rowe, and Michael Hirohama. “The implementation of POSTGRES”. In: *IEEE Transactions on Knowledge & Data Engineering* 1 (1990), pp. 125–142.

[172] Michael Stonebraker and Greg Kemnitz. “The POSTGRES next generation database management system”. In: *Communications of the ACM* 34.10 (1991), pp. 78–92.

[173] Apache. *Welcome to Apache™ Hadoop!* `https://hadoop.apache.org/`. 2014.

[174] Tom White. *Hadoop: The definitive guide*. ” O’Reilly Media, Inc.”, 2012.

[175] Alex Holmes. *Hadoop in practice*. Manning Publications Co., 2012.

[176] Federal Aviation Administration. *FAA to Rise to the Cloud*. `https://www.faa.gov/news/updates/?newsId=83609`. 2015.

[177] Federal Aviation Administration. *FAA Cloud Services DTFACT-13-R-00013*. `file:///C:/Users/e.mangortey/Downloads/FCS_L-6_Program_Objectives_Amend14.pdf`. 2015.

[178] Sanjeev Thakur et al. “Mitigating and patching system vulnerabilities using ansible: A comparative study of various configuration management tools for iaas cloud”. In: *Information Systems Design and Intelligent Applications*. Springer, 2016, pp. 21–29.

[179] Lorin Hochstein and Rene Moser. *Ansible: Up and Running: Automating Configuration Management and Deployment the Easy Way*. " O’Reilly Media, Inc.", 2017.

[180] Madhurranjan Mohaan and Ramesh Raithatha. *Learning Ansible*. Packt Publishing Ltd, 2014.

[181] Anil Jain, Karthik Nandakumar, and Arun Ross. “Score normalization in multimodal biometric systems”. In: *Pattern Recognition* 38.12 (2005), pp. 2270 –2285.

[182] S Patro and Kishore Kumar Sahu. “Normalization: A preprocessing stage”. In: *arXiv preprint arXiv:1503.06462* (2015).

[183] Ian Jolliffe. *Principal component analysis*. `https://link.springer.com/book/10.1007%2Fb98835`. Springer, 2011.

[184] Svante Wold, Kim Esbensen, and Paul Geladi. “Principal component analysis”. In: *Chemometrics and intelligent laboratory systems* 2.1-3 (1987), pp. 37–52.

[185] Hervé Abdi and Lynne J Williams. “Principal component analysis”. In: *Wiley interdisciplinary reviews: computational statistics* 2.4 (2010), pp. 433–459.

[186] Emmanuel J. Candès et al. “Robust Principal Component Analysis?” In: *J. ACM* 58.3 (June 2011), 11:1–11:37.

[187] Yuan Tang, Masaaki Horikoshi, and Wenxuan Li. “ggfortify: Unified Interface to Visualize Statistical Results of Popular R Packages”. In: *The R Journal* 8.2 (2016), pp. 474–485.

[188] Emily Mankin. “Principal Components Analysis: A How-To Manual for R”. In: *Desde http://psych. colorado. edu/wiki/lib/exe/fetch. php* (2008).

[189] Analytics Vidhya. *Practical Guide to Principal Component Analysis (PCA) in R & Python*. `https://www.analyticsvidhya.com/blog/2016/03/practical-guide-principal-component-analysis-python/`. 2016.

[190] James C Bezdek, Richard J Hathaway, and Jacalyn M Huband. “Visual assessment of clustering tendency for rectangular dissimilarity matrices”. In: *IEEE Transactions on fuzzy systems* 15.5 (2007), pp. 890–903.

[191] Jacalyn M Huband, James C Bezdek, and Richard J Hathaway. “bigVAT: Visual assessment of cluster tendency for large data sets”. In: *Pattern Recognition* 38.11 (2005), pp. 1875–1886.

[192] Richard J Hathaway, James C Bezdek, and Jacalyn M Huband. “Scalable visual assessment of cluster tendency for large data sets”. In: *Pattern Recognition* 39.7 (2006), pp. 1315–1324.

[193] James C Bezdek and Richard J Hathaway. “VAT: A tool for visual assessment of (cluster) tendency”. In: *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN’02 (Cat. No. 02CH37290)*. Vol. 3. IEEE. 2002, pp. 2225–2230.

[194] A. Banerjee and R. Dave. “Validating clusters using the Hopkins statistic”. In: *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No.04CH37542)* (2004).

[195] James Bezdek, Richard Hathaway, and Jacalyn Huband. “Visual Assessment of Clustering Tendency for Rectangular Dissimilarity Matrices”. In: *IEEE Transactions on Fuzzy Systems, vol. 15, no. 5, pp. 890-903* (2007).

[196] Richard G. Lawson and Peter C. Jurs. “New index for clustering tendency and its application to chemical problems”. In: *Journal of Chemical Information and Computer Sciences* 30.1 (1990), pp. 36–41. eprint: `https://pubs.acs.org/doi/pdf/10.1021/ci00065a010`.

[197] DataNovia. *Assessing Clustering Tendency*. `https://www.datanovia.com/en/lessons/assessing-clustering-tendency/`. 2018.

[198] Statistical Tools For High-Throughput Data Analysis. *Assessing clustering tendency: A vital issue - Unsupervised Machine Learning*. 2008.

[199] Per-Erik Danielsson. “Euclidean distance mapping”. In: *Computer Graphics and image processing* 14.3 (1980), pp. 227–248.

[200] Liwei Wang, Yan Zhang, and Jufu Feng. “On the Euclidean distance of images”. In: *IEEE transactions on pattern analysis and machine intelligence* 27.8 (2005), pp. 1334–1339.

[201] Dhanamma Jagli, Seema Purohit, and N Subash Chandra. “Fuzzy clustering algorithm intended for mining software services on the cloud”. In: *2017 International Conference on Advances in Computing, Communication and Control (ICAC3)*. IEEE. 2017, pp. 1–4.

[202] Tamás Mátrai and János Tóth. “Cluster Analysis of Public Bike Sharing Systems for Categorization”. In: *Sustainability* 12.14 (2020), p. 5501.

[203] Guy Brock et al. “clValid: An R Package for Cluster Validation”. In: *Journal of Statistical Software, Articles* 25.4 (2008), pp. 1–22.

[204] Alain Guénoche, Pierre Hansen, and Brigitte Jaumard. “Efficient algorithms for divisive hierarchical clustering with the diameter criterion”. In: *Journal of classification* 8.1 (1991), pp. 5–30.

[205] Tengke Xiong et al. “DHCC: Divisive hierarchical clustering of categorical data”. In: *Data Mining and Knowledge Discovery* 24.1 (2012), pp. 103–135.

[206] K Sasirekha and P Baby. “Agglomerative hierarchical clustering algorithm-a”. In: *International Journal of Scientific and Research Publications* 83 (2013), p. 83.

[207] Feng Luo et al. “A dynamically growing self-organizing tree (DGSOT) for hierarchical clustering gene expression profiles”. In: *Bioinformatics* 20.16 (2004), pp. 2605–2617.

[208] Huai-Chun Wang, Joaquin Dopazo, and Jose Maria Carazo. “Self-organizing tree growing network for classifying amino acids.” In: *Bioinformatics (Oxford, England)* 14.4 (1998), pp. 376–377.

[209] Jian Feng Cui and Heung Seok Chae. “Applying agglomerative hierarchical clustering algorithms to component identification for legacy systems”. In: *Information and Software technology* 53.6 (2011), pp. 601–614.

[210] Shibing Zhou, Zhenyuan Xu, and Fei Liu. “Method for determining the optimal number of clusters based on agglomerative hierarchical clustering”. In: *IEEE Transactions on Neural Networks and learning systems* 28.12 (2016), pp. 3007–3017.

[211] Aruna Bhat. “K-medoids clustering using partitioning around medoids for performing face recognition”. In: *International Journal of Soft Computing, Mathematics and Control* 3.3 (2014), pp. 1–12.

[212] Lamiaa Fattouh Ibrahim and Manal Hamed Al Harbi. “Using Modified Partitioning Around Medoids Clustering Technique in Mobile Network Planning”. In: *arXiv preprint arXiv:1302.6602* (2013).

[213] S Salvador and P Chan. “Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms”. In: *6th IEEE International Conference on Tools with Artificial Intelligence,pp. 576-584,doi: 10.1109/ICTAI.2004.50* (2004).

[214] Anja Struyf, Mia Hubert, and Peter Rousseeuw. “Clustering in an Object-Oriented Environment”. In: *Journal of Statistical Software, Articles* 1.4 (1997), pp. 1–30.

[215] Ashish Kumar Patnaik, Prasanta Kumar Bhuyan, and K.V. Krishna Rao. “Divisive Analysis (DIANA) of hierarchical clustering and GPS data for level of service criteria of urban streets”. In: *Alexandria Engineering Journal* 55 (2016), pp. 407 –418.

[216] Javier Herrero, Alfonso Valencia, and Joaquın Dopazo. “A hierarchical unsupervised growing neural network for clustering gene expression patterns ”. In: *Bioinformatics* 17.2 (Feb. 2001), pp. 126–136. eprint: `https://academic.oup.com/bioinformatics/article-pdf/17/2/126/441664/170126.pdf`.

[217] William HE Day and Herbert Edelsbrunner. “Efficient algorithms for agglomerative hierarchical clustering methods”. In: *Journal of classification* 1.1 (1984), pp. 7–24.

[218] Daniel Defays. “An efficient algorithm for a complete link method”. In: *The Computer Journal* 20.4 (1977), pp. 364–366.

[219] F Murtagh. “Clustering Algorithms”. In: *Physica-Verlag, W urzburg* (1985).

[220] Robin Sibson. “SLINK: an optimally efficient algorithm for the single-link cluster method”. In: *The computer journal* 16.1 (1973), pp. 30–34.

[221] Clark Olson. “Parallel algorithms for hierarchical clustering”. In: *Parallel Computing 21, pp. 1313-1325* (1995).

[222] Fionn Murtagh and Pierre Legendre. “Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm”. In: *arXiv e-prints* (2011).

[223] Tapas Kanungo et al. “An efficient k-means clustering algorithm: Analysis and implementation”. In: *IEEE transactions on pattern analysis and machine intelligence* 24.7 (2002), pp. 881–892.

[224] Khaled Alsabti, Sanjay Ranka, and Vineet Singh. “An efficient k-means clustering algorithm”. In: (1997).

[225] Paul S Bradley and Usama M Fayyad. “Refining initial points for k-means clustering.” In: *ICML*. Vol. 98. Citeseer. 1998, pp. 91–99.

[226] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. “The global k-means clustering algorithm”. In: *Pattern recognition* 36.2 (2003), pp. 451–461.

[227] Guo-fu ZHAO and Guo-qing QU. “Analysis and implementation of CLARA algorithm on clustering”. In: *Journal of Shandong University of Technology (Science and Technology)* 2 (2006), pp. 45–48.

[228] Tanvi Gupta and Supriya P Panda. “A comparison of k-means clustering algorithm and clara clustering algorithm on iris dataset”. In: *International Journal of Engineering & Technology* 7.4 (2018), pp. 4766–4768.

[229] Ka Yee Yeung et al. “Model-based clustering and data transformations for gene expression data”. In: *Bioinformatics* 17.10 (2001), pp. 977–987.

[230] Chris Fraley, Adrian Raftery, and Luca Scrucca. “mclust: Normal mixture modeling for model-based clustering, classification, and density estimation”. In: *R package version* 4.7 (2014).

[231] Chris Fraley et al. *mclust version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation*. Tech. rep. Technical report, 2012.

[232] Chris Fraley and Adrian E Raftery. “Model-based clustering, discriminant analysis, and density estimation”. In: *Journal of the American statistical Association* 97.458 (2002), pp. 611–631.

[233] Tilman Lange et al. “Stability-Based Validation of Clustering Solutions”. In: *Neural Computation Vol. 16, 1299-1323* (2004).

[234] Sunghae Jun. “An Ensemble Method for Validation of Cluster Analysis”. In: *IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 6, No 1* (2011).

[235] Y. Liu et al. “Understanding of Internal Clustering Validation Measures”. In: *2010 IEEE International Conference on Data Mining, Sydney, NSW, 2010, pp. 911-916.* (2010).

[236] Statistical tools for high-throughput data analysis (STHDA). *How to choose the appropriate clustering algorithms for your data? - Unsupervised Machine Learning*. 2019.

[237] J.C. Dunn. “A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters”. In: *J. Cybern.,vol. 3,pp. 32-57,1973* (2008).

[238] Peter Rousseeuw. “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis”. In: *Journal of Computational and Applied Mathematics Volume 20, Pages 53-65* (1987).

[239] Leo Breiman. “Random forests”. In: *Machine learning* 45.1 (2001), pp. 5–32.

[240] GÃŠrard Biau. “Analysis of a random forests model”. In: *Journal of Machine Learning Research* 13.Apr (2012), pp. 1063–1095.

[241] Adele Cutler, D Richard Cutler, and John R Stevens. “Random forests”. In: *Ensemble machine learning*. Springer, 2012, pp. 157–175.

[242] J. Wang et al. “Classification of Imbalanced Data by Using the SMOTE Algorithm and Locally Linear Embedding”. In: *2006 8th international Conference on Signal Processing*. Vol. 3. Nov. 2006.

[243] R. C. Bhagat and S. S. Patil. “Enhanced SMOTE algorithm for classification of imbalanced big-data using Random Forest”. In: *2015 IEEE International Advance Computing Conference (IACC)*. June 2015, pp. 403–408.

[244] Fabian Pedregosa et al. “Scikit-learn: Machine learning in Python”. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[245] Juan D Rodriguez, Aritz Perez, and Jose A Lozano. “Sensitivity analysis of k-fold cross validation in prediction error estimation”. In: *IEEE transactions on pattern analysis and machine intelligence* 32.3 (2009), pp. 569–575.

[246] Tadayoshi Fushiki. “Estimation of prediction error by using K-fold cross-validation”. In: *Statistics and Computing* 21.2 (2011), pp. 137–146.

[247] Yoshua Bengio and Yves Grandvalet. “No unbiased estimator of the variance of k-fold cross-validation”. In: *Journal of machine learning research* 5.Sep (2004), pp. 1089–1105.

[248] Gareth James et al. *An introduction to statistical learning*. Vol. 112. Springer, 2013.

[249] K. H. Brodersen et al. “The Balanced Accuracy and Its Posterior Distribution”. In: *2010 20th International Conference on Pattern Recognition*. Aug. 2010, pp. 3121–3124.

[250] García V. and Mollineda R.A. and Sánchez J.S. *Index of Balanced Accuracy: A Performance Measure for Skewed Class Distributions*. Springer, Berlin, Heidelberg, 2009.

[251] Koehrsen, Will. *An Implementation and Explanation of the Random Forest in Python*. [Online; accessed August 9, 2020]. 2018.

[252] Scott Moeller et al. "Routing without routes: The backpressure collection protocol". In: *Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks*. 2010, pp. 279–290.

[253] Amit Dvir and Athanasios V Vasilakos. "Backpressure-based routing protocol for DTNs". In: *Proceedings of the ACM SIGCOMM 2010 conference*. 2010, pp. 405–406.

[254] Matthias Brantner et al. "Building a database on S3". In: *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*. ACM. 2008, pp. 251–264.

[255] MuleSoft LLC. *MuleSoft*. `https://www.mulesoft.com/`. 2020.

[256] Richard Wesley, Matthew Eldridge, and Pawel T Terlecki. "An analytic data engine for visualization in tableau". In: *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*. 2011, pp. 1185–1194.

[257] TABLEAU SOFTWARE. *Tableau*. `https://www.tableau.com/`. 2020.

[258] Ansible, Red Hat. *Ansible Tower — Ansible.com*. [Online; accessed August 11, 2020]. 2020.

[259] Apache NiFi. *GenerateFlowFile*. `https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-standard-nar/1.6.0/org.apache.nifi.processors.standard.GenerateFlowFile/`. 2020.

[260] Apache NiFi. *UpdateAttribute*. `http://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-update-attribute-nar/1.11.3/org.apache.nifi.processors.attributes.UpdateAttribute/additionalDetails.html`. 2020.

[261] Apache NiFi. *ExecuteSQL*. `https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-standard-nar/1.5.0/org.apache.nifi.processors.standard.ExecuteSQL/index.html`. 2020.

[262] Varun Singh, Uday V Sawanat, and Naresh G Deshaveni. *Method and system for transparent database connection pooling and query queuing*. US Patent 8,484,242. July 2013.

[263] CHEN Jie. "Self-optimized Configuration of DBCP [J]". In: *Computer and Modernization* 12 (2010).

[264] Apache NiFi. *DBCPConnectionPool*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-dbcp-service-nar/1.5.0/org.apache.nifi.dbcp.DBCPConnectionPool/index.html. 2020.

[265] Deepak Vohra. “Apache avro”. In: *Practical Hadoop Ecosystem*. Springer, 2016, pp. 303–323.

[266] Nurzhan Nurseitov et al. “Comparison of JSON and XML data interchange formats: a case study.” In: *Caine* 9 (2009), pp. 157–162.

[267] Apache NiFi. *SplitAvro*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-avro-nar/1.5.0/org.apache.nifi.processors.avro.SplitAvro/. 2020.

[268] Apache NiFi. *EvaluateJsonPath*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-standard-nar/1.6.0/org.apache.nifi.processors.standard.EvaluateJsonPath/index.html. 2020.

[269] Apache NiFi. *RouteOnAttribute*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-standard-nar/1.6.0/org.apache.nifi.processors.standard.RouteOnAttribute/index.html. 2020.

[270] Apache NiFi. *PutS3Object*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-aws-nar/1.5.0/org.apache.nifi.processors.aws.s3.PutS3Object/index.html. 2020.

[271] Apache NiFi. *PutElasticsearchHttp*. https://nifi.apache.org/docs/nifi-docs/components/org.apache.nifi/nifi-elasticsearch-nar/1.5.0/org.apache.nifi.processors.elasticsearch.PutElasticsearchHttp/. 2020.

[272] Federal Aviation Administration. *Air Route Traffic Control Centers (ARTCC)*. https://www.faa.gov/about/office_org/headquarters_offices/ato/service_units/air_traffic_services/artcc/. 2020.