Degree in Management Computer Engineering and Information Systems

Dashboard for outbound phase at Schiphol Airport

Pol Rivas Balagueró
TUTOR: Rosa Herrero Antón

ACADEMIC YEAR

2022/2023

Acknowledgments

Abstract

This project aims to develop a dashboard tool for Schiphol International Airport in Amsterdam. Dashboard must allow the display of data provided by the same airport, perform filtering and display anomalies that are presented within the same dataset.

Resum

Aquest projecte té com a objectiu el desenvolupament d'una eina dashboard per a l'aeroport internacional de Schiphol a Amsterdam. El dashboard ha de permetre la visualització de les dades proporcionades pel mateix aeroport, realitzar filtratges i visualitzar anomalies que es presentin dins del mateix conjunt de dades.

Resumen

Este proyecto tiene como objetivo el desarrollo de una herramientadashboardpara el aeropuerto internacional de Schiphol en Ámsterdam. El dashboard tiene que permitir la visualización de los datos proporcionados por el mismo aeropuerto, realizar filtrajes y visualizar anomalías que se presenten dentro del mismo conjunto de datos.

Index

Figure Index

Table index

Glossary of terms

AOBT   Actual Off-Block Time

ASRT   Actual Start-Up Approval Time

ATOT   Actual Takeoff Time

CTOT   Calculated Takeoff Time

Dashboard   Graphical interface

DAX   Data Analysis Expressions

EOBT   Estimated Off-Block Time

ETOT   Estimated Takeoff Time

Outbound   Time that passes from the time an aircraft leaves its parking lot until it reaches
its destination

SOBT   Scheduled Off-Block Time

TOBT   Target Off-Block Time

TSAT   Target Start-Up Approval Time

Turnaround   Time spent by an airplane blocking the same parking parcel spot

TTOT   Target Takeoff Time

ACDM   Airport Collaborative Decision Making

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

1

1. Introduction

This project aims to achieve the creation of a dashboard tool for the turnaround process at Schiphol International Airport in Amsterdam (Netherlands), the realization of this project is motivated by the constant need for optimization and detection of anomalies in a process that can cause significant losses in an organism such as an international airport and can seriously impact its operation.

With the elaboration of this project, a tool is achieved that allows to improve the efficiency of the turnaround process in that it can identify where delays or jams occur and thus optimize the process and reduce the time required to prepare an aircraft for its next flight.

The dashboard also gives a clearer and more understandable view of the entire process so that the airport can better understand what is always happening in such sensitive operations.

In addition, this tool serves to make an easier transmission of information with everyone who requires it, such as the airlines themselves or other companies involved in turnaround, thus making communication and cooperation with the different agents involved easier.

 It also means an improvement in security having more control over the situation of passengers and their luggage.

2. Theoretical framework and analysis of referents

Amsterdam Schiphol International Airport is one of the largest airports in Europe and operates millions of flights each year. The turnaround process, which includes unloading, loading and servicing aircraft between flights, is one of the most critical tasks for the airport's success. The time it takes to complete this process can affect flight timelines, customer satisfaction, and the airport's ability to operate efficiently. In this context, this project aims to analyse data related to the Amsterdam airport turnaround process to identify possible anomalies and develop an interactive dashboard that allows airport managers to monitor the process and make informed decisions to improve its efficiency.

2.1. Context

The "Turnaround" process refers to the set of operations involved in preparing an aircraft for its next flight, including disembarking and boarding passengers, refuelling, cleaning, restocking, and performing maintenance and inspections, among others. In an airport, the resting process is essential for the efficient and safe operation of flights, as it determines the amount of time an aircraft is on the ground and, therefore, the number of flights an airport can handle in a day.

This is a critical process that has a great impact on many areas of work at the airport, after reading the documentation provided for the airport itself these are the areas that we determine as most fundamental and, therefore, of greater importance.

Time management: The process must be completed in a specific time frame to ensure that the aircraft leaves on time and does not interrupt subsequent flight schedules.

Safety: The process includes security, maintenance and inspection checks to ensure that the aircraft is aviation-fit and ready for its next flight.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

3

Cost efficiency: The process should be carried out as efficiently as possible to reduce costs for airlines and the airport, and to maximize the utilization of resources such as personnel, ground equipment and facilities.

Passenger satisfaction: The process also involves passenger-oriented activities, such as boarding and disembarkation, which can affect passenger satisfaction and overall airport experience.

Environmental concern: This includes factors such as noise pollution, air quality, wildlife management, environmental regulations, and weather conditions. Efforts are made to minimize noise impact on communities, reduce emissions, mitigate wildlife hazards, comply with regulations, and optimize fuel efficiency.

2.1.2. Background

Previously, more tools have been made that offer a service like that offered by this project, companies such as APIJET [1], Levarti [2] ,Amadeus [3], pivotchain [4] and Assaia [5] offer solutions for the management and analysis of turnaround using video analysis and artificial intelligence to not only perform real-time analysis, but also to predict where anomalies and jams may occur.

2.1.2.1 APIJET

APiJET Turnaround Management is an innovative industry solution designed to enhance operational efficiency and minimize delays in the aviation industry. By leveraging real-time aircraft data and predictive analytics, APiJET provides comprehensive situational awareness and status updates throughout the turnaround process.

This solution seamlessly integrates with your existing back-office systems, enabling the collection and analysis of real-time data from various sources. By applying advanced analytics techniques, APiJET offers objective insights into the turnaround process, allowing for informed decision-making and proactive management.

With APiJET Turnaround Management, operators can gain real-time situational awareness of their aircraft turnaround activities. The system generates alerts and notifications based on the analyzed data, highlighting potential issues or bottlenecks that may affect the turnaround process.

By harnessing the power of predictive analytics, APiJET helps optimize turnaround operations, leading to increased efficiency and reduced delays. The solution's objective and descriptive approach ensure that operators have a clear understanding of the current status of turnarounds and can take appropriate actions to minimize disruptions.

The main difference between this project and the rest is that, while the other tools elaborate the analysis in real time using video analysis tools and interface intelligence, this project offers the possibility of doing the analysis a posteriori according to data already provided by Schiphol International Airport.

2.1.2.2. Assaia

Assaia uses cameras installed at aircrafts stands, apron in aerobridges that process video footage either on premise or in the cloud, so it generates automated timestamps prediction and alerts to help improve operations.

2.2. Theoretical foundations

In this section, the theoretical foundations that underpin the methodology and tools used in the development of this project will be presented. First, the key concepts related to aeronautics, data analysis and anomaly detection will be addressed, including the different techniques and algorithms available for this task. Next, the features and advantages of the main Python libraries used in this project will be described, along with their specific functionalities for data analysis and visualization. Finally, the operation and capabilities of Power BI will be explained, a Business Intelligence platform that has been used to create the interactive dashboard for monitoring the turnaround process at Amsterdam airport.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

5

2.2.1. Aeronautics

Airport Collaborative Decision Making (ACDM) is a concept and framework aimed at enhancing the efficiency and performance of airport operations. It involves the collaborative efforts of various stakeholders, including airlines, air traffic control, ground handlers, airport operators, and others.

The primary objective of ACDM is to improve the flow of information and coordination among these stakeholders, enabling them to make informed decisions and optimize airport operations. This is achieved using advanced technologies and automated systems that facilitate data exchange and support the decision-making process.

ACDM focuses on four key goals. First, it seeks to enhance predictability by sharing real-time information and operational data. This helps reduce delays and improves resource allocation. Second, ACDM aims to improve efficiency by coordinating activities such as aircraft turnarounds, resource allocation, and traffic management.

The third objective is to enhance capacity utilization by optimizing operations and minimizing delays, thereby maximizing the airport's capacity to handle air traffic. Finally, ACDM contributes to enhanced safety by improving coordination and communication among stakeholders, reducing the risk of errors and misunderstandings.

To implement ACDM, airports often rely on dedicated software platforms and systems. These tools facilitate data sharing, communication, and collaborative decision making. They typically include modules for flight planning, resource management, slot coordination, and real-time information sharing.

In summary, Airport Collaborative Decision Making aims to improve the overall efficiency, capacity, and safety of airport operations by promoting collaboration and information sharing among stakeholders.
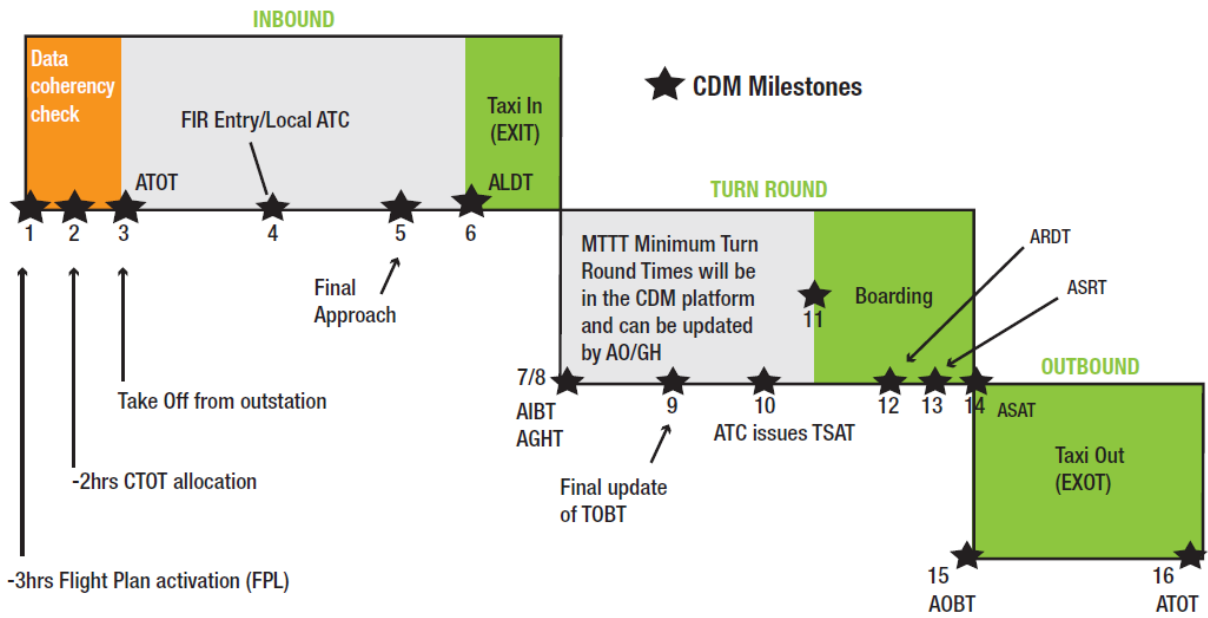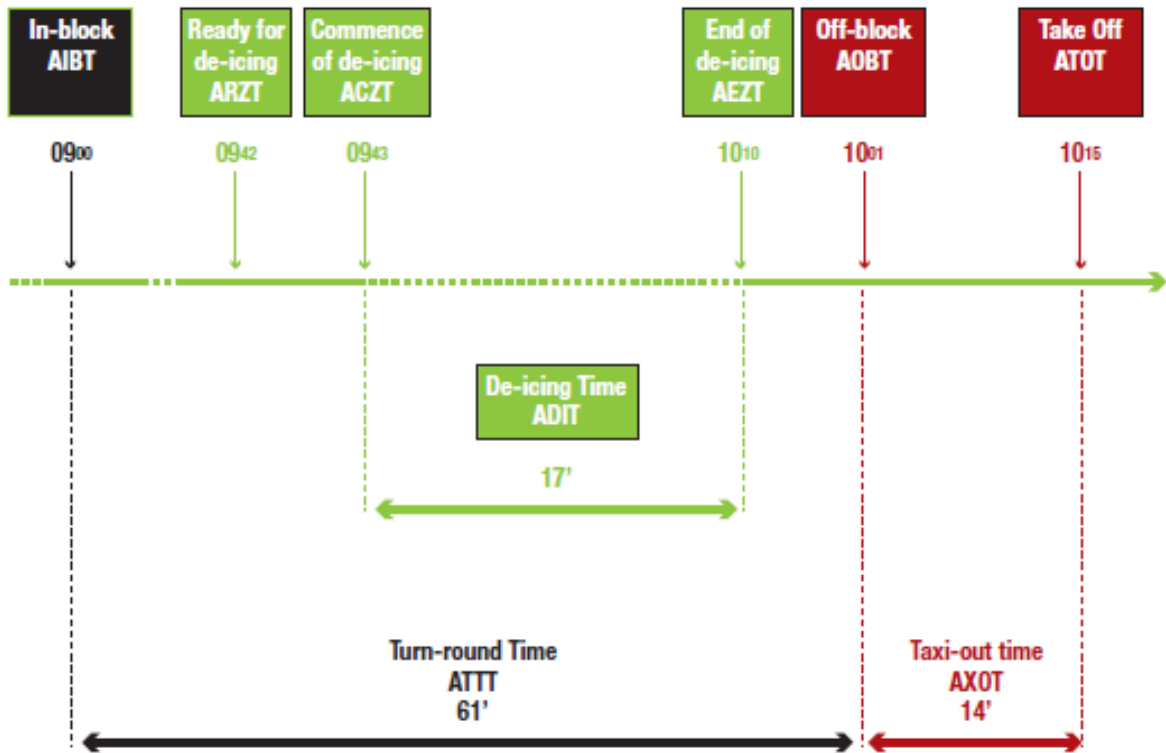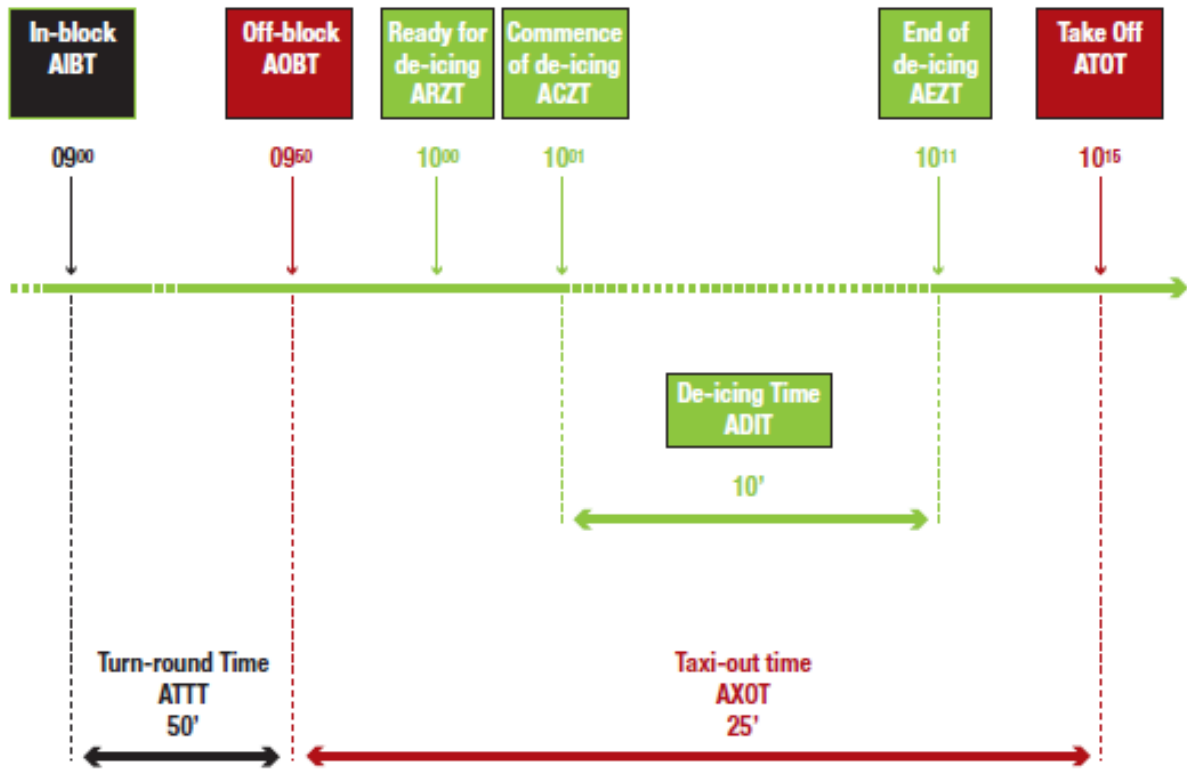
Fig 2.2.1.1 CDM Milestones

Fig 2.2.1.2 Milestones V1

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

7

Fig 2.2.1.3 Milestones V2
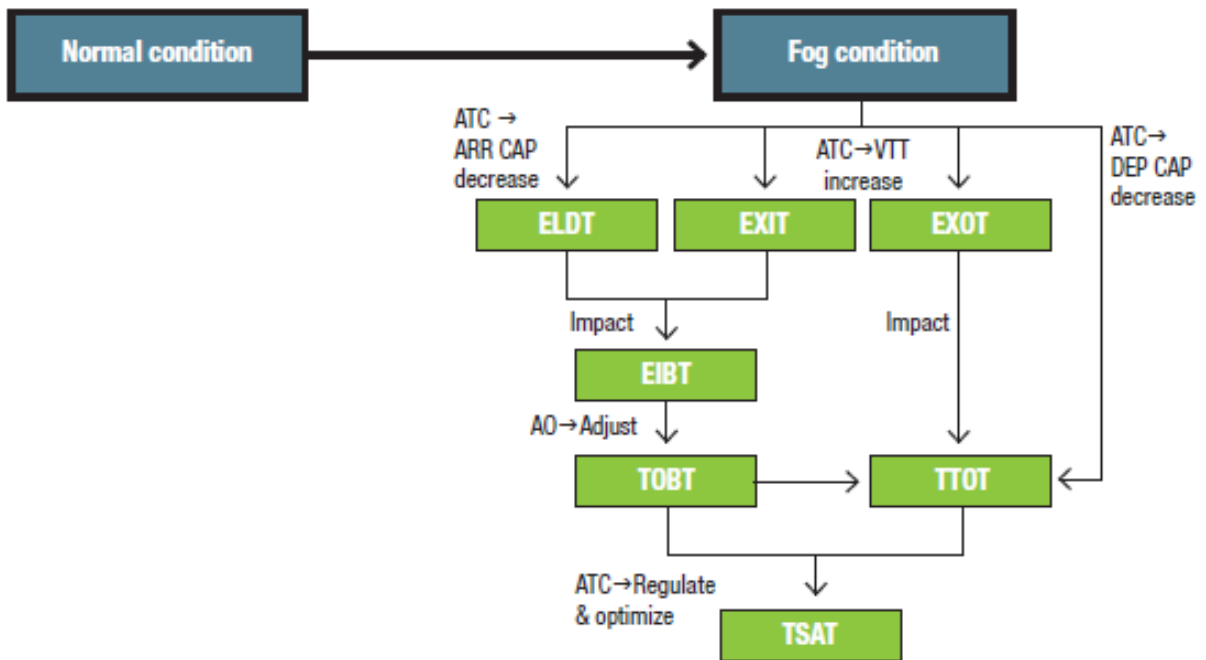
Fig 2.2.1.4 Normal Condition

2.2.1.1 Inbound

Phase when an aircraft is approaching or arriving at an airport. It involves the aircraft's descent, landing, and taxiing to the designated gate or parking area.

2.2.1.2 Turndaround

The process of preparing an aircraft for its next flight after it has arrived at an airport. It includes activities such as refuelling, cleaning, restocking, maintenance checks, and passenger and cargo handling, with the goal of ensuring a safe and efficient departure of the aircraft.

2.2.1.3 Outbound

Phase when an aircraft is departing from an airport. It involves activities such as taxiing to the runway, take-off, and climbing to the designated cruising altitude.

2.2.1.4. SOBT (Scheduled Off-Block Time):

The scheduled time when an aircraft is expected to push back from the gate or parking position, signalling the beginning of the departure process.

2.2.1.5. EOBT (Estimated Off-Block Time):

The estimated time at which an aircraft is expected to push back from the gate or parking position, considering any delays or changes to the original schedule.

2.2.1.6. TOBT (Target Off-Block Time):

The target time for an aircraft to push back from the gate or parking position, which is typically adjusted based on operational factors such as air traffic control restrictions, taxiway congestion, or other considerations.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

9

2.2.1.7. AOBT (Current Off-Block Time):

The actual time when an aircraft physically pushes back from the gate or parking position, as recorded by ground personnel or automated systems.

2.2.1.8. TSAT (Target Start-Up Approval Time):

The target time at which an aircraft is expected to receive start-up clearance from air traffic control, allowing the engines to be started.

2.2.1.9. ASRT (Current Start-Up Approval Time):

The actual time when an aircraft receives start-up clearance from air traffic control, indicating that the engines can be started.

2.2.1.10. CTOT (Calculated Take-off Time):

The calculated time at which an aircraft is expected to begin its take-off roll, considering various factors such as traffic flow, runway availability, and airspace congestion.

2.2.1.11. ETOT (Estimated Take-off Time):

The estimated time at which an aircraft is expected to start its take-off roll, considering any delays or changes to the original schedule.

2.2.1.12 TTOT (Target Take-off Time):

The target time for an aircraft to start its take-off roll, typically adjusted based on operational factors and air traffic control instructions.

2.2.1.13. ATOT (Current Take-off Time):

The actual time when an aircraft initiates its take-off roll and lifts off from the runway, as recorded by ground-based systems or pilot reports.

2.2.2. Data analysis

2.2.2.1 Exploratory data analysis:

Exploratory Data Analysis (AED) [6] is a methodology that involves conducting a detailed statistical study of the data to obtain a better understanding of their characteristics. This process involves identifying patterns, trends, relationships and exceptions in the data to discover important information that can be used for decision making. The AED can be applied to different types of data, including numerical, categorical, and text data.

The main objective of the AED is to explore data systematically and find patterns and relationships that can be useful for decision making. This process helps to gain a deeper understanding of the data and identify problems or anomalies in them. This can help improve data quality and make more informed decisions.

Some of the common techniques used in AED include visualizing data with summary graphs, tables and statistics, identifying outliers or outliers, detecting correlations between different variables and identifying patterns or trends. The AED may also include data cleansing, allocation of missing values, and other processes to ensure data reliability and integrity.

2.2.2.2. Normal distribution theorem:

The normal distribution theorem [7], also known as Gauss's law or normal law, is a key concept in statistics and other fields related to data processing and probabilities. This theorem states that many natural phenomena follow a normal or Gaussian bell distribution. In this distribution, most of the data are concentrated around the mean, and are distributed symmetrically along a curve that flattens out as it moves away from the mean.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

11

The normal distribution is characterized by two parameters: the mean and the standard deviation. The mean indicates the central value of the distribution, while the standard deviation measures the variability of the data. In addition, the normal distribution is symmetric with respect to the mean and can be standardized to obtain the standard normal distribution, which has an average of zero and a standard deviation of 1.

2.2.2.3. Isolation Forest:

A decision tree-based technique for anomaly identification is called Isolation Forest [8]. Data is divided into random trees, and the number of partitions required to isolate each instance is calculated. Anomalies will require fewer partitions and have shorter routes in the decision tree since they are uncommon exceptions. Based on this metric, the system awards anomaly values, with occurrences with lower values being more unusual. Although it may struggle with low-density data or with undifferentiable anomalies of normal data, the Isolation Forest is successful in identifying anomalies in big and high-dimensional data sets. It is a reliable and adaptable method for finding abnormalities.

2.2.2.4. Z-Score:

The Z-score [9], also known as the Z score or standard, is a statistical measure used to determine the relative position of an observation with respect to the mean and standard deviation of a data set. The Z-score indicates how many standard deviations an observation is above or below average.

To calculate the Z-score of an observation, the formula is used:

$$Z = (x - \mu) / \sigma$$

Where:
x is the individual assessment or observation.
μ is the average of the dataset.
σ is the standard deviation of the dataset.

The Z-score provides a measure of the relative distance of an observation from the distribution of the data. A positive Z-score indicates that the observation is above average, while a negative Z-score indicates that the observation is below average.

2.2.2.2.5. Data pre-processing:

The process of cleaning, transforming and integrating data to ensure its quality, coherence and suitability for analysis.

2.2.3. Anomaly detection

2.2.3.1. Clustering algorithms:

Clustering algorithms [10] are a technique used to group a set of data into groups or clusters based on similarities between them. Clustering algorithms are useful when you want to discover natural patterns or groupings in data without having a prior idea of groups or categories.

Algorithms work by finding similarities between observations in a database and grouping them into clusters or groups that share similar characteristics. This is done using different distance metrics to measure the similarity between observations.

In many fields, including data processing, artificial intelligence, data mining, biology, medicine, and other scientific disciplines, clustering algorithms are often employed. They are a useful tool for discovering underlying patterns and organic groupings in data. As a result, businesses may be better able to interpret their data and base choices on more reliable information.

2.2.3.2. Filtering techniques:

Filtering techniques are a set of methods used to process and analyse data and reduce its complexity to focus only on data relevant to a specific problem. These techniques are often used to reduce the amount of data being manipulated and improve the quality and accuracy of results.

Filtering techniques include different methods such as variable selection, sample selection and dimensionality reduction, i.e., selecting only the variables or characteristics most relevant to the task, selecting only part of the data, or reducing the number of dimensions or variables.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

13

The use of filtering strategies may enhance the calibre and effectiveness of analytical procedures, making them crucial for data management and analysis. They can also aid in speeding up and saving money on data analysis.

2.2.3.3. Unsupervised anomaly detection techniques:

Techniques for detecting anomalous or uncommon data in a data set without the requirement for previous categorization are known as unsupervised anomaly detection techniques.

Techniques for unsupervised anomaly detection locate data points that deviate statistically from the rest of the dataset. Different techniques, such as the detection of outliers, the detection of abnormal clusters, and the detection of aberrant patterns, can be used to accomplish this.

A handy approach for finding abnormalities in a dataset without the requirement for previous categorization is unsupervised anomaly detection. This can assist organizations in discovering problems and making decisions based on more knowledgeable and precise information.

2.2.3.4 Supervised anomaly detection techniques:

Supervised anomaly detection techniques are a set of methods used to detect abnormal or unusual data in a data set but require the availability of a dataset labelled with anomaly information for detection.

In contrast to unsupervised anomaly detection techniques, which can detect anomalies without prior knowledge of labels, supervised anomaly detection techniques use classification algorithms to detect anomalies. The labelled dataset is divided into normal and anomalous data, and thus, the model can be trained with these data to be able to classify new data.

Supervised anomaly detection techniques are applicable in many fields, such as the detection of machinery failures, the identification of financial fraud, or the detection of malicious emails.

One of the advantages of supervised anomaly detection techniques is that they are more accurate than unsupervised techniques, as they are based on a set of labelled data. This helps to avoid the identification of false positives, since only data that is abnormal is classified.

2.2.4. Python and libraries

2.2.4.1. Python:

Python [11] is an interpreted and open-source programming language that was created in the late 80s by Guido van Rossum. Its design focuses on the simplicity and readability of the code, making it easy to learn and use for developers at all levels.

Python has become a popular tool for data analysis for several reasons. First, Python is a flexible tool that allows users to process and manipulate data from many different sources. In addition, Python is a general-purpose programming language, meaning it can be used for a wide variety of different applications, including data analysis.

In addition, Python has a large community of developers who have created many libraries and packages for data analysis. These libraries and packages are designed specifically for data analysis and include predefined and optimized functions for tasks such as data visualization, statistics, machine learning.

2.2.4.2 Bookstores:

2.2.4.2.1 Scikit:

Scikit-learn [12] is an open-source library for Python that offers tools for machine learning and is one of the most popular for data analysis. It includes different algorithms for supervised and unsupervised learning, as well as tools for data visualization and processing. Scikit-learn has a simple syntax and clear interface and includes tools for feature selection and model validation. It also offers features for data visualization, which can help you better understand data and identify trends and patterns.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

15

2.2.4.2.2. SciPy:

SciPy [13] is an open-source library for Python that offers features for optimization, linear algebra, statistics, and other scientific and engineering tasks. It is a very useful library for data analysis, as it offers different functions for manipulating data and solving numerical problems. It also includes specialized submodules for data visualization and machine learning.

2.2.4.2.3. Pyod:

Pyod [14] is an open-source library for Python that focuses on data anomaly detection. It offers supervised and unsupervised algorithms and includes tools for data visualization, feature selection, and model validation. Pyod is compatible with other popular libraries for machine learning such as scikit-learn, making it easier to integrate into larger projects.

2.2.4.2.4 Pandas:

Pandas [15] is an open-source library for Python that focuses on structuring and processing data. The library offers flexible and efficient data structures for manipulating numerical and time series data. Pandas is very popular in data analysis because it makes it easy to import data from different formats, clean and transform data, manipulate missing data, and create graphics and visualizations. The library also offers tools for grouping and aggregation of data, as well as the fusion and concatenation of datasets.

2.2.4.2.5 NumPy:

NumPy [16] is an open-source library for Python that offers tools for manipulating and efficiently processing numerical data. The library focuses on the creation and manipulation of multidimensional matrices and vectors, which makes it a very useful tool for data analysis. In addition, NumPy includes a set of advanced mathematical operations that can be applied to data, such as linear algebra, Fourier transforms, and statistical operations. This makes NumPy a fundamental tool for creating machine learning models and exploring data.

2.2.4.2.6 Matplotlib:

Matplotlib [17] is an open-source library for Python that offers tools for visualizing data in the form of graphs and submissions. This library is a very useful tool for data analysis, as it allows you to create personalized graphs and submissions to display data concisely. Matplotlib includes a wealth of features and customization options for creating charts, including bar charts, line charts, scatter charts and many others. In addition, the library also includes tools for creating interactive graphics and animations, which makes it a very versatile tool for data exploration and presentation of results.

2.2.4.2.7 Seaborn:

Seaborn [18] is an open source Matplotlib-based data visualization library for Python. It is designed to create attractive and informative statistical graphs in a single code. Seaborn offers a variety of chart types and colour palettes for displaying statistical data, including heatmaps, line, bar, scatter and regression charts, among others. Its syntax is simple and easy to use and includes features for customizing graphics.

2.2.5 Development Tools

2.2.5.1. Power BI:

Power BI [19] is a Business Intelligence tool developed by Microsoft that allows data visualization and the creation of interactive dashboards and reports. Users can connect to multiple data sources, transform and combine data, and create interactive visualizations for data exploration and discovery. Power BI offers a variety of graphical visualizations and tools for data analysis, such as the ability to create custom measurements and use the DAX query language. You can also share reports and dashboards with other users and collaborate in real time. Power BI is widely used in the enterprise for data analysis and data-driven decision making.

2.2.5.2. Power Query:

Power Query [20] is a Microsoft Excel and Power BI data transformation tool that allows you to connect to different data sources and perform transformations such as filtering, sorting, adding or deleting columns, combining and grouping data, among others. These

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

17

transformations can be applied in a graphical user interface and applied in a previous step to modelling and visualizing data in Power BI.

Power Query is integrated into Power BI as part of its suite of Business Intelligence tools and allows users to connect to multiple data sources for a more complete and accurate view of their businesses. Transformations performed with Power Query can be used in conjunction with other Power BI tools, such as Power Pivot and Power View, to create interactive reports and dashboards.

2.2.5.3. Google Collab:

Google Collab [21] is a cloud-based Python development environment by Google. It allows users to write, run, and share code through their browser without any setup. Collab offers a notebook interface, free access to GPUs/TPUs, collaboration features, preinstalled libraries, integration with Google services, and different code execution environments. It's a user-friendly platform for coding, experimentation, and collaboration in data science and machine learning.

2.2.5.4. Google Drive:

Google Drive [22] is a cloud-based storage and collaboration platform by Google. It offers free and affordable storage plans for files of all types. Users can organize, share, and collaborate on files seamlessly. Google Drive integrates with Google Workspace, supports offline access, and includes a powerful search function. It's a versatile platform for file management and collaboration.

3. Objectives and Scope

The scope of the project is framed in the creation of a dashboard tool for the analysis of data provided by Schiphol airport.

This has certain minimums established for the client, the preparation of three informative screens, the detection of anomalies and the possibility of filtering results based on chronology or other parameters such as the size of the plane of the airline to which it belongs.

3.1. Product objectives

- Display three screens with the records of the milestones of the turnaround process.
- Detect which of the records are anomalies.
- Realize a data filtering from entered parameters.

3.2. Customer objectives

- Receive a dashboard to do a post-analysis of the departure phase from Schiphol airport, with three screens.
- Show the duration times of the departure phase during a given time range.
- To show the situation of the departure phase for a specific time and to be able to move quickly.
- Be able to do anomaly analysis.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

19

4. Methodology

This project follows agile methodologies, the SCRUM software process, with weekly sprints that determine the tasks during the project through the results obtained at the end of each of the sprints This methodology allows, as the project is developed , to limit the viability of the objectives that are marked as optional and, therefore, are not a priority, both for its contribution to the project, for its complexity or viability.

4.1. Project definition

First phase of the project, research phase. The project to be carried out is defined, planned and analysed to see its viability from a technical, economic and environmental point of view, the market and competition are studied.

4.2. Deepening of the project syllabus

Familiarization with the world of Schiphol Airport and the process of turnaround, milestones, how it occurs and impacts the operation of the airport.

4.3. Technical knowledge

Research others language to use, Python, SQL and DAX libraries that are useful within them. Also, environments where   the project is developed and tools for graphic development such as Power Bi and cloud or local environments.

4.4. Graphic design

Graphic design is included within the technical knowledge, to provide a good user experience and make the dashboard as accessible as possible.

4.5 Implementation

Where to apply everything mentioned in the previous phases and get a product from this project that meets the maximum requirements of the client.

4.6 Testing

Over a period of two weeks the product is tested by three people from the main focus groups of the project, a manager from Schiphol airport, an analyst, and a person with a variety of colour blindness to test the choice of colours and the overall user experience of the product.

4.7 Presentation

At the last step, the final product is presented once the testing period has finished and the aspects detected during the testing phase have been corrected .

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

21

5. Development

The development phase of this project is divided into 3 parts: the pre-treatment, the ELT process and the design, each of the stages has been carried out chronologically and sequentially.

CSV, short for Comma-Separated Values, is a common file format used to store tabular data. It consists of seven CSV files, each containing over six thousand rows and forty-five columns. This dataset consists of seven files of forty-five columns each that makes just over six-thousand rows and focuses on flight-related information, where each row represents a specific flight. The columns cover various aspects of the flights, such as timestamps, unique identifiers, call signs, destination airport codes, flight rules, aircraft types, wake turbulence categories, engine types, number of engines, gate and runway assignments, standard instrument departures, collaborative decision making flight states, calculated take-off time states, estimated and actual times for different stages of the flights, target start-up and take-off times, delay information, aerodrome information providers, direction control stations, clearance states, and more.

Analysing this dataset can provide valuable insights into flight patterns, performance metrics, and operational efficiency. By examining the recorded data, researchers can gain a better understanding of how flights are scheduled, managed, and executed, potentially leading to improvements in air traffic management and overall aviation operations.

5.1. Python processing

The pre-processing of datasets is an important initial step in project development. It involves data cleaning, filtering, and transforming raw data into a more precise and valuable format. In this case, Python and Python libraries are used within the Google Collab IDE, which can also connect with Google Drive for accessing data stored in the cloud.

Name	Dtype	Conversion	Description
SOBT	Object	Datetime	The scheduled time when an aircraft is expected to push back from the gate or parking position, signalling the beginning of the departure process.
EOBT	Object	Datetime	The estimated time at which an aircraft is expected to push back from the gate or parking position, considering any delays or changes to the original schedule.
TOBT	Object	Datetime	The target time for an aircraft to push back from the gate or parking position, which is typically adjusted based on operational factors such as air traffic control restrictions, taxiway congestion, or other considerations.
AOBT	Object	Datetime	The actual time when an aircraft physically pushes back from the gate or parking position, as recorded by ground personnel or automated systems.

TSAT	Object	Datetime	The target time at which an aircraft is expected to receive start-up clearance from air traffic control, allowing the engines to be started.
ASRT	Object	Datetime	The actual time when an aircraft receives start-up clearance from air traffic control, indicating that the engines can be started.
CTOT	Object	Datetime	The calculated time at which an aircraft is expected to begin its take-off roll, considering various factors such as traffic flow, runway availability, and airspace congestion.
ETOT	Object	Datetime	The estimated time at which an aircraft is expected to start its take-off roll, considering any delays or changes to the original schedule.
TTOT	Object	Datetime	The target time for an aircraft to start its take-off roll, typically adjusted based on operational factors and air traffic control instructions.
ATOT	Object	Datetime	The actual time when an aircraft initiates its take-off roll and lifts off from the runway, as recorded by ground-based systems or pilot reports.

Table 5.1.1 Columns used

The first task in this process is to handle null values in the relevant columns. These null values are removed to ensure the data is complete and accurate. Additionally, irrelevant columns that do not contribute to the project are also eliminated. The aim is to focus only on the columns that are essential for analysis.

Next, the dataset is examined to identify any instances where the airport may have reached its maximum working capacity or potential. This may involve searching for specific patterns or outliers in the data. By identifying these situations, it becomes possible to understand the factors contributing to the airport's workload and performance.

Name	Column 1	Column 2	Dtype	Description
EOBT-SOBT	EOBT	SOBT	datetime	The output is the difference between the estimated off block time and the scheduled block off time
TOBT-TTOT	TOBT	TTOT	datetime	The output is the difference between the target off-block time and the take-off target time
EOBT-AOBT	EOBT	AOBT	datetime	The output is the difference between the estimated off block time and the actual block off time
SOBT-AOBT	SOBT	AOBT	datetime	The output is the difference between the scheduled off block time and the actual block off time
EOBT-TOBT	EOBT	TOBT	datetime	The output is the difference between the estimated off block time and the target block off time

Table 5.1.2 Calculated columns

Once the initial cleaning and filtering are completed, calculated columns are added to the dataset. These calculated columns are based on specific criteria and will be used in the next step of the pre-processing. They provide additional information and insights into the data.

__Error! Use the Home tab to apply Título 1 to the text that you want to appear here.__

23

The primary objective of this pre-processing phase is to determine the time during which abnormal behaviour occurred within the available data. By analysing the dataset, patterns of abnormal behaviour can be identified. This allows the focus to be placed on that specific period for further analysis and investigation.

To pinpoint the abnormal behaviour, the Z-score is applied to the rows within intervals of 10 minutes. This means that every 10 minutes, all available rows are used to calculate their respective Z-scores based on the previously calculated columns. The Z-score helps in identifying deviations from the expected behaviour and highlights potentially abnormal situations.

The Z-score is a statistical measure used to determine how far a data point deviates from the mean of a distribution, measured in terms of standard deviations. In the context of the provided code snippet, the Z-score is utilized to assess the deviation of various time-related metrics within each data fragmentation from their respective average values.

To calculate the Z-score for a particular metric within a fragmentation, the following steps are typically followed:

1. Compute the mean of the metric within the fragmentation.
2. Determine the standard deviation of the metric within the fragmentation.
3. For each data point of the metric, subtract the mean and divide the result by the standard deviation. This yields the Z-score for that specific data point.
4. The Z-scores provide a standardized measure of how much each data point differs from the average, expressed in terms of standard deviations.

By calculating Z-scores for the selected columns ('EOBT-SOBT', 'TOBT-TTOT', 'EOBT-AOBT', 'SOBT-AOBT', 'EOBT-TOBT', 'EOBT-SOBT') within each fragmentation, the code can identify the fragmentation with the highest maximum Z-score. This highest Z-score indicates the extent to which the corresponding metrics deviate from their average values, highlighting potential areas of interest or concern.

5.2. ELT

ELT (extraction, loading and transformation) applied to the data provided by Schiphol, in this project the extraction phase is carried out by Schiphol thus reducing the complexity of the process, the loading phase consists of loading this data from its origin to a folder in Google Drive to its destination, the development tool chosen for this project, Power BI using the same tool offered by the environment for data upload in this case. CSV, manually and with the access credentials to Drive, all files are loaded one by one, thus completing the upload process.

Once the data has been loaded, the transformation process begins, the data in its previous state are separated into several files, all with the same format and structure, the analysis that is intended is of the data as a single set, therefore, the files must be combined in a single query that allows us to do this analysis, this process is carried out using the Power Query tool, a data editor included within Power Bi itself that bases its operation on the M language, the same editor offers the possibility of attaching queries.

Once the data are structured as a single query, it is necessary, depending on what you want to analyse, to make the relevant calculations, to achieve the main objectives of the project they are carried out using DAX language (Data Analysis Expressions).

For the first visualizations you want to work with columns within the set that contain a date with hours minutes and seconds, therefore, the same calculation is applied to each view, the chosen function is DATEDIFF([data1], [data2], interval), where the date fields are filled with the columns with the dates, and the range parameter is assigned a value MINUTE, to return the difference in minutes.

Once the necessary calculations have been made following this formula, the ETL process is finished, and the design phase is entered.

5.3 Design

The Power BI report has been divided into 3 pages to ensure more options are provided to assure a comprehension of what's presented, those being the "Home" page, the "Week" page and the

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

25

"Day" page, each of those providing different insights of the data and different experiences for the final user.
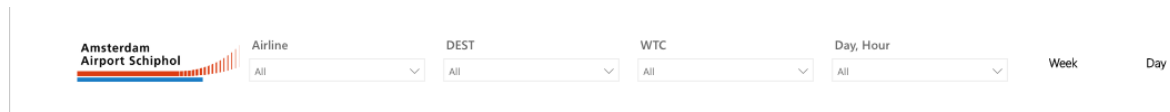

Fig 5.3.1 Filtering elements common to every page

Every page in this report shares a common element on top see Fig 5.3.1 that contains filtering items and the navigation buttons that allow the user to move through the pages of the report on the top left side.

From left to right the filters are "Airline" which enables to sort data based on the airline of the aircraft, "Destination" based on where that flight was going , "WTC" the size of the aircraft, if its large medium or small and "Day, Hour" which zooms on specific days and hours.

Each of this filtering items displays in cascade the elements inside them also enabling the user to type and search for the element they are looking for, when each of these elements are selected the charts of that page instantly are modified applied the restrictions of the filter, that can be added one upon the other in order to each time make the filter more accurate.
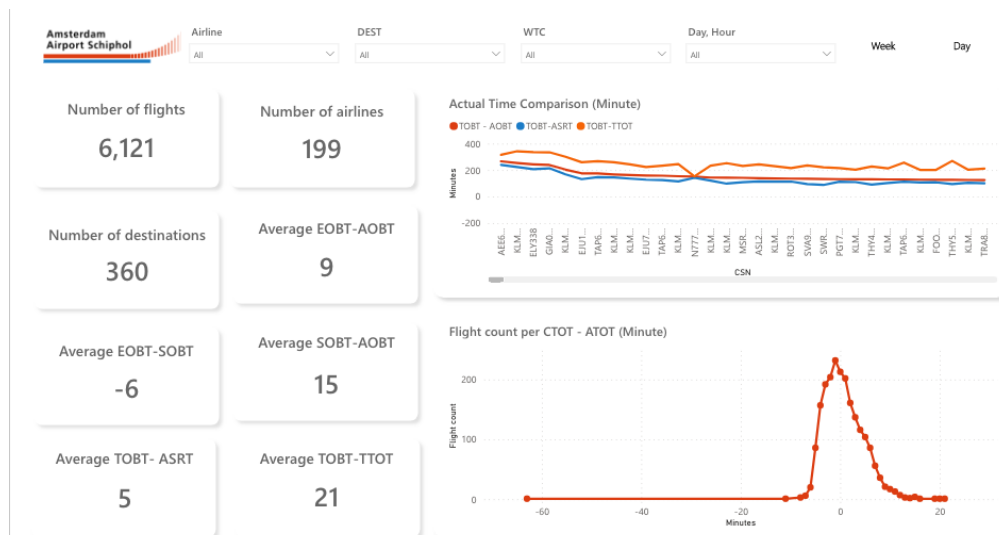
5.3.1 "Home"

Fig 6.3.1.1 "Home" page

Home page can be seen in Fig 5.3.1.1, it acts as a landing page for the user and provides easy and instant information about the whole dataset, presented as numerical cards that give quick key points. The layout of the page is divided into three parts, the one on top, common to all three pages, contains filter elements to help sort the data and give a closer look into certain aspects of it, some navigation buttons to allow the user to move between the pages.
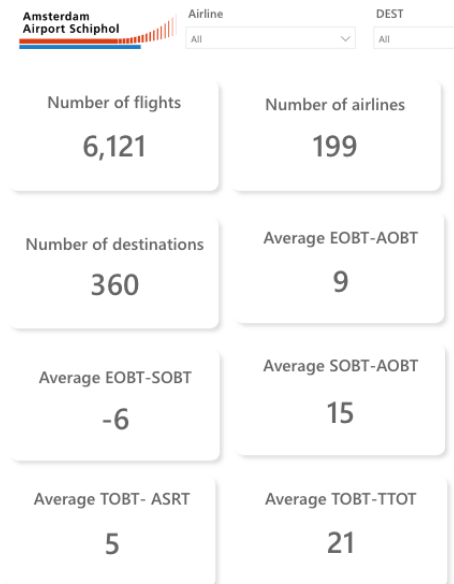

Fig 7.3.1.2 Numeric cards

The numeric cards are made with the "card" option that based on a selected field of the data set creates this sort of numeric elements that provide information that seems to fit perfectly to what a landing page should be used to, give a quick inside without going into very much detail into what the dataset is or what it can tell.
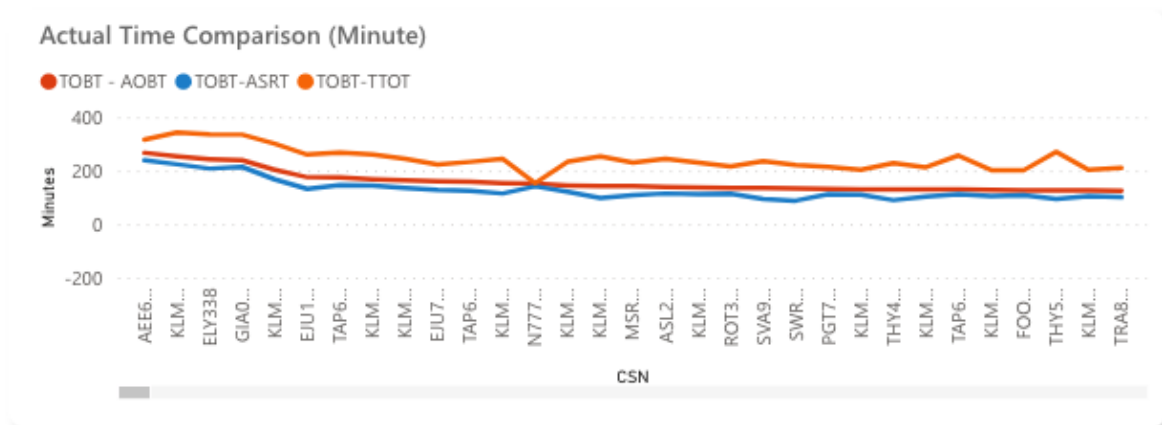

Fig 8.3.1.3 Actual Time Comparison

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

27

Fig 5.3.1.3 is a chart sorted descending by the y axis used to display "AOBT", "ASRT" and "TTOT" in comparison to "TOBT", the x axis uses column "CSN" tfor each aircraft that departed from Schiphol, this comparison is chosen to see how the based on the target off block time the next steps are determined and if maybe according to these values a time where the airport was congested with flights.
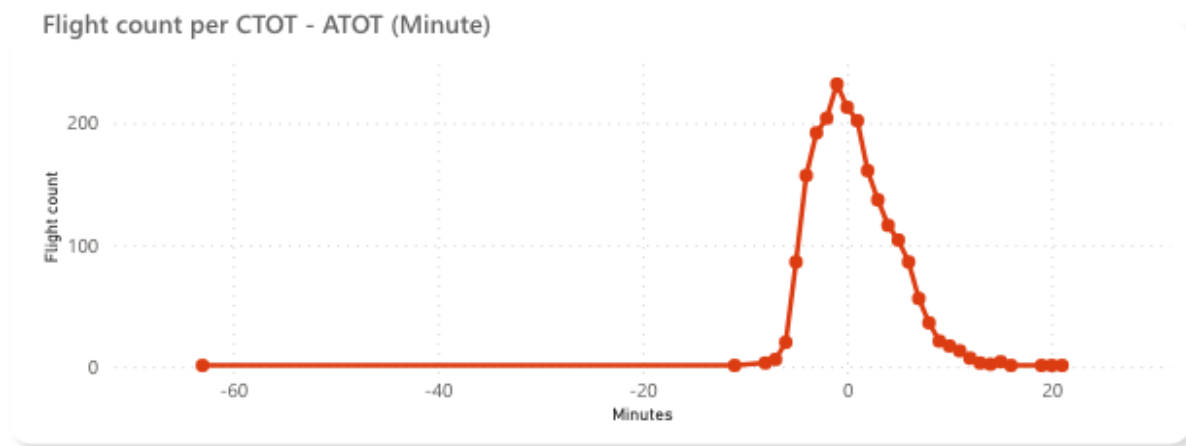

Fig 9.3.1.4 CTOT-ATOT chart

CTOT- ATOT is duo that is used very frequently in this report, because is on the most strict restrictions airports have to manage since it influence from where the aircraft should depart, or if it can depart at all, this value is determined by the Air traffic control according to, in this case European regulation based on the use and transit of flight routes, any values above plus or minus five minutes is considered a situation that should not be happening and should be tracked. Fig. 5.3.1.4 shows the number of flights or frequency per minutes.
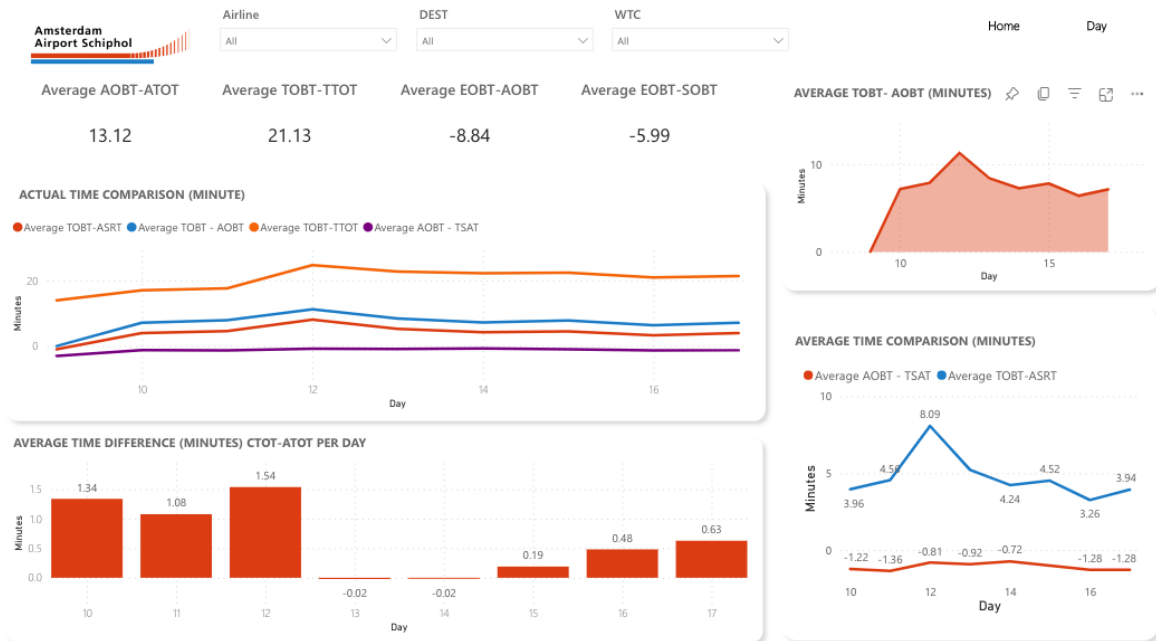
5.3.2 "Week"

Fig 10.3.2.1 "Week" page

Fig 5.3.2.1 shows the week page, this second page objective is to give a more inside look into what the week of data has been, reduces the number of numeric cards in favour of the charts, prioritising comprehension over quantity of information, the reduction of cards slightly modifies the layout allowing charts take part of the top space of the page making them more readable.

Fig 11.3.2.2 Numeric cards "Week" page

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.
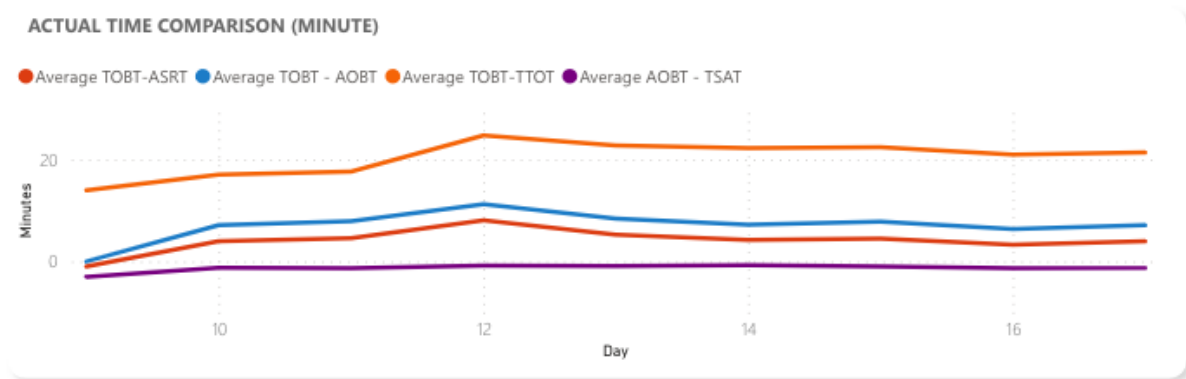
29

Fig 12.3.2.3 Actual Time Comparision "Week" page

Fig 5.3.2.3 is closely related to Fig 4, but the focus is made on the average per day so have a look on what the days of the week have been and based on that have a proper look on the day in the "Page" day.
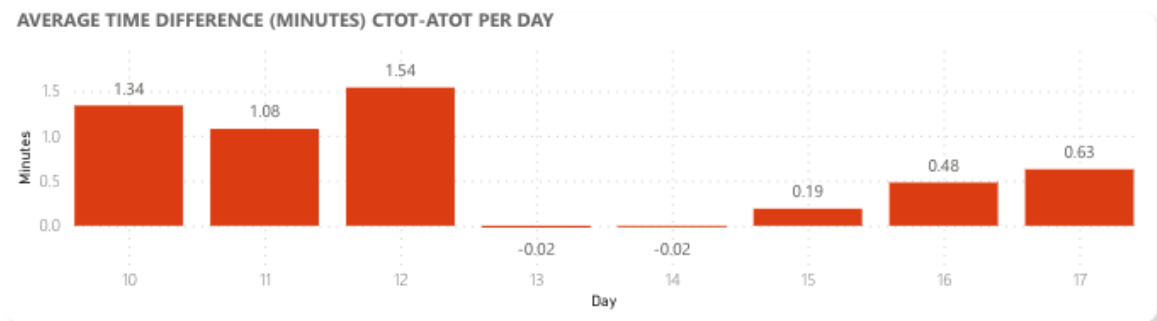

Fig 13.3.2.5 Average Time difference per day

In Fig 5.3.2.5 again CTOT-ATOT is chosen to be examined to spot the day where the more possible anomalous behaviour took place.
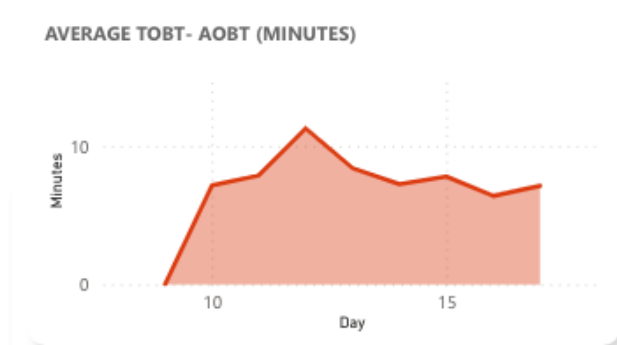

Fig 14.2.5 TOBT-AOBT per day chart

Figure 5.3.2.5 illustrates the disparity between the target and the actual Off Block Time (TOBT). TOBTs require updating if they differ by 5 minutes from the previously declared TOBT. It can be adjusted as many times as necessary until the Target Start-up Approval Time (TSAT) has been issued. After the TSAT has been issued, it can still be updated up to three times without losing priority in the PreDeparture Sequencer. However, if a fourth update is required, the flight's TSAT will be removed from the sequence, and the flight will be re-sequenced.
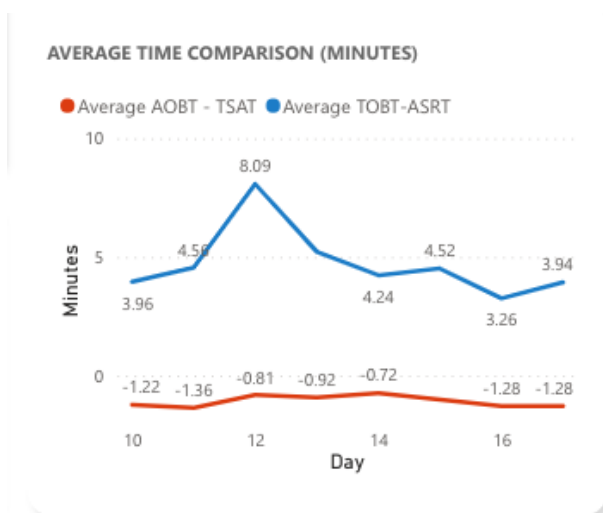

Fig 15.2.6 Average Time Comparison per day

Fig 5.3.2.5 and 5.3.2.6 are to be understood as one entity since they display information about the same part of the process but divided into two to compare and also make them more readable and understandable.

5.3.3 "Day"

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.
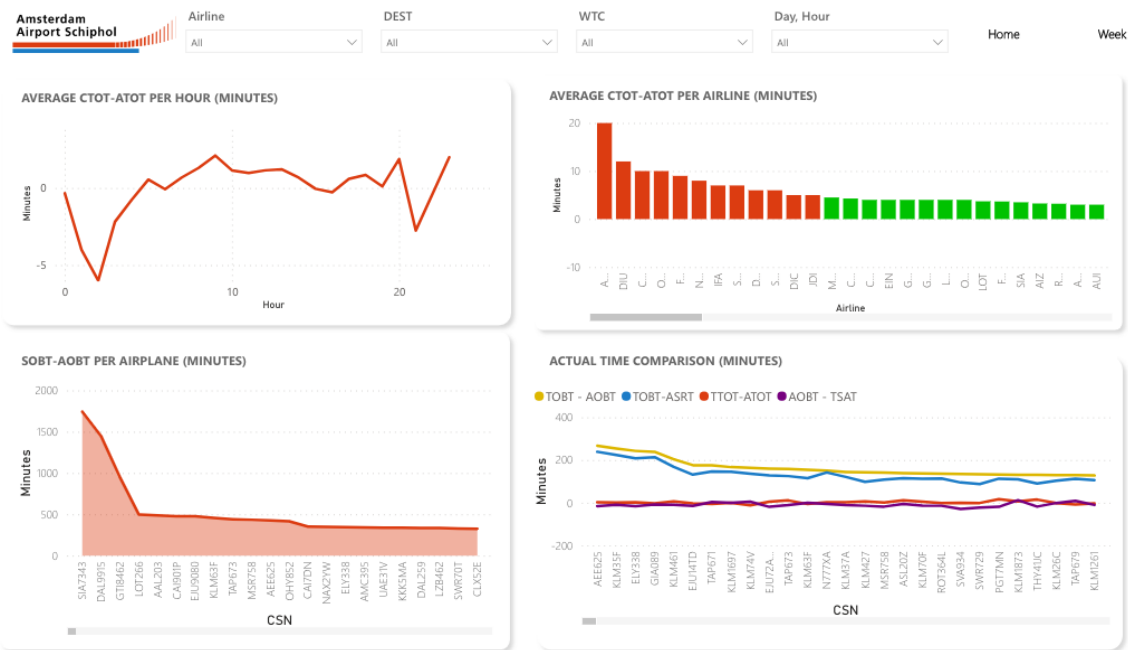
31

Fig 16.3.1 "Day" page

This final page gets completely rid of any numerical card and divides its space equally into four charts and adds a new element to the filter space on top that enables it to zoom into a specific day and hour.
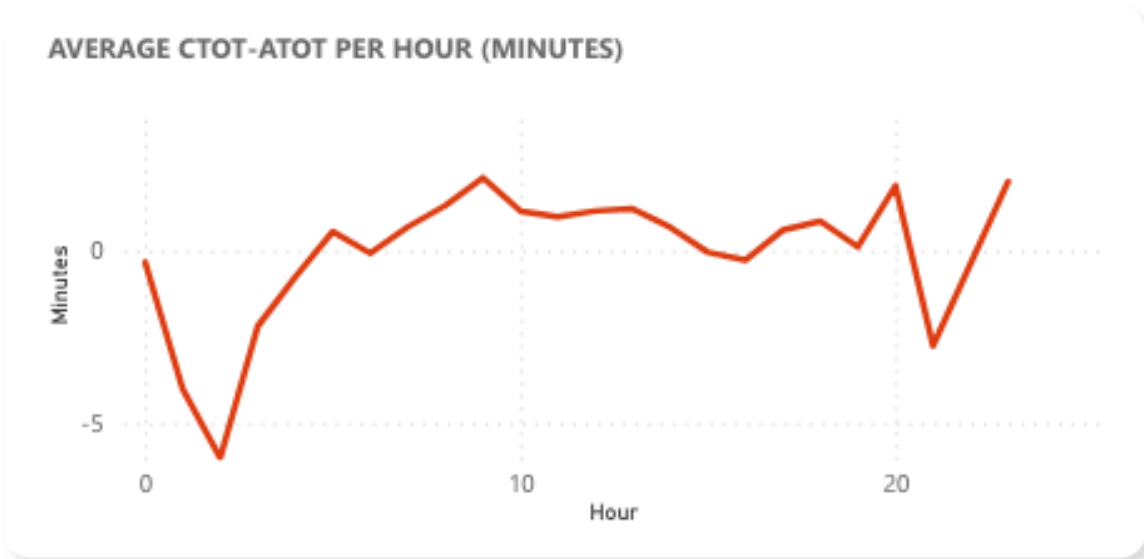

Fig 17.3.2 Average CTOT-ATOT per hour

Once again CTOT-ATOT occupies the chart but broken down per hours so in collaboration with the information provided in the "Week" page the user know in which day should focus and look for uncommon situations.
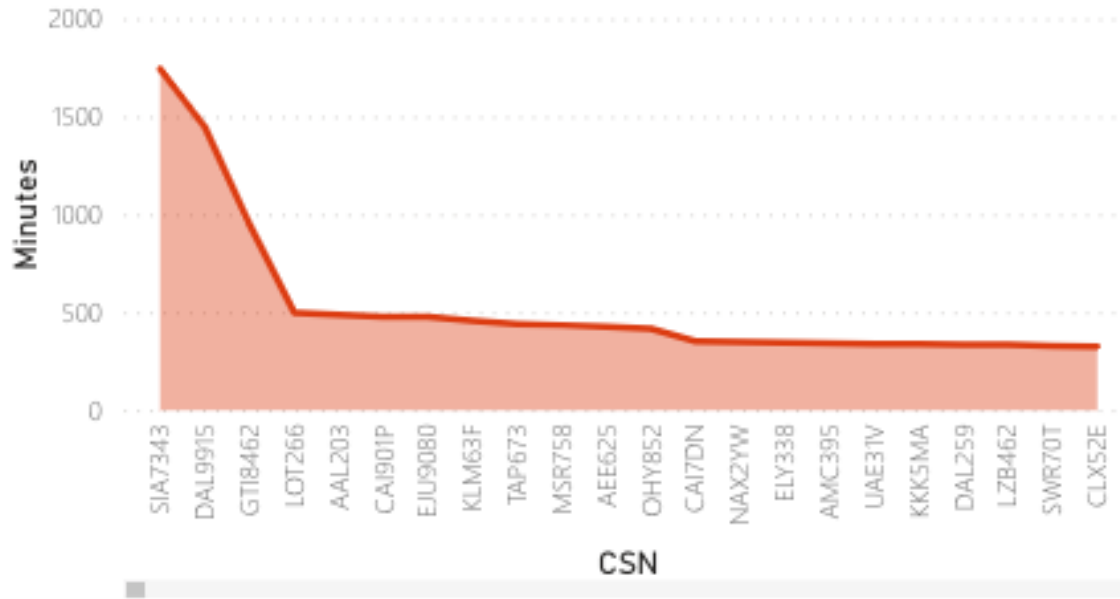
Fig 18.3.3 SOBT-AOBT per airplane

Fig 5.3.3.3 compares the first estimation made on what the block off time was going to be to what it actually was so it highlights where the first estimations made by ATC started go wrong.
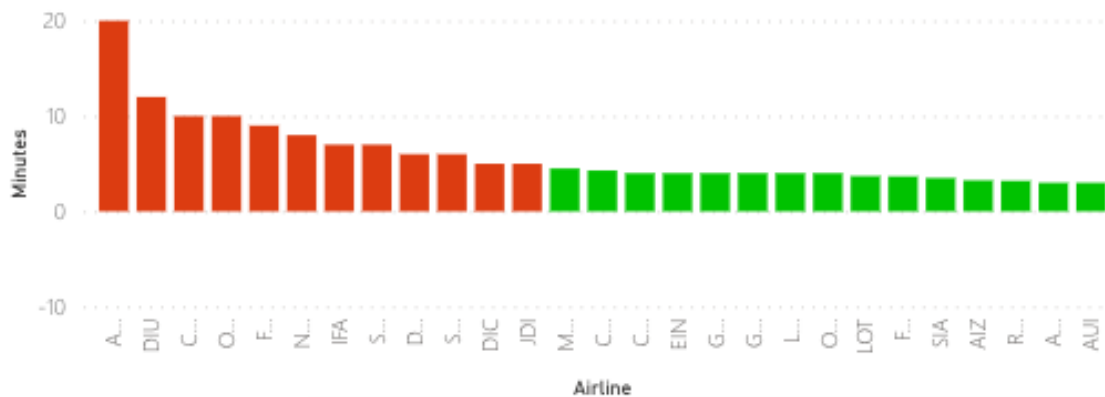

Fig 19.3.4 CTOT-ATOT per airline

Fig 5.3.3.4 again is about CTOT-ATOT but this time the spotlight is on the airlines to see if there's an airline that is systematically having a problematic behaviour for the airport.
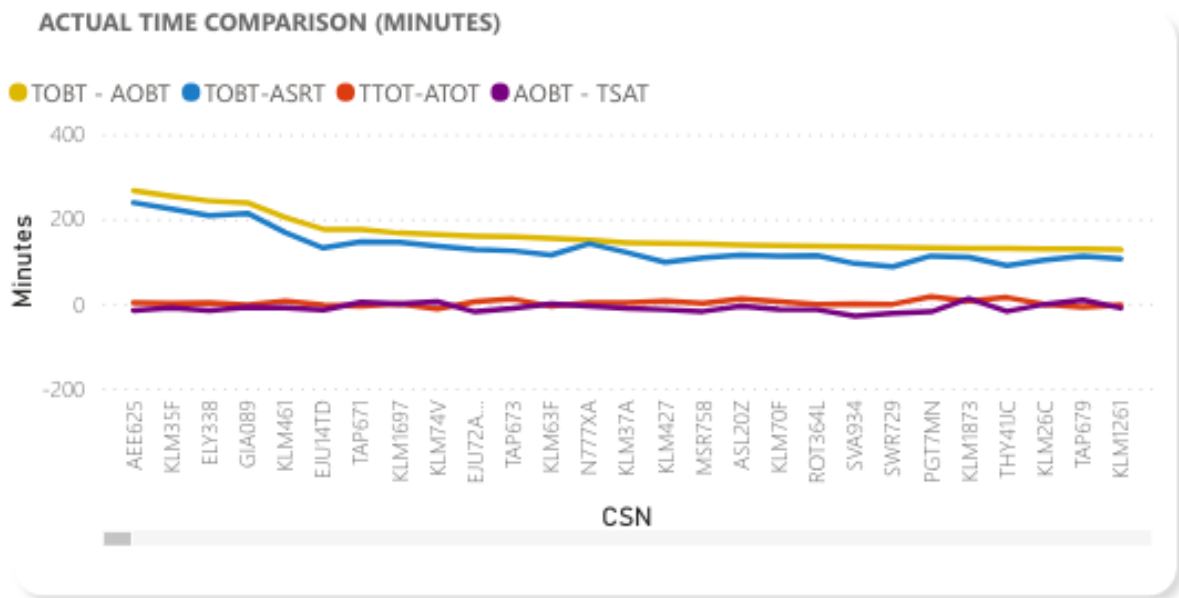
Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

33

Fig 20.3.5 Actual Time Comparison

6. Analysis of results

After completing the entire phase of tool development, it is time to analyse the results of the project and determine which objectives have been achieved as outlined in the project definition.

In order to thoroughly evaluate the obtained results and draw meaningful conclusions, the tool undergoes a final stage review conducted by Dr. Catya Zuñiga Alcaraz from the Amsterdam Hogeschool. Dr. Zuñiga Alcaraz has been closely involved in the project, attending all meetings throughout the development process. She acted as a liaison between the Barcelona-based development team and the intended users of the dashboard, the professionals working in various departments at Schiphol International Airport in Amsterdam. Her active participation at every stage of development ensured that the tool aligns with the expectations of the end-users and provided valuable insights from an aviation expert's perspective. These insights were taken into consideration during the design phase to ensure the visual and functional aspects of the tool catered to the needs of the final users.

Overall, it is fair to say that the final tool is the successful outcome of the development stage. It meets the initial requirements but one, this being the use of python into for the detection of possible bottlenecks due tue delays that thook place in the meetings and and feed back from people at the airport, With the completion of this project, there is now a clear path to expand its scope, enabling it to not only work with the provided data but also establish a new process for extracting, transforming, and loading data, as well as generating reports automatically.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

35

7. Conclusions

The development of a dashboard for the Amsterdam International Airport using Power BI and Python has proven to be a highly valuable and successful endeavor. Through the integration of these powerful tools, we have gained valuable insights and streamlined operations for the airport management.

Firstly, the dashboard has provided a comprehensive overview of various critical aspects of airport operations. By consolidating data from multiple sources, including flight schedules, passenger information, and security measures, we have been able to create a unified platform that offers real-time insights. This has enabled the airport management to make informed decisions, optimize resource allocation, and enhance overall operational efficiency.

Secondly, the integration of Python into the dashboard has unlocked advanced analytical capabilities. Through Python's extensive libraries and data manipulation capabilities, we have been able to perform in-depth analysis on key metrics such as passenger traffic patterns, flight delays, and baggage handling efficiency. These insights have allowed the airport management to identify bottlenecks, proactively address issues, and improve the overall passenger experience.

Furthermore, the Power BI platform has proven to be an excellent choice for visualizing and presenting data in an intuitive and interactive manner. The dashboard's user-friendly interface has enabled stakeholders to easily access and interpret complex information. This has fostered collaboration among different departments, facilitating data-driven discussions and fostering a culture of continuous improvement.

Overall, the development of the dashboard for the Amsterdam International Airport has brought significant benefits to the airport management. It has empowered decision-makers with data-driven insights, enabling them to optimize operations, enhance passenger experience, and ensure the smooth functioning of the airport. The integration of Power BI and Python has been instrumental in unlocking the full potential of the data, providing advanced analytics capabilities, and creating a powerful tool for efficient airport management.

8. Possible extensions.

As mentioned in the conclusions, the project is progressing at a good pace and seems to have a solid base that gives possibilities to different extensions that could lead to a tool that was more complete than first estimated, if more data is provided, more analyses can be carried out than were initially established, like an environmental study and this could lead to a more complete final product, this is a possibility that depends on external agents such as Schiphol.

It also opens the possibility of establishing a cloud system with scheduled updates that keeps the tool updated and working in the long term so that as more data is provided it can update the measurements and visualizations made at first.

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

37

9. Bibliography

[1]   «ApiJet,» [En línea]. Available: https://www.apijet.com. [Último acceso: 10 1 2023].

[2]   «Levarti,» [En línea]. Available: https://www.taconnections.com. [Último acceso: 10 1 2023].

[3]   «Amadeus,» [En línea]. Available: https://amadeus.com/en/industries/airlines/airline-platform. [Último acceso: 10 1 2023].

[4]   «PivotChain,» [En línea]. Available: https://pivotchain.com/raven-airport/. [Último acceso: 10 1 2023].

[5]   «Assaia,» [En línea]. Available: https://assaia.com/turnaround-control. [Último acceso: 10 1 2023].

[6]   «Exploratory Data Analysis | US EPA,» [En línea]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjJ0amWi8P_AhXSVKQEHUY-CTsQFnoECHIQAQ&url=https%3A%2F%2Fwww.epa.gov%2Fcaddis-vol4%2Fexploratory-data-analysis&usg=AOvVaw3eceBbwt37pFQw_9hoeeKt. [Último acceso: 17 3 2023].

[7]   «Central Limit Theorem - SPH,» [En línea]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi1jpLUi8P_AhV1Y6QEHSOmCD4QFnoECDYQAQ&url=https%3A%2F%2Fsphweb.bumc.bu.edu%2Fotlt%2Fmph-modules%2Fbs%2Fbs704_probability%2FBS704_Probability12.html&usg=AOvVaw1LKd_iXtFgJ9QsEE7MgF96. [Último acceso: 18 3 2023].

[8]   «An Introduction to Isolation Forests,» [En línea]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjnoJP1i8P_AhV4UaQEHX3UDxcQFnoECC4QAQ&url=https%3A%2F%2Fcran.r-project.org%2Fweb%2Fpackages%2Fisotree%2Fvignettes%2FAn_Introduction_to_Isolation_Forests.html&usg=AOvVaw2jTvLl6G6it7fS. [Último acceso: 20 3 2023].

[9]   «What is a Z-Score? — Mathematics & statistics - DATA SCIENCE,» [En línea]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjUzfeVjMP_AhVgXaQEHQ4VD2sQFnoECHQQAQ&url=https%3A%2F%2Fdatascience.eu%2Fmathematics-statistics%2Fwhat-is-a-z-score%2F&usg=AOvVaw27vibTT_RFTSqoHDOvwspA. [Último acceso: 25 3 2023].

[10]   «What is Clustering in Machine Learning: Types and Methods,» [En línea]. Available: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwjV_qaujMP_AhVRXaQEHe9jDG8QFnoECCEQAQ&url=https%3A%2F%2Fwww.analytixlabs.co.in%2Fblog%2Ftypes-of-clustering-algorithms%2F&usg=AOvVaw1u612FKnnwS1naGdmFEFQM. [Último acceso: 5 4 2023].

[11] «Welcome to Python.org,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwjBpuPPjMP_AhXBVaQEHVZFBi0QFnoECBQQAQ&url=htt
ps%3A%2F%2Fwww.python.org%2F&usg=AOvVaw0QREvGsjwHKp2GtoYvs1JH.
[Último acceso: 10 4 2023].

[12] «scikit-learn: machine learning in Python — scikit-learn 1.2.2 ...,» [En línea].
Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwj9vaHsjMP_AhVtUqQEHZg7AwQQFnoECAgQAQ&url=http
s%3A%2F%2Fscikit-learn.org%2F&usg=AOvVaw3pidYsGhglQXGDh_4GMetL.
[Último acceso: 10 4 2023].

[13] «SciPy,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwixnayEjcP_AhWZVKQEHbIqB3IQFnoECAgQAQ&url=https
%3A%2F%2Fscipy.org%2F&usg=AOvVaw1JwxlDaFnouRisaTkMOXjv. [Último
acceso: 10 4 2023].

[14] «pyod 1.0.9 documentation,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwiHgMWdjcP_AhVWVKQEHWYKAt0QFnoECBAQAQ&url
=https%3A%2F%2Fpyod.readthedocs.io%2F&usg=AOvVaw3T-
k07RbpmuwgMygTuXO5g. [Último acceso: 10 4 2023].

[15] «pandas - Python Data Analysis Library,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwi70pCxjcP_AhVxQaQEHW4yA2YQFnoECAYQAQ&url=htt
ps%3A%2F%2Fpandas.pydata.org%2F&usg=AOvVaw3cD5ulu4AnZcNusojIyttY.
[Último acceso: 10 4 2023].

[16] «NumPy,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwiKh5nLjcP_AhWUVKQEHbNUBeEQFnoECAYQAQ&url=ht
tps%3A%2F%2Fnumpy.org%2F&usg=AOvVaw3L2i9HVc9ZeynETpNrPxO-.
[Último acceso: 10 4 2023].

[17] «Matplotlib — Visualization with Python,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwizsJ_hjcP_AhUsRKQEHWC8DTIQFnoECAYQAQ&url=http
s%3A%2F%2Fmatplotlib.org%2F&usg=AOvVaw0YgZr7XYZzco6IDaGIE2sw.
[Último acceso: 10 4 2023].

[18] «seaborn: statistical data visualization — seaborn 0.12.2 ...,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwjwrafxjcP_AhWTWqQEHY-
hDFIQFnoECA8QAQ&url=https%3A%2F%2Fseaborn.pydata.org%2F&usg=AOvV
aw0GhFbgBpk8OwHfgpvMyxJm. [Último acceso: 10 4 2023].

[19] «Microsoft Power BI: Data Visualisation,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwjh542EjsP_AhUEVKQEHW4SA6oQFnoECA0QAQ&url=http
s%3A%2F%2Fpowerbi.microsoft.com%2Fen%2F&usg=AOvVaw0dFgor3kPvEFIm6
QlkG9Pd. [Último acceso: 10 4 2023].

Error! Use the Home tab to apply Título 1 to the text that you want to appear here.

39

[20]   « Microsoft Power Query,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwi4ncq5jsP_AhXhUaQEHZI4C2kQFnoECB8QAQ&url=https%
3A%2F%2Fpowerquery.microsoft.com%2Fen-
us%2F&usg=AOvVaw2PjmceJZF_6Ve2nDRzqybX. [Último acceso: 11 4 2023].

[21]   «Welcome To Colaboratory,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwjH_v3IjsP_AhUeVKQEHX0zCU4QFnoECAwQAQ&url=http
s%3A%2F%2Fresearch.google.com%2Fcolaboratory%2F&usg=AOvVaw38J01zt_Dl
b6pQ1fe6FGrI.

[22]   «Personal Cloud Storage & File Sharing Platform - Google,» [En línea]. Available:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&u
act=8&ved=2ahUKEwjvganYjsP_AhV2WqQEHVr2D0QQFnoECAYQAQ&url=http
s%3A%2F%2Fwww.google.com%2Fdrive%2F&usg=AOvVaw1IGG1uG7O5zL8XK
dZ_uZIJ. [Último acceso: 14 4 2023].