



UCD Michael Smurfit Graduate Business School

Report - Tableau Assignment

Master of Science in Business Analytics

**Module: Business Decision Support System
(MIS41040)**

Date of submission – 16th April 2023

Submitted to:

Assoc Professor Peter Keenan

Submitted by:

Assignment Team - 20

Anirudh Muralidharan – 22200946

Atharv Chitre - 22200118

Pooja Padmcharan Dash – 22200862

1. Introduction

The aviation industry has a vital role in the global economy as it enables people and goods to be transported worldwide. However, flight delays and cancellations present a significant challenge to airlines as they can lead to unhappy customers, decreased revenue, and higher operational costs. Therefore, it is essential to analyze flight delays and determine the factors that contribute to them to enhance airline performance (ICAO, 2022).

This report analyzes the performance of different airlines that operated in the United States in 2018. It is based on raw data provided in the given [link](#) (Mulla, 2022). The Flight Delay Performance dataset provides insights into the causes of flight delays and cancellations, including factors such as airlines, airports, weather conditions, and operational issues. The information in this report can be useful for airlines and airports to identify areas that require improvement in their operations and reduce costs associated with flight delays.

Our team has analyzed a dataset and identified parameters that are associated with delayed flights. Our goal is to help airlines improve their flight schedules, increase their on-time performance, and ultimately enhance customer satisfaction. Through our analysis, we have gained valuable insights and we are able to provide practical recommendations for airlines to improve their performance.

In summary, this analysis provides valuable insights into the factors that contribute to flight delays and offers practical solutions for airlines to improve their on-time performance and customer satisfaction.

2. Business Understanding

This report is intended to assist stakeholders in the airline industry by providing an understanding of the factors behind flight delays. It aims to aid decision-making with respect to airline operations, scheduling, and customer service. The report outlines the primary causes of delays and analyzes their impact on airlines, passengers, and the economy. It also compares the performance of various airlines and offers recommendations for reducing delays. The ultimate goal of this report is to increase airline efficiency and improve the overall customer experience.

3. Data Understanding

The dataset contains approximately six million records and provides details on flight delays, cancellations, diversions, and other factors that impact on-time performance. Additionally, it includes information on weather conditions, airports, and airlines.

4. Data Requirements

To analyze flight delays in the airline industry, we need information about airlines, airports, operational factors, and the factors that contribute to flight delays. Flight delays can be caused by various types of delays, including arrival and departure delays. However, these delays can be attributed to five factors: Carrier Delay, Weather Delay, NAS Delay, Security Delay, and Late Aircraft Delay. It is also important to consider seasonal delays, which can be identified by the quarter, month, and weekday. Additionally, geographical data is necessary to understand how delays are affected by states and cities.

5. Data Preparation

5.1. Python

In order to conduct our analysis, we utilized the Python programming language to merge the raw datasets from all months, and we eliminated extraneous columns. We specifically extracted columns related to delays and seasonal information, and organized the data into a dataframe. To consolidate airline and airport information, which included airline names and the geographical locations of flight routes and airports, we performed a left join on the relevant datasets. We then merged the dataframe with the Airports dataset by left joining on the "Origin" and "IATA_CODE" columns to obtain geographical information on the origin airport. Subsequently, we merged the dataframe with the Airports dataset again by left joining on the "Dest" and "IATA_Code" columns to extract geographical information on the destination airport. We carefully verified that all necessary fields were present for further analysis before exporting the dataframe to a CSV file.

5.2. Tableau Prep Builder

The CSV file was imported into Tableau Prep builder to clean the data further. To prepare the flight delay dataset for analysis, various data cleaning techniques were implemented to handle null or missing values. Null values in the columns "Airport_origin", "Latitude_origin", "Airport_dest", and "Latitude_dest" were entirely removed as they do not provide any additional information for analysis. The columns "DepDelayMinutes" and "DepartureDelayGroups" with null values were replaced with 0 as null values would be an outlier for these fields. For the columns "TaxiIn," "TaxiOut," "WheelsOn," and "WheelsOff," the average of the non-null values was used to replace any null values. Null values in the columns "ArrDelay" and "ArrDelayMinutes" were also replaced with 0. For the "ArrDel15" column, the average of non-null values was used to replace any null values. Lastly, null values in the "CRSArrTime" and "ArrTime" columns were replaced with 0 to ensure no data was lost during analysis. By using these data cleaning techniques, the dataset is now complete and accurate, which allows for more reliable analysis and insights.

6. Data Visualization

The dataset was imported into Tableau Desktop, and visualizations were created to provide better recommendations to the stakeholders.

Link: [Flight Delay Dashboard](#)

6.1. Dashboards:

The dashboard includes several interactive visualizations that allow users to explore different aspects of the dataset. The dashboard has been designed to be user-friendly, with easy-to-understand visualizations and controls.

6.1.1. Summary:

The Summary dashboard offers a clear and concise overview of 2018's delays, including various graphs and statistics such as a bar chart depicting the average arrival delay per weekday and key performance indicators. According to the dashboard, 5,471,559 flights flew in 2018, with 1,241,830 hours of arrival delays and 1,216,095 hours of departure delays. The dashboard reports that 81.24% of flights arrived within 15 minutes of their scheduled time,

and 81.92% of departures were on time. The "Avg. Arrival Delay per Weekday" metric demonstrates that Fridays had the highest average delay at 15.45 minutes.

The line chart on the dashboard showcases the average departure delay and the average arrival delay by month, indicating that August experienced the highest average delay in both arrival and departure, while October had the lowest delay. Another chart, "Avg. Arrival Delay per Airline," provides valuable information to travellers in selecting an airline with a desirable on-time performance. "Commuter aka Champlain Enterprises" recorded the highest average arrival delay of 29.11 minutes, whereas "Hawaiian Airlines" reported the lowest average arrival delay of just 5.94 minutes.

The average departure delay per airport measures the average time that planes take to depart after their scheduled departure time, calculated by dividing the total departure delay time for all flights by the total number of departures from that airport. The geographical map depicts dark purple indicating the highest delay, purple indicating moderate, and light purple indicating the lowest delay. The "Barnstable Municipal Airport" had the highest average departure delay of 36.20 minutes, while "Lewiston Nez Perce County Airport" had the lowest average departure delay of 4.31 minutes.

6.1.2. Causes of Delay:

The Causes of Delay Dashboard represents the five major reasons for flight delays in 2018: Air System, Carrier, Security, Late Aircraft, and Weather. The dashboard provides various visualizations, including the total delay for each delay type, as well as the variation in delay types by weekday and month.

Additionally, the dashboard features Delay Type by State and most and least impacted airports and airlines, enabling users to filter and view data by delay type and months. This information can assist airlines and airports in identifying the most problematic delay types and implementing measures to improve those areas. The dashboard also includes average delay information by type and month / weekday, which can help the airline industry prepare for periods with high delays and take measures to reduce them.

For example, when the delay type is selected as Weather Delay on the dashboard, it shows that "Mesa Airlines Inc" was the most impacted airline and "Houghton County Memorial Airport" was the most impacted airport. Furthermore, In the year 2018, the industry experienced weather delay highest in the month of August.

6.1.3. Airport Delay:

The Airport Delay dashboard analyzes delays experienced by US airports in 2018, displaying key performance indicators such as "Departures without Delay," "Average Departure Delay," and "Flights departed" for each airport. The dashboard presents visual representations such as stacked bar graphs, bar charts, and line charts, including important visualizations like "Avg. Departure Delay per Weekday" and "Seasonality: Avg. Departure Delay & Number of Flights." Users can select specific airport codes and month ranges with the dashboard's filters.

The dashboard's information can aid the airport authority in improving their flight scheduling, implementing proactive maintenance checks by performing regular maintenance checks and

fixing issues before they cause delays. Improving communication with passengers by providing updates on flight status and offering alternative travel options can also help. Furthermore, the implementation of technology solutions, such as automated check-in and boarding processes, real-time flight status updates, and predictive analytics, can enhance delay management and reduce passenger wait times at the airport.

For instance, if we choose “Waco Regional Airport” (ACT) as our point of interest, we can see that it managed 667 flights in 2018, with 77.06% of them departing on time. When there were delays, the average departure delay was 3.162 minutes. Looking at the bar graphs, we can see that Thursdays had the highest average departure delay on weekdays, reaching 14.92 minutes. Moreover, October had the highest number of airlines operating, totalling 138, with an average delay of 15.65 minutes. We recommend ACT airport to increase staff during peak hours to reduce delay time on Thursdays. Consider adjusting airline scheduling to spread out the number of flights across different months to avoid congestion.

6.1.4. Airline Delay:

The Airline Delay Dashboard presents an in-depth analysis of delays experienced by various airlines, categorized by parameters such as Air System Delay, Carrier Delay, Security Delay, Late Aircraft Delay, and Weather Delay. It also includes average arrival and departure delay statistics. The dashboard offers a range of visualizations, including a stacked bar graph that displays the total delay and number of flights delayed each month of the year, a donut chart that shows the delay faced by each airline by delay parameter, and other critical visualizations such as "Avg. Departure and Arrival Delay per airline," "Most Delayed Route by Airlines," and "Seasonality: Avg. Departure Delay & Number of Flights."

Furthermore, the dashboard features filters that enable users to select data by airline and delay type, such as average arrival or departure delay. This can assist the airline industry in identifying the primary reasons for delays and taking appropriate steps to improve their services.

For instance, in 2018, “Jet Blue Airlines” reported an average arrival delay of 37.76 minutes for the “Illinois-New York” route. Armed with this information, the airline can analyze the data and explore alternate routes to minimize delays. Additionally, Jet Blue Airlines experienced a total delay of 5,596,192 minutes, out of which Late Aircraft Delay accounted for 2,308,204 minutes. To address this issue, they can consider investing in more aircraft or leasing options to accommodate potential delays. Moreover, they can focus on enhancing their maintenance practices and reducing the turnaround time between flights.

6.1.5. Airline Benchmark Dashboard:

The Airline benchmark dashboard enables airlines to compare their performance with that of other airlines in the industry. The dashboard typically displays key performance indicators (KPIs), such as average delays and performance metrics, to help airlines identify areas where they can improve their operations and increase profitability. Furthermore, there are filters that allow for a side-by-side comparison of the performance and delays of two airlines.

One such KPI is “Departures without delay”, which measures the proportion of flights that depart exactly at their scheduled time. A high rate is desirable as it allows airlines to maintain

schedules, reduce delays, and minimize the cost of disruptions. For example, a rate of 73% means that 73% of an airline's flights leave on time according to schedule. Another KPI is "Arrival within 15 min", which measures the proportion of flights that reaches the destination within 15 minutes of their scheduled arrival time. A high rate indicates that an airline is able to operate more efficiently, improve customer satisfaction, and experience fewer delays.

The average weather delay is determined by dividing the number of weather-related flight delays by the total number of flights and calculating the average extra time incurred. The average security delay is calculated by dividing the number of flights by the total security delay duration. The average security delay for all airlines is 0.01777 minutes.

The average late aircraft delay is determined by subtracting the actual arrival time from the scheduled arrival time and any additional delays due to weather or other factors. The average late aircraft delay for all airlines is 5.121 minutes. A carrier delay occurs when an airline is responsible for a flight's delayed arrival or departure. The overall average carrier delay is 3.679 minutes. The National Air System (NAS) is a network of airports, airspace, and air navigation facilities in the US. The average NAS delay in the airline industry is due to factors such as airport services, high traffic volume, non-extreme weather conditions, and air traffic control.

Overall, the Flight Delay Dataset and dashboard provide a wealth of information about flight delays in the United States. This information can be used by airlines, airports, and other stakeholders to make data-driven decisions and improve the overall performance of the industry. The visualizations are user-friendly and provide a quick and easy way to explore the data. The dataset is valuable for anyone interested in studying flight delays and their impact on air travel.

References

ICAO, 2022. *ICAO*. [Online]

Available at: <https://www.icao.int/sustainability/Documents/AVIATION-BENEFITS-2019-web.pdf>

[Accessed 5 April 2023].

Mulla, R., 2022. *Kaggle*. [Online]

Available at: <https://www.kaggle.com/datasets/robikscube/flight-delay-dataset-20182022>

[Accessed 20 March 2023].