

Technical Report for Case Study:

Airport Authority Data Analysis

ALGONQUIN COLLEGE

CST2102-Database Analytics

Instructor - Yasser Jafer

By:

Prajna Ganji, Shiva Shiva, Neetika Upadhyay and Maruta Zalane

07.03.2025

Abstract

In this report we have presented the findings of our data analysis of flight operations, delays and of course the passenger feedback. The main objective of this analysis was to look at the results from year 2018, but our team decided that the best way to really understand the data we must also compare it to the years before and see whether there are positive or negative changes. By including data from 2014 to 2018 we believe it would give the business a clearer vision of the reality. The main goal of this analysis is to provide deeper understanding of the airport performance and flight dynamics. The key metrics for this analysis were the total number of flights, delays, and ground processing times and most importantly the analysis of passenger feedback. An important part of this project was dashboard interface which would allow the user to see the data and make well informed business decisions.

Introduction

The main objective of Airport Authority Data Analysis is to make a deep and thorough analysis of dataset found from Kaggle. The project aims to help the airport improve the operational efficiency as well as make well informed, data driven decisions. Thanks to the data that was available on Kaggle we were able to analyse incoming and departing flights, see a clearer picture of delays and their reasons, ground processing times, and most importantly the feedback from the customers.

In this report you will find the methodology that was used to work with and analyse the data. We will discuss the findings from the visualised data and what conclusions could be made. We will also discuss our findings and look how the information shown on dashboard correlates and what conclusions could be made.

Methodology

For the project we used data from Kaggle's "Airline Delay and Cancellation Data (2009-2018)" To analyse the data, it had to undergo the following processes:

1. Getting to know the dataset:

Before we could start working with the data, we first needed to understand what we were working with. Understanding the columns gave insight into the types of data available and its range, this allowed us to make informed decisions about how to clean and process it. In this stage we had to note what had to be changed, fixed or improved.

2. Data cleaning and transformation:

For this project it was important that we cleaned the data properly so that we would be able to successfully and correctly visualise it on Power BI. When we imported the data in SQL Server the first thing, we noticed was that all columns had the same data type- Varchar. We immediately understood that it will cause a lot of trouble in the visualization stage if it is not properly dealt with. The first tasks were to give each column its correct data type. We changed the time to HH:MM:SS format, removed decimal from all columns, calculated the duration values where it was needed. We also dropped columns that were not needed, we changed all blanks in delay into NULL.

3. Dashboard development:

For this step we used Power BI to create dashboards for our project. It was important for us to create something that is user-friendly (various filtering and drill-down options), clear and concise.

Results

For the first task we had to analyze and visualize the total number of flights during the period 2014-2018. In our visualization we made a separate dashboard for this task where we showcase data about total scheduled flights, total operational flights and count of cancelled flights. From the line chart visualizations there is a noticeable change in the number of flights from 2018. During the time period of 2014-2017 we can see that during the year there were around 5.7 million flights in the USA, and it noticeably increased in 2018. This trend could be explained by economic growth and the expansion of rout networks (Dan Reed, April 04, 2018, "Airlines' Domestic Growth Plans Will Keep Fares Low (And Passengers Crammed) In 2018", www.forbes.com). We also made several card visualizations to show total scheduled flight, total operational flights and count of cancelled flights. From these cards we see that during five years from 30.1 million scheduled flights there have only been around 482 thousand cancelations.

When working with visualization of delays we found out that during the period of 2014 to 2018 the total delay flights during this was 14 million, and total flights was 30.13 million. The average arrival delay flights were highest in 2014 and lowest in 2016, whereas same trend was seen in the average departure flights. A closer look at quarter and monthly patterns reveals a consistent trend with June, July, August and December showing the highest delay for both arrival and departure, which aligned with peak holiday travel periods.

Ground processing time: "Ground processing time" refers to the time required for an aircraft to complete all necessary procedures on the ground before its next flight. This includes Turnaround Time, Passenger Processing, Baggage Handling, Maintenance & Refueling.

In the given data, ground processing time is determined by summing the values of the **TAXI IN**, **TAXI OUT**, and **Departure Delay** fields.

The average ground processing time for airlines over five years is **59.1 minutes**, which is considered reasonable.

Observation: The ground processing time has increased during the summer vacation and Christmas vacation

Comparison to Industry Standards:

"An average of 59.1 minutes is relatively favorable when compared to industry standards, where typical ground processing times can range from 45 to 90 minutes, depending on various factors."

References: ansperformance.eu

Impact of Ground Processing Time:

"Minimizing ground processing time can contribute to better on-time performance, higher customer satisfaction, and more efficient use of aircraft, ultimately leading to cost savings for airlines."

The passenger feedback section:

Process

Passenger satisfaction plays a critical role in evaluating airline performance. However, since our dataset did not include a direct feedback column, we developed a proxy satisfaction score using available data on delays, cancellations, and other operational factors. We devised a method to estimate passenger satisfaction based on measurable factors. For this we created a scoring system from 1 to 5 (where 1 is lowest and 5 is the highest).

The first step was to identify key factors affecting satisfaction. We selected the following variables from the dataset as an indicators of passenger experience:

- **Flight Delays:** Longer delays generally lead to decreased satisfaction.
- **Cancellations:** Flight cancellations negatively impact passenger experience.
- **Delay Reasons:** Delays and cancellations caused by the airline and NAS tend to frustrate passengers more than weather or security related delays.
- **Duration of Delays:** Longer delays (more than 30 mins) frustrate passengers more than shorter delays (less than 30 mins)

Results and Insights

After applying the proxy satisfaction model, we gained several insights:

1. Delays and Satisfaction Correlation

- Flights delayed by **over 30 minutes** had significantly lower satisfaction scores.
- Short delays (**under 30 minutes**) had minimal impact on passenger sentiment.

2. Impact of Cancellations & Delay Reason

- Airline-related cancellations & delay led to the lowest satisfaction scores.
- Weather-related cancellations and delay had slightly higher scores, suggesting that passengers were more understanding in these cases.

Discussions

Doing this project has provided a valuable insight into operational aspects of flights from 2014 to 2018. The visualizations that were created could offer a clear picture of important factors in airport and airline performance, such as flight volumes, delays, cancellations, ground processing time, and passenger satisfaction. In this part of the technical report, we will discuss the key findings.

Increase of flights

In our analysis we noticed that there is rather sharp increase in flights in 2018 in comparison to previous years. Without taking data from previous years, we would not have known that. Context is very important and without it we cannot truly understand and analyse the data. And even though the number of flights have increased there hasn't been a sharp decrease in customer satisfaction or increase of delays, which could indicate that airports are able to maintain a high level of operational efficiency despite the increase of volume.

Seasonal influences

During our analysis we noticed a trend that the peak seasons of traveling have noticeable impact in airport operations. There is an increase in airport delays, cancelations and ground processing time which understandably impacts customer satisfaction score. Although these findings are not surprising, the airport management should look into improving operations during high-traffic months.

Challenges

While working with such a high-volume dataset we encountered the following challenges:

- When analysing departures and arrivals into cities we quickly noticed that were missing airport codes in our airport table, so we had to fix this by researching missing values.
- Although it gave us great insight in the data, the decision to look into 2014-2018 data and not just 2018 made our work more complex and challenging.

- Confusion in ground processing time calculation. Had to use the data that we have, to figure out the best way to calculate it. Some records had missing values for taxi times or delays, making it difficult to calculate ground processing time accurately.
- While developing and visualizing the proxy satisfaction score, we had to calculate, the proxy score. It was based on assumptions about how passengers perceive delays and cancellations as we did not have direct passenger feedback.
- Additionally, the score could not capture qualitative factors such as customer service experience or communication from airlines.
- Some flights lacked delay reasons, making it difficult to assess the true impact on passengers.

Conclusions

- There is a noticeable increase of number of flights in 2018 in comparison to 2014-2017. This could be due to economic improvement as well as increase of network routes.
- There is a noticeable trend when we look more closely at differences between months in delays. June, July, August and December are showing the highest delay for both arrival and departure
- Over the span of 5 years the average ground processing time is 59.1 minutes. As we mentioned earlier peak travel seasons also impact ground processing time.
- There are a lot of factors that could impact that customer satisfaction is lower or higher, like, delays, cancellations and other factors.

References

<https://www.kaggle.com/datasets/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018>

<https://www.forbes.com/sites/danielreed/2018/04/04/airlines-subtle-domestic-growth-plans-for-2018-likely-mean-fares-wont-be-rising-this-year/>

Ansperformance.eu