Table of Contents

Link to the dashboard:

https://public.tableau.com/app/profile/ankit.vohra5766/viz/MIS41040Team40/Story1?publish=yes

I. Introduction

Modern transportation is significantly influenced by the airline sector. Each year, millions of people depend on air travel to get to their destinations, making it a crucial part of international trade and communication. However, flight delays are one of the biggest problems the sector must deal with. Staffing challenges, air traffic control, aircraft inspections, security concerns, and weather conditions are just a few of the many causes that might cause flight delays. These delays might interfere with travel plans and have an impact on the economy.

1. Objective

To understand the root cause of aircraft delays, we developed a Decision Support System (DSS), which analyses the flight data and offers information about delays.
This DSS's goal is to offer information to stakeholders that will enable them to make wise decisions about infrastructure planning, flight operations, and scheduling. We set out to provide answers to important issues, such as what delays are caused by, how long they last on average, if they are seasonal, how long each airline delays departure relative to the destination, and which airport experiences the most delays. We also did a narrowed-down analysis on airlines, airports, types of delays, and time of delays.

2. Stake Holders

Airline companies, airport administrations, and travelers are the analysis stakeholders.

- **Airlines:** In order to take action to reduce delays, airlines must be able to recognize the reasons that cause them. They must also be able to anticipate delays and make appropriate plans.
- **Airport officials:** In order to streamline airport operations and cut down on delays, airport authorities must be able to pinpoint the causes of delays.
- **Passengers:** Passengers must be able to plan their travel in advance, taking into account the possibility of delays.

3. Information Requirements

The domestic US flight statistics airport and airline data for the year 2018 were used for the development of this system. The flight data includes details on more than 5 million inter-state and intra-state flights, such as the time of departure and arrival, the operating airline, the origin and destination airports, the length of the delay, the types of delays, and other pertinent data. The airport data includes details of 323 US airports, such as their name, IATA code, and location. Over 1500 airlines are listed in the airline data, of which 28 were relevant for our analysis. It provided information about the airlines, including name and code.

4. Approach

Tableau, a data visualization program that offers a potent platform for data exploration, analysis, and visualization, was leveraged to develop this DSS. To understand the data, we used a variety of analytical methods, including counts, ratios, averages, and

correlations. We performed the data cleaning and transformation task using Python, where in we fixed the missing data and dropped the corrupted data. Additionally, we added interactive elements that let users change how information is presented, such as filters and interactions over various visual representations present in the dashboards. To ensure simplicity of use and navigation through the system, operational guidelines have also been provided in the later section.

II. Data cleaning and transformation

For understanding the data at hand, we did our exploratory data analysis (EDA) using Python. The EDA was done to formalise the columns we will be needing for our Key Performance Indicators (KPI).

1. Airline data

All the raw files containing US flight domestic flight records from Jan 2018 to Dec 2018 were used for this analysis.

- **EDA**
  It involved understanding the data types, data structures, and fundamental statistics like mean, median, mode, range, standard deviation, and variance. With the aid of this data, we were able to spot potential issues such as outliers, data conflicts, and missing values. It also aided us in determining the proper data cleaning and pre-processing methods that should be used on the data.

- **Dimensionality Reduction**

  It involved reducing the number of features or columns in the dataset. The strategy involved was to find the columns that could support our KPIs. 49 columns out of 120 available were kept for further cleaning and transformation.

- **Data deletion**
  Nan-value data was dropped which included:
  1. Flights that were cancelled without the presence of any type of delay. Since these columns could not provide a reason for cancellation.
  2. Flights with zero departure delays, which had null departure time and delay causes since neither factor could validate the delay.

- **Fixing missing values**
  The following columns which initially had null values were appended:
  1. Departure delay of flights in which “CRS Departure time” and “Actual Departure time” were present.
  2. Arrival delay of flights in which “CRS Arrival time” and “Actual Arrival time” were present.

3. Arrival delay in minutes and departure delay in minutes were calculated using departure delay and arrival delay columns.
4. CRS Elapsed time was calculated using CRS arrival time and departure time.

- **No scope for fixing missing values**

  Missing values in DepDelay: Since there is no date column distinguishing the scheduled departure and actual departure. So, to treat null values we cannot identify whether it is an early departure or delayed departure, while calculating DepDelay.

  Example:

  CRSDepTIme = 2302
  DepTime = 102

It can be either 2 hours delay of 10 hours early departure. Here, its evident based on the numbers that 2 hours delay makes more sense but when one of the below case occurs:

Index	DepTime	CRSDepTime	DepDelay
5688952	105	2200	185

This calls for more information about the date of the flight took off, actual departure and scheduled departure.

- **Using nulls information as is**
  The following columns with null values were used as it is:
  1. Flights that had actual elapsed time as null and were diverted had diverted actual elapsed time which was relevant in this case. Hence, the nulls of actual elapsed time were kept as it is.
  2. All the diverted parameters for the records which didn't have a diverted flight were accepted as null and were kept as it is throughout the analysis.

2. Airport Data

The final generated file for flight records of 2018 was joined with the Airport data file using the left join on the columns Origin/Dest with IATA codes to identify the name and location of origin and destination airports.

3. Airline Data

Since the operational airline name and code were provided in the raw files, no join was made with the airline data file. The analysis identified a total of 28 airline companies that were operational within the US.

4. Feature Engineering

The following parameters were calculated in Tableau to aid the development of the intended system:
a. **Inter_Intra:** When ORIGIN_STATE = DESTINATION_STATE, this helps to identify the flights within a state and between 2 statees.
b.

III. Key Performance Indicators

The following are the key performance indicators for the decision support system:

No.	KPI	Description
1.	Flight Count	The Flight Count measures the total number of flights, arrival delay flights, and departure delay flights. This aids in determining the most common forms of delays as well as how frequently planes are delayed.
2.	Average Delays	The Average Delays calculates the average delay in minutes for all types of delays, including arrival delay, departure delay, carrier delay, weather delay, NAS delay, security delay, and late aircraft delay. The average delays are also calculated for interstate and intrastate flights. This KPI helps to understand the overall delay situation and identify the main factors contributing to delays.
3.	Flight Delay count %	The Flight Delay count % calculates the percentage of flights getting delayed for all types of delays. This helps to understand the magnitude of delays and their impact on flight operations.
4.	Busiest Airport	The Busiest Airport KPI is based on the total number of flights arriving at the airports of each state. This helps to identify the busiest airports and the traffic volume of each state.
5.	On-time flights and their percentage	The On-time flights and its percentage KPI measures the count of on-time flights and its percentage share out of total flights. This helps to understand the performance of airlines in terms of delivering flights on time.
6.	Delay Hotspots	The Delay Hotspots identify arrival and departure state hotspots based on their respective average delays in minutes. This aids to pinpoint the states with the biggest average delays and the causes behind them. It also helps to prioritize the states for improvement in terms of reducing delays.
7.	Average Ground Flight time	The time between Elapsed time and Airtime of a flight is treated as Ground Time to identify the

IV. Conclusion

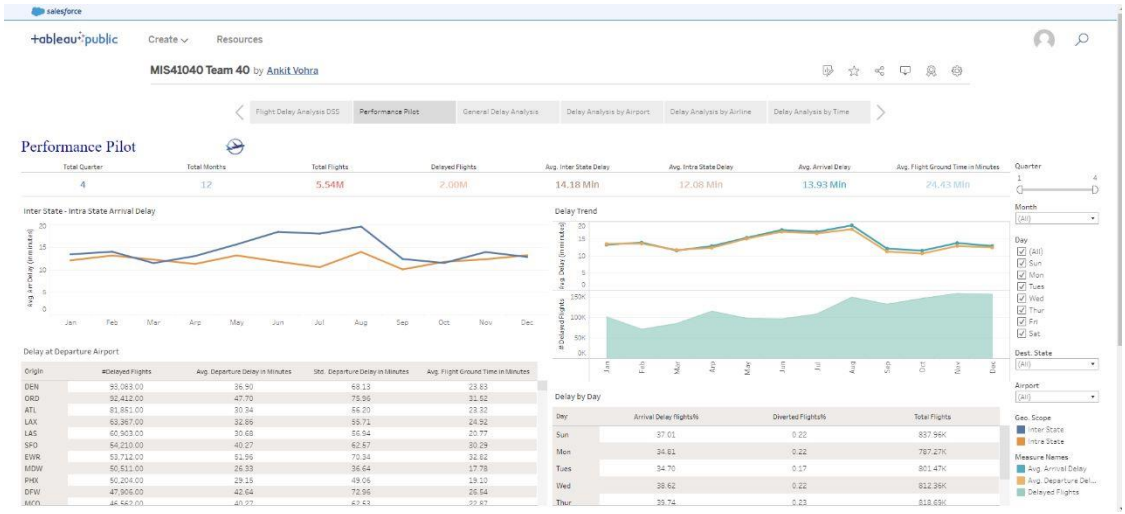
1. On average, Inter-state delay of flights arrival is greater than intra-state, but during months March, October, and December the gap reduced significantly.
2. Denver(DEN) Airport has maximum number of flights delayed in 2018 followed by Orchard(ORD) Field Airport but the departure delay time and deviation from the mean indicates different factors responsible for delay.
3. After the month of August, although the overall number of delayed flights went up but the average delay time continued to fall down and stabilize, probably because of large number of flights with small delays during the peak travelling season.
4. There is close positive correlation between arrival delay time and departure delay time, which is obvious as the external factors attributing to the delay like weather, staffing, NAS, carrier, and security are comparatively less frequent.

5. New Jersey, NewYork, South Dacota, Puerto Rico are few of the states with highest flight delays given the large traffic that go through these airports can be a red flag as this could be critical to business of airlines and airports.

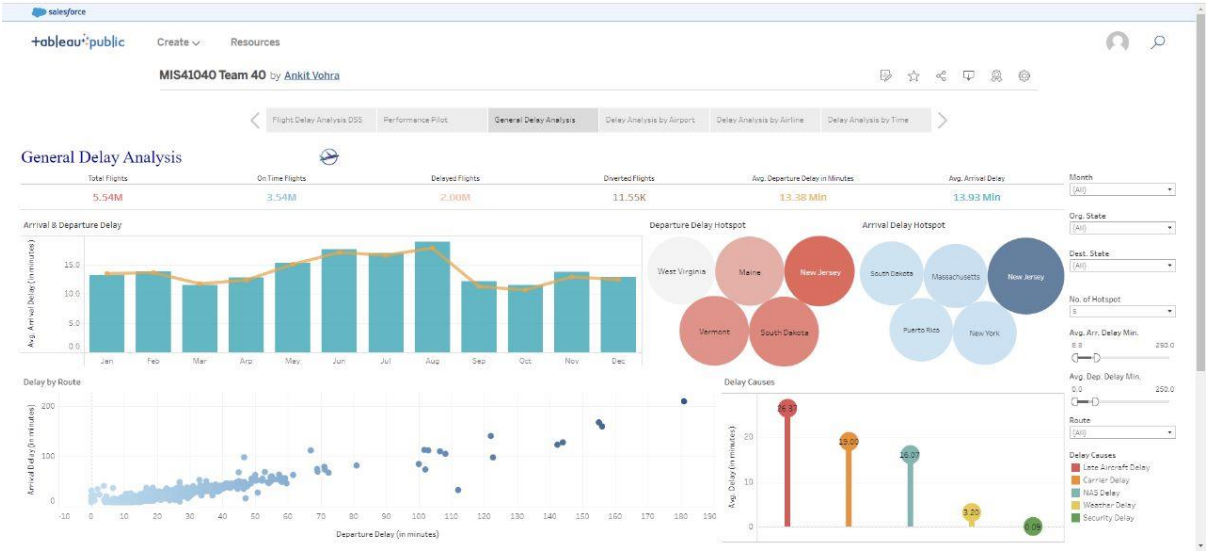
V. Operational Guidelines

1. First Page

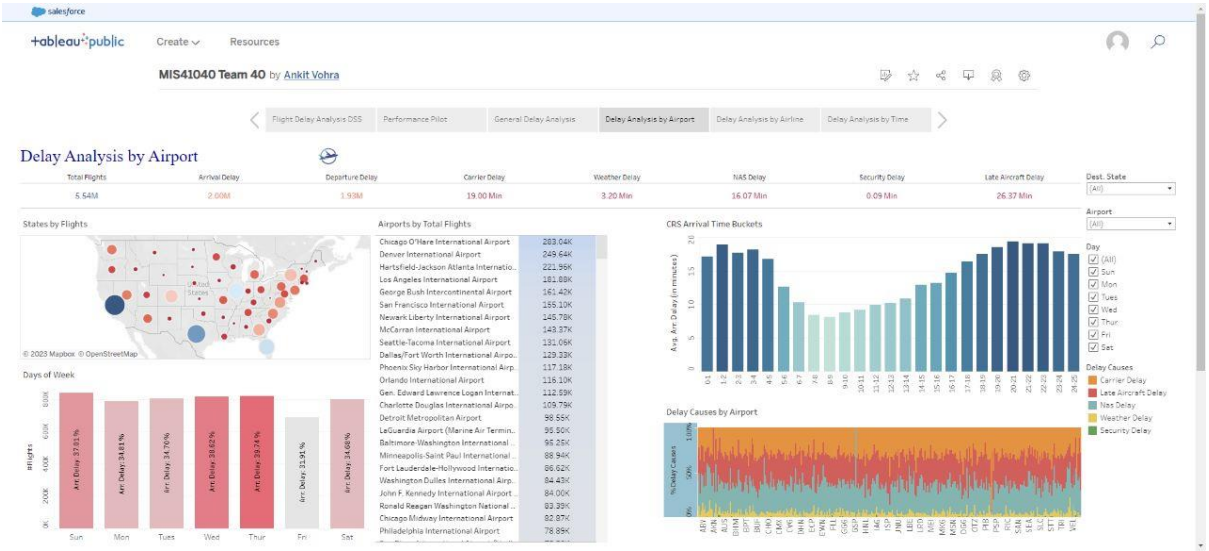
2. **Performance Pilot :** The performance pilot dashboard provides an overview of the delay analysis for inter-state and intra-state flights within US. By analyzing and comparing the arrival delay trends for inter-state versus intra-state flights, we can potentially identify areas for improvement in airline operations and management, as well as potential solutions to reduce delays and improve customer satisfaction. The filters provided are on Quarter, month, day of the week, airport, and destination state. All the charts provided support interactions.



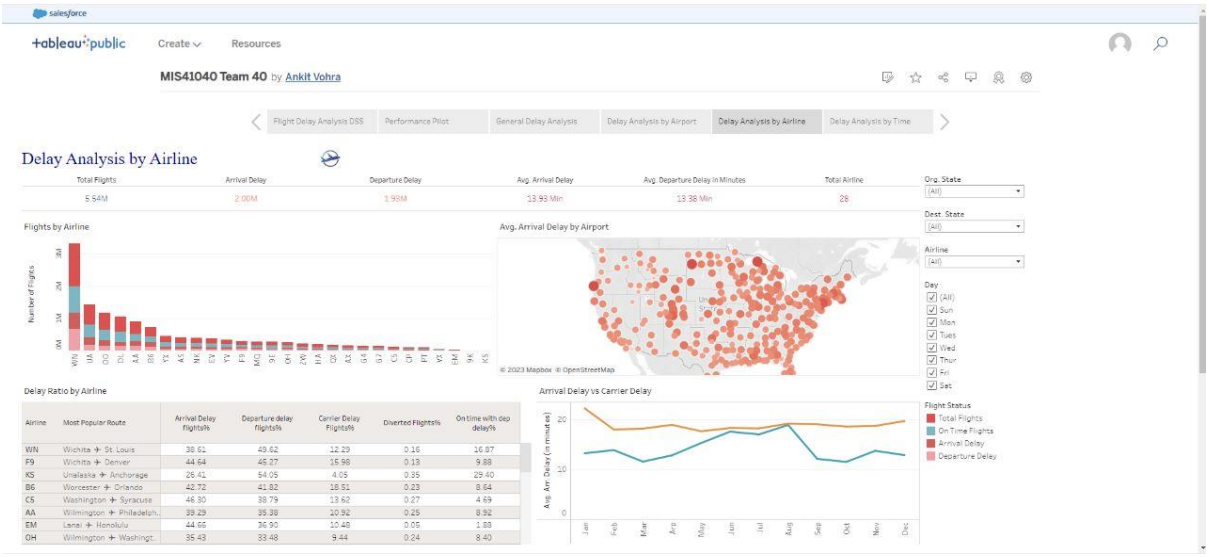
3. **General Delay Analysis:** This dashboard provides general delay analysis for different months and delay hotspots which are based on states. By analyzing the "Departure Delay Hotspot" and "Arrival Delay Hotspot" charts, airlines and airports can develop targeted strategies to address the root causes of delays at specific locations or routes. For example, if an airport consistently experiences departure delays due to airport congestion, the airport may consider expanding their facilities or implementing new procedures to alleviate congestion and reduce delays. The filters provided are on month, day of the week, airport, state, average delays and route.



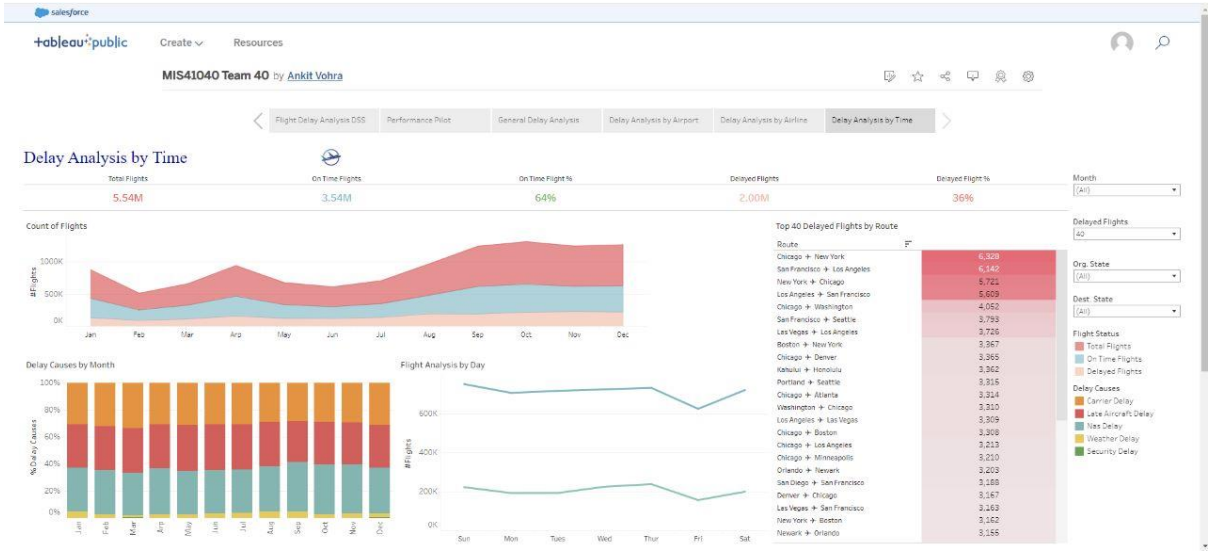
4. **Delay Analysis by Airport:** This dashboard provides an overview of the delay analysis for arrival delay airports. Overall, a "CRS Arrival Time Buckets on Arrival Delay Bar Chart" can help to identify specific time periods that are most prone to delays and inform decision-making to improve operational efficiency and reduce delays for airlines and airports. The filter provided are on airport, destination state and day of the week. All the charts provided support interactions.



5. **Delay Analysis by Airline:** The performance pilot dashboard provides an overview of the delay analysis for airlines. An "Arrival Delay vs Carrier Delay" chart typically shows the relationship between the total arrival delay and the portion of that delay that is attributed to the carrier. This chart can help airlines and airports to identify the extent to which delays are caused by factors within their control, such as operational issues or scheduling, versus external factors such as weather or air traffic control. By analyzing the "Arrival Delay vs Carrier Delay" chart, airlines can identify the proportion of delay time that is attributable to the airline itself, which can help in understanding the causes of the delay. For example, if the chart shows that a significant portion of the total arrival delay is attributed to the carrier, it may indicate that the airline needs to improve its operational efficiencies, such as its scheduling or maintenance procedures, to reduce its own delay times. All the charts provided support interactions. Overall, an "Arrival Delay vs Carrier Delay" chart can help airlines and airports to identify the root causes of delays and develop targeted strategies to improve operational efficiency and reduce delay times.



6. **Delay Analysis by Time:** This dashboard provides an overview of the delay analysis by time. By analyzing the "Delay by Day" chart, we can identify the root causes of delays and take corrective actions to prevent or minimize them in the future. For example, the chart shows that delays consistently occur on Fridays, we may investigate whether there are any operational issues or bottlenecks that need to be addressed on Fridays to prevent delays. Overall, a "Delay by Day" chart can help to visualize and understand the extent and impact of delays, and inform decision-making to improve processes and reduce delays in the future. The filter provided are on month, delayed flights count ,origin and destination state. All the charts provided support interactions.


salesforce
+ableau public
Create
Resources
MIS41040 Team 40 by Ankit Vohra
Flight Delay Analysis DSS
Performance Pilot
General Delay Analysis
Delay Analysis by Airport
Delay Analysis by Airline
Delay Analysis by Time
Delay Analysis by Time
Total Flights
5.54M
On Time Flights
3.54M
On Time Flight %
64%
Delayed Flights
2.00M
Delayed Flight %
36%
Count of Flights
#Flights
1000K
500K
0K
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Top 40 Delayed Flights by Route
Route
Chicago → New York 6,328
San Francisco → Los Angeles 6,142
New York → Chicago 5,721
Los Angeles → San Francisco 5,609
Chicago → Washington 4,052
San Francisco → Seattle 3,793
Las Vegas → Los Angeles 3,726
Boston → New York 3,367
Chicago → Denver 3,365
Kahului → Honolulu 3,362
Portland → Seattle 3,315
Chicago → Atlanta 3,314
Washington → Chicago 3,310
Los Angeles → Las Vegas 3,309
Chicago → Boston 3,308
Chicago → Los Angeles 3,213
Chicago → Minneapolis 3,210
Orlando → Newark 3,203
San Diego → San Francisco 3,188
Denver → Chicago 3,167
Las Vegas → San Francisco 3,163
New York → Boston 3,162
Newark → Orlando 3,158
Month
(All)
Delayed Flights
40
Org. State
(All)
Dest. State
(All)
Flight Status
Total Flights
On Time Flights
Delayed Flights
Delay Causes
Carrier Delay
Late Aircraft Delay
Nas Delay
Weather Delay
Security Delay
Delay Causes by Month
% Delay Causes
100%
80%
60%
40%
20%
0%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Flight Analysis by Day
#Flights
600K
400K
200K
0K
Sun
Mon
Tues
Wed
Thur
Fri
Sat

References

1. Hati.Royani (2019) “Download Aircraft, airplane, airline logo label. Journey, air travel, airliner symbol. Vector illustration for free,” Vecteezy [Preprint]. Available at: https://www.vecteezy.com/vector-art/620372-aircraft-airplane-airline-logo-label-journey-air-travel-airliner-symbol-vector-illustration.

2. Owobu, F. (2021). Flight Delays and Cancellations in 2015 [Tableau visualization]. Tableau Public. https://public.tableau.com/app/profile/faith.owobu/viz/FlightDelaysandCancellations1_16539176411580/2015USAFLIGHTDELAY?publish=yes

3. Tableau: Analyzing Flight Delays and Cancellations – Anastasiia Demydenko (2018). Available at: https://www.analyticsgirl.com/2018/04/21/tableau-airline-delays-analysis/

4. Thakrani, S. (n.d.). Flight Delay Analysis.[Tableau visualization] Tableau Public. Retrieved April 16, 2023, from https://public.tableau.com/app/profile/saurabh.thakrani/viz/FlightDelayAnalysis-2/Story1?publish=yes

5. Thakrani, S. (2021) “FLIGHT DELAY ANALYSIS - Saurabh Thakrani - Medium,” Medium, 6 December. Available at: https://medium.com/@saurabhthakrani/analyzing-flight-delays-d8cba808bcd3