



MIS41040-Business Decision Support Systems - 2022/2023

Report of Assignment Team - 14

| Student Name | Student Number |
|-----------------|----------------|
| Arundhati Ghosh | 22200223 |
| Roshmi Dutta | 22200033 |

Flight delays in the airline industry can be caused by various factors such as staffing issues, air system delays, air traffic control, aircraft checks, security concerns, aircraft management, weather, and more. The delay is defined as the duration of time by which a flight is postponed or runs late, which can result in cascading issues such as missed connecting flights.

For this assignment, we were provided with the following datasets:

1. *Flights_2018.csv*,
2. *airlines.csv*, and
3. *airports.csv*

Data Cleaning

The basis of any informative DSS begins with data cleaning. Upon observation of the datasets, our initial approach was to address inconsistencies in the raw data. In the process of identifying missing or null values, we made the decision to eliminate columns with over 90% of missing values. This was done by using Python coding in jupyter notebook. After the initial cleaning was done, we imported the dataset into Tableau Prep for further exploration and cleaning. From there, we determined which were the relevant fields needed for data-driven decision-making out of the 120 fields and removed the unwanted ones. We also checked for data-type mismatches in Tableau Prep and ruled out any inconsistencies present.

An important note to be observed here was that several columns like WeatherDelay, CarrierDelay and similar ones had null values, but that was not removed because it is anticipated that those null values arose due to the absence of any delay.

To make informed decisions about flight delays, the airline industry relies on specific types of information, which must be accurate and timely. These include real-time flight status updates such as departure and arrival times, gate assignments, and cancellations, as well as weather information, airport capacity details, and air traffic control updates on airspace congestion and rerouting options. Passenger information, such as the number of affected travellers, is also crucial for airlines to decide on rebooking or compensation. In summary, access to accurate and timely information on flight status, weather, airports, air traffic control, and passengers is essential for effective decision-making in the airline industry.

Data Visualization

Using Tableau, we visually analysed and interpreted the data to identify the factors contributing to flight delays and determine the peak season when delays occur.

After cleaning the *Flights_2018.csv* file, we joined it with the *airlines.csv* and *airports.csv*. We have created some dashboards with dynamic filters for the user to navigate with ease and get a rough estimation of the factors affecting delays.

Departure delays for each airport and city

This geographical heatmap gives the user an idea about the average departure delays arising at each airport and city. The red regions experience higher delays. Users can filter based on quarter.

Quarterly delay causes

It shows the various delays arising on an average seasonally.

Airline Delays

Here, we have shown the most significant airlines that are affected by delays. The delays here have been gathered by constructing a 'calculated field' in Tableau that adds up all the different delays. The bigger the circle, the greater the delay. Again, the user has the choice to filter based on quarter.

Total delays for each airport

This packed bubble graph shows the airport-based delay.

Central Tendency for each airline

Here a treemap portrays the various airlines under operation and the average, median and standard deviation of departure-related delays for them.

Diverted and cancelled flights for each destination

A combined graph has been plotted to describe the linear progression of diverted and cancelled flights across all 12 months.

Delay cause for each airline

This is a complete breakdown of all the delays, grouped together on the basis of each operating airline. Users can filter each airline to get an overview of the delays.

Average delay minutes for each state

This geographical map shows the average arrival and departure delays for each state.

These are our observations from the dataset:

1. There are 4 main reasons for delay across the geographic region, namely:
 - a. Arrival Delay
 - b. Carrier Delay
 - c. Late Aircraft delay
 - d. NASDelaySecurity delay is the lowest across all quarters and late aircraft delay is the highest, mostly in Q3.
2. The maximum number of delays have been seen to occur in the third quarter and specifically in August month. The reasons for this can be accounted to summer vacations and more tourists travelling worldwide.
3. Region-wise, the state of Vermont has seen the highest number of delays due to its and airport-wise, the highest number of delays is JFK, New York.
4. Airline JetBlue Airways has the highest number of delays across the year and connects some busy airports across the USA like Chicago, San Diego and Denver.
5. Comparatively, diverted and cancelled flights do not seem to occur with much frequency, although 51 flights got cancelled in January and 7 diversions in March. It drops towards the end of the year.

Based on our observations, we hope to provide a DSS that will help users to make an informed decision about the right time to board a flight and the delays to be concerned about.

Link to Tableau Public: [MIS41040 Team 14 Tableau Dashboard Link](#)

