Business Decision Support System
Tableau Assignment

Student Name	Pranjal Agarwal	Sonal Kashyap	Vineet Paulson
Student Number	22201084	22200178	22200117
Assessment Title	US Flights Tableau Assignment		
Module Code	MIS41040		
Module Title	Business Decision Support System 2022/23-Spring		
Module Co-Ordinator	Peter Keenan		
Date Submitted	16/04/2023		
Date Received			
Grade/Mark			

A SIGNED COPY OF THIS FORM MUST ACCOMPANY ALL SUBMISSIONS FOR ASSESSMENT. STUDENTS SHOULD KEEP A COPY OF ALL WORK SUBMITTED.

Procedures for Submission and Late Submission
Ensure that you have checked the school's procedures for the submission of assessments.

Note: There are penalties for the late submission of assessments. For further information please see the University's ***Policy on Late Submission of Coursework,*** (http://www.ucd.ie/registrar/)

Plagiarism: the unacknowledged inclusion of another person's writings or ideas or works, in any formally presented work (including essays, examinations, projects, laboratory reports or presentations). The penalties associated with plagiarism are designed to impose sanctions that reflect the seriousness of the University's commitment to academic integrity. Ensure that you have read the University's ***Briefing for Students on Academic Integrity and Plagiarism*** and the UCD ***Plagiarism Statement, Plagiarism Policy and Procedures***,(http://www.ucd.ie/registrar/)

Declaration of Authorship

I declare that all material in this assessment is my own work except where there is clear acknowledgement and appropriate reference to the work of others.

Signed: Vineet Paulson, Pranjal Agarwal, Sonal Kashyap
Date: 16/04/2023

Table of Contents

Introduction

This project aims to build simple and interactive data dashboards using Tableau, which will display visual representations of a dataset sourced from Kaggle. The dataset contains US Aviation information on various airports and airlines, including delays, cancellations, dates, origin, and destination airports of 2018. The primary objective of these dashboards is to analyze the dataset and identify the airlines that experience the most delays and investigate the possible reasons behind these delays. The insights obtained from these dashboards can be utilized to make crucial decisions within the airline industry that can improve performance and enhance customer satisfaction.

To prepare the data for analysis, Tableau Prep is used for data pre-processing and cleaning. The dashboards are constructed using Tableau Desktop. Additionally, this information can be used to optimize and improve the airline industry, which could ultimately enhance the industry's branding and customer experience/satisfaction. The dashboards prepared are highly interactive and a data-driven system in which users/stakeholders can view and control the data using the dedicated filters on each screen. Navigation between dashboards is also made possible for the users. Along with these flexibilities, quantitative approaches like linear regression, central measures, and standard deviation are calculated and plotted on visualizations.

DSS Decision-making and its key fields/metrics

- **Operational Decisions**
  The dashboard can assist airlines in making operational decisions by monitoring flight delays, allowing them to make informed decisions. For instance, if a specific flight path experiences regular delays caused by weather conditions, airlines can adjust their schedules or modify flight routes accordingly.
- **Resource allocation**
  the dashboard can help airlines allocate their resources more efficiently. For example, airlines can utilize historical data to determine the most suitable times for maintenance on their aircraft.
- **Supply chain decision**
  Flight delays can affect supply chain operations, which include the transportation of goods by air. Companies can utilize the dashboard to recognize the airports and airlines that are most prone to experiencing delays. This information enables them to modify their supply chain operations suitably.
- **Performance monitoring**
  The dashboard can serve as a tool for monitoring the performance of airlines and other entities involved in the aviation industry. Companies can use it to monitor their own performance over time and to make comparisons with the performance of their competitors.
- **Financial decision**
  The financial performance of airlines and other entities in the aviation industry can be impacted by flight delays. The dashboard can be utilized by companies to evaluate the financial consequences of delays, such as the expenses incurred in providing compensation to passengers or the effect on revenue generated by ticket sales.

The key fields (columns) from the datasets provided that contribute to decision-making are as follows:

- Arrival & Departure Delay
- Delay Cost (Derived-metric)
- Airport
- Airline
- Flight Date
- Geo-location of airports
- Airline route

Datasets used

1. Flights2018.csv
    - This is the master dataset that contains information on primary fields like flight_date, origin, destination, airline, and airport along with other relevant fields such as arrival and departure delays, number of flights, flight route etc.
2. Airports.csv
    - This dataset is used to extract the latitude and longitude of the US airports by joining (Left join) it with Airlines.csv on IATA_CODE.
3. Airlines.csv
    - The dataset is used to extract the Airline name based on Airline code by joining (Left join) it with master dataset.

Data Flow

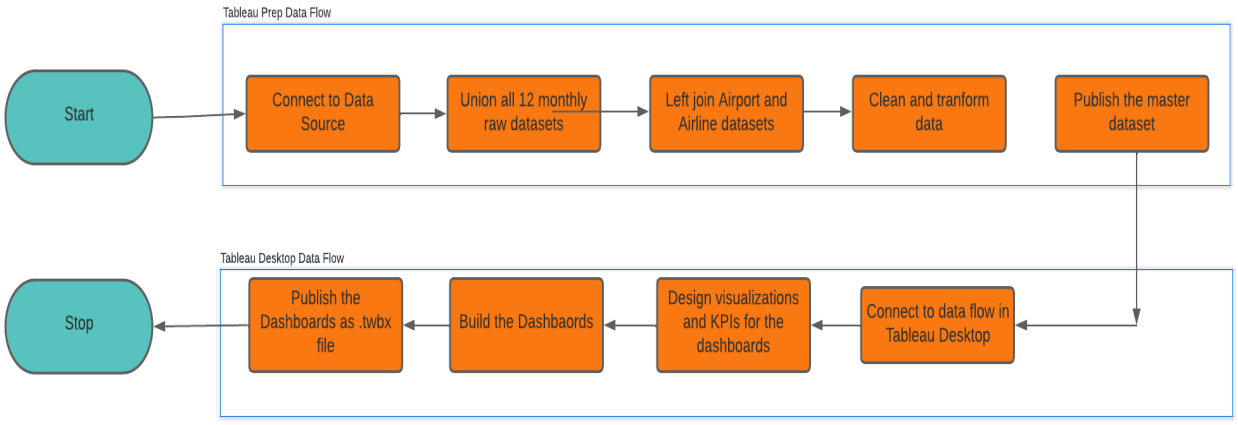

Fig. 1: Data Flow Diagram

Data Pre-processing and Data Cleaning

1. **Uniqueness/Duplicate Check**
    - All the 5.6 million records are checked for any duplicate records and no duplicates are found.
2. **Null/Missing Value Validation**

- The ArrDelay, DepDelay, TaxiIn, and TaxiOut columns contains <1% of Null values which are fixed with 0 value in the master dataset.
- Origin and Destination Airport columns along with corresponding latitude and longitude columns have less than 1% of null values which are removed as they cannot be replaced with 0 or any default value.

3. **Outliers Removal**
   - Extreme values are removed from ArrDelay and DepDelay using scatter plot visualization and Interquartile range.
   - 24 data points are removed as outliers which are extreme values.
4. **Data type Manipulation**
   - Columns namely Year, Month, and DayOfMonth are converted from Integer to String to avoid them being treated as aggregated numbers.
5. **Consistency Check:**
   - Columns namely AirTime, ArrTime, and DepTime are made consistent in hh:mm format across all the records which were given as 3 or 4-digit number (like 430 → 04:30).
6. **Trim Check**
   - Any unwanted spaces are removed from IATA code and Airline code before making join operation of the Airport and Airline dataset with the master dataset to avoid any drop of records.
7. **Left padding**
   - Left padding with '0' is applied on time columns namely ArrTime, DepTime, and AirTime to make their length consistent to 5 characters (hh:mm like 945 to 09:45) wherever needed.
8. **Removing unwanted columns**
   - Unwanted columns like Flight_Number_Marketing_Airlin, OriginStateFips, OriginWac, etc. are removed which are not required for visualization.
9. **Null data pattern observed**
   - Delay reason columns namely CarrierDelay, NASDelay, WeathreDelay, SecurityDelay and LateAircraftDelay contains over 80% of null values that are fixed with '0' value.
   - The reason is the fact that the aircraft which has an acceptable arrival delay of <15 mins have null delay reason columns.
10. **Removing repeated columns**
    - Repeating and unwanted columns are removed to prepare the master dataset.
    - Columns like origin and destination states information is repeating in OriginCityName and DestCityName and hence removed.

Statistical Measures for Data Analysis

- Central measures line such as mean is plotted in '*Top and Bottom 5 Airlines by Departure Delay*' on the '*Summary*' Dashboard to visualize the Average departure delay. Mean line is also plotted on '*Arrival – Departure Delay Comparison*' in '*Airport Analysis'* dashboard. This line denotes that the average delay is ~10.6k (mins) on an hourly basis.
- Linear regression line is plotted on '*Relation between Departure and Arrival Delay*' in the '*Airline Analysis*' dashboard. The line depicts that the increase in arrival delay is linearly proportional to departure delay. P-value <0.0001 signify that there is a strong statistical relationship between arrival and departure delay. Moreover, R value = 0.95 depicts a strong correlation between both variables.
- Median is also plotted on '*Severely and least impacted Airlines' and 'Severly and least impacted Airports*' representing that it separates the dataset into two equal halves, with 50% of the data points falling below the median value and 50% falling above it.

Tableau Dashboards

In this Decision Support System, there are 4 dashboards in total which helps a user to analyse the patterns and reasons causing flight delays.

1. Summary Dashboard

Summary dashboard provides an overall metrics and visualizations for the delays and depicts the impact of flight delay in US in 2018.

What KPIs mean in the dashboard?

1. **Total flights:** The KPI indicates the total number of flights that were operated by all airlines collectively in United States in 2018. From the dashboard, we found out that there were approximately 5.6 million flights that helped passengers to travel within US.
2. **A15% Arrival on Time:** The percentage of flights that arrive at their destination within 15 minutes of their planned arrival time is referred to as "A15% Arrival on Time" in the airline business. This measurement is frequently referred to as "15-minute arrival factor" or A15.
   From the dashboard, the overall A15% of airlines in USA was 81.2% in 2018, which means out of every 1000 flights, 812 flights arrived within 15 minutes of their planned arrival time.
3. **D0% Departure on Time:** "D0% Departure on Time" refers to the proportion of flights that leave their origin airport at or before the scheduled departure time.
   It is crucial metric in Decision Support System for airlines since it shows how well they operate and how well they can keep to their flight schedules as delays at the point of departure can lead to further operational problems and possibly financial losses for airlines.
4. **CF% Cancellation Factor:** The proportion of scheduled flights that are cancelled is referred to as the "CF% Cancellation Factor" in the airline business.
   The users of this DSS (specifically airlines) can track the CF% Cancellation Factor to look for patterns and trends that can point to operational opportunities for improvement.
   Overall, a low CF% Cancellation Factor is desirable for airlines, as it indicates a high level of reliability and operational efficiency. vitable due to safety issues or other unanticipated events.
5. **Total Delay:** The metric gives total delay in minutes, i.e. total departure and arrival delays. The metric helps to analyze the total amount of loss incurred by airlines due to delays and therefore impacting customers to pay higher fares.
6. **Total Delay Cost:** Total delay cost in the airline industry refers to the monetary cost incurred by an airline as a result of flight delays. This cost includes additional airport fees, labor charges, and fuel expenditures, among other things. To recover from the huge pile of delay cost, which was more than 9 billion US dollars in 2018, it mostly impacts the customers who have to pay higher fares for the travel. Therefore, the cost is a reflection of what needs to be improved in order to perform more efficiently and improve profitability.

Design and Filters in dashboard-

All the metrics and graphs presented in the dashboard can be filtered out on basis of below mentioned metrics:

- Origin state name- To navigate through the departure flight delays state-wise, the user has an option to choose one or more US state at a time. This helps user to find out metrics per state/states.

- Airline name- This filter allows user to sort the visualizations as per airlines, thus giving user the flexibility to get insights for flight departure analysis airport-wise and state-wise.

The heatmap provides average departure delay of flights state-wise in USA. It helps to highlight the states with most-delayed departures and therefore diving deep to find out specific airports, airlines, and reasons for it.

This bar chart helps to analyze the best and worst airlines in respect to departure delays. If flights are departed late, it causes subsequent flights to be delayed due to runway congestion, gates unavailability, baggage belts occupied, among others. Thus, it helps user to analyze which airlines need more corrective measures to avoid customer dissatisfaction and related costs.

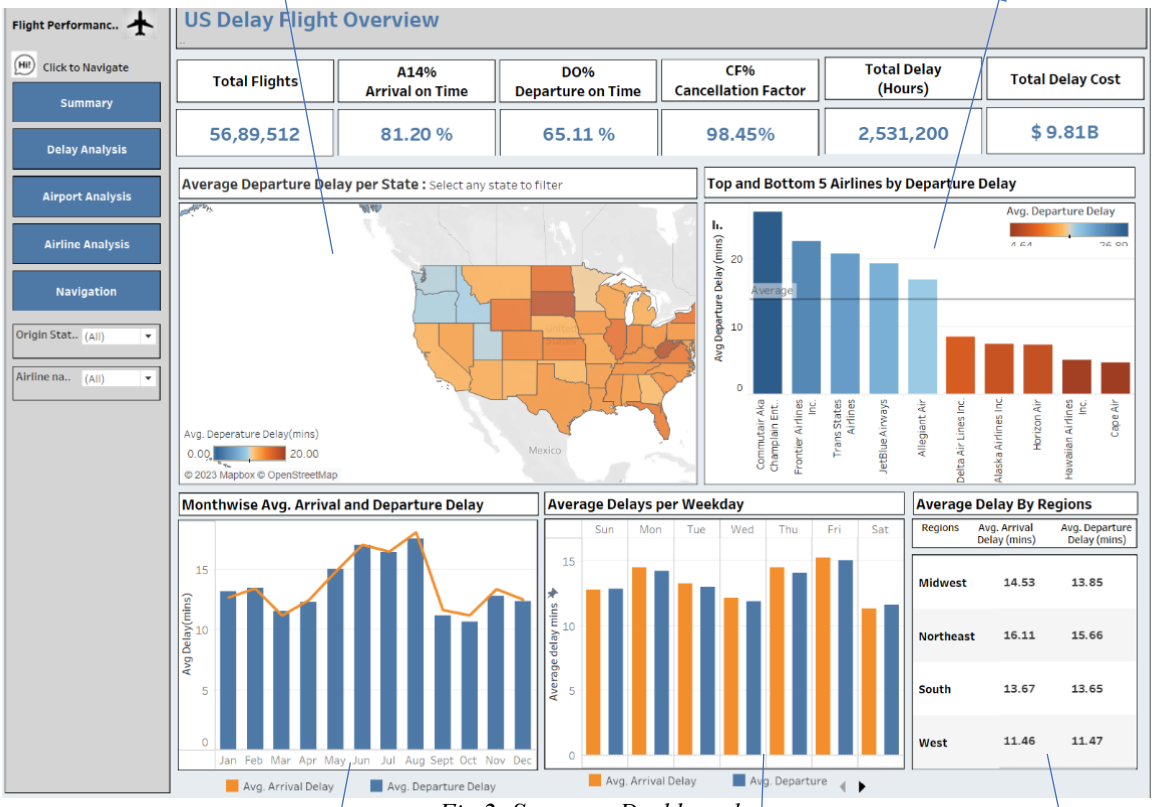

Fig 2: Summary Dashboard

The tabular sheet compares average delays based on four regions in USA, that comprises of Midwest, Northeast, South and West.

The bar chart helps user to analyze and find out the relation between departure and arrival delays, on the monthly basis. As can be observed from the graph, the average departure delay is always slightly higher than average arrival delay in first six months of 2018, which is reversed in remaining six months.

This bar graph compares the average departure and arrival delay on daily basis. The same can be filtered for different states and airlines in USA using filters.

2. Delay Analysis Dashboard

What KPIs mean in this dashboard-

1. **Flights arrived on time-** It refers to the proportion of flights that arrived at their destination airport on or before the scheduled arrival time.
2. **Flights departed on time-** It refers to the proportion of flights that departed from their origin airport at or before the scheduled departure time.
3. **Total delay-** The total delay in hours caused due to different reasons of delay impacting different airlines.
4. **Average delay-** It represents the average arrival and departure delay in minutes due to different reasons of delay impacting different airlines.
5. **Cancel Rate-** The proportion of scheduled flights that were cancelled from the filtered airport is referred to as the cancel rate.
6. **15+ min delay-** The figures mean the proportion of total flights that arrived after 15 minutes of their scheduled arrival and total flights that departed after 15 mins of their scheduled departure as per the selected airport by user.
7. **Diversion Rate-** The proportion of flights that were diverted to/from the selected airport.

Design and Filters in dashboard-

- **Reason for delay-** This filter allows user to filter the dashboard as per delay reason, so that the user can find out the metrics and analyse the graphs as per their need.
- **Airline filter-** The filter helps user to analyse the dashboard on the basis of different airlines and different reasons for delay.

This combination of line graph and stack bar chart provides an insight of the duration of total delay in hours on different days due to different reasons such as delay due to weather, late aircraft, etc. The line chart depicts total departure delay hours due to all delay reasons day-wise.

The geo bubble map of the United States provides airport-wise total average delay, i.e. sum of arrival and departure delays.

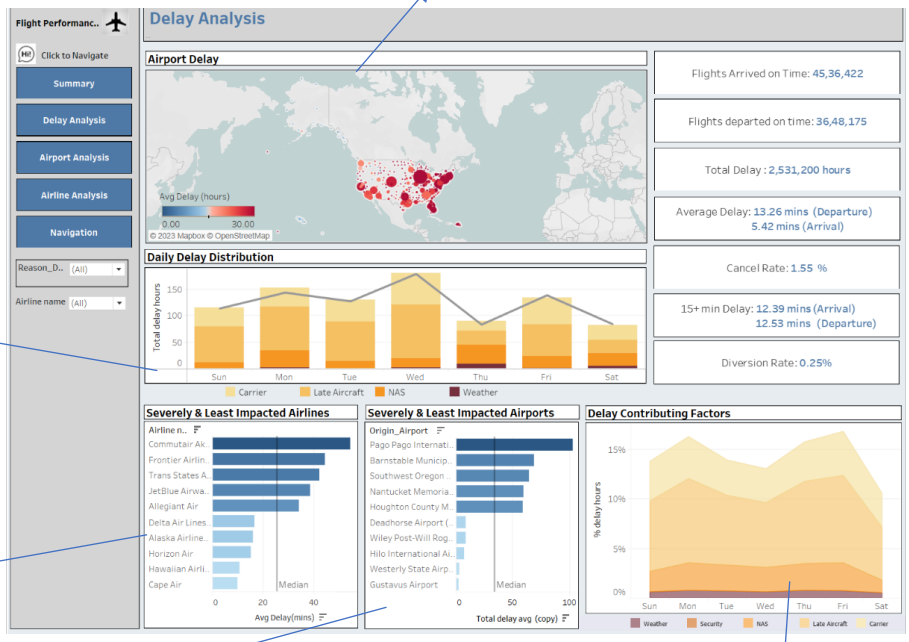

Fig 3: Delay Analysis Dashboard

The column chart provides an insight about the most and least impacted airlines due to all five types of delay reasons.

The column chart provides an insight about the most and least impacted airports due to different delay reasons.

The area chart provides an understanding about the contribution of each delay reason on the daily basis. As can be observed from the screenshot, late aircraft contributed to the highest amount of delay on any given day while security delay contributed least.

3. Airport Analysis Dashboard

What KPIs mean in this dashboard-

1. **Flights departed on time-** The figures represent the total amount of flights that departed on time from each airport.
2. **Flights arrived on time-** The figures represent the total amount of flights that arrived on time from each airport.
3. **Average delay-** The KPI gives the average arrival and departure delay in minutes at each airport.
4. **15+ min delay-** The figures mean the proportion of total flights that arrived after 15 minutes of their scheduled arrival and total flights that departed after 15 mins of their scheduled departure as per the selected airport by user.

Design and Filters in dashboard-

- **Origin Airport selection-** The user has an option to filter the KPIs and visualization charts on the basis of different origin airports.

The geo connection map depicts the average departure delay from the origin airport to all the destination airports where there is an existing route. The origin airport can be chosen from the filters list. The higher is the delay, the darker is the connection line on red color.

The daily traffic air distribution graph provides a relationship between the total amount of flights that departed from origin airport each day in 2018 and total departure delay in hours.

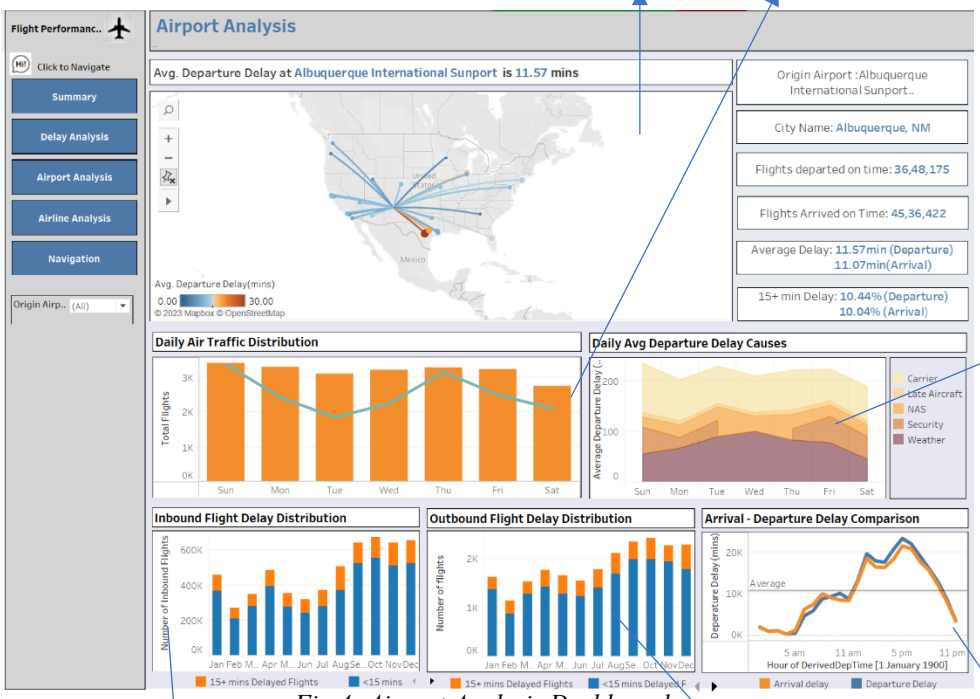

Fig 4: Airport Analysis Dashboard

The daily average departure delay causes gives an idea about the average departure delay (in minutes) due to all five reasons of delay. The area chart is plotted for day-to-day findings which can be filtered for specific origin airport from airport filter.

This line chart gives comparison of duration of arrival delay and departure delay from each airport on hourly basis.

The stacked bar chart provides flight delay distribution for outbound flights. It shows the number of flights that departed on-time v/s flights that departed more than 15 minutes after scheduled departure time.

The stacked bar chart provides flight delay distribution for inbound flights. It shows the number of flights that arrived on-time v/s flights that arrived more than 15 minutes after scheduled arrival time.

4. Airline Analysis Dashboard

Design and Filters in dashboard-

- **Airline-** The filter helps user to analyse the dashboard on the basis of different airlines operating in United States.
- **State name-** To understand all the metrics and visualizations better, the user has an option to choose one US state at a time. This helps user to analyse which airlines operate from each state and what are the routes the airline is operating on.

The geo connection map illustrates all the connection flights from one state airports to all the remaining ones. The map can be filtered for a choice of airline. It also shows the average departure delay between the routes.

The tabular sheet provides a list of most frequent routes in a state with associated average departure delay and average arrival delay in minutes.

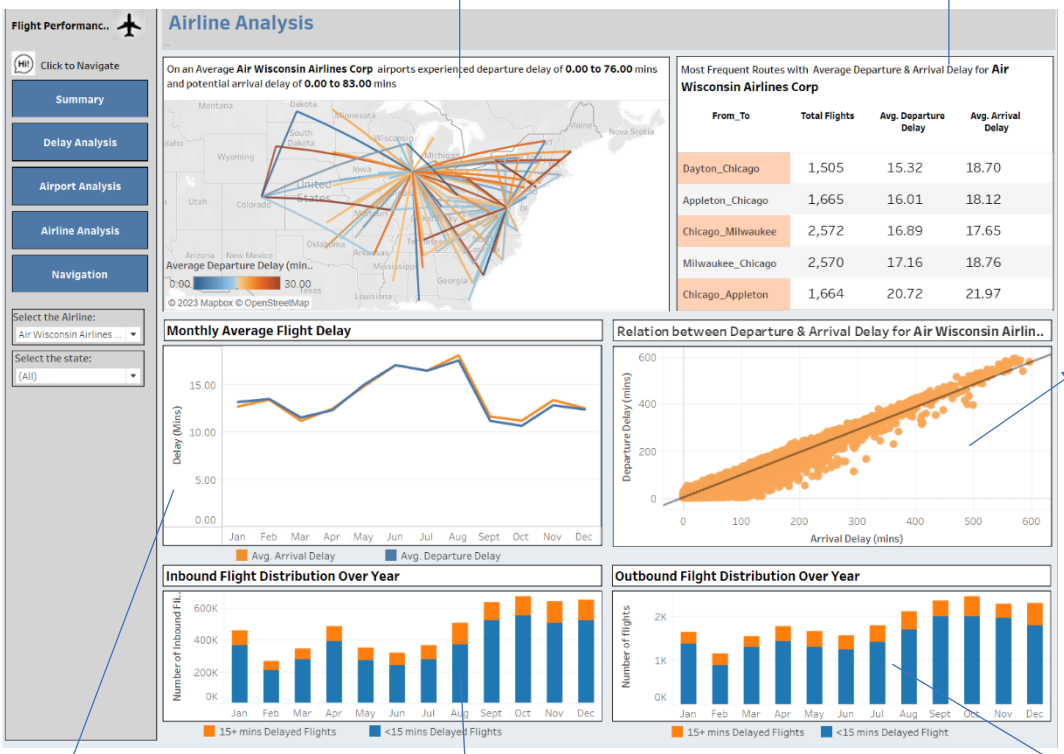

Fig 5: Airlines Analysis Dashboard

The scatter plot provides an overview of how departure delays and arrival delays are related. It also shows a best fitting regression line which shows the linear relation between the two.

The monthly average flight delay gives a comparison of arrival and departure delay on daily basis depending upon the airline and state chosen by DSS user.

The stacked bar chart illustrates the proportion of number of inbound flights that arrived more than 15 minutes late from scheduled arrival to the flights that were late by less than 15 mins.

The stacked bar chart illustrates the proportion of number of outbound flights that were departed more than 15 minutes late from scheduled departure to the flights that were late by less than 15 mins.

5. Performance Score Dashboard

The performance score dashboard provides user to compare flights metrics of different airlines on the basis of airports in all the states in US.

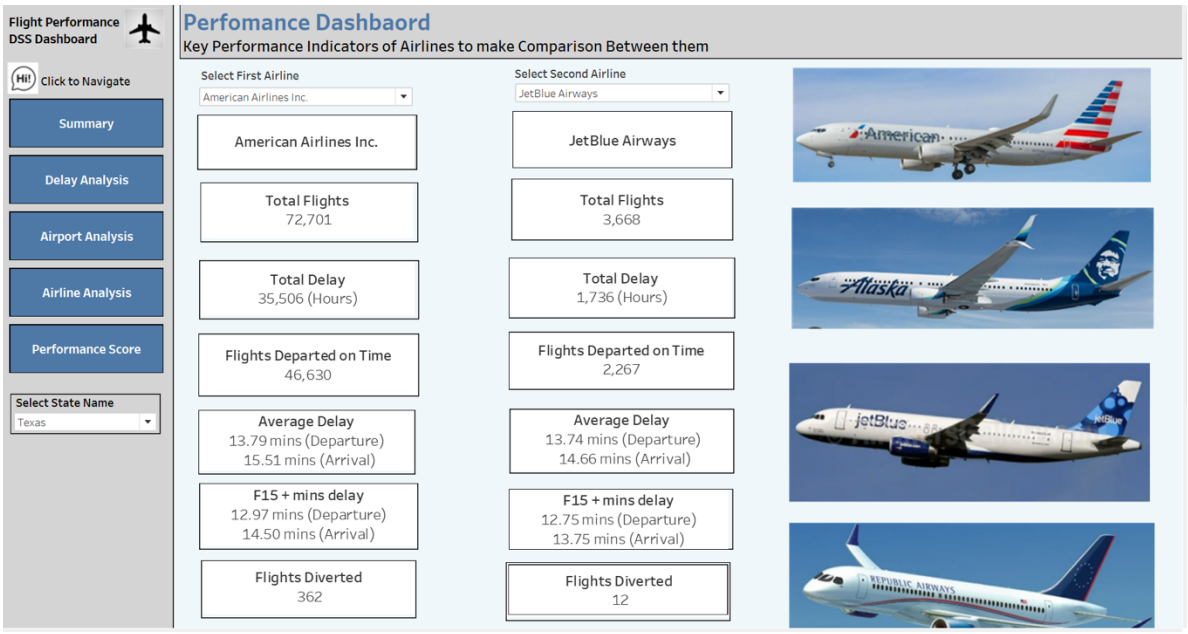

Fig 6: Performance Score Dashboard

Navigation Instructions

1. Click on the Login Button the cover page to access the flight delay analysis dashboard.
2. After logging in, you will be automatically navigated to Summary/Overview dashboard.
3. On the left grey panel, you can find the dashboard navigation buttons to navigate to your required dashboard.
4. Please note that the Airline Analysis dashboard might take some time to load the data due to high volume and computation. Kindly wait for a couple of minutes to load or refresh otherwise.
5. You can use the different filters such as Origin Airport, Delay reason, Airline Name, Origin State Name etc on each dashboard to view the data as per user requirement. Kindly note that you might see some blank charts which is due to the fact that the data for the same filters is not available in the raw file itself. Try changing the filter options if you face such issues.

Link for the online version of Flight DSS

https://public.tableau.com/app/profile/sonal.kashyap/viz/MIS41040Team5/Front_Dashboard?publish=yes

References

[1] Tableau Dashboards (no date) *public.tableau.com* Available at: https://public.tableau.com/app/search/vizzes/mis41040 (Accessed: 16 April 2023)

[2] *77 Essential KPIs Airlines Track For Improved Operations.* Available at: https://www.id1.de/2020/10/02/77-airline-operations-kpis-to-track-for-improved-operations/ (Accessed 28 March 2023)

[3] *6 Most Important KPIs For Airline Operations.* Available at: https://www.id1.de/2019/10/25/6-most-important-kpis-for-airline-operations-and-performance-analysis/ (Accessed 27 March 2023)