UCD Michael Smurfit Graduate Business School

Group Assignment – Team 12

Assignment Topic	Tableau Assignment		
Members Name	***Name***	***Student No.***	***Programme***
	Disha Shetty	22200931	MSc. Business Analytics
	Richa Gopal Sinha	22200101	MSc. Business Analytics
	Anushka Jain	22200246	MSc. Business Analytics
Module Title	Business Decision Support System - 2022/23 Spring		
Module Code	MIS41040		
Lecturers	Prof. Peter Keenan		
Submission Deadline	16th April 2023		
Session	2022/23 Spring		
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.

Name : Disha Shetty, Richa Gopal Sinha, Anushka Jain

Date : 16th April 2023

Table of Contents

Link to our DSS Tableau Public

https://public.tableau.com/app/profile/disha.shetty7122/viz/USFlightDelay2018/Story1?publish=yes

Analysis of the information requirements for decision-making.

We were provided with 2 sets of data: Assignment and Raw Data.

First, we attempted to comprehend the assignment data using a data dictionary to determine whether we had all of the criteria required to answer the questions. Further investigation and comprehension revealed that the assignment data lacked the following parameters, which were critical for visualizing distinct types of delay:

1. **Weather Delay**: caused by severe weather on departure, in route, or arrival.
2. **Carrier Delay**: It is within the control of the air carrier. Examples are Airline control. Aircraft cleaning, aircraft damage, awaiting connecting passengers or crew, baggage, bird strike, cargo loading, catering, computer, outage-carrier equipment, crew legality (pilot or attendant rest), hazardous goods damage, engineering inspection, fueling, handling disabled passengers, late crew, lavatory servicing, maintenance, oversales, potable water servicing, removal of unruly passenger, (kaggle.com, n.d.).
3. **NAS Delay**: The NAS prevents delays from non-extreme weather, airport operations, heavy traffic volume, air traffic management, etc.
4. **Security Delay**: Security delays include evacuation of a terminal or concourse, reboarding an aircraft after a security breach, faulty screening equipment, and waits longer than 29 minutes at screening stations.
5. **Late Aircraft Delay**: This is due to the same aircraft's delayed arrival at a previous airport. Delay propagation describes downstream airport delays.

When we learned that the raw data contained this information, we started to examine, understand, and clean the entire collection of raw data. To cut down on duplicate, incorrectly labeled, or redundant values, we dug deeper into the meaning and values of each column. MS Excel was used throughout the process.

Data Exploration:

1. From the shortlisted columns, we analyzed the values of each of those columns in MS Excel itself for 3 months: February, May and December. Adding snapshots for the same (Figure 1).

Headers	Details	Feb	May	December
IATA_Code_Operating_Airline	Code assigned by IATA and commonly used to identify a carrier. As the same code may have been assigned to different carriers over time, the code is not always unique. For analysis, use the Unique Carrier Code.	ok	ok	ok
Tail_Number	Tail Number	have blank values	have blank values for	have blank values
Flight_Number_Operating_Airline	Flight Number	ok	ok	ok
OriginAirportID	Origin Airport, Airport ID. An identification number assigned by US DOT to identify a unique airport. Use this field for airport analysis across a range of years because an airport can change its airport code and airport codes can be reused.	ok	ok	ok
Origin	Origin Airport	ok	ok	ok

Headers	Details	Feb	May	December	
DepTime	Actual Departure Time (local time: hhmm)	null for cancelled	null for	null for	
DepDelay	Difference in minutes between scheduled and actual departure time. Early departures show negative numbers.	null for cancelled flights	null for cancelled flights	null for cancelled flights	Note: There are few cancellations that have departure time
DepDelayMinutes	Difference in minutes between scheduled and actual departure time. Early departures set to 0.	null for cancelled flights	null for cancelled flights	null for cancelled flights	
DepDel15	Departure Delay Indicator, 15 Minutes or More (1=Yes)	null for cancelled flights	null for cancelled flights	null for cancelled	
DepartureDelayGroups	Departure Delay intervals, every (15 minutes from <-15 to >180)	null for cancelled flights	null for cancelled flights	null for cancelled	
DepTimeBlk	CRS Departure Time Block, Hourly Intervals	Note sure why needed but have	Removed	Removed	
TaxiOut	Taxi Out Time, in Minutes	null for cancelled flights and no data	null for cancelled flights	null for cancelled	
WheelsOff	Wheels Off Time (local time: hhmm)	null for cancelled flights and no data	null for cancelled flights	null for cancelled	

Headers	Details	Feb	May	December
ArrDelayMinutes	Difference in minutes between scheduled and actual arrival time. Early arrivals set to 0.	null for cancelled and diverted flights	null for cancelled and diverted flights	null for cancelled and diverted flights
ArrDel15	Arrival Delay Indicator, 15 Minutes or More (1=Yes)	null for cancelled and diverted	null for cancelled and diverted flights	null for cancelled and diverted flights
ArrTimeBlk	CRS Arrival Time Block, Hourly Intervals	ok	ok	ok
Cancelled	Cancelled Flight Indicator (1=Yes)	ok	ok	ok
Diverted	Diverted Flight Indicator (1=Yes)	ok	ok	ok
CRSElapsedTime	CRS Elapsed Time of Flight, in Minutes	ok	ok	ok
ActualElapsedTime	Elapsed Time of Flight, in Minutes	null for cancelled	null for cancelled	null for cancelled
AirTime	Flight Time, in Minutes	null for cancelled	null for cancelled	null for cancelled

Fig 1: Data Exploration

2. Cleaning of Airlines.csv:
   We joined the extract and Airlines.csv with a left join within Tableau Desktop. While creating various charts and filters for the DSS, we were picking the Airline Name directly from the Airlines.csv instead, which was giving us null values when we were trying to plot it against delays. Thus, we removed the Airlines for which their code did not exist in the raw data.

Invalid Data Types:

The columns “DepTime” and “ArrTime” are whole numbers and not in Time Format. This is a significant data issue for anyone wishing to use this field. Fortunately, we opted not to fix these columns because these values were not being used.

Fixing Missing Values:

Three of the airports in Airports.csv lacked latitudinal and longitudinal information. By obtaining the values from Google, we were able to fix them. Associated screenshot is attached (Figure 2a and 2b).

	A	B	C	D	E	F	G
1	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
98	ECP	Northwest Florida Bea	Panama City	FL	USA		
236	PBG	Plattsburgh Internatior	Plattsburgh	NY	USA		
315	UST	Northeast Florida Regi	St. Augustine	FL	USA		

Fig 2a: Missing Data

After applying the fix:

Initially we had kept the longitudinal values as positive and saw the respective airports on map landing in India and China on Tableau Desktop. Post which we realized our mistakes and then refixed our fix.

1	IATA_CODE	AIRPORT	CITY	STATE	COUNTRY	LATITUDE	LONGITUDE
98	ECP	Northwest Florida Bea	Panama City	FL	USA	30.3549	-85.7995
236	PBG	Plattsburgh Internatior	Plattsburgh	NY	USA	44.6521	-73.4679
315	UST	Northeast Florida Regi	St. Augustine	FL	USA	29.9544	-81.3429

Fig 2b: Fixing Missing Data

To remove the unnecessary columns from each of the 12 Excel Sheets, we used Power-Query after narrowing the selection of parameters down to about 35. We discovered that the generated CSV files weren't compatible with Tableau and were causing a number of problems, such as 14% of the **Flight_Date** column showed NULL values despite the raw data having proper information. After making multiple attempts to resolve the problem using tools like "Create Calculated Fields" and other solutions mentioned in Google and Tableau documentation, we made the decision to clean the raw data again, by using Tableau Prep only.

Tableau Prep:

We have attached a small snippet from Tableau Prep where we used Tableau Prep to include all 12 Excel Sheets of Raw Data, added a clean step and then finally used the output from it as our extract for Tableau Desktop.

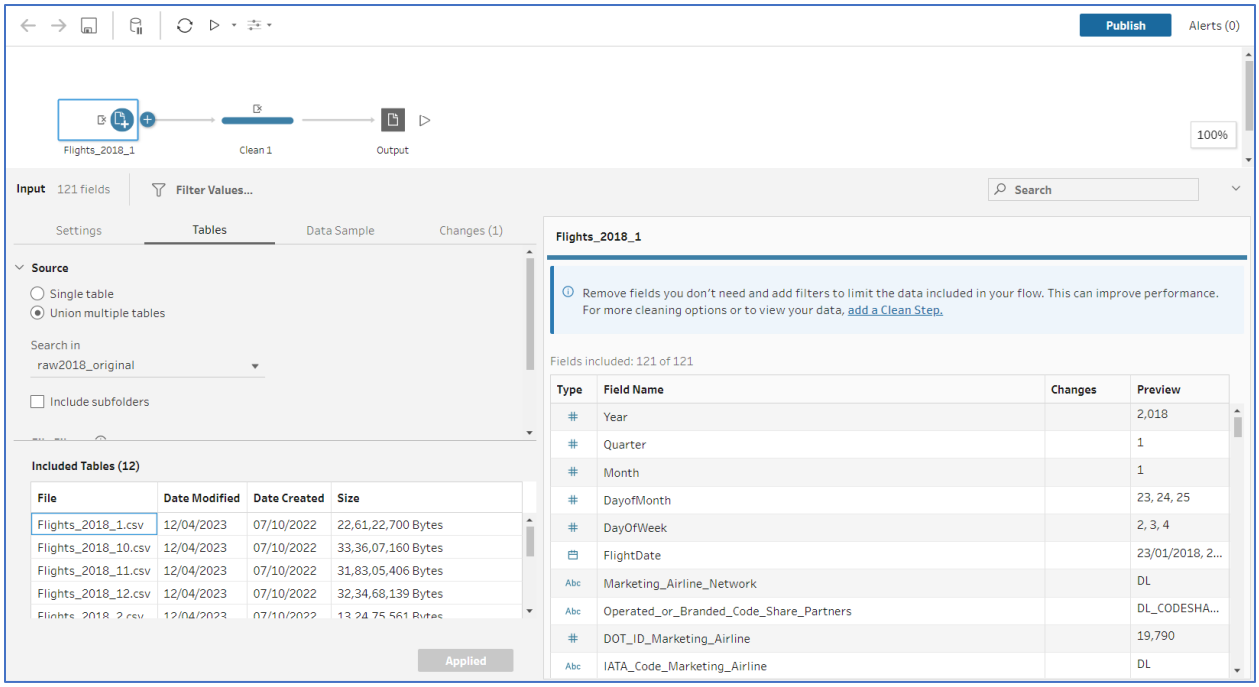

Fig 3: Data Combining and Cleaning

Next section, will give more information on our DSS.

Dashboard 1: Delay Group Analysis

Flight data provides information about the status of flights arriving at airports. In this case, the data includes the number of flights in different categories, namely on time, slightly delayed, medium delayed, highly delayed, cancelled, and diverted.

- On-time flights: Flights that arrive within 15 minutes of their scheduled time.
- Slightly delayed flights: Flights that arrive between 15 and 60 minutes after their scheduled time.
- Medium delayed flights: Flights are those that arrive between 60 and 120 minutes after their scheduled time.
- Highly delayed flights: Flights are those that arrive more than 120 minutes after their scheduled time.
- Cancelled flights are those that are cancelled before departure, and
- Diverted flights are those that are redirected to another airport due to unforeseen circumstances.

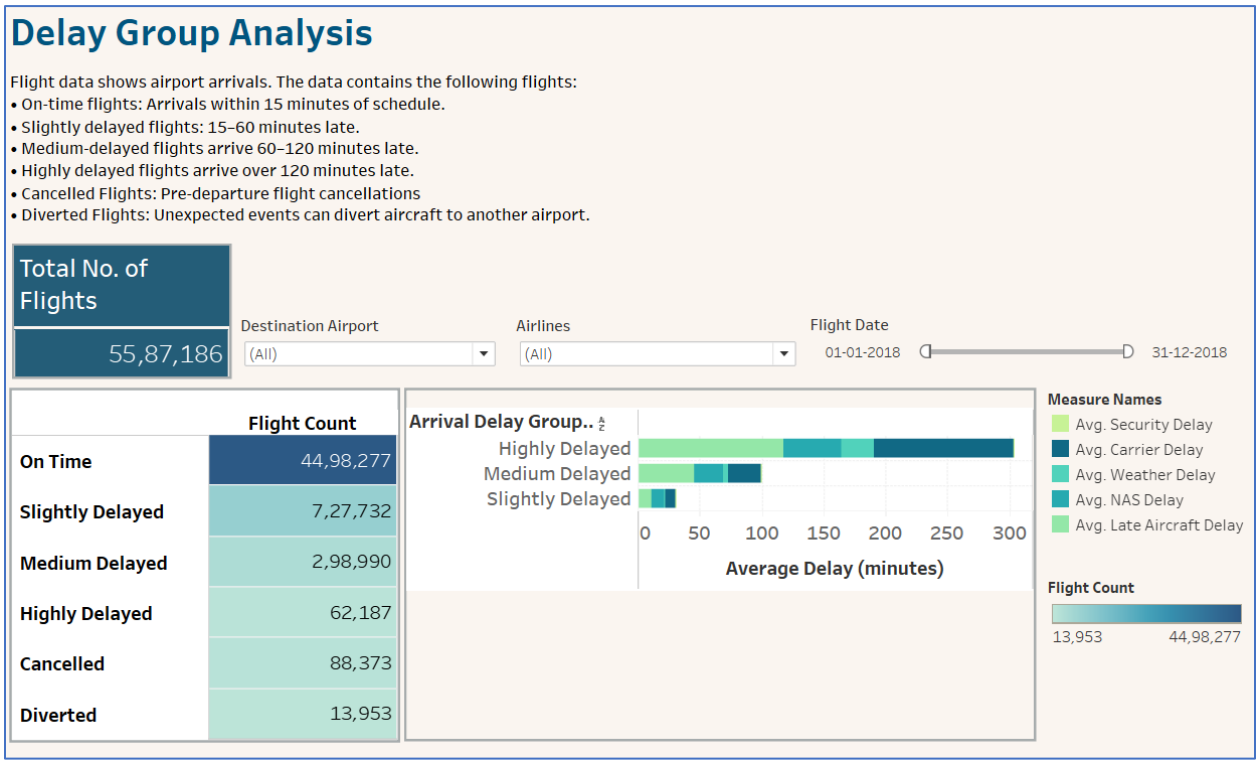

Fig 4: Dashboard 1 – Delay Group Analysis

Analyzing the delays with respect to airports and airlines can provide valuable insights into the performance of different airports and airlines. For example, an airport or airline that consistently has a high number of highly delayed or cancelled flights may indicate issues with their operational efficiency or reliability. By identifying these patterns, airport and airline management can take corrective actions to improve their performance and enhance customer satisfaction.

The 5 Causes of delays have been discussed in detail in Dashboard 5: Causes of Delay.

Dashboard 2: Arrival Delay Analysis

Delay analysis with respect to arrival refers to the amount of time that flights are delayed when arriving at their destination airport. In order to analyze this delay, we can calculate the average delay time for all flights arriving at a particular airport. Through our dashboard 2:

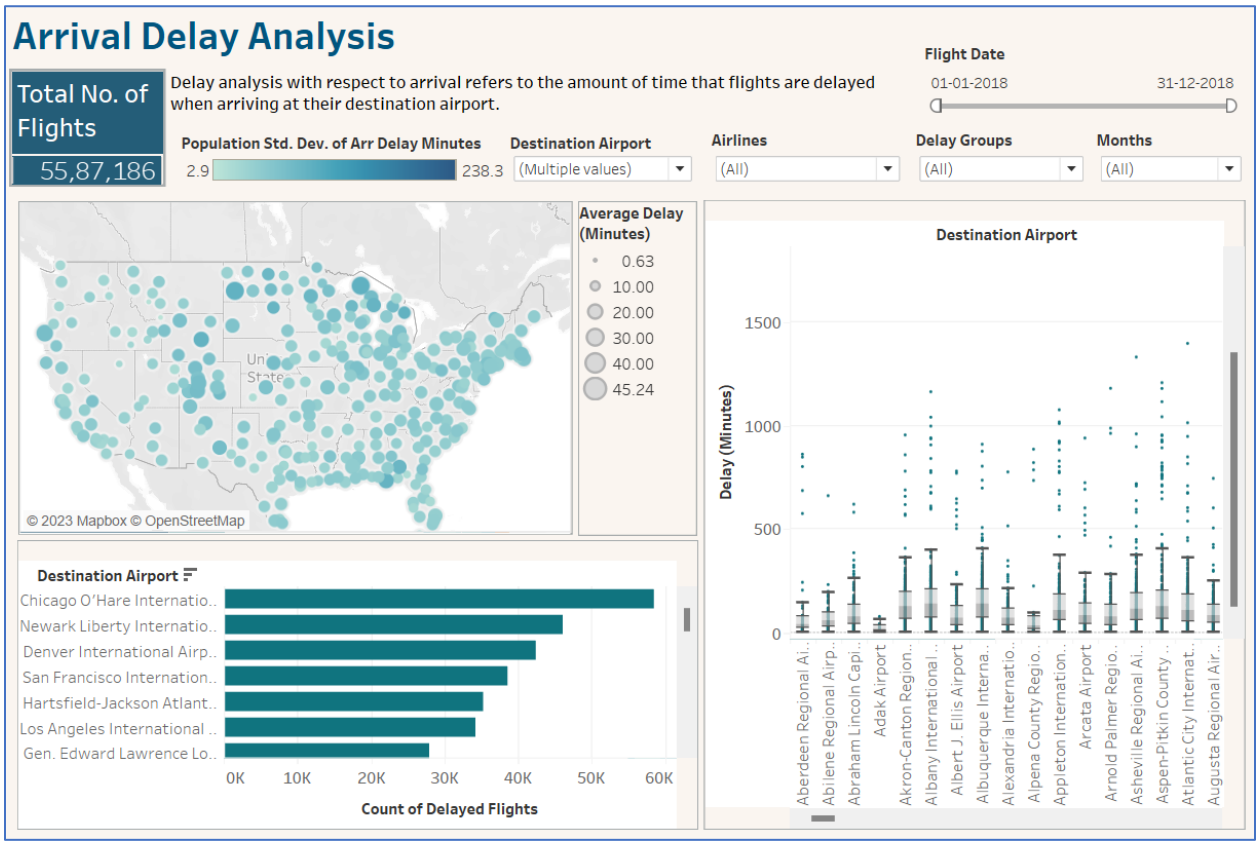

Fig 5: Dashboard 2 – Arrival Delay Analysis

Users can use one or more filters (such as Airports, Airlines, Flight Delay Groups, Months, and Flight Date) to view the following parameters:

1. Map representation of Population Std. Deviation of Arrival delay in minutes:

   This can help airport authorities and airlines identify which airports require additional resources to improve on-time performance.

2. With the size of the tip depicting Average Delay in Minutes:

   This can help identify airports with longer average delay times that require additional attention from airport authorities.

3. Horizontal Bar Graph depicting Count of Delayed Flights for all Airports in the decreasing order of Delayed Flight Count:

   This can help airport authorities and airlines prioritize improvements for airports with the highest number of delayed flights.

4. Box and Whisker Plot of Arrival Delay in Minutes w.r.t Airports:

   This can help airport authorities and airlines identify which airports have the highest variability in arrival delay times and may require additional resources to improve on-time performance.

Overall, by providing filters on the dashboard, users can interactively explore the data and tailor the graphs to their specific needs, making the dashboard an effective Decision Support System (DSS) by:

- Identifying airports with the highest arrival delay times
- Comparing arrival delay times across airports

This can help airport authorities and airlines identify areas for improvement and make data-driven decisions to improve on-time performance and passenger satisfaction.

Dashboard 3: Departure Delay Analysis

Departure delay analysis is a type of analysis that focuses on the delay time for flights departing from an airport. The analysis aims to identify the reasons behind the departure delays and provide insights that can help airlines and airport authorities improve their operations and reduce the delay times.

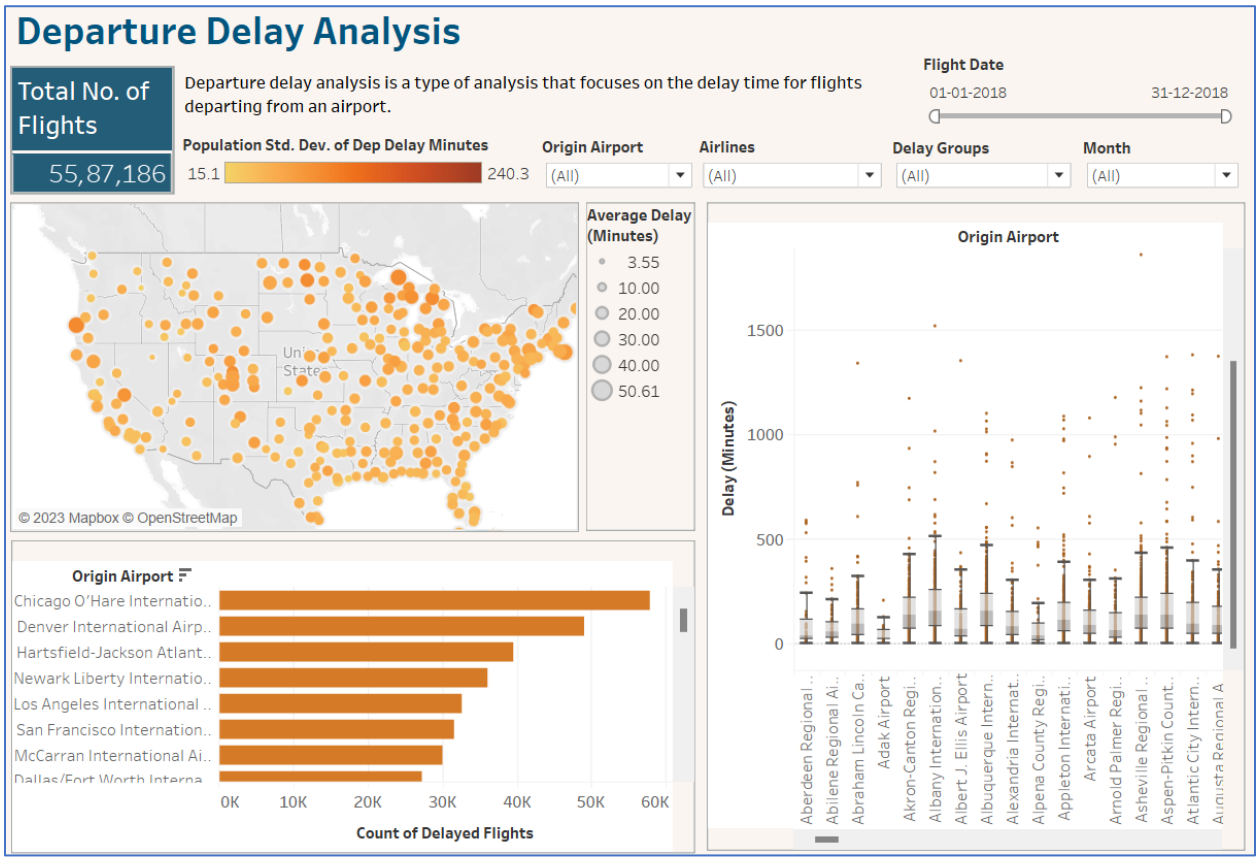

Fig 6: Dashboard 3 – Departure Delay Analysis

Similar to previous Dashboard, users can use one or more filters (such as Airports, Airlines, Flight Delay Groups, Months, and Flight Date) to view the following parameters:

1. Map representation of Population Std. Deviation of Departure delay in minutes:

The map representation of the population standard deviation of departure delay in minutes shows how the departure delays vary across different airports in the United States. The darker shades of blue indicate a higher standard deviation, which means that the departure delays are more dispersed and unpredictable. The lighter shades of blue indicate a lower standard deviation, which means that the departure delays are more consistent and predictable.

2. With the size of the tip depicting Average Delay in Minutes:

The size of the tip represents the magnitude of the average delay, with larger tips indicating longer delays. The graph allows for easy comparison between different airports and can help identify airports with the highest average delays.

3. Horizontal Bar Graph depicting Count of Delayed Flights for all Origin Airports in the decreasing order of Delayed Flight Count:

This can help identify the airports with the highest number of delayed flights. This information can be used to prioritize efforts to reduce delays at these airports.

4. Box and Whisker Plot of Departure Delay in Minutes w.r.t Origin Airports:

The box and whisker plot shows the distribution of departure delay times for each origin airport in the United States. The box represents the middle 50% of the data, with the median indicated by a line in the middle. The whiskers represent the range of data, with outliers shown as individual points. The plot allows for easy comparison of the delayed distributions across different airports and can help identify airports with high variability in departure delay times.

Overall, the departure delay analysis provides valuable insights into the delay times and distributions at different airports in the United States. The analysis can help airlines and airport authorities identify the factors contributing to delays and develop strategies to reduce them. By adding filters such as airports, airlines, months, and flight dates on a Tableau dashboard, we have made the previously mentioned graphs a user interactive decision support system (DSS) in the following ways:

- Identification of the airports with the highest departure delays.
- The user can compare the departure delays of different airports and identify the airports with the highest delays.
- The user can compare the delay times across different months and identify months with the highest delays.
- The user can compare the performance of different flights using a date filter and identify the flights with the highest delays.
- Comparison of the performance of different airlines: By filtering the data by airlines, one can compare the departure delays of different airlines. This information can help travelers to choose airlines that have a better track record of on-time departures.

This information can provide valuable insights into the performance of departing airports, airlines, and seasons, which can help airlines and airport operators to reduce delays, improve customer satisfaction, and increase operational efficiency.

Dashboard 4: Airline Delay Analysis

Airline delay analysis provides insights into the performance of different airlines in terms of departure delays. By incorporating filters like destination airports, airlines, delay groups, months, and flight dates on the Tableau dashboard, we have made a user-interactive Decision Support System (DSS) that enables users to slice and dice the data to gain deeper insights into the performance of airlines in terms of departure delays:

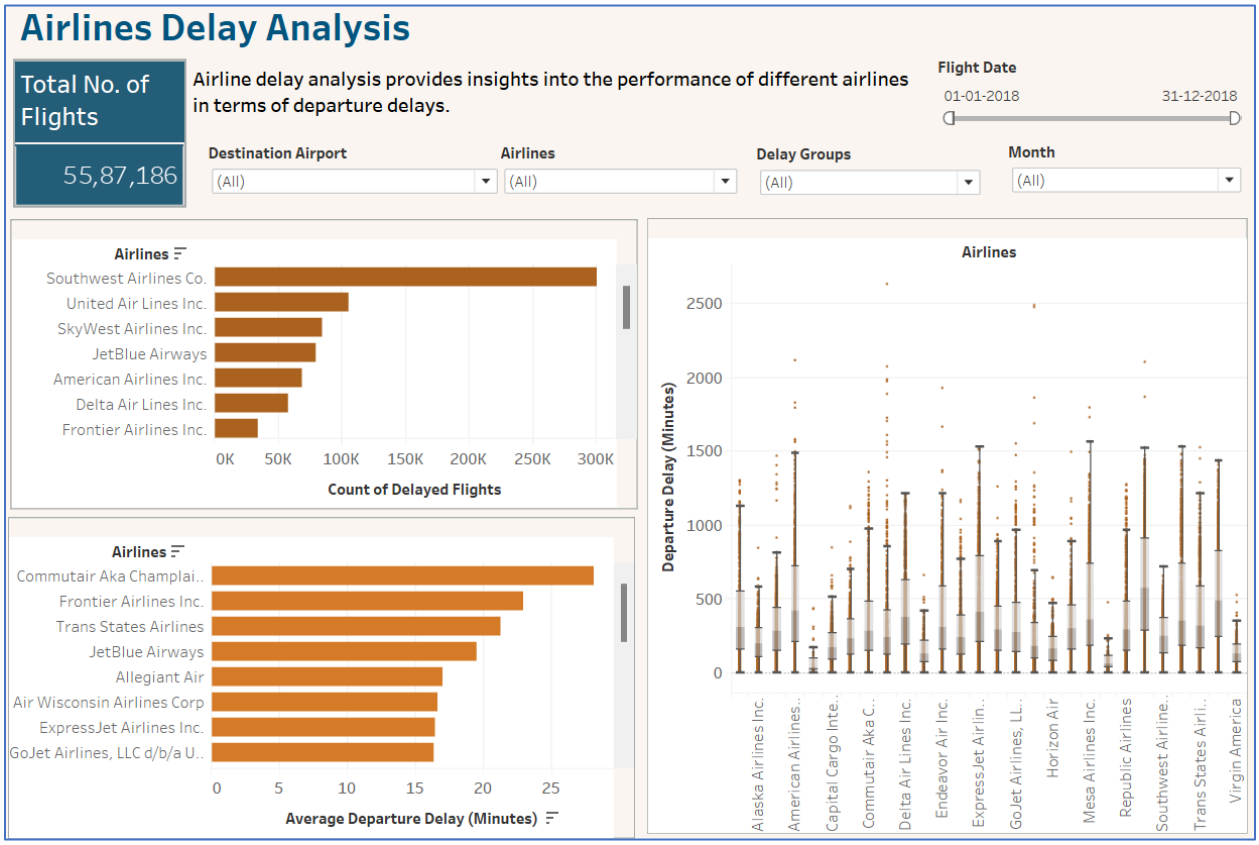

Fig 7: Dashboard 4 – Airlines Delay Analysis

1. Delay groups filter: Allowing users to filter data on the basis of slightly delayed, medium delayed and highly delayed flights.
2. Bar Graphs: Its crucial to identify how for a flight having the highest count of delayed flights still has a lower average delayed flights.
3. Understanding the Causes of Delay: If the median delay is high, it indicates that the majority of flights of that airline are delayed. In contrast, if the median delay is low, it indicates that the majority of flights of that airline are on time.
4. Improvement Opportunities: The Airlines vs. Departure Delay analysis can help airlines identify areas for improvement in their operations, such as increasing the number of staff, improving maintenance, or revising their flight schedules.
5. What is the departure delay for each airline with respect to destination (average, median, stdev etc).

By analyzing the data and identifying the root cause of delay, airlines can take corrective actions and improve their on-time performance. This can lead to better customer satisfaction, increased revenue, and improved operational efficiency.

Dashboard 5: Causes of Delay

We chose raw data specially for this particular dashboard. As mentioned earlier, we have 5 causes of delay and have been used in this dashboard. We are only considering a flight as a delayed one if it arrives 15 minutes after its scheduled arrival time. We have the following 2 charts:

1. A horizontal stacked bar chart is used to map the causes of flight delays with airlines. The x-axis represents the airlines, and the y-axis represents the count of delayed flights. The bars are stacked with different colors representing different causes of delay, such as Carrier, Weather, NAS, Security, and Late Aircraft. By analyzing the chart, we can identify the airlines that have the maximum number of delayed flights due to each cause.
2. A horizontal bar chart is used to represent the airlines with the maximum arrival delays. The x-axis represents the average arrival delay in minutes, and the y-axis represents the 5 causes of delays for all the airlines selected in the filter. By analyzing the chart, we can identify the airlines that have the maximum arrival delays.

The analysis from the above charts can help airlines and airport stakeholders to identify the causes of flight delays and take appropriate measures to minimize them as:

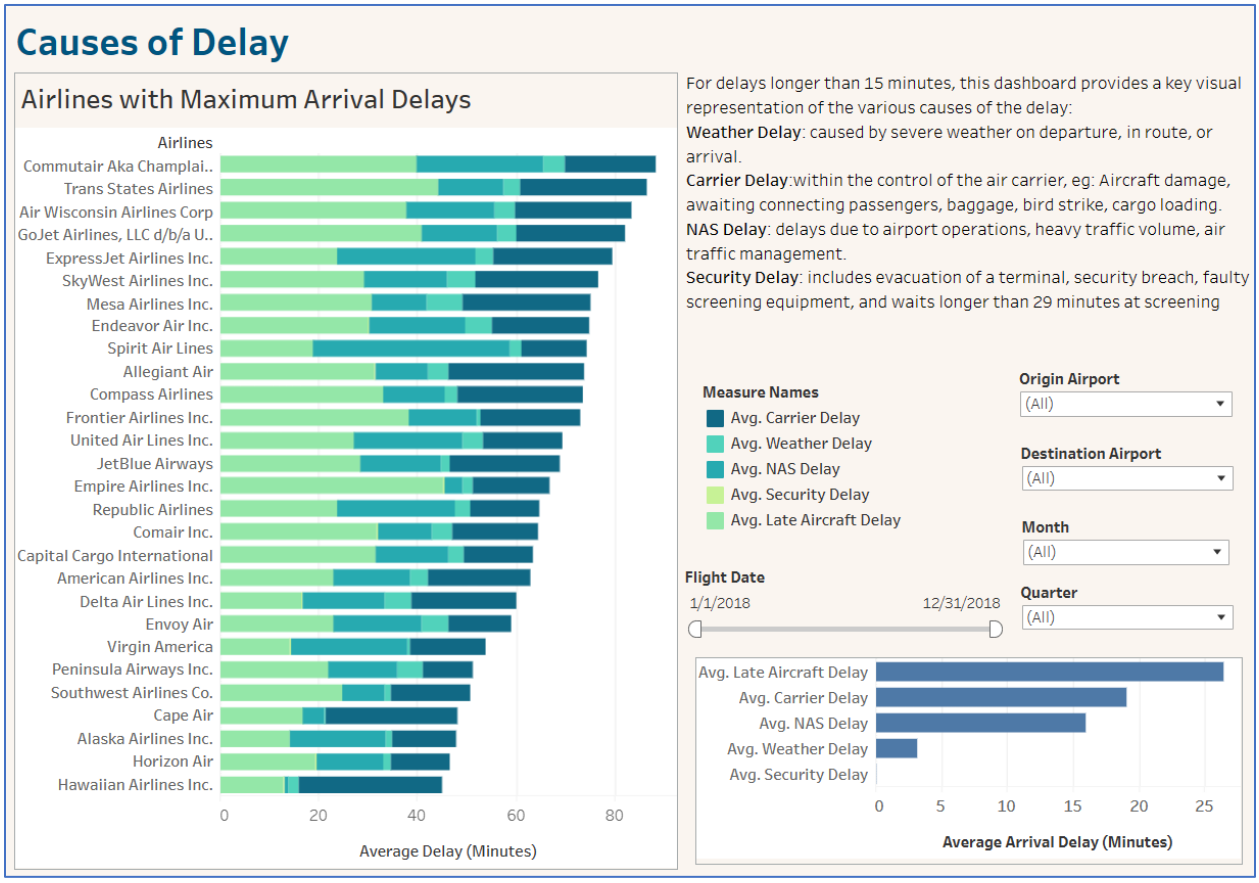

Fig 8: Dashboard 5 – Causes of Delay

Dashboard 6: Seasonal Analysis

Seasonal delay analysis can provide insights into how flight delays vary over different time periods such as months, quarters, weeks, and days. We have 3 charts in this part of BDSS which can be used by the stakeholders as:

1. The first graph shows the delayed flight count for arrival delays, with data aggregated by quarter, month, week, and day. This can help identify patterns in delays over time and can inform strategies to mitigate delays during peak seasons.
2. The second graph is a heatmap that shows the average departure delay for each airline by month. The colors in the heatmap indicate the severity of the delay, with darker colors indicating longer delays. This can provide can provide insights into which airlines are experiencing the most significant delays during specific months, and this information can help airlines to identify and address any underlying issues that may be causing these delays.
3. The third graph is a heatmap that shows the average arrival delay for each airline by month. Similar to the previous graph, the colors in the heatmap indicate the severity of the delay. This can help identify which airlines tend to have more delays during certain months, allowing for more targeted solutions to be implemented.

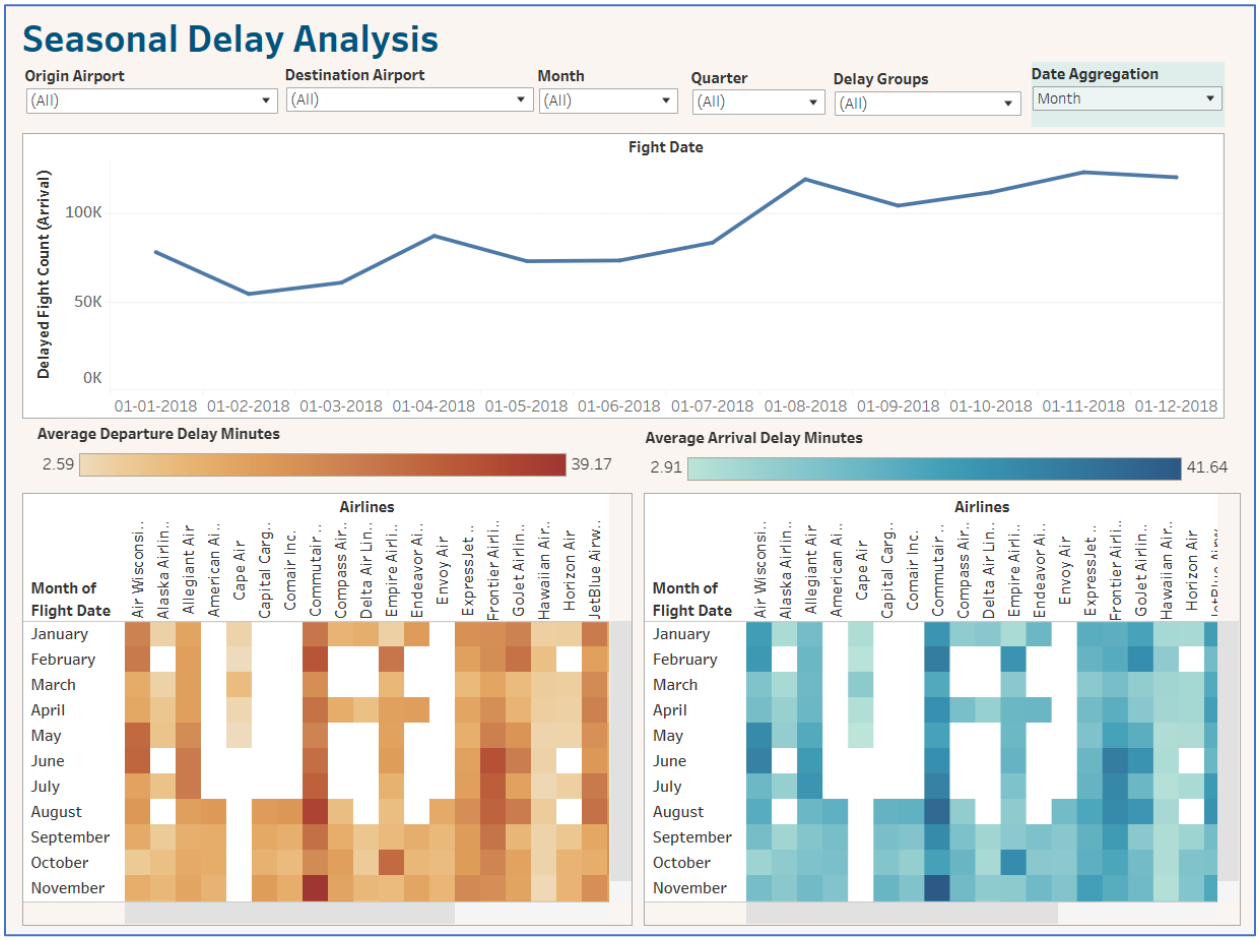

Fig 9: Dashboard 6 – Seasonal Delay Analysis

Overall, the seasonal delay analysis can help airlines to understand the pattern of delays over time and make informed decisions to improve their operations, reduce delays, and enhance customer satisfaction. The analysis can help airlines plan for peak travel seasons and adjust their schedules accordingly to minimize delays. It can also help airlines identify specific routes or airports that are prone to delays during certain times of the year and take proactive measures to reduce delays such as increasing staffing or adjusting flight schedules. By understanding the seasonal patterns of delay, airports can schedule maintenance and repairs during periods of lower traffic. This can minimize disruptions to passengers and reduce the risk of delays.

CONCLUSION:

In our effort to create a Decision Support System (DSS) using the Tableau Dashboard on Delay Analysis of US Flights with respect to Airlines and Airports provides a comprehensive view of the performance of different airlines and airports in terms of arrival and departure delays. The interactive interface with filters allows users to obtain specific information about their chosen airlines, airports, and time frames.

The various graphs and visualizations provide valuable insights into the delays experienced by airlines and airports, the causes of the delays, and the performance of airlines in terms of delays. The delay counts, delay time, and the frequency of delay times are presented in various formats such as bar graphs, heat maps, box and whisker plots, and line graphs, which enable users to compare and contrast the data easily.

These insights can be used by airlines to identify areas of improvement in their operations and to implement strategies to reduce delays. The stakeholders can also use this information to evaluate the performance of airlines and airports and make informed decisions. Overall, this Dashboard can serve as a valuable decision support system for the aviation industry, aiding in the optimization of operations and ultimately enhancing customer satisfaction.

REFERENCES:

kaggle.com. (n.d.). Flight Delay EDA (Exploratory Data Analysis). [online] Available at: https://www.kaggle.com/code/adveros/flight-delay-eda-exploratory-data-analysis/notebook#2.-Exploratory-Data-analysis [Accessed 8 Apr. 2023].