**US INTERNATIONAL AIR TRAFFIC FOR FLIGHT DEPARTURE FLIGHTS**

Bharanidharan M P

AJ23SYD018

Student: MAIB23

SP Jain School of Global Management, Sydney

Australia.

**Contents:**

US INTERNATIONAL AIR TRAFFIC FOR FLIGHT DEPARTURE……………………………………………………………….1

[1. Problem Statement: 3](#_Toc133296381)

[2. Dataset: 3](#_Toc133296382)

[3. Data Cleaning and Feature Engineering: 4](#_Toc133296383)

[4. Dashboard: 4](#_Toc133296384)

[5. Conclusion: 8](#_Toc133296385)

[6. Reference: 8](#_Toc133296386)

# Problem Statement:

We have been told to find out a data set and perform data analysis using a tool (Dashboard) or develop Python-based software (regression-based predictor). I have decided to use data about US International Air Traffic for Flight Departures. " US International Air Traffic for Flight Departures " refers to the study and analysis of departure of flights in the United States of America, mentioning the US and Foreign airports too. This includes data on the number of departures, the year range, and different flights, the types of flights involved, and the departures in US and Foreign airports. I have used a dashboard to analyze patterns and trends in the frequency departures over time. This type of analysis can provide insights into the effectiveness of traffic control and efficient use of existing airlines and help identify areas where more airlines are/may be needed.

# Dataset:

Dataset Link- <https://www.kaggle.com/datasets/parulpandey/us-international-air-traffic-data>

This dataset consists of **16 features and 930808 rows.** Also there are **11 numerical and 5 categorical features.** The features in the dataset include:

Data\_Dte, Year, Month, Usg\_Apt\_Id, Usg\_Apt, Usg\_Wac, Fg\_Apt\_Id, Fg\_Apt, Fg\_Wac, AirlineId, Carrier, CarrierGroup, Type, Scheduled, Charter, Total.

* + DATA\_DTE: Represents the date for which the data pertains to.
  + YEAR: Represents the year portion of the date.
  + MONTH: Represents the month portion of the date.
  + USG\_APT\_ID: Refers to the unique identifier for the usage airport.
  + USG\_APT: Represents the name or code of the usage airport.
  + USG\_WAC: Refers to the World Area Code of the usage airport.
  + FG\_APT\_ID: Refers to the unique identifier for the forecast airport.
  + FG\_APT: Represents the name or code of the forecast airport.
  + FG\_WAC: Refers to the World Area Code of the forecast airport.
  + AIRLINEID: Represents the unique identifier for the airline.
  + CARRIER: Represents the name or code of the airline.
  + CARRIERGROUP: Represents the group or category to which the airline belongs.
  + TYPE: Represents the type of flight, whether it is scheduled or charter.
  + SCHEDULED: Represents the number of scheduled flights.
  + CHARTER: Represents the number of charter flights.
  + TOTAL: Represents the total number of flights (both scheduled and charter) for the given data date.

Reason for selecting this dataset:

Insights into the patterns and trends of US International Air Traffic for Flight Departures, as well as aspects like usage and forecast airport, carrier details, and flight type, can be gained through the examination of this dataset. The analysis can aid in the creation of plans for better implementation of necessary and suitable flights and enhancing air traffic safety.

# Data Cleaning and Feature Engineering:

After performing initial data exploration and analyzing the features in-depth, the column’s features were cleaned.

1. I checked for the Null Values in the entire dataset column wise.
2. The column having null values was identified and was cross checked for the appropriate values to replace the null. It was found that airline id of 20414 and 20415 were having null values and OW and XG are the suitable values for replacing.
3. Decoded the binary values into their respective values to have a meaningful data explanation i.e, under the feature carriergroup, 1 and 0 were available which means US Domestic Air Carriers and Foreign Air Carriers respectively.
4. New column was added namely Type of Flight which categorised the departures as Scheduled or Chartered.
5. Columns removed:

* Type – This feature had the value of Departure. Since the whole dataset is on departure, the feature was dropped.
* Usg\_Apt\_Id – The name of the feature was useful because of which this feature was dropped.
* Fg\_Apt\_Id – The name of the feature was useful because of which this feature was dropped.

# Dashboard:

A data dashboard is a commonly used tool by businesses to monitor, analyze, and present data with the aim of gaining insights into the overall performance of an organization, department, or process. The power of a dashboard lies in its ability to provide real-time information and context, aiding in informed decision-making and empowering employees. For example, an IT team can utilize a dashboard to detect signs of a security breach, or a business may embed a dashboard into a mobile app for field workers to access data on the go.

Dashboards can be customized to display different types of data based on user requirements and can be interactive, allowing users to explore data in-depth or perform specific tasks. They are widely used in various industries, including business, finance, marketing, and others, to help managers and decision-makers track organizational performance, identify patterns, and make data-driven choices.

Overall, the main objective of a dashboard is to provide users with a quick and user-friendly way to access and analyze complex data, enabling them to make informed decisions and improve decision-making processes.

* 1. **How to build and deploy a dashboard in Python**

There are various libraries available for users to build a dashboard. The libraries are- Dash, Bokeh, Streamlit, Plotly, etc.

In this project work, I have used the Plotly dash for the purpose of Dashboard creation.

The steps to build a dashboard in Python are:

1. Import Data and Libraries
2. Define Layout
3. Create Visual Components
4. Define The Call Back Function
5. Create Visualization
6. Run App
7. Deploy

The infographics used in this dashboard are slider, line graph, pie chart, histogram and bar chart.



Figure 1

The above figure show cases the slider function from the min and max year as per the available dataset. Here, the year ranges from 1990 to 2020. The slider can be adjusted and the depiction of other infographics changes accordingly.

* 1. **Business Insights:**

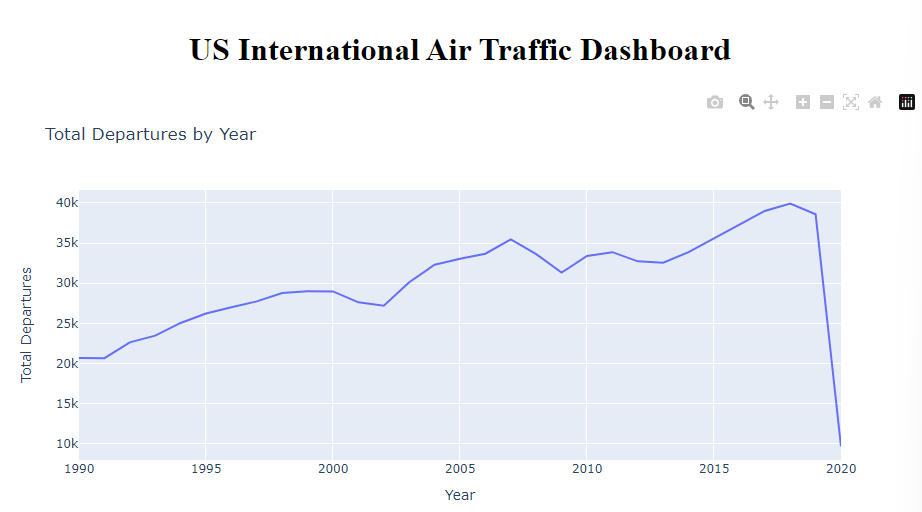
****

Figure 2

The above figure depicts the line graph depicting the total departures across the years. From the above graph, we can get the variation in departure across the range of years. As we can see a sudden decrease in departures considerably due to covid 19 situation. Apart from this, there seems to be more or less a steady flow in departures. This information can be used to follow up a regression on the future study of departures.

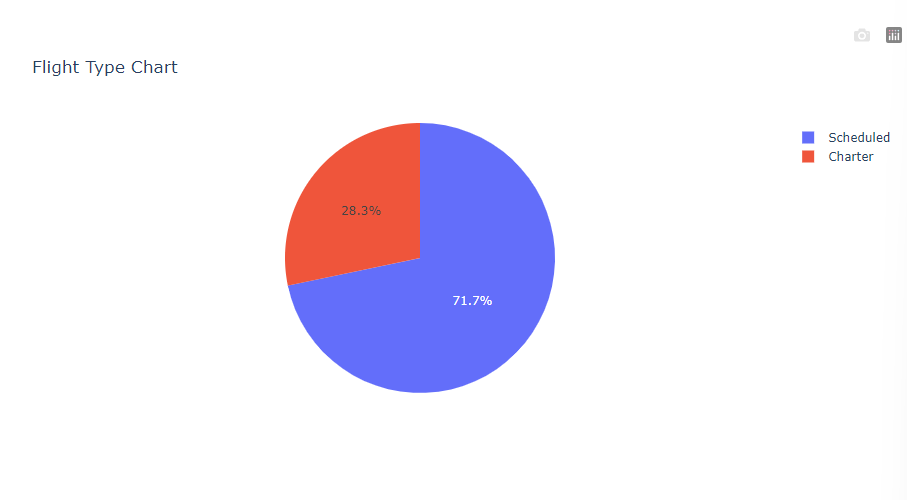


Figure 3

The above figure depicts the pie chart for the Flight Type feature. There are two types of flights according to the data namely Scheduled and Chartered. Scheduled flights are regular flights and the chartered flights are intermittently scheduled. From the chart it can be interpreted that majority of the departures are of Scheduled flights. The company can focus more on the Scheduled flights and the shifts of pilots can be concentrated on rotation basis to get more output from them.

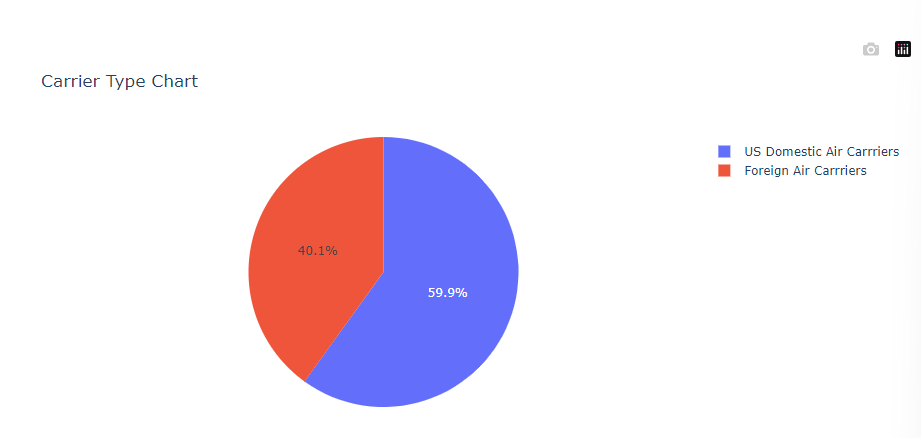


Figure 4

The above figure depicts the pie chart for the Carrier Type feature. There are two types of carriers according to the data namely US Domestic Air Carriers and Foreign Air Carriers. US Domestic Air Carrier flights are regular flights within US domestic range and the Foreign Air Carrier flights are internationally scheduled/flown. From the chart it can be interpreted that majority of the departures are of US Domestic Air Carrier flights. The company should focus almost equally between domestic and international clients.

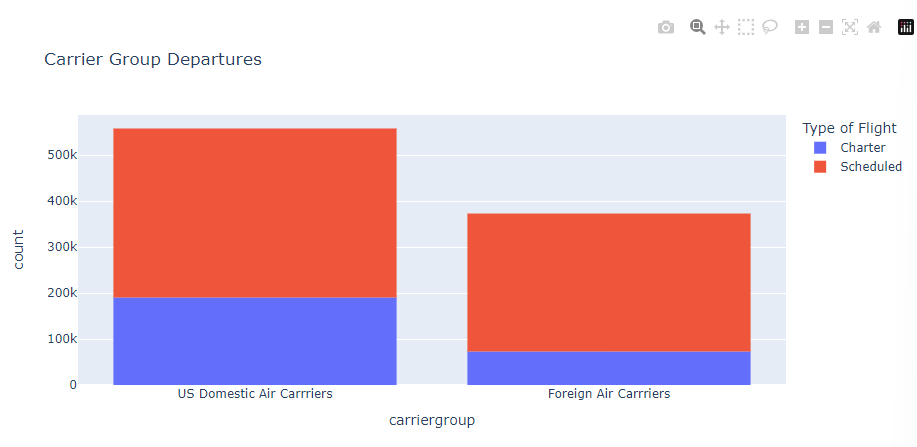


Figure 5

The above figure depicts the histogram for the Carrier Type feature. There are two types of carriers according to the data namely US Domestic Air Carriers and Foreign Air Carriers. US Domestic Air Carrier flights are regular flights within US domestic range and the Foreign Air Carrier flights are internationally scheduled/flown. From the figure it can be interpreted that majority of the departures are of US Domestic Air Carrier flights. The reason might be because of more departures of scheduled flights. The company should focus almost equally between domestic and international clients.

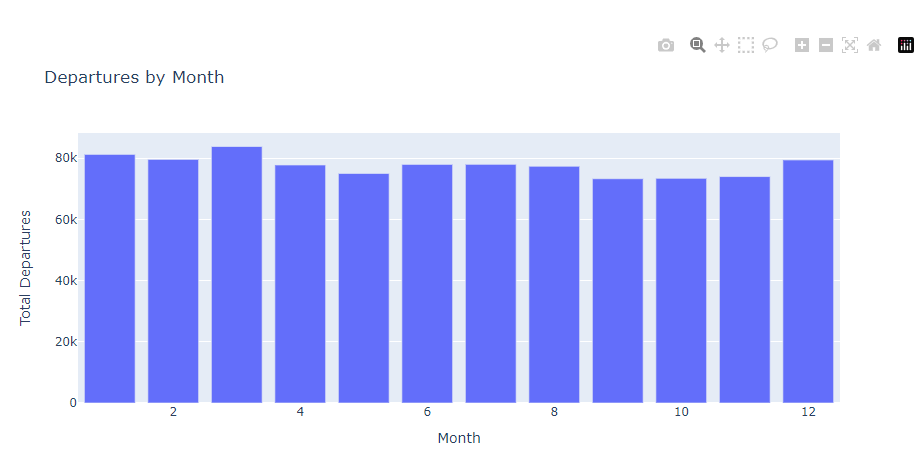


Figure 6

The above figure depicts the total flight departures based on the month and year range of the data. We can see that on an average the top three departures are made during March, January and February respectively. The reason can be for studies or any other specifics. The company can increase the flight charges to see through a profit or provide offers to their specific or frequent customers.

The dashboard can be accessed from the link: <http://127.0.0.1:8050/>

If any issue faced, I kindly request to run the file from the supported jupyter notebook.

**Note:** The slider option can be used to shuffle through the period available (1990-2020) to check with the variations in data. This helps in avoiding time consuming repetitive codes and labour.

# Conclusion:

In conclusion, dashboards are effective tools that provide businesses with a quick and straightforward way to visualize and analyze their data. By presenting key performance indicators and metrics in a clear and concise manner, dashboards help companies make data-driven decisions and identify areas for improvement. The ability to customize and update dashboards in real-time allows businesses to remain agile and adapt to changes in their operations or market conditions. Overall, dashboards can provide a competitive advantage by empowering businesses to leverage their data and make informed choices.

# Reference:

* <https://www.kaggle.com/datasets/parulpandey/us-international-air-traffic-data?select=International_Report_Passengers.csv>
* <https://www.kaggle.com/datasets/jboysen/us-perm-visas>
* <https://www.kaggle.com/code/pavlofesenko/interactive-titanic-dashboard-using-bokeh>
* <https://www.kaggle.com/code/thebrownviking20/intermediate-visualization-tutorial-using-plotly>
* <https://www.kaggle.com/code/tavoosi/suicide-data-full-interactive-dashboard/report>
* <https://www.justintodata.com/python-interactive-dashboard-with-plotly-dash-tutorial/>
* <https://chat.openai.com/>
* <https://stackoverflow.com/search?q=Dashboard+using+plotly>