Name: Pranav Vinod

Course: POM 681 – Business Analytics and Data Mining

Project 1 – Airline Delay Analysis Dashboard in R

Supervisor: Professor Bharatendra Rai

This is a project report which explains each page on the dashboard and the thinking behind them.

**Page 1: Interactive Data Visualization**

This page is the first page of the dashboard. It has a short summary section on the top and 4 interactive visualizations on the bottom 2 rows.

The 2 row contains a bar chart containing the states and the number of departure delays in minutes in each of them during the month of observation. We can see that California and Texas are the states with the greatest number of delays and that seems likely because they are two of the biggest states with busy cities and airports.

Then we can see a pie chart with the states with most arrival delays in minutes. Again, we can see that California and Texas are the top 2 states.

In the next row, there are 2 plots. The first is a scatter plot between Arrival Delay and Departure Delays in minutes. We can see that there is a linear relation between them.

Finally, there is a box plot of the top 5 states with the most Arrival Delay to check if there are any outliers skewing the data. We find that there are no apparent outliers.

**Page 2: Map**

Page 2 contains a geospatial map which colors each state according to the average departure delay of the flights originating in that state. We find that Illinois, Arkansas, and Vermont are states with the most delay on average. Also, to note is that despite being the busiest states, Texas and California have low average departure delay for flights originating there.

**Page 3: Data Table**

On this page is a data table which gives everyone a glimpse into the dataset which has been used to create this dashboard. At a time 25 columns are visible out of 49101.

**Page 4: Carrier Delay Comparison**

Page 4 has a bar chart which compares all the unique carriers and the departure delay on their flights. It has been done to get an idea about the performance of airlines and compare that metric on a single glance. We can see that carrier MQ has the highest average departure delay on its flights.

**Page 5: Delay Cause Analysis**

This page deals with the various causes of airline delays on a daily basis. It contains a line chart which graphs the total delay in minutes due to each factor. We can notice that on most days Late Aircraft Delay is the major reason for airline delays and Security delay being the reason which causes the least delay.

**Page 6: Summary of Data**

On this page, we find a summary of the dataset.

It contains information about the average departure delays during the month of observation.

Also in a value box is the average difference in the CRS Estimated flying time and the actual flying time. We find that on average, a flight had an airtime of 25.59 minutes less than its estimated flying time.

It contains information about the average arrival delays during the month of observation.

**Page 7: About Report**

Page 7 is the final page of the dashboard and contains the name of the creator of the dashboard and the confidentiality status of the dashboard.

Also, on the top right corner of the dashboard, the source code has been embedded and the option to share it on various social media websites has been enabled.

**NOTE:**

1. Because of the large size of the dataset, the HTML page is slow to load.
2. The dashboard is best viewed by maximizing the browser screen size, so all the elements are properly visible.