POM 681- BUSSINESS ANALYTICS

AND DATA MINING

**Project Title:** Interactive Dashboard for Airline Data



**Presented By: KIRTI BENDIGERI**

**Student ID: 01967822**

**Submitted Date: 02/28/2022**

**ABSTRACT:**

Flight delay is a serious and widespread problem in the United States. Increasing flight delays place a significant strain on the US air travel system and cost airlines, passengers, and society at many billions of dollars each year. This report summarizes the findings from airline.csv file focusing on data from October,2014. The data contained in the file (airline.csv) has been extracted from the On-Time Performance data table of the "On- Time" database from the Tran Stats data library. I have tried to analyze the data and build a dashboard which depicts some important highlights regarding the trends observed in delays of flights, number of flights for that month, comparison between then different delays etc. Using R language we are able to extract the key findings from the dataset and represent it in form of different interactive charts and graphs which helps in better understanding of the data file.

# This dataset is composed by the following variables:

1. **Year** 2014
2. **Month** October
3. **DayofMonth** 1-31
4. **DayOfWeek** 1 (Monday) - 7 (Sunday)
5. **DepTime** actual departure time (local, hhmm)
6. **CRSDepTime** scheduled departure time (local, hhmm)
7. **ArrTime** actual arrival time (local, hhmm)
8. **CRSArrTime** scheduled arrival time (local, hhmm)
9. **UniqueCarrier** unique carrier code
10. **FlightNum** flight number
11. **TailNum** plane tail number: aircraft registration, unique aircraft identifier
12. **ActualElapsedTime** in minutes
13. **CRSElapsedTime** in minutes
14. **AirTime** in minutes
15. **ArrDelay** arrival delay, in minutes: **A flight is counted as "on time" if it operated less than 15 minutes later the scheduled time shown in the carriers' Computerized Reservations Systems (CRS).**
16. **DepDelay** departure delay, in minutes
17. **Origin** origin IATA airport code
18. **Dest** destination IATA airport code
19. **Distance** in miles
20. **TaxiIn** taxi in time, in minutes
21. **TaxiOut** taxi out time in minutes
22. **CancellationCode** reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
23. **Diverted** 1 = yes, 0 = no
24. **CarrierDelay** in minutes: Carrier delay is within the control of the air carrier. Examples of occurrences that may determine carrier delay are: aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, baggage, bird strike, cargo loading, catering, computer, outage-carrier equipment, crew legality etc.
25. **WeatherDelay** in minutes: Weather delay is caused by extreme or hazardous weather conditions that are forecasted or manifest themselves on point of departure, enroute, or on point of arrival.
26. **NASDelay** in minutes: Delay that is within the control of the National Airspace System (NAS) may include: non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control, etc.
27. **SecurityDelay** in minutes: Security delay is caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.
28. **LateAircraftDelay** in minutes: Arrival delay at an airport due to the late arrival of the same aircraft at a previous airport. The ripple effect of an earlier delay at downstream airports is referred to as delay propagation

### OBSERVATIONS BASED ON DASHBOARD:

#### We have created an interactive dashboard with four graphs and two charts along with data table, pivot table, value boxes and summary to depict the key features observed in the dataset.

Some of the key observations are listed below:

#### The Value boxes represent the total number of delays caused in some of the states like California, Texas, New york And New Jersey and also the total delays in US for all airlines.

1. In the univariate bar graph comparing the number of flights taking of for each day in the particular months, we can infer that on a given date there are not many abnormalities except on 4th,11th,18th and 25th date of the month where in the flight frequency is least. Since all thesedate are falling on Sundays, we can conclude that the business on Sundays is relatively low compared to other days of the week.
2. In the violin graph, only when Arrival Delay is longer than 15 minutes there's data about what caused the delay. With F9 and FL airlines showing almost close to no delays with respect to weather. The majority of delays are short, and the longer delays, while unusual, are more heavy loaded in time.
3. For the 4th graph of bar graph , late Aircraft Delay seems to be the most important variable, not only due to its longer wait periods, but it's affected by the others due to the delay propagation. Its distribution is almost parallel to NAS Delay, and seems to be the compounded effect of the other delays minus the delay reduction from the alloted Elapsed time > Actual Elapsed Time in the flight operations.  If carriers allowed aircrafts a higher speed and fuel consumption on planes that departed late, the effect wouldn't spread along the flight chain. This would decrease the NAS Delay itself and the Late Aircraft Delay, decreasing the number of planes being delayed.

#### In the scatterplot , we can observe from the graph that a good correlation exits between these two variables. Also we can see that most of the data points lie between 0 to 200 minutes hence most of flights have an average delay in-between these data points. But overall the plots seems to be aligned and has a synchronous pattern.Also we can infer that when planes take off on time, landing usually takes place before the expected time so that also allows to decrease the subsequent delay of net scheduled flights. One more likely interpretation is that the delays generated by previous flights either increase or decrease in each following trip.

1. The geo map depicts distribution of flight delays State wise on the US map. We can infer that state with a darker colored shade has greater number of delays than a state with lighter color shade. Since it’s an interactive plot, we can hover above the state by placing cursor on a particular state, for example , Arizona has 1465 no. of delayed flights for October , 2014.
2. Using the pie chart we could infer that the top states to have the highest carrier delays in Texas followed by California and other states. Texas alone contributes to around 21.6% of the total carrier delays. It can be seen how states with lower volume of flights tend to have a less Carrier Delay then the states with more frequency of flights, so it seems like number of flights per state matters for the delay to cause it.
3. In order for users to see the data from which the above references were made, we have created a separate tab in which the whole data is loaded with 35 entries showing for each page.
4. Using Pivot tables we have done cross tabulation of different variables and displayed values as either sums, counts, averages, etc. Here in I have taken heat map to project the influence of security delay on all the states of US. California accounts to most of the security delays and hence its represented with darker shade followed by other states with lighter shades.

## **CONCLUSION**:

## In the summary report in dashboard, we can see the average delays for each type of delay where in highest average delay is caused by Arrival delay and major reason for the delays is late aircraft delay. While the dataset itself doesn't offer a solution to the delay issue, it's clear that delay reduction has many intertwined variables. The solution to Late arrival Delay lies within the carriers. If carriers allowed aircrafts a higher speed and fuel consumption on planes that departed late, the effect wouldn't spread along the flight chain. This would decrease the Late arrival Delay which would in turn significatively reduce the number of planes being delayed.

#### In conclusion we can say that a solution applicable to one type of delay will heavily affect the others, resulting in a compounded effect that will allow more efficient operations which benefit the passengers, airports, carriers and even the world as a whole.