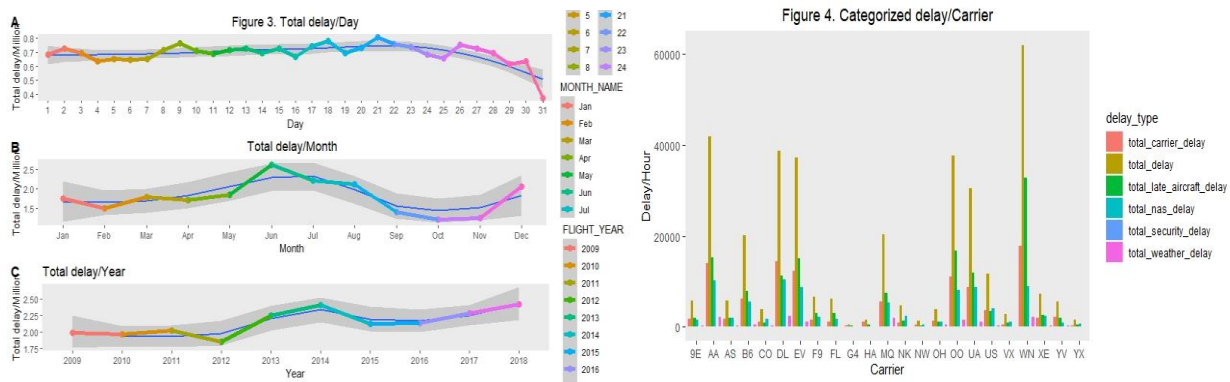




## BIG DATA REPORT-I

Findings: Flight delays happen due to various reasons such as bad weather, security delays, late aircraft delays, carrier delays, etc. It can be seen from Figure 1. graph that late aircraft delay contributes the most and security delay conditions contribute the least to the flight delays. Findings: From the Figure 2. bar plot we can notice that most of the carriers facing the late aircraft delay more often followed by carrier delay followed by the NAS delay followed by the weather delay followed by security delay.

### 6.2 Visualization 2:

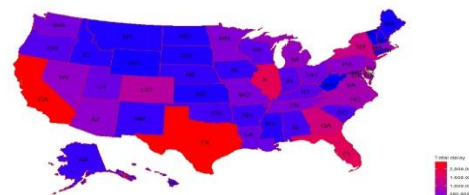


Objective: To find out which are the best and worst days to travel in a month, to find out which month in a year contributes the most /least to the flight delays, to find out which year contributed the most /least to the flight delays and, to analyze which delays are influencing the most per each carrier between the year 2009 and 2018

Findings: It can be seen that from the Figure 3. line graph “A” the 18<sup>th</sup> and 21<sup>st</sup> day of all the months in all the years between 2009 to 2018 has the highest amount of delay which proves that they are the worst days to travel whereas the 4<sup>th</sup> and 29<sup>th</sup> days are the best days to travel as on those days flight delays are experienced less. The day 31<sup>st</sup> has the least amount of delay(needs to make a note that not all months have 31 days in a year). It can be seen from the Figure 3. Line Graph “B” that October has the least delays followed by November whereas June has the highest delays followed by July in all years together. It can be seen from the Figure 3. line graph “C” that the 2012 year had the least delay(less than 1 million minutes) whereas the year 2018 had the highest flight delays (approx. 2.5 million minutes), the year 2014 had the second-highest flight delay(approx. 2.4millionminutes)

From the above Figure 4 bar plot, we can notice that most of the carriers facing the late aircraft delay more often followed by carrier delay followed by weather delay followed by the NAS delay followed by security delay.

Figure 5: Total Delay for Each State in a US Map



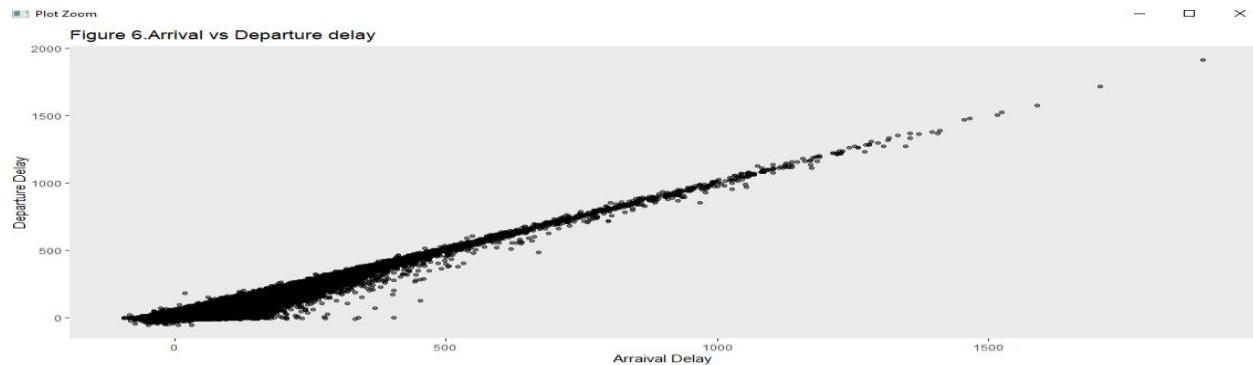
## BIG DATA REPORT-I

US Map that Texas (TX) and California State(CA) has the most flight delays followed by Illinois State(IL). The darker

the red color the greater the flight delay. The darker the blue color the lesser the flight delay.

### 6.4 Visualization 4:

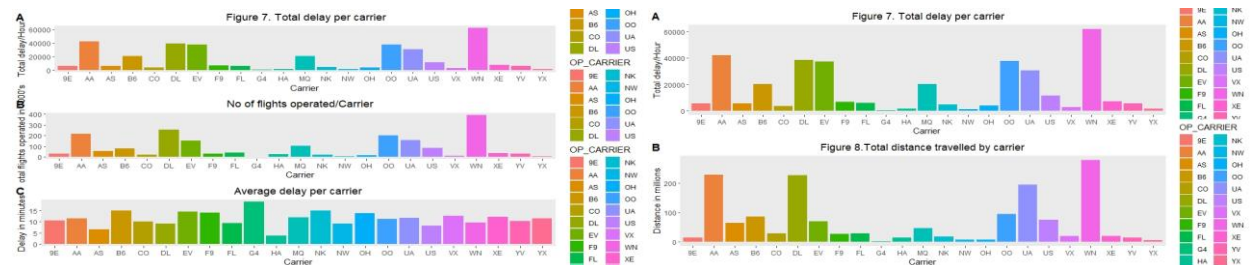
Objective: A plot between the Arrival Delay and Departure Delay.



Findings: From this Figure 6. Scatter plot it can be seen that there exists a Straight linear line between Arrival and Departure delays which means if the flight arrives early to the Airport then it will take off early whereas if an airline arrives late to the airport then it will take off late. ( ie departure delay)

### 6.5 Visualization 5:

Objective: A plot to analyze to get the insights from total delay, total flights operated, and the average flight delay per carrier. Also to visualize does distance influence the total delay per carrier?



Findings: From the Figure 7 and Figure 8., we can notice that the carrier WN has the highest flight delay but it has also operated the most number of flights close to 400,000 between the years 2009 and 2018. Hence, the average flight delay for the carrier WN is not as much as it would have thought to be based on the total delay. With the above comparison, we can conclude that the frequency of flight delay for carrier WN and DL is quite less. It is the other way around for the carriers 9E, AS, B6, CP, G4, F9, etc. which have operated the least but have the more average delays. We can also conclude that the flights that have operated the most have the highest amount of delays.

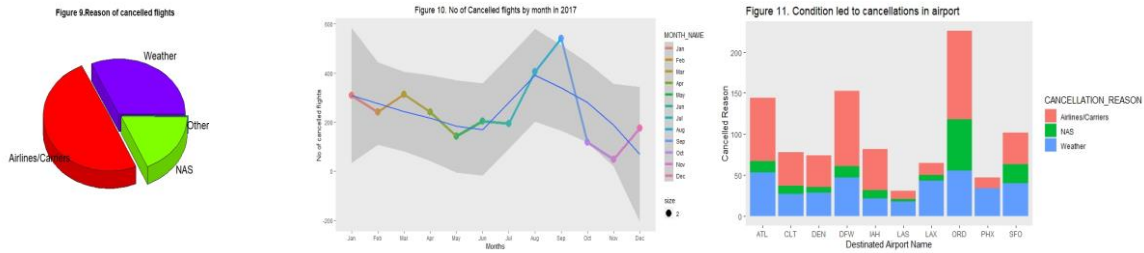
From the Figure 8, we can notice that WN traveled close to 300 million km which is the most followed by AA close to 228 million km followed by the carrier DL close to 226 million km, etc. The carriers that traveled the most have greater delays but carriers EV, MQ, and OO have traveled

## BIG DATA REPORT-I

less but have a greater delay. The carrier US is the only carrier that has traveled more but has less delay.

### 6.6 Visualization 6:

Objective: To visualize the cancellation reason in a 3D Pie Chart and to find out which airport in the USA has the highest/lowest cancellations with a cancellation reason. visualize flight cancellations per each month in the year 2017. To find out which airport in the USA has the highest/lowest cancellations with a cancellation reason.



Findings: We can note from the Figure 9. 3D Pie chart that approximately 50% of the flights were canceled with a reason Airlines/Carriers, followed by Weather, followed by NAS, followed by the cancellation reason other which contributes to the least amount of flight cancellations. We can note from the Figure 10. graph that November experienced fewer flight cancellations followed by Oct followed by May etc in the year 2017. We can note that September experienced higher flight cancellations followed by July followed by January etc in the same year.

It can be seen from the Figure 11. that Chicago O' Hare (Airport code -ORD ) had the highest number of cancellations among all the Airports in the USA, whereas McCarran International Airport(Airport Code -LAS) had the least Flight Cancellations. Furthermore, the four different cancellation reasons are:1) Airlines/Carriers cancellations, 2)NAS cancellations and 3)Other Cancellations 4)Weather Cancellations Respectively.

### 6.7 Visualization 7:

Objective: To visualize which are the best and worst days to travel based on the data plotted from the years 2009 and 2018 and to visualize which are the best and worst days to travel based on the data plotted from the years 2009 and 2018 for the state of Texas.

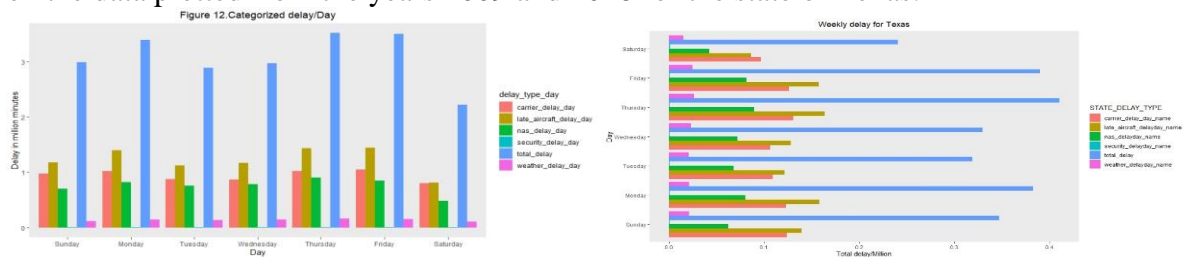


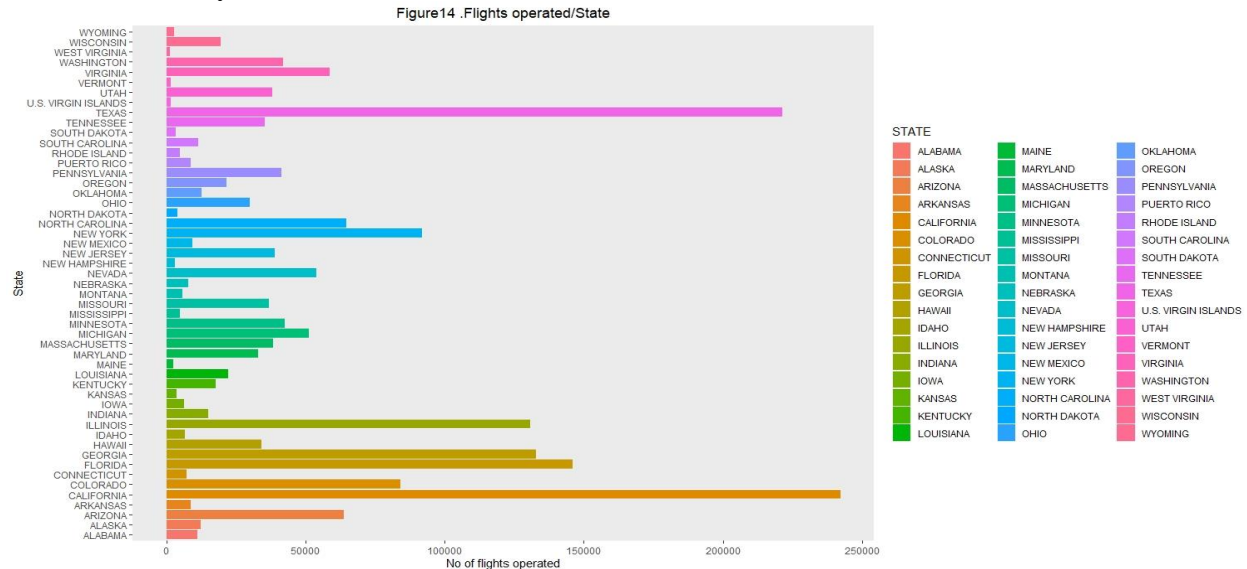
Figure 13.

Findings: From the Figure 12, we can notice that Saturday is the best day to travel as the total delay for Saturday is close to 2.5 million minutes while Thursday and Friday are the worst days to travel as the total delay for these days is close to 4 million minutes. From the Figure 13. plot we can notice still Saturday is the best day to travel in Texas which has a delay close to 0.22

## BIG DATA REPORT-I

million minutes while Thursday and Friday are the worst days to travel as the total delay for these days is close to 0.4 million minutes.

### 6.8 Visualization 8: Objective: To visualize how many flights are operated by each state in the US between the year 2009 and 2018



Findings: From the above Figure 14, bar plot we can notice that the California state operated the most number of flights around 250,000 flights followed by Texas close to 230,000 flights followed by Florida followed by Georgia etc.

### 7. R Code:



Report\_1\_Code.pdf

### 8. Conclusions and discussion:

Airline Delay is an unpleasant condition that has caused the huge cost to the Airline Networks, Air Passengers in many Ways. Airline Delays is a crucial subject because of its economic impact. Airline Delays might cause the hike in Flight Ticket charges. Furthermore also causes an increase in the Operational /Maintenance costs for the AirLine Companies. Till date, a huge effort has been made to minimize the AirLine Delays in Airports. In our project, some Data Analysis and Data Visualization techniques have been used to find Various factors responsible for AirLine Delay Analysis at Airports. Also in which years the Flight delay was maximum and minimum was found out. Furthermore data visualization in Which Airports Airline Cancellations have been maximum, minimum were found out. Taking all these factors into considerations some new solutions have to be found to minimize the AirLine delays in the Future. Minimizing the Airline delays is not only going the reduce Air Ticket costs and other costs to the Air Passengers but also it improves the economy of the Country. Moreover minimizing the AirLine Delays brings huge Reputation and profits to the AirLine Network Companies.



## 9. Bibliography:

1. <https://medium.com/@saurabhthakrani>. (2019, March 13). *FLIGHT DELAY ANALYSIS*. Medium. [Link](#)
2. <https://medium.com/@mosesandersonong>. (2018, February 17). *Data Visualisation Of Flight Delays With Tableau*. Medium. [Link](#)
3. Ferguson, J., Kara, A. Q., Hoffman, K., & Sherry, L. (2013). Estimating domestic US airline cost of delay based on the European model. *Transportation Research Part C: Emerging Technologies*, 33, 311–323. [Link](#)
4. Ionescu, L., Gwiggner, C., & Kliwer, N. (2016). Data Analysis of Delays in Airline Networks. *Bus Inf Syst Eng*, 58(2), 119–133. [Link](#)
5. Wu, W., Wu, C.-L., Feng, T., Zhang, H., & Qiu, S. (2018). Comparative Analysis on Propagation Effects of Flight Delays: A Case Study of China Airlines. *Journal of Advanced Transportation*, 2018, 1–10. [Link](#)
6. Hsiao, C.-Y., & Hansen, M. (2006). Econometric Analysis of U.S. Airline Flight Delays with Time-of-Day Effects. *Transportation Research Record*, 1951(1), 104–112. [Link](#)
7. Package 'tidyverse.' (2019, November 21). Cran.Rstudio.Com. [Link](#)
8. Package 'lubridate.' (2020, June 8). Cran.r-Project. [Link](#)
9. Package 'caret.' (2020, March 20). Cran.r-Project. [Link](#)
10. Package 'dplyr.' (2020, August 12). Cran.r-Project. [Link](#)
11. Package 'ggpubr.' (2020, June 27). Cran.r-Project. [Link](#)
12. Package 'usmap' (2020, October 7). Cran.r-Project. [Link](#)
13. Package 'corrplot' (2017, October 17). Cran.r-Project. [Link](#)
14. Santos, B. F., Wormer, M. M. E. C., Achola, T. A. O., & Curran, R. (2017). Airline delay management problem with airport capacity constraints and priority decisions. *Journal of Air Transport Management*, 63, 34–44. [Link](#)
15. Mueller, E., & Chatterji, G. (2002). *Analysis of Aircraft Arrival and Departure Delay Characteristics*. [Link](#).
16. Sherry, L., Calderon-Meza, G., & Samant, A. (2010). *Trends in airline passenger trip delays (2007 – 2009)*. [Link](#)

## Bibliography:

1. <https://medium.com/@mosesandersonong/data-visualisation-of-flight-delays-with-tableau-40aa6abf676b>.
2. <https://medium.com/@saurabhthakrani/analyzing-flight-delays-d8cba808bcd3>.
3. Ferguson, J., Kara, A. Q., Hoffman, K. & Sherry, L. Estimating domestic US airline cost of delay based on the European model. *Transportation Research Part C: Emerging Technologies* **33**, 311-323 (2013).
4. Ionescu, L., Gwiggner, C. & Kliwer, N. Data analysis of delays in airline networks. *Business & Information Systems Engineering* **58**, 119-133 (2016).
5. Wu, W., Wu, C., Feng, T., Zhang, H. & Qiu, S. Comparative analysis on propagation effects of flight delays: a case study of China airlines. *J. Adv. Transport*. **2018** (2018).

