ECO520 Business Analytics Tools II FLIGHT DELAY DATA

1 Group Members

- 1. Abhishek Manohar
- 2. Neeraj Tank
- 3. Preeti Lazarus

2_Data Sources and Topics

Sources:-

https://bigblue.depaul.edu/jlee141/econdata/eco520/

- Flight Delay Data (CSV FILE)

4.1 Motivation or Main Business Idea (1 to 2 pages)

The purpose of performing data analysis on airline data related to delayed flights is to identify the underlying factors that contribute to flight delays and to develop strategies to reduce their occurrence. By analyzing the data, airlines can identify patterns and trends in flight delays, such as specific routes or times of day that are more prone to delays, and can use this information to make operational changes or improvements to their scheduling.

Furthermore, data analysis can help airlines identify the root causes of flight delays, such as weather conditions, technical issues, or staffing problems, and take proactive measures to mitigate these issues. This can include strategies such as scheduling additional crew members or implementing maintenance programs to prevent technical problems from arising.

Overall, data analysis can help airlines improve their operational efficiency, reduce costs associated with flight delays, and ultimately provide a better experience for their passengers by reducing the frequency and duration of flight delays.

Motivation: Flight delays can cause significant inconvenience to passengers and can have financial implications for airlines. Therefore, understanding the underlying causes of delays and developing strategies to reduce them is of great importance for both airlines and their customers.

Question: What are the key factors contributing to flight delays in our airline, and how can we use this information to improve our scheduling and reduce the frequency and duration of delays?

Method: To answer this question, we would collect and analyze data on our airline's flights, including factors such as departure and arrival times, weather conditions, aircraft maintenance, crew availability, and passenger load. We could use statistical methods such as regression analysis to identify which factors have the greatest impact on flight delays, and then develop strategies to mitigate those factors.

Results: The results of our analysis could inform operational changes such as adjusting flight schedules or increasing staffing levels during peak periods, as well as implementing preventative maintenance programs to reduce technical issues. By

reducing the frequency and duration of flight delays, we could improve the overall customer experience and reduce costs associated with delayed flights.

4.2 Data and Empirical Methodology (1 to 2 pages)

Data: We will use a dataset of flight records for our airline. The data includes information on departure and arrival times, flight durations, aircraft types, route information, and weather conditions. We will use this data to investigate the factors that contribute to flight delays.

Summary statistics: We will present summary statistics of the data, such as the average delay time, the percentage of flights delayed, and the distribution of delay times across different factors such as route and time of day. We may also present graphs or charts to illustrate the trends in the data over time and highlight any historical events or changes in airline operations that may have affected delay rates.

Estimating equation: We will use a multiple linear regression model to estimate the factors that contribute to flight delays. The regression equation will take the form:

Delay time = β 0 + β 1Weather conditions + β 2Aircraft type + β 3Route information + β 4Time of day + ϵ

where $\beta 0$ is the intercept term, $\beta 1$ -4 are the coefficients for the different factors, and ϵ is the error term.

Methodology: The multiple linear regression model will allow us to identify the relative importance of different factors in contributing to flight delays. We can compare the results from the regression model to simple descriptive statistics or other methods such as correlation analysis to gain a deeper understanding of the underlying relationships in the data. The regression model will also allow us to control for the effects of different factors and estimate their individual contributions to flight delays, which would not be possible with simpler methods.

4.3 Results (3 to 4 pages)

- Descriptive Analytics
- Proc mean, Proc summary, Proc Univariate, Proc sgplot of ggplot, and Maps

#Summary

fd %>% summary(security_delay)

fd %>% summary(weather_delay)

fd %>% summary(arr_del15)

fd %>% summary(arr_cancelled)

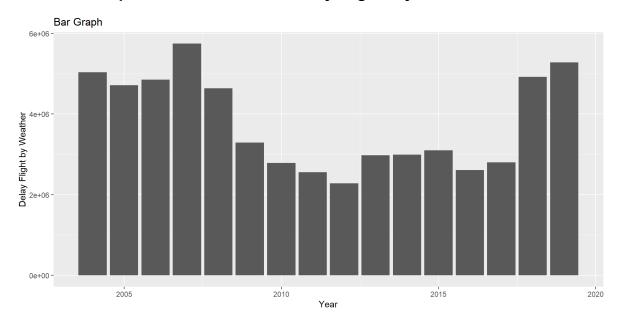
fd %>% summary(nas_delay)

	rr_del15)					
vear		arrier (carrier name	airport	airport_name	arr_flights
	n. : 1.000 Len	ath:265047	enath:265047	Length:265047	Length: 265047	Min. : 1.0
1st Qu.:2007 1s	t Ou.: 4.000 Cla	ss :character (lass :character	Class :character	Class :character	1st Ou.: 61.0
	dian : 7.000 Mod	le :character M	Mode :character	Mode :character	Mode :character	Median : 124.0
Mean :2011 Me	an : 6.507					Mean : 396.3
3rd Qu.:2015 3r	d Qu.: 9.000					3rd Qu.: 284.0
Max. :2019 Ma	x. :12.000					Max. :21977.0
						NA's :366
arr_del15	carrier_ct	weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled
Min. : 0.00	Min. : 0.00	Min. : 0.000	Min. : -0.01	Min. : 0.0000	Min. : 0.00	Min. : 0.000
1st Qu.: 10.00	1st Qu.: 3.58	1st Qu.: 0.000	1st Qu.: 2.03	1st Qu.: 0.0000	1st Qu.: 2.00	1st Qu.: 0.000
Median : 25.00	Median : 9.00	Median : 0.680	Median : 6.19	Median : 0.0000	Median : 6.82	Median : 1.000
Mean : 78.07	Mean : 21.95	Mean : 2.758	Mean : 25.94	Mean : 0.1772	Mean : 27.23	Mean : 6.771
3rd Qu.: 60.00	3rd Qu.: 20.76	3rd Qu.: 2.170	3rd Qu.: 16.66	3rd Qu.: 0.0000	3rd Qu.: 18.82	3rd Qu.: 4.000
Max. :6377.00	Max. :1792.07	Max. :717.940	Max. :4091.27	Max. :80.5600	Max. :1885.47	Max. :1969.000
NA's :422	NA's :366	NA's :366	NA's :366	NA's :366	NA's :366	NA's :366
arr_diverted	arr_delay	carrier_delay	weather_delay	nas_delay	security_delay	late_aircraft_delay
Min. : 0.0000	Min. : 0	Min. : 0	Min. : 0.0	Min. : -1	Min. : 0.000	Min. : 0
1st Qu.: 0.0000	1st Qu.: 513	1st Qu.: 173	1st Qu.: 0.0	1st Qu.: 71	1st Qu.: 0.000	1st Qu.: 105
Median : 0.0000	Median : 1330	Median : 476	Median : 30.0	Median : 231	Median : 0.000	Median : 410
Mean : 0.9181	Mean : 4482	Mean : 1323	Mean : 228.8	Mean : 1196	Mean : 7.026	Mean : 1727
3rd Qu.: 1.0000	3rd Qu.: 3303	3rd Qu.: 1154	3rd Qu.: 171.0	3rd Qu.: 656	3rd Qu.: 0.000	3rd Qu.: 1225
Max. :256.0000	Max. :433687	Max. :196944	Max. :57707.0	Max. :238440	Max. :3194.000	Max. :148181
NA's :366	NA's :366	NA's :366	NA's :366	NA's :366	NA's :366	NA's :366
. Ed 00:00						
> fd %>% summary(a						
year	month c		carrier_name	airport	airport_name	arr_flights
year Min. :2004 Mi	month c n. : 1.000 Len	gth:265047	_ength:265047	Length:265047	Length:265047	Min. : 1.0
year Min. :2004 Mi 1st Qu.:2007 1s	month c n. : 1.000 Len t Qu.: 4.000 Cla	gth:265047 l ss:character (ength:265047 Class :character	Length:265047 Class :character	Length:265047 Class :character	Min. : 1.0 1st Qu.: 61.0
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me	month c n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod	gth:265047 l ss:character (_ength:265047	Length:265047	Length:265047	Min. : 1.0 1st Qu.: 61.0 Median : 124.0
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me	month c n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507	gth:265047 l ss:character (ength:265047 Class :character	Length:265047 Class :character	Length:265047 Class :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r	month con. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000	gth:265047 l ss:character (ength:265047 Class :character	Length:265047 Class :character	Length:265047 Class :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me	month con. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000	gth:265047 l ss:character (ength:265047 Class :character	Length:265047 Class :character	Length:265047 Class :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma	month	gth:265047 I ss:character (e:character N	ength:265047 Class :character Mode :character	Length:265047 Class :character Mode :character	Length:265047 Class :character Mode :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15	month	gth:265047 I ss :character (le :character N weather_ct	ength:265047 lass :character Mode :character	Length:265047 Class :character Mode :character	Length:265047 Class :character Mode :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00	month c n. : 1.000 Len t Qu.: 4.000 Cla ddan : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00	gth:265047 I ss:character (le:character M weather_ct Min.: 0.000	ength:265047 Class :character Mode :character nas_ct Min. : -0.01	Length:265047 Class :character Mode :character security_ct Min. : 0.0000	Length:265047 Class :character Mode :character	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00	month n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 xx. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58	gth:265047 Iss :character (e :character N	_ength:265047 Tlass :character Mode :character 	Length:265047 Class :character Mode :character security_ct Min. : 0.0000 1st Qu.: 0.0000	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_dell5 Min. : 0.00 1st Qu.: 10.00 Median : 25.00	month	gth:265047 I ss :character (le :character N weather_ct Min. : 0.000 1st Qu.: 0.000 Median : 0.680	ength:265047 Tlass :character Mode :character nas_ct Min. : -0.01 1st Qu.: 2.03 Median : 6.19	Length:265047 Class :character Mode :character security_ct Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07	month n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95	gth:265047 Iss:character (e:character N	_ength:265047 Tlass :character Mode :character 	Length:265047 Class :character Mode :character security_ct Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Median : 0.1772	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00	month n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76	gth:265047 Iss :character (e :character N	_nas_ctna	Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2019 Ma .:2019 Ma	month : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07	gth:265047 I ss :character (e :character N weather_ct Min. : 0.000 1st Qu. : 0.000 Median : 0.680 Mean : 2.758 3rd Qu. : 2.170 Max. :717.940		Length:265047 Class :character Mode :character security_ct Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000 Max. :80.5600	Length:265047 Class:character Mode:character late_aircraft_ct Min.: 0.00 1st Qu.: 2.00 Median: 6.82 Mean: 27.23 3rd Qu.: 18.82 Max.:1885.47	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422	month n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366	gth:265047 1.5s :character (e :character N N N N N N N N N	nas_ct n	Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000 Max. :80.5600 NA's :366	Length: 265047 Class : character Mode : character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82 Max. :1885.47 NA's :366	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted	month . : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod dian : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366 arr_delay	gth:265047 Iss :character (e :character N		Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000 Max. :80.5600 NA's :366 nas_delay	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82 Max. :1885.47 NA's :366 security_delay	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 Median : 1.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000	month : 1.000 Len t. 0.: 1.000 Len t. 0.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366 arr_delay Min. : 0	gth:265047 1 1 1 1 1 1 1 1 1	ength:265047 -lass :character node :character Min. : -0.01 1st Qu.: 2.03 Median : 6.19 Mean : 25.94 3rd Qu.: 16.66 Max. :4091.27 NA's :366 weather_delay Min. : 0.0	Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000 Max. :80.5600 NA's :366 nas_delay Min. : -1	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82 Max. :1885.47 NA's :366 security_delay Min. : 0.000	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000 1st Qu.: 0.0000	month n. : 1.000 Len t Qu.: 4.000 Cla ddan : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366 arr_delay Min. : 0 1st Qu.: 513	gth:265047 1.5s : character (e : character N	nas_ct n	Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.1772 3rd Qu.: 0.0000 Max. :80.5600 NA's :366 nas_delay Min. : -1 1st Qu.: 71	Length:265047 Class:character Mode:character late_aircraft_ct Min.: 0.00 1st Qu.: 2.00 Median: 6.82 Mean: 27.23 3rd Qu.: 18.82 Max.:1885.47 NA's:366 security_delay Min.: 0.000 1st Qu.: 0.000 1st Qu.: 0.000	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Median : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0 1st Qu.: 105
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2019 Ma arr_dell5 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Median : 0.0000	month : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod dian : 7.000 Mod dian : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366 arr_delay Min. : 0 1st Qu.: 513 Median : 1330	gth:265047 Iss:character (e:character N weather_ct Min. : 0.000 Ist Qu.: 0.000 Median : 0.680 Mean : 2.758 3rd Qu.: 2.170 Max. :717.940 NA's :366 carrier_delay Min. : 0 Ist Qu.: 173 Median : 476		Length:265047 Class :character Mode :character Min. : 0.0000 1st Qu.: 0.0000 Median : 0.0000 Mean : 0.1772 3rd Qu.: 0.0000 Max. : 80.5600 NA's : 366 nas_delay Min. : -1 1st Qu.: 71 Median : 231	Length: 265047 Class : character Mode : character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82 Max. : 1885.47 NA's : 366 security_delay Min. : 0.000 1st Qu.: 0.000 Median : 0.000	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0 1st Qu.: 105 Median : 410
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000 Median : 0.0000 Median : 0.0000 Median : 0.9081	month : 1.000 Len t : 1.000 Len t. Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. : 12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. : 1792.07 NA's : 366 arr_delay Min. : 0 1st Qu.: 513 Median : 1330 Mean : 4482	weather_ct Min. : 0.000 1st Qu.: 0.000 Median : 0.680 Mean : 2.758 3rd Qu.: 2.170 Max. : 717.940 NA's : 366 carrier_delay Min. : 0 1st Qu.: 173 Median : 476 Median : 1323	_nas_ct _nas_c	Length:265047 Class:character Mode:character Security_ct Min.:0.0000 1st Qu.:0.0000 Median:0.0000 Mean:0.1772 3rd Qu.:0.0000 Max.:80.5600 NA's:366 nas_delay Min.:-1 1st Qu.: 71 Median: 231 Mean: 1196	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu. : 18.82 Max. :1885.47 NA's :366 security_delay Min. : 0.000 Median : 0.000 Median : 0.000 Median : 7.026	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0 1st Qu.: 105 Median : 410 Mean : 410 Mean : 1727
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me 3rd Qu.:2015 3r Max. :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000 Median : 0.0001 Median : 0.9181 3rd Qu.: 1.0000	month n. : 1.000 Len t Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. :12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. :1792.07 NA's :366 arr_delay Min. : 0 1st Qu.: 513 Median : 1330 Mean : 4482 3rd Qu.: 3303	weather_ct Min. : 0.000 1st Qu.: 0.000 Median : 0.680 Mean : 2.758 3rd Qu.: 2.170 Max. :717.940 NA's :366 carrier_delay Min. : 0 1st Qu.: 173 Median : 476 Mean : 1323 3rd Qu.: 1154	nas_ct nas_ct min. : -0.01 1st Qu.: 2.03 Median : 6.19 Mean : 25.94 3rd Qu.: 16.66 Max. :4091.27 NA's :366 weather_delay Min. : 0.0 1st Qu.: 0.0 Median : 30.0 Median : 228.8 3rd Qu.: 171.0	security_ct Min. : 0.0000 1st Qu.: 0.0000 Median : 0.1772 3rd Qu.: 0.0000 Max. : 80.5600 NA's : 366 nas_delay Min. : -1 1st Qu.: 71 Median : 231 Mean : 1196 3rd Qu.: 6566	Length: 265047 Class : character Mode : character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu.: 18.82 Max. :1885.47 NA's :366 security_delay Min. : 0.000 1st Qu.: 0.000 Median : 0.000 Median : 7.026 3rd Qu.: 0.000	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0 1st Qu.: 105 Median : 410 Mean : 1727 3rd Qu.: 1225
year Min. :2004 Mi 1st Qu.:2007 1s Median :2011 Me Mean :2011 Me Mean :2019 Ma arr_del15 Min. : 0.00 1st Qu.: 10.00 Median : 25.00 Mean : 78.07 3rd Qu.: 60.00 Max. :6377.00 NA's :422 arr_diverted Min. : 0.0000 Median : 0.0000 Median : 0.0000 Median : 0.9081	month : 1.000 Len t : 1.000 Len t. Qu.: 4.000 Cla dian : 7.000 Mod an : 6.507 d Qu.: 9.000 x. : 12.000 carrier_ct Min. : 0.00 1st Qu.: 3.58 Median : 9.00 Mean : 21.95 3rd Qu.: 20.76 Max. : 1792.07 NA's : 366 arr_delay Min. : 0 1st Qu.: 513 Median : 1330 Mean : 4482	weather_ct Min. : 0.000 1st Qu.: 0.000 Median : 0.680 Mean : 2.758 3rd Qu.: 2.170 Max. : 717.940 NA's : 366 carrier_delay Min. : 0 1st Qu.: 173 Median : 476 Median : 1323	_nas_ct _nas_c	Length:265047 Class:character Mode:character Security_ct Min.:0.0000 1st Qu.:0.0000 Median:0.0000 Mean:0.1772 3rd Qu.:0.0000 Max.:80.5600 NA's:366 nas_delay Min.:-1 1st Qu.: 71 Median: 231 Mean: 1196	Length:265047 Class :character Mode :character late_aircraft_ct Min. : 0.00 1st Qu.: 2.00 Median : 6.82 Mean : 27.23 3rd Qu. : 18.82 Max. :1885.47 NA's :366 security_delay Min. : 0.000 Median : 0.000 Median : 0.000 Median : 7.026	Min. : 1.0 1st Qu.: 61.0 Median : 124.0 Mean : 396.3 3rd Qu.: 284.0 Max. :21977.0 NA's :366 arr_cancelled Min. : 0.000 1st Qu.: 0.000 Median : 1.000 Mean : 6.771 3rd Qu.: 4.000 Max. :1969.000 NA's :366 late_aircraft_delay Min. : 0 1st Qu.: 105 Median : 410 Mean : 410 Mean : 1727

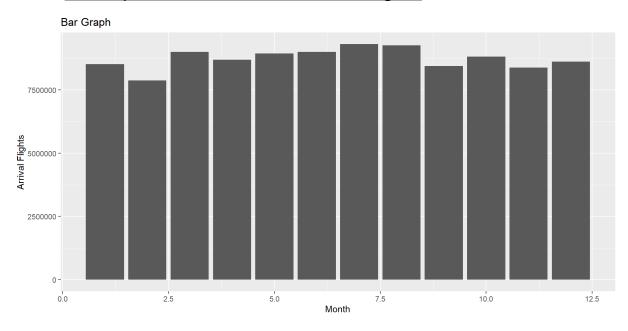
> fd %>% summary(nas_delay	y)				
year month	carrier	carrier carrier_name		airport airport_name	
Min. :2004 Min. : 1	1.000 Length:265047	Length:265047	Length:265047	Length:265047	Min. : 1.0
1st Qu.:2007 1st Qu.: 4	4.000 Class :character	Class :character	Class :character	Class :character	1st Qu.: 61.0
Median :2011 Median : 7	7.000 Mode :character	Mode :character	Mode :character	Mode :character	Median : 124.0
Mean :2011 Mean : 6	6.507				Mean : 396.3
3rd Qu.:2015 3rd Qu.: 9	9.000				3rd Qu.: 284.0
	2.000				Max. :21977.0
					NA's :366
arr_del15 carri	ier_ct weather_ct	nas_ct	security_ct	late_aircraft_ct	arr_cancelled
Min. : 0.00 Min.				Min. : 0.00	Min. : 0.000
1st Qu.: 10.00 1st Qu.					1st Ou.: 0.000
Median : 25.00 Median		30 Median : 6.19	Median : 0.0000		Median : 1.000
	: 21.95 Mean : 2.75				Mean : 6.771
	.: 20.76 3rd Qu.: 2.17	70 3rd Qu.: 16.66	3rd Qu.: 0.0000		3rd Ou.: 4.000
	:1792.07 Max. :717.94				Max. :1969.000
NA's :422 NA's	:366 NA's :366	NA's :366	NA's :366	NA's :366	NA's :366
arr_diverted arr_	_delay carrier_delay	weather_delay	nas_delay	security_delay	late_aircraft_delay
Min. : 0.0000 Min.			Min. : -1	Min. : 0.000	Min. : 0
1st Qu.: 0.0000 1st Qu	u.: 513 1st Qu.: 173	3 1st Qu.: 0.0	1st Qu.: 71	1st Qu.: 0.000	1st Qu.: 105
Median: 0.0000 Median	n : 1330 Median : 476	5 Median : 30.0	Median : 231	Median : 0.000	Median : 410
Mean : 0.9181 Mean	: 4482 Mean : 1323	Mean : 228.8	Mean : 1196	Mean : 7.026	Mean : 1727
3rd Qu.: 1.0000 3rd Qu	u.: 3303 3rd Qu.: 1154	3rd Qu.: 171.0	3rd Qu.: 656	3rd Qu.: 0.000	3rd Qu.: 1225
Max. :256.0000 Max.	:433687 Max. :196944		Max. :238440	Max. :3194.000	Max. :148181
NA's :366 NA's	:366 NA's :366	NA's :366	NA's :366	NA's :366	NA's :366
. 1					

Proc sgplot of gplot, and Maps:

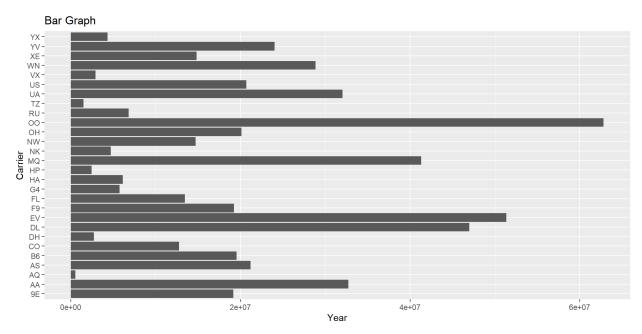
1. Bar Graphs between Year and Delay Flights by Weather.



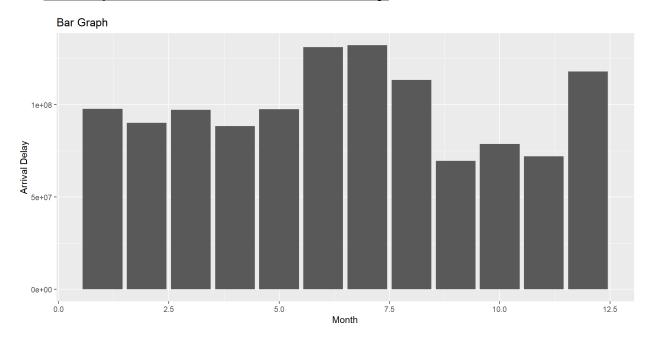
2. Bar Graphs between Month and Arrival Flights.



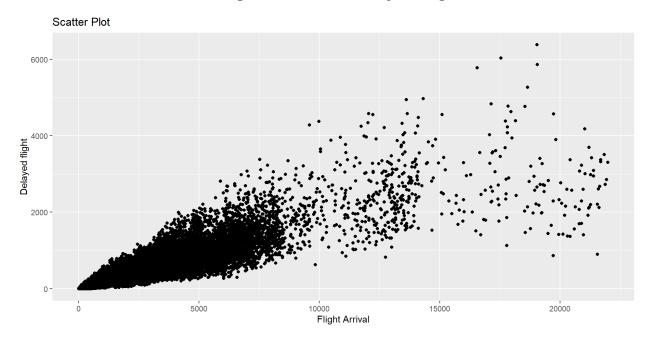
3. Bar Graphs between Year and Carrier.



4. Bar Graphs between Month and Arrival Delay.



5. Scatter Plot between Flight Arrival and Delayed Flight.

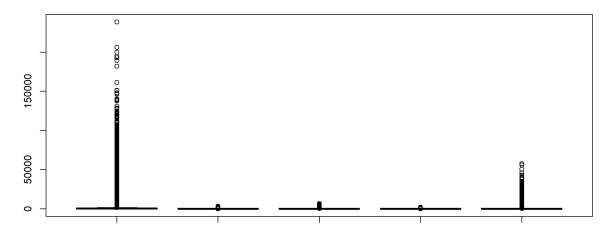


6. Line Graphs between Arrival Canceled and Carrier Control.

Arrival cancelled due to Carrier control

7. <u>Boxplot between NAS Delay, Security Delay, Arrival Delay by 15 min, Arrival Canceled and Weather Delay(To Find Outliers).</u>

Distribution of NAS Delay, Security Delay, Arrival Delay by 15min, Arrival Cancelled and Weather Delay



Correlation analysis and Analysis of Variance (ANOVA)

```
> cor_mat1
arr_del15 security_ct
arr_del15 1.0000000 0.4902167
security_ct 0.4902167 1.0000000
>
```

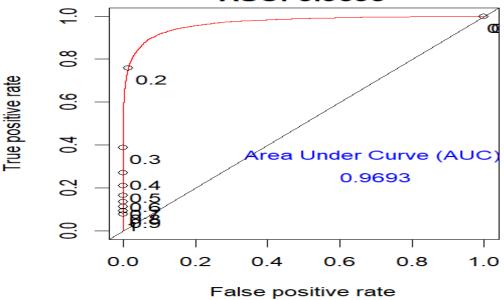
- Other relative analyses including histograms and statistics

- Simple, Multiple Regression on linear or nonlinear models
- Linear Regression Model :
 - Linear Regression Model to understand the impact of late_aircraft_ct on check_delay

```
call:
lm(formula = check_delay ~ late_aircraft_ct, data = train)
Residuals:
    Min
              10 Median
                               3Q
                                       Max
-3.9470 -0.1415 -0.1259 -0.1179 0.8821
Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
                                                    <2e-16 ***
(Intercept)
                  1.179e-01 7.658e-04
                                           153.9
late_aircraft_ct 2.561e-03 9.124e-06
                                           280.7
                                                    <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 0.3333 on 211698 degrees of freedom
Multiple R-squared: 0.2712,
                                  Adjusted R-squared: 0.2712
F-statistic: 7.88e+04 on 1 and 211698 DF, p-value: < 2.2e-16
> conf_table(yhat1,testy,"LPM-1")
 estname prob true_total truepos falsneg detection_rate false_total falspos trueneg false_pos_rate
1 LPM-1 0.1
                                                 43023 43023
                9902
2 LPM-1 0.2
                9902
                             2380
                                       0.7596
                                                 43023
                                                              42396
                                                                        0.0146
                      7522
                                                         627
                                                          0 43023
0 43023
                                                 43023
  LPM-1 0.3
                9902
                      3850
                             6052
                                       0.3888
                                                                            0
4 LPM-1 0.4
                9902
                      2665
                             7237
                                      0.2691
                                                 43023
                                                          0 43023
                                                                            0
5 LPM-1 0.5
                9902
                      2066
                             7836
                                      0.2086
                                                 43023
                                                         0 43023
                                                                            0
6
  LPM-1 0.6
                9902
                      1639
                             8263
                                       0.1655
                                                 43023
                                                          0
                                                              43023
                                                                            0
                     1325
  LPM-1 0.7
                                                          0 43023
                                                 43023
                9902
                             8577
                                      0.1338
                                                                            0
 LPM-1 0.8
                9902
                      1099
                             8803
                                       0.111
                                                 43023
                                                          0
                                                              43023
                                       0.0919
9
  LPM-1 0.9
                9902
                     910
                             8992
                                                 43023 0
                                                              43023
                                                                             0
```

The p value of late_aircraft_ct is less than 0.05. Hence it has a significant impact on check delay.

LPM-1 Receiver Operating Characteristic (ROC AUC: 0.9693



The AUC is 96% impling that it is a good fit.

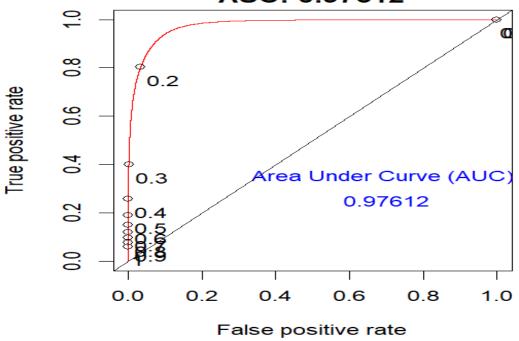
2. Linear Regression Model to understand the impact of carrier_delay, weather delay, nas delay, security delay on check delay.

```
Call:
lm(formula = check_delay ~ carrier_delay + weather_delay + nas_delay +
    security_delay, data = train)
Residuals:
             10 Median
                             3Q
-9.2213 -0.1449 -0.1257 -0.1145 0.9230
Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
                                               <2e-16 ***
(Intercept)
                1.135e-01 7.819e-04
                                      145.10
                                               <2e-16 ***
carrier_delay
               4.848e-05
                          3.593e-07
                                      134.93
                                               <2e-16 ***
weather_delay
               -2.402e-05
                          1.239e-06
                                      -19.39
                                               <2e-16 ***
                7.650e-06 2.079e-07
                                       36.80
nas_delay
              8.989e-04 2.130e-05
                                       42.20
                                               <2e-16 ***
security_delay
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Residual standard error: 0.3361 on 211695 degrees of freedom
Multiple R-squared: 0.2588,
                                Adjusted R-squared: 0.2587
F-statistic: 1.847e+04 on 4 and 211695 DF, p-value: < 2.2e-16
```

The p value of carrier_delay, weather_delay, nas_delay, security_delay is less than 0.05. Hence it has a significant impact on check_delay.

>	conf_tal	ole(ył	nat2,testy,	"LPM-2")						
	estname	prob	true_total	truepos	falsneg	detection_rate	false_total	falspos	trueneg	false_pos_rate
1	LPM-2	0.1	9902	9900	2	0.9998	43023	42973	50	0.9988
2	LPM-2	0.2	9902	7948	1954	0.8027	43023	1450	41573	0.0337
3	LPM-2	0.3	9902	3965	5937	0.4004	43023	66	42957	0.0015
4	LPM-2	0.4	9902	2548	7354	0.2573	43023	9	43014	2e-04
5	LPM-2	0.5	9902	1886	8016	0.1905	43023	4	43019	1e-04
6	LPM-2	0.6	9902	1478	8424	0.1493	43023	1	43022	0
7	LPM-2	0.7	9902	1176	8726	0.1188	43023	1	43022	0
8	LPM-2	0.8	9902	959	8943	0.0968	43023	1	43022	0
a	I DM_2	Λ Q	9902	762	9140	0.077	43023	1	43022	0

LPM-2
Receiver Operating Characteristic (ROC AUC: 0.97612



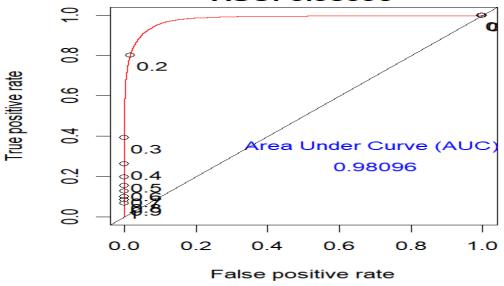
The AUC is 97% implying that it is a good fit.

3. Linear Regression Model to understand the impact of arr_cancelled, arr_diverted, arr_delay, late_aircraft_delay on check_delay.

```
lm(formula = check_delay ~ arr_cancelled + arr_diverted + arr_delay +
    late_aircraft_delay, data = train)
Residuals:
    Min
              1Q Median
                                3Q
                                        Max
-5.2143 -0.1428 -0.1268 -0.1183 1.2532
Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
                       1.180e-01 7.682e-04 153.62 <2e-16 ***
(Intercept)
                      -5.922e-04 3.599e-05
                                                          <2e-16 ***
                                               -16.46
arr_cancelled
                      -7.631e-03 2.524e-04 -30.23
9.891e-06 1.797e-07 55.03
                                                         <2e-16 ***
arr_diverted
                                                55.03 <2e-16 ***
arr_delay
late_aircraft_delay 2.100e-05 4.088e-07 51.37 <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.334 on 211695 degrees of freedom
Multiple R-squared: 0.2681,
                                  Adjusted R-squared: 0.2681
F-statistic: 1.939e+04 on 4 and 211695 DF, p-value: < 2.2e-16
| > conf_table(yhat3,testy,"LPM-3")
  estname prob true_total truepos falsneg detection_rate false_total falspos trueneg false_pos_rate
                                                                     0.9955
 1 LPM-3 0.1
                 9902
                       9882
                             20 0.998 43023 42830 193
    LPM-3 0.2
                 9902
                        7941
                              1961
                                        0.802
                                                  43023
                                                         726 42297
                                                                          0.0169
                                                        0 43023
0 43023
0 43023
0 43023
 3 LPM-3 0.3
                 9902 3888 6014
                                       0.3926
                                                 43023
                                       0.2618
                 9902 2592 7310
                                                 43023
 4 LPM-3 0.4
                 9902
                       1965
                              7937
                                       0.1984
                                                  43023
    LPM-3 0.5
                9902 1965
9902 1541
                            8361
 6 LPM-3 0.6
                                       0.1556
                                                 43023
7 LPM-3 0.7 9902 1254 8648
8 LPM-3 0.8 9902 986 8916
9 LPM-3 0.9 9902 829 9073
                                                43023 0 43023
43023 0 43023
43023 0 43023
                                      0.1266
                                      0.0996
0.0837
                                                                            0
```

The p value of arr_cancelled, arr_diverted, arr_delay, late_aircraft_delay is less than 0.05. Hence it has a significant impact on check delay.

LPM-3
Receiver Operating Characteristic (ROC AUC: 0.98096



The AUC is 98% implying that it is a good fit.

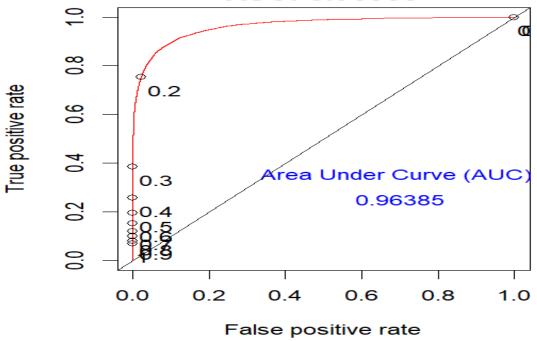
- Discrete Probability Model : Logistic Model
 - Logistic Regression Model to understand the impact of late_aircraft_ct on check_delay

```
Call:
glm(formula = check_delay ~ late_aircraft_delay, data = train)
Deviance Residuals:
    Min
             1Q
                  Median
                               3Q
-4.7968
       -0.1433 -0.1281 -0.1215
                                    0.8785
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
                                                 <2e-16 ***
(Intercept)
                   1.215e-01 7.714e-04
                                          157.5
late_aircraft_delay 3.830e-05 1.416e-07
                                          270.4
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for gaussian family taken to be 0.1133011)
    Null deviance: 32269 on 211699 degrees of freedom
Residual deviance: 23986 on 211698 degrees of freedom
AIC: 139762
Number of Fisher Scoring iterations: 2
```

> conf_table(yhat4,testy,"LRM-1")										
	estname	prob	true_total	truepos	falsneg	detection_rate	false_total	falspos	trueneg	false_pos_rate
1	LRM-1	0.1	9902	9902	0	1	43023	43023	0	1
2	LRM-1	0.2	9902	7464	2438	0.7538	43023	962	42061	0.0224
3	LRM-1	0.3	9902	3805	6097	0.3843	43023	3	43020	1e-04
4	LRM-1	0.4	9902	2539	7363	0.2564	43023	0	43023	0
5	LRM-1	0.5	9902	1921	7981	0.194	43023	0	43023	0
6	LRM-1	0.6	9902	1514	8388	0.1529	43023	0	43023	0
7	LRM-1	0.7	9902	1193	8709	0.1205	43023	0	43023	0
8	LRM-1	0.8	9902	980	8922	0.099	43023	0	43023	0
9	LRM-1	0.9	9902	791	9111	0.0799	43023	0	43023	0

The p value of late_aircraft_ct is less than 0.05. Hence it has a significant impact on check_delay.

LRM-1
Receiver Operating Characteristic (ROC
AUC: 0.96385



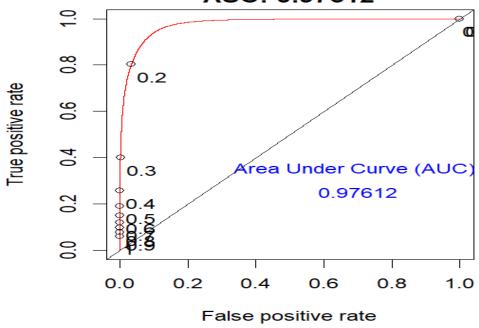
The AUC is 96% implying that it is a good fit.

2. Logistic Regression Model to understand the impact of carrier_delay, weather_delay, nas_delay, security_delay on check_delay.

```
glm(formula = check_delay ~ carrier_delay + weather_delay + nas_delay +
    security_delay, data = train)
Deviance Residuals:
                    Median
    Min
               10
                                  30
                                           Max
-9.2213 -0.1449 -0.1257 -0.1145
                                        0.9230
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
                 1.135e-01 7.819e-04 145.10
                                                  <2e-16 ***
(Intercept)
carrier_delay
                 4.848e-05 3.593e-07 134.93
                                                  <2e-16 ***
weather_delay -2.402e-05 1.239e-06 -19.39 <2e-16 ***
                 7.650e-06 2.079e-07 36.80 <2e-16 ***
nas_delay
security_delay 8.989e-04 2.130e-05
                                        42.20
                                                 <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
(Dispersion parameter for gaussian family taken to be 0.1129906)
    Null deviance: 32269 on 211699
                                        degrees of freedom
Residual deviance: 23920 on 211695 degrees of freedom
AIC: 139184
Number of Fisher Scoring iterations: 2
> conf_table(yhat5, testy, "LRM-2")
estname prob true_total truepos falsneg detection_rate false_total falspos trueneg false_pos_rate
1 LRM-2 0.1
               9902
                     9900
                             2
                                     0.9998
                                               43023 42973
                                                            50
                                                                      0.9988
                                               43023
               9902
                     7948
                           1954
                                     0.8027
  LRM-2 0.2
                                                      1450
                                                           41573
                                                                      0.0337
3 LRM-2 0.3
               9902
                           5937
                                     0.4004
                                               43023
                                                      66 42957
                     3965
                                                                      0.0015
              9902
                                                       9 43014
4 LRM-2 0.4
                     2548
                          7354
                                     0.2573
                                               43023
                                                                       2e-04
 LRM-2 0.5
LRM-2 0.6
               9902
                     1886
                           8016
                                     0.1905
                                               43023
                                                           43019
                                                                       1e-04
              9902
                     1478
                                                       1 43022
                           8424
                                     0.1493
                                               43023
                                                                         0
6
                                                       1 43022
             9902
                                              43023
  LRM-2 0.7
                     1176 8726
                                    0.1188
                                                                          0
  LRM-2 0.8
LRM-2 0.9
                                                       1 43022
1 43022
                                    0.0968
               9902
                      959
                           8943
                                               43023
                                                                          0
              9902
                    762 9140
                                               43023
                                      0.077
```

The p value of carrier_delay, weather_delay, nas_delay, security_delay is less than 0.05. Hence it has a significant impact on check delay.

LRM-2 Receiver Operating Characteristic (ROC AUC: 0.97612



The AUC is 97% implying that it is a good fit.

3. Logistic Regression Model to understand the impact of arr_cancelled, arr_diverted, arr_delay, late_aircraft_delay on check_delay.

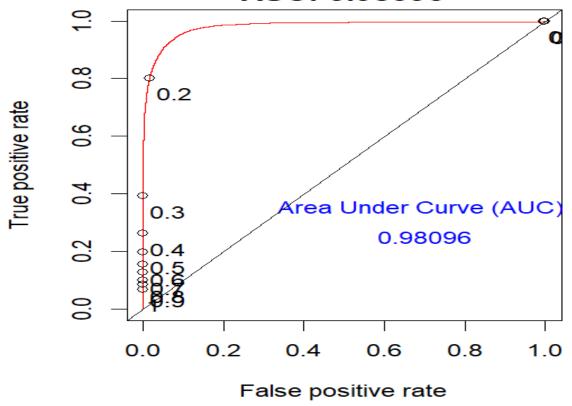
```
Call:
glm(formula = check_delay ~ arr_cancelled + arr_diverted + arr_delay +
   late_aircraft_delay, data = train)
Deviance Residuals:
             1Q
   Min
                  Median
                               3Q
                                       Max
-5.2143 -0.1428 -0.1268 -0.1183
                                    1.2532
Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)
                    1.180e-01
                              7.682e-04
                                                   <2e-16 ***
                                          153.62
                               3.599e-05
arr_cancelled
                   -5.922e-04
                                          -16.46
                                                   <2e-16 ***
arr_diverted
                   -7.631e-03 2.524e-04 -30.23
                                                   <2e-16 ***
arr_delay
                    9.891e-06 1.797e-07
                                           55.03
                                                   <2e-16 ***
late_aircraft_delay 2.100e-05 4.088e-07
                                           51.37
                                                   <2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for gaussian family taken to be 0.1115592)
                                    dearees of freedom
   Null deviance: 32269
                         on 211699
Residual deviance: 23617
                         on 211695
                                    degrees of freedom
AIC: 136485
```

Number of Fisher Scoring iterations: 2

```
> conf_table(yhat6,testy,"LRM-3")
 estname prob true_total truepos falsneg detection_rate false_total falspos trueneg false_pos_rate
  LRM-3 0.1
                   9902
                          9882
                                   20
                                              0.998
                                                         43023 42830
                                                                         193
                                                                                    0.9955
  LRM-3 0.2
                   9902
                          7941
                                 1961
                                              0.802
                                                         43023
                                                                  726
                                                                       42297
                                                                                    0.0169
3
  LRM-3 0.3
                   9902
                          3888
                                 6014
                                             0.3926
                                                         43023
                                                                       43023
                                                                                         0
  LRM-3 0.4
                   9902
                          2592
                                 7310
                                             0.2618
                                                         43023
                                                                       43023
  LRM-3 0.5
                   9902
                          1965
                                 7937
                                             0.1984
                                                         43023
                                                                   0
                                                                       43023
                                                                                         0
6
  LRM-3 0.6
                   9902
                          1541
                                 8361
                                            0.1556
                                                         43023
                                                                   0
                                                                       43023
                                                                                         0
   LRM-3 0.7
                                                                    0
                  9902
                          1254
                                 8648
                                             0.1266
                                                         43023
                                                                       43023
                                                                                         0
 LRM-3 0.8
                 9902
                          986
                                 8916
                                             0.0996
                                                         43023
                                                                       43023
                 9902
                                                                   0 43023
9 LRM-3 0.9
                         829
                                 9073
                                            0.0837
                                                         43023
```

The p value of arr_cancelled, arr_diverted, arr_delay, late_aircraft_delay is less than 0.05. Hence it has a significant impact on check_delay.

LRM-3
Receiver Operating Characteristic (ROC AUC: 0.98096

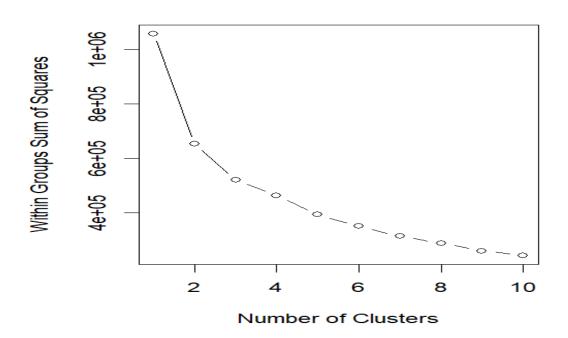


The AUC is 98% implying that it is a good fit.

Predictive Analytics (All required to apply to your model)

Clustering Analysis

```
[1] 1058496.0 652307.1 520114.7 461868.3 392247.0 349545.3 312936.0 287166.7 259202.6 242376.1
  year month carrier
                              carrier_name airport
                                                                                                 airport_name
                                               PBI West Palm Beach/Palm Beach, FL: Palm Beach International
1 2004
           1
                  DL Delta Air Lines Inc.
2 2004
           1
                  DL Delta Air Lines Inc.
                                               PDX
                                                                        Portland, OR: Portland International
3 2004
           1
                  DL Delta Air Lines Inc.
                                               PHL
                                                                Philadelphia, PA: Philadelphia International
4 2004
                  DL Delta Air Lines Inc.
                                                               Phoenix, AZ: Phoenix Sky Harbor International
                                               PHX
5 2004
           1
                  DL Delta Air Lines Inc.
                                               PIT
                                                                    Pittsburgh, PA: Pittsburgh International
6 2004
           1
                  DL Delta Air Lines Inc.
                                               PNS
                                                                      Pensacola, FL: Pensacola International
  arr_flights arr_del15 carrier_ct weather_ct nas_ct security_ct
                                                                   late_aircraft_ct arr_cancelled
                                          6.44
                                                51.58
                              21.06
                                                                              45.92
          650
                    126
                                                34.25
                                                                                                30
2
                     61
                              14.09
                                                                 0
                                                                              10.05
          314
                                          2.61
3
          513
                     97
                              27.60
                                          0.42
                                                51.86
                                                                 0
                                                                              17.12
                                                                                                15
4
          334
                     78
                              20.14
                                          2.02
                                                39.39
                                                                 0
                                                                               16.45
                                                                                                 3
5
          217
                     47
                                          0.44
                                                21.89
                                                                 0
                                                                              16.59
                              8.08
                     42
                                          1.06 11.87
                                                                 0
                                                                               18.58
          181
                              10.48
  arr_diverted
               arr_delay carrier_delay weather_delay nas_delay security_delay late_aircraft_delay
                                                   397
                                                            2016
             0
                    5425
                                    881
                                                                             15
                                                                                                2116
2
                                    478
                                                            1365
                                                                              0
             3
                    2801
                                                   239
                                                                                                 719
3
             0
                    4261
                                   1150
                                                   16
                                                            2286
                                                                              0
                                                                                                 809
4
                     3400
                                   1159
                                                            1295
                                                                               0
5
             1
                    1737
                                    350
                                                   28
                                                             522
                                                                              0
                                                                                                 837
6
             0
                    1814
                                    469
                                                  195
                                                             365
                                                                               0
                                                                                                 785
  check_delay predicted_arr_del15
1
            1
2
            0
                                 1
3
            1
                                 1
4
            0
                                 1
5
            0
                                 1
            0
```



4.4 Summary of Project (1 to 2 pages)

• Summarize everything briefly (i.e. in one paragraph you should be able to state your project question, empirical approach, and results).

The project involves analyzing airline data related to flight delays and scheduling, with the goal of identifying patterns and factors that contribute to delays. This analysis may involve examining historical data on flight schedules and delays, as well as real-time data on current flights. The ultimate aim is to use this analysis to improve the accuracy of airline scheduling and reduce delays, which can have a significant impact on customer satisfaction and profitability. The project may also involve developing predictive models that can anticipate potential delays and help airlines take proactive measures to avoid them.

Potential shortcoming of your project and desirable future works.

One potential shortcoming of this project is the availability and quality of data. Airlines may not always provide complete and accurate data on flight delays and scheduling, which can limit the accuracy of any analysis or models developed. Additionally, external factors such as weather, air traffic control, and security issues can also impact flight delays, and it may be difficult to account for all of these factors in the analysis.

Desirable future works could involve exploring ways to address these data limitations and external factors. For example, airlines could be encouraged to provide more complete and standardized data on flight delays and scheduling, and machine learning algorithms could be developed to better account for external factors that impact flight delays. Additionally, more research could be done on the impact of flight delays on customer satisfaction and revenue, to better inform the development of strategies for reducing delays and improving airline performance.

4.5 Bibliography (1 page)

4.6 Appendix: SAS or R command and Data Files

Include all SAS or R commands used to generate the output. Codes and Data need to be included in separate files. Make sure all submitted SAS or R codes without any errors as a .txt file. *There will be a very high penalty if they are not working with errors or not completed*.

- Is shared with the report.