

DS 600 DATA MINING PROJECT- AIRLINE INDUSTRY

BY

BHAGYASHREE CHAVAN

PROBLEM STATEMENT

Flight delayed can be huge problem to airline companies and in people's daily life. Therefore **The study focus on which are the best airlines has most reliable carriers.**

Data Set

```
1 df.head()
```

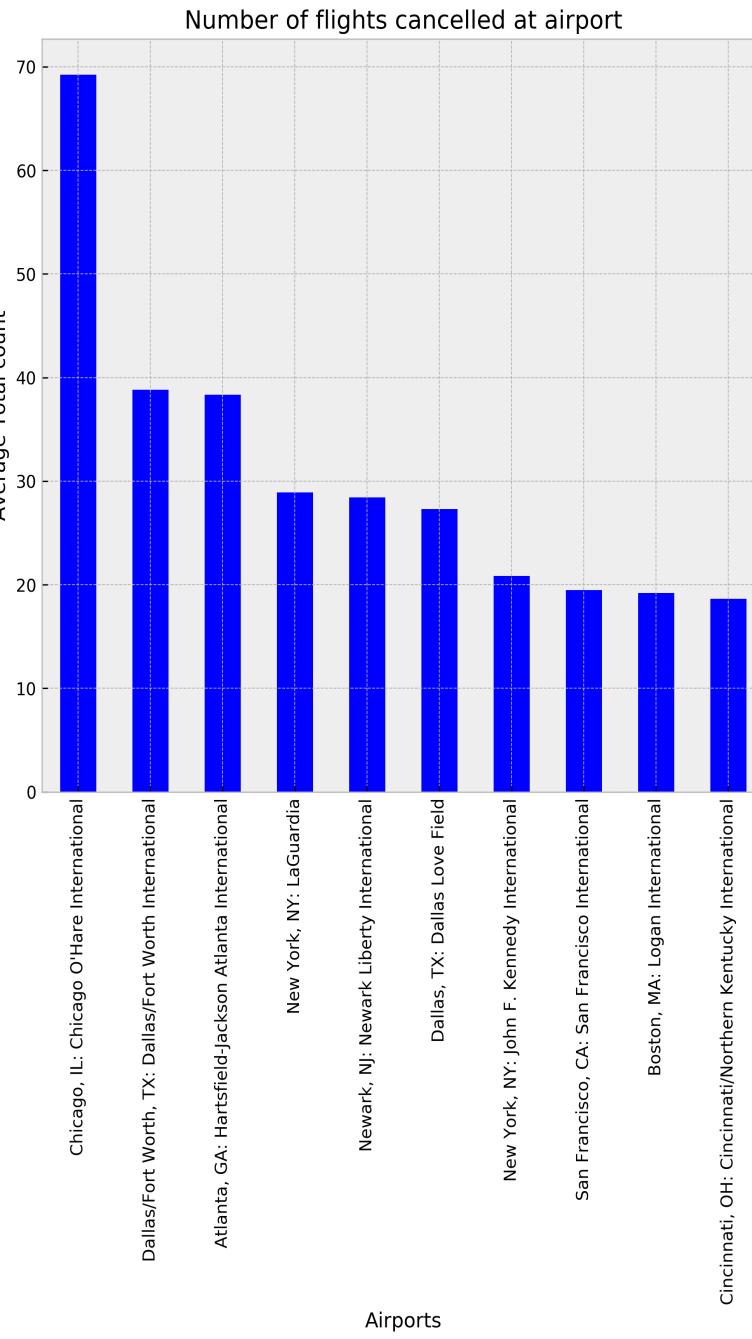
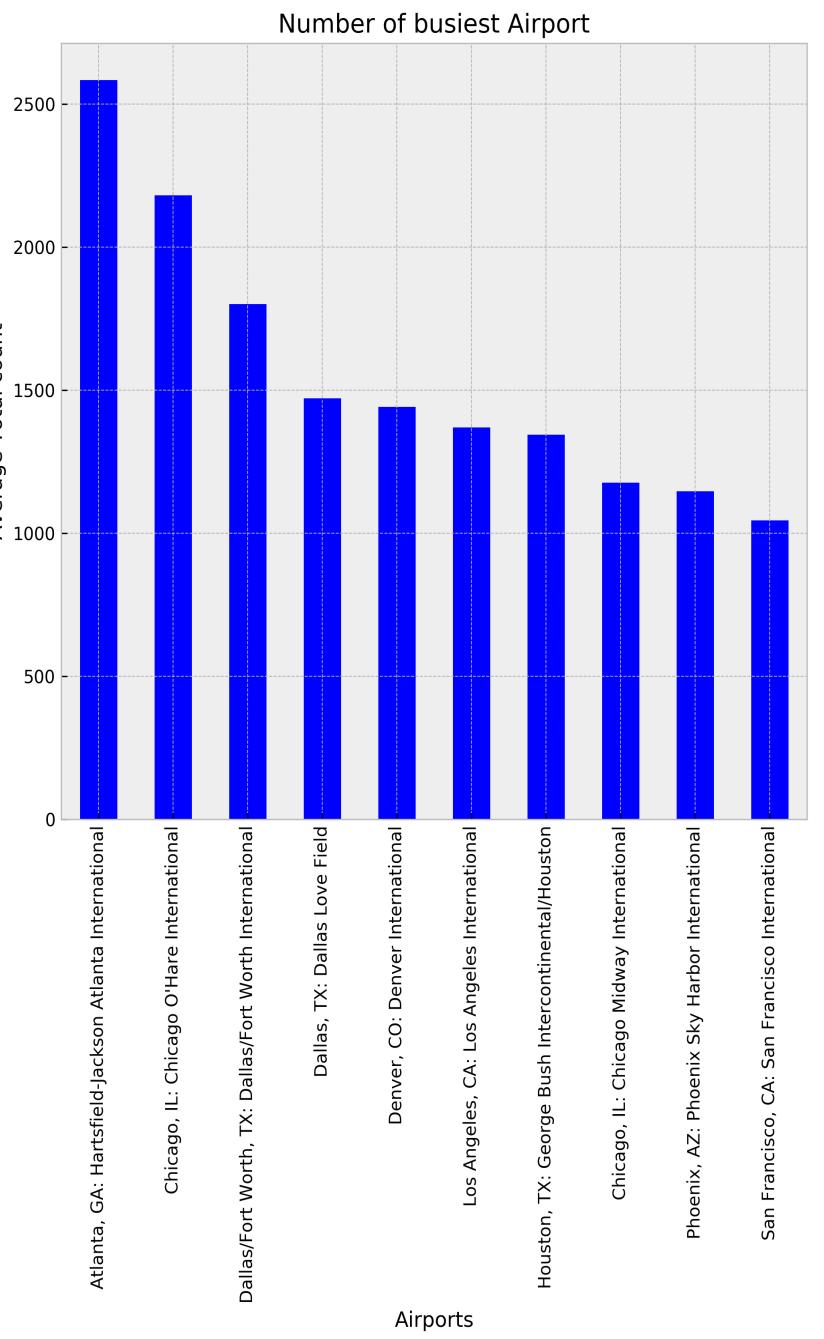
	year	month	carrier	carrier_name	airport	airport_name	arr_flights	arr_del15	carrier_ct	weather_ct	...	security_ct	late_aircraft_ct	arr_cancelled	arr_div
0	2004	1	DL	Delta Air Lines Inc.	PBI	West Palm Beach/Palm Beach, FL: Palm Beach Int...	650.0	126.0	21.06	6.44	...	1.0	45.92	4.0	
1	2004	1	DL	Delta Air Lines Inc.	PDX	Portland, OR: Portland International	314.0	61.0	14.09	2.61	...	0.0	10.05	30.0	
2	2004	1	DL	Delta Air Lines Inc.	PHL	Philadelphia, PA: Philadelphia International	513.0	97.0	27.60	0.42	...	0.0	17.12	15.0	
3	2004	1	DL	Delta Air Lines Inc.	PHX	Phoenix, AZ: Phoenix Sky Harbor International	334.0	78.0	20.14	2.02	...	0.0	16.45	3.0	
4	2004	1	DL	Delta Air Lines Inc.	PIT	Pittsburgh, PA: Pittsburgh International	217.0	47.0	8.08	0.44	...	0.0	16.59	4.0	

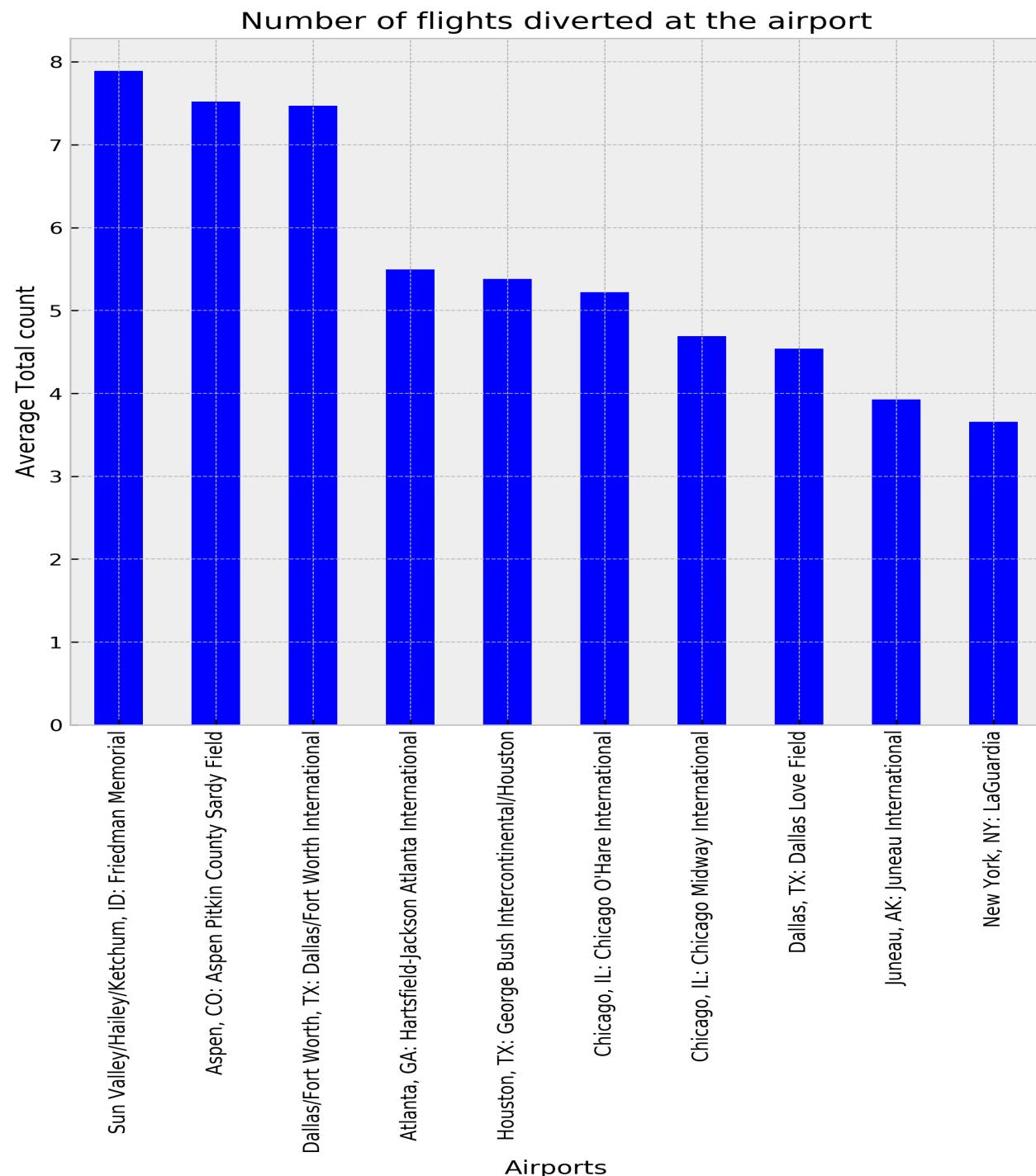
5 rows × 21 columns

Attributes About Data

1. Source is United States Department of Transportation
2. This is data set which has more than ~200,000 rows and 20 features.
3. The data has contained date, airport names, arrival flights count, delayed flights counts, late aircraft counts and so on.
4. Data set contain categorical information, and numeric information.

Exploratory Data Analysis





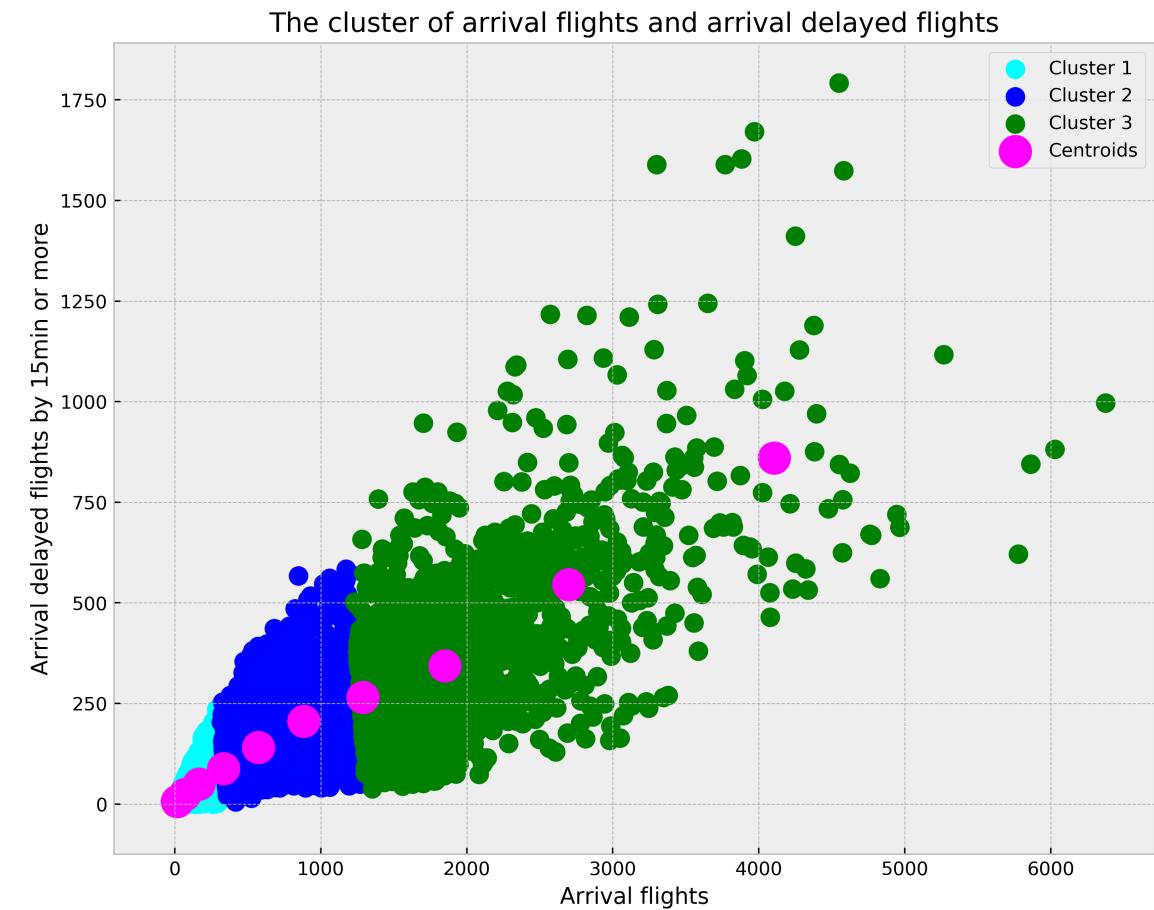
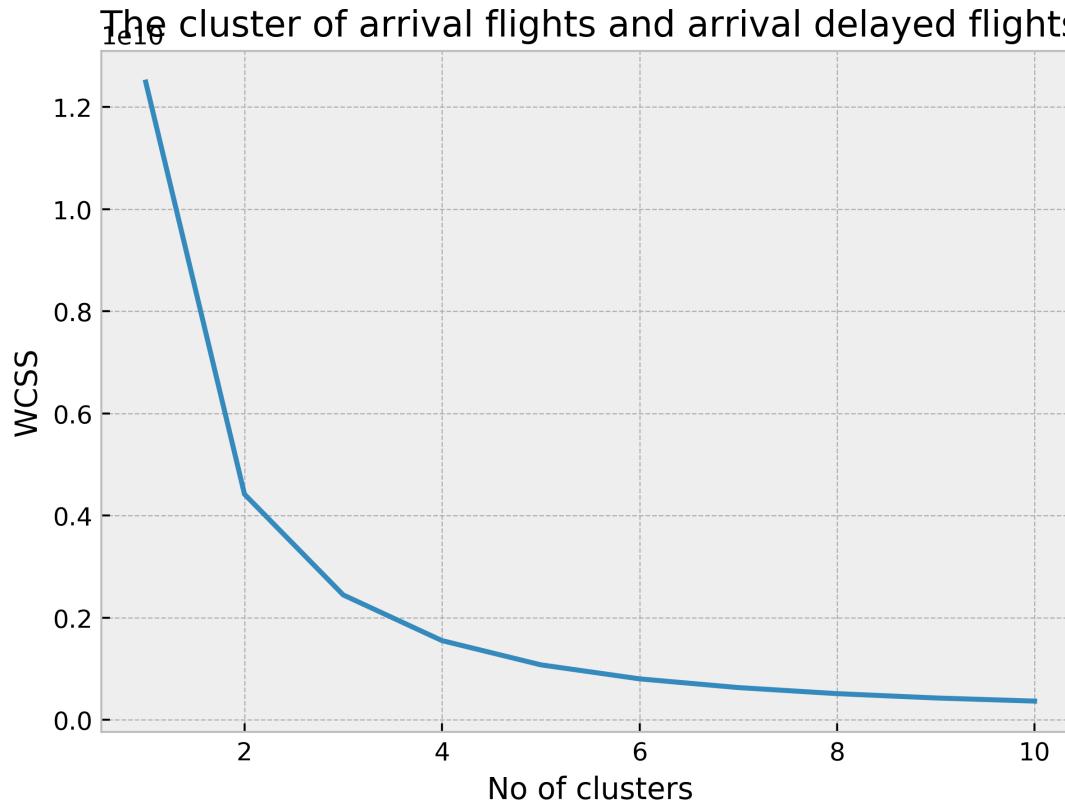
Exploratory Data Analysis

Atlanta is the busiest airport amongst top 10 airport in United States.

At **Chicago O'Hare international** airport has been most flights cancelled.

At **Sun Valley, Aspen and Dallas fort worth international** airport most flights has been diverted

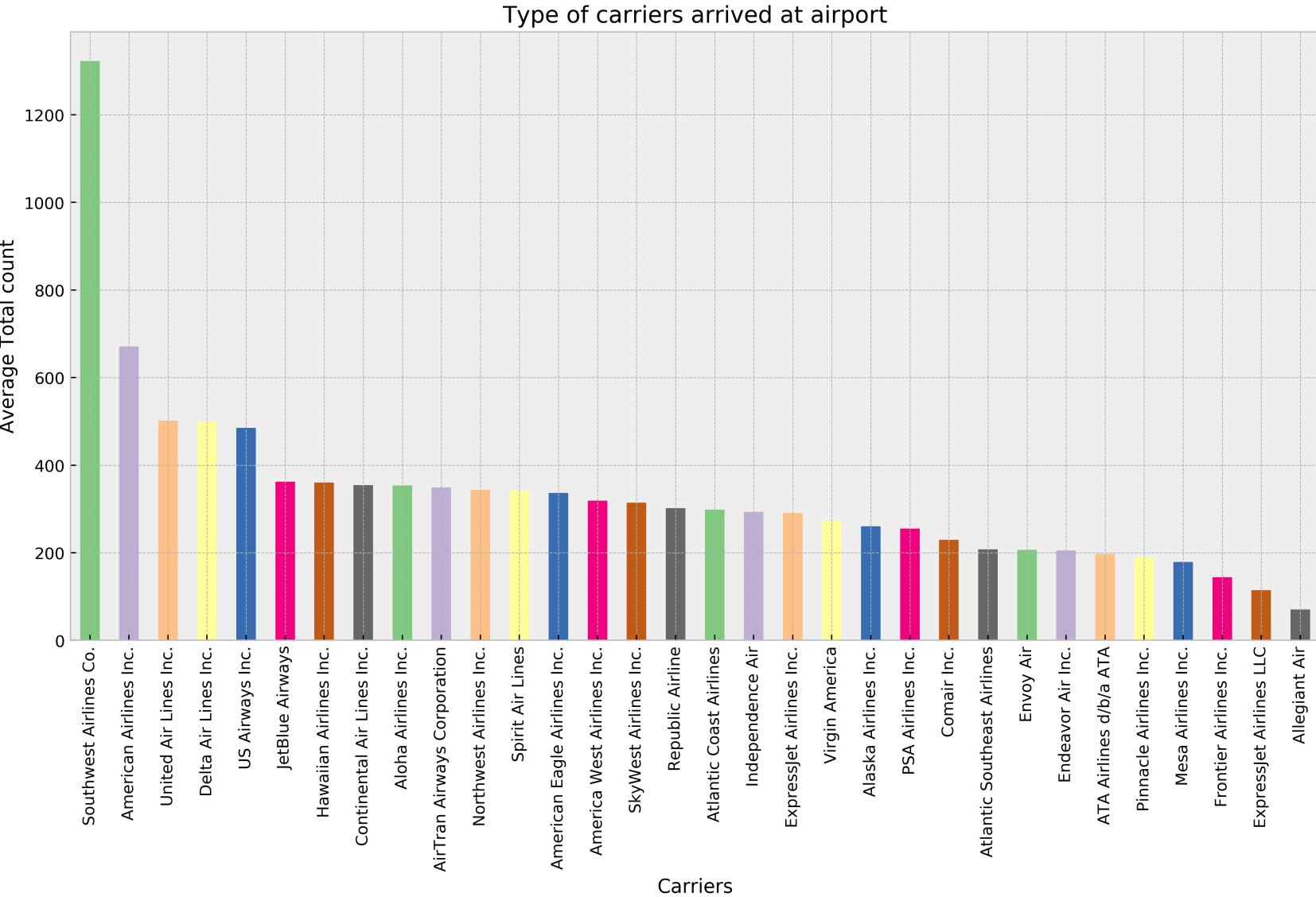
Cluster Method



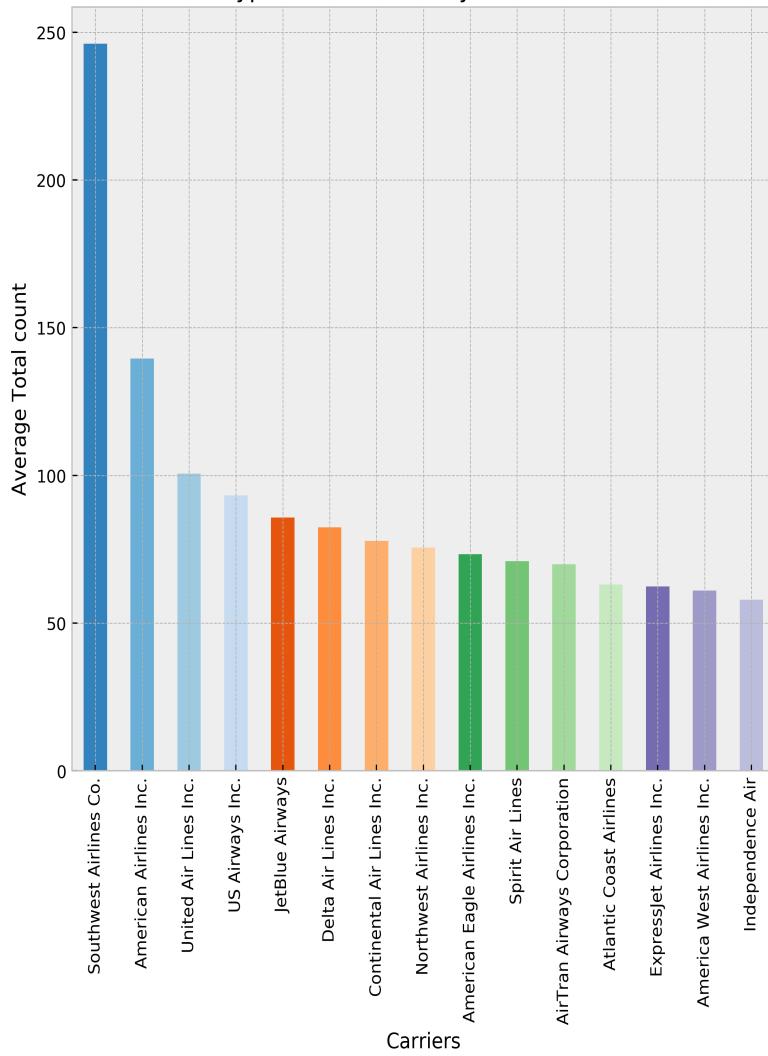
Data Insights

Out of 32 carriers I have selected top 15 busiest carriers for Analysis.

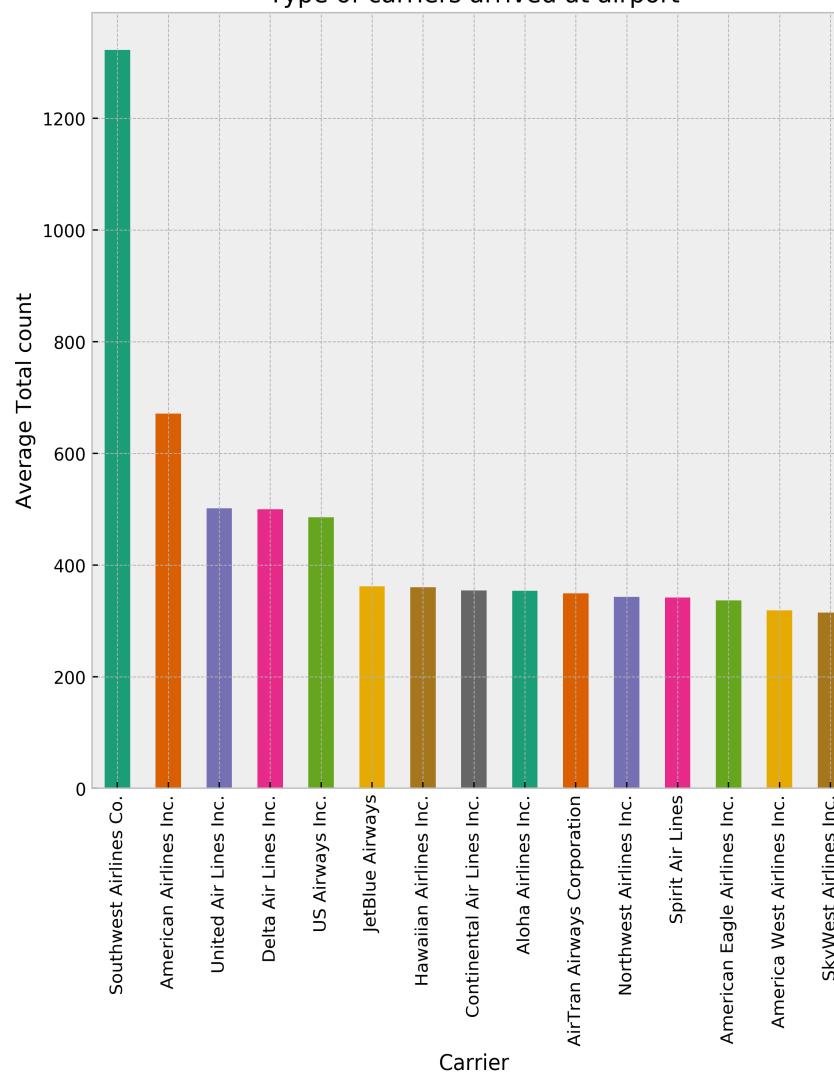
Southwest Airline company has a greater number of arrivals compared to its competitors.



Type of carriers Delayed >= 15min

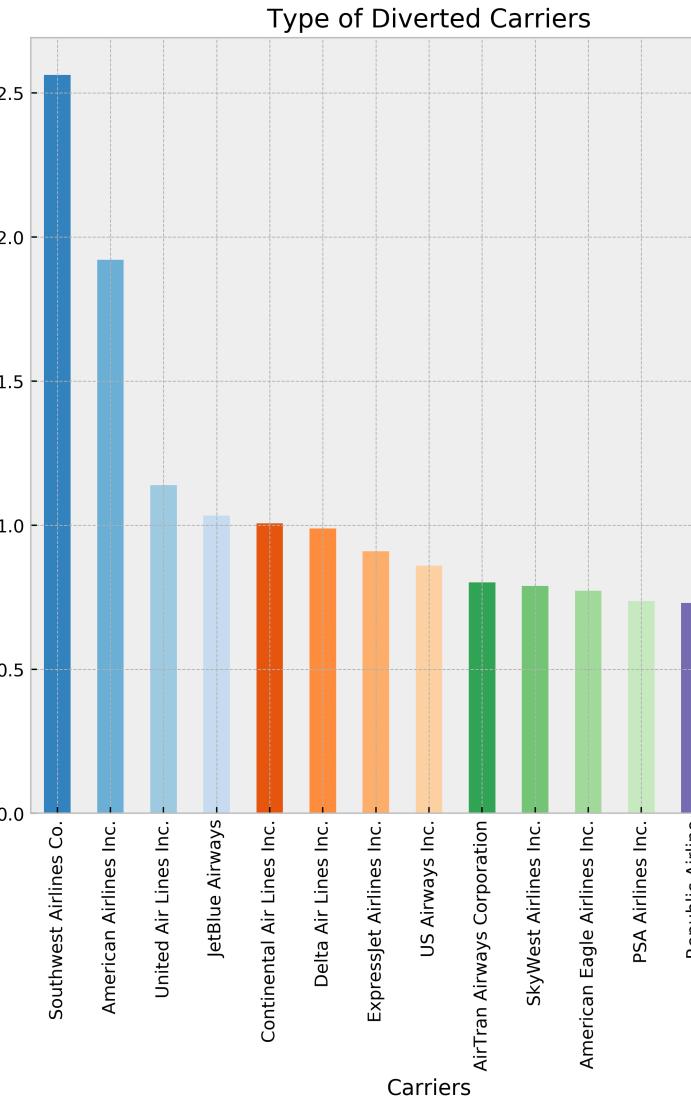
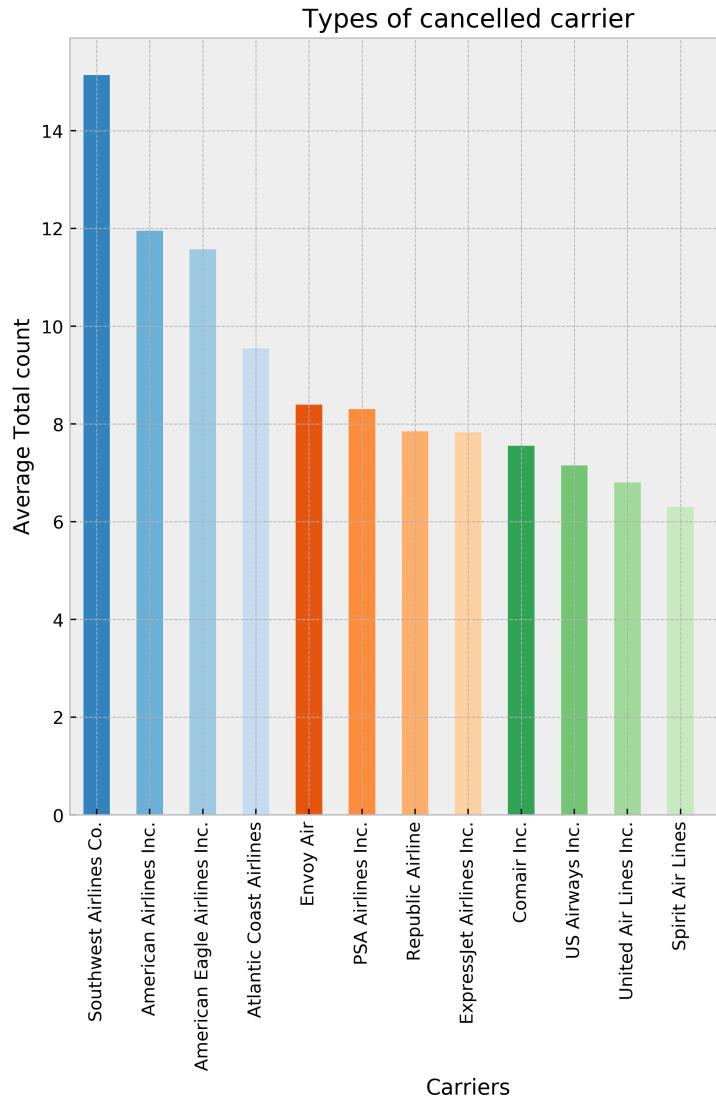


Type of carriers arrived at airport



Further Analysis

In this part of analysis I observed southwest airlines has most number of delayed compared to its competitors because they have most number of flying and arrivals count. Whereas only flights are listed above delayed by greater than equal to 15 minutes.

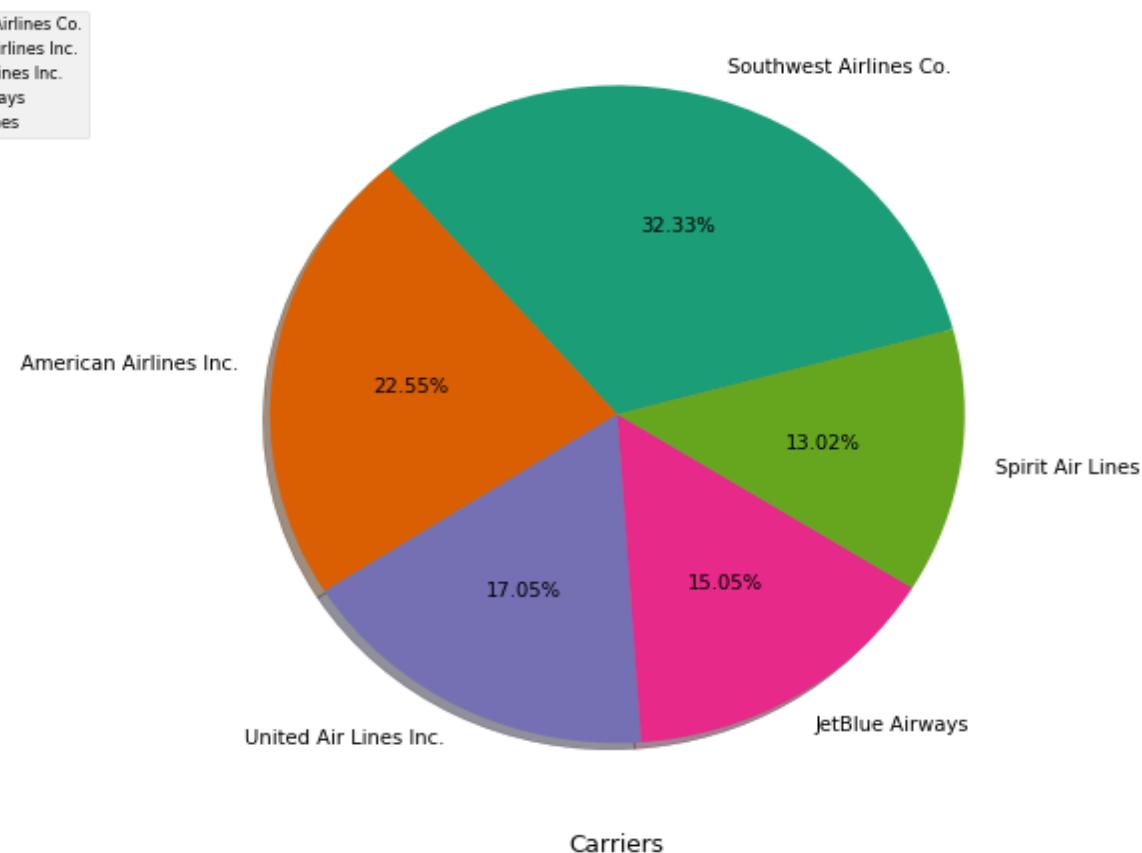


Analysis

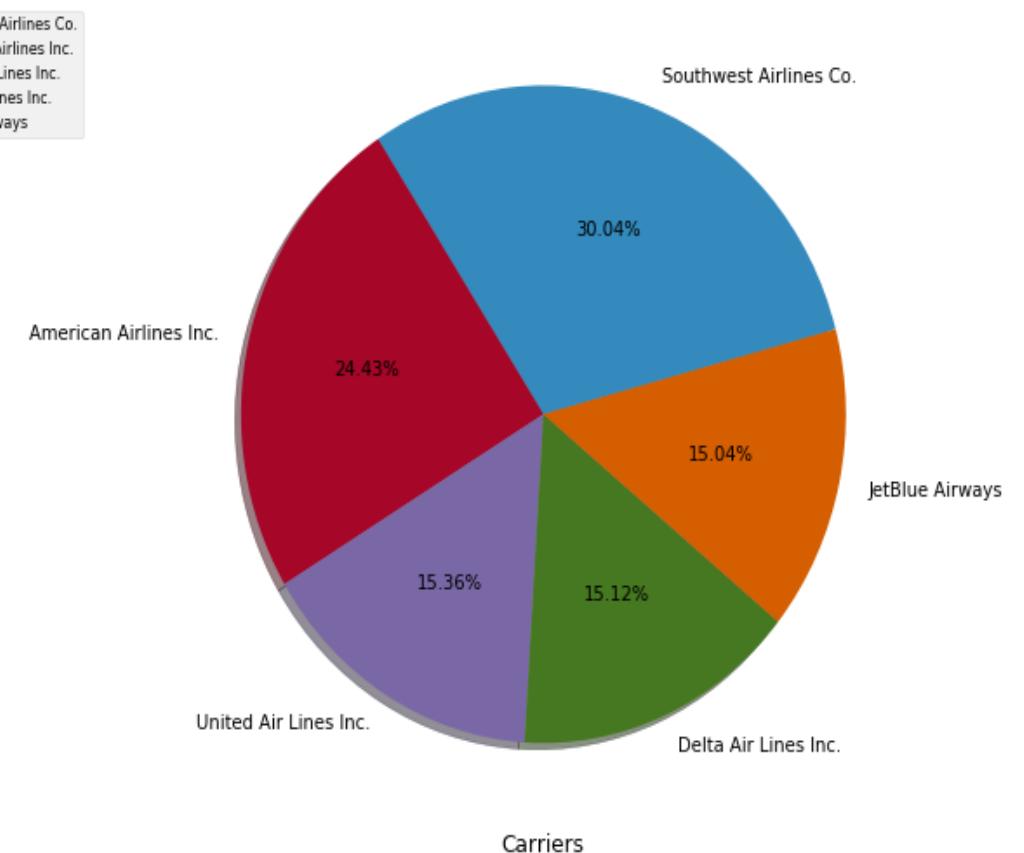
So far, we have seen arrival count, delayed count, cancelled count and diverted count in which we can see usual trend for top 2 carrier those are not reliable, whereas trend form Jet Blue to Spirit Airlines are steady.

Delayed Causes

Percentage of delayed due to arrival delayed



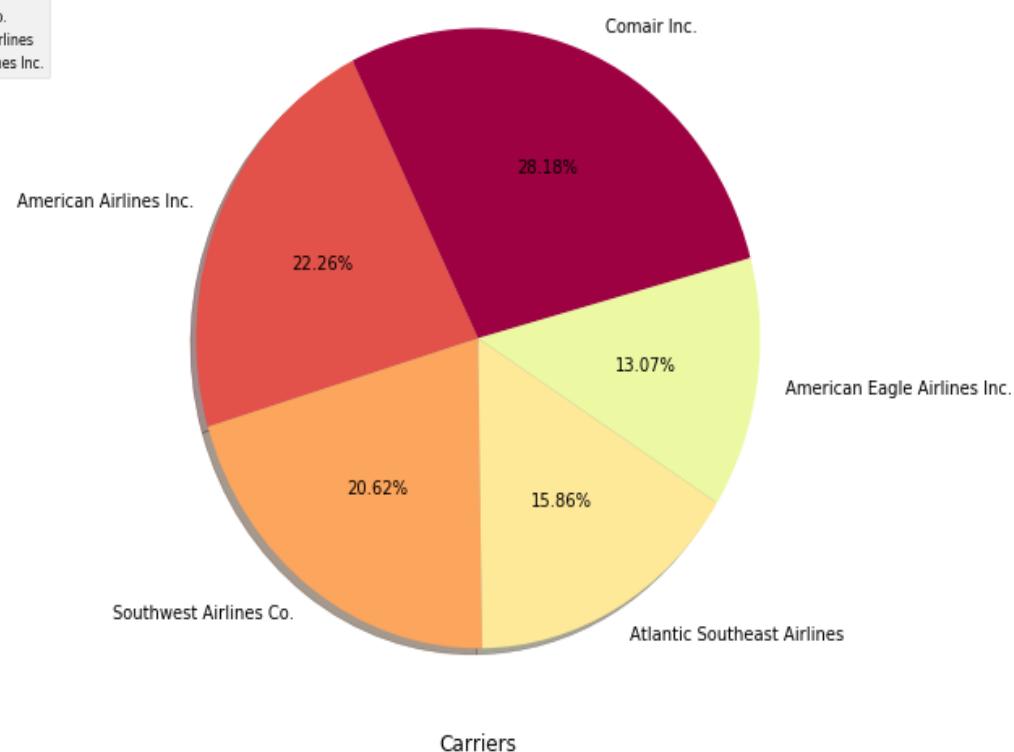
Percentage of delayed due to Air carrier delayed



Delayed Causes

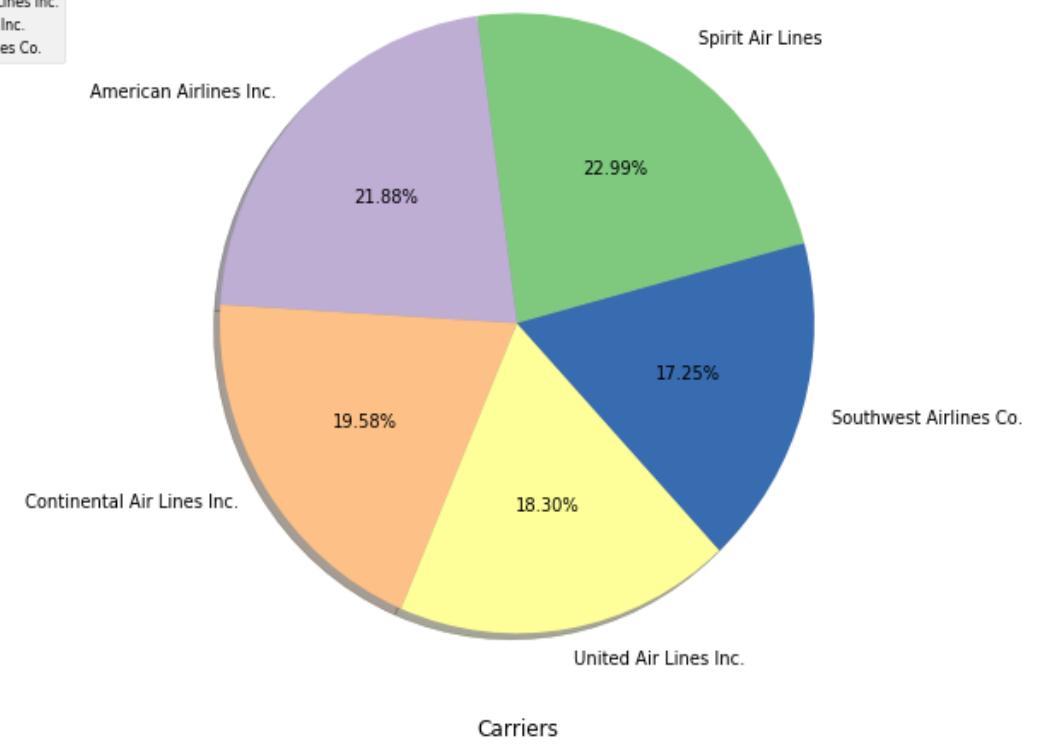
Percentage of delayed due to Weather delayed

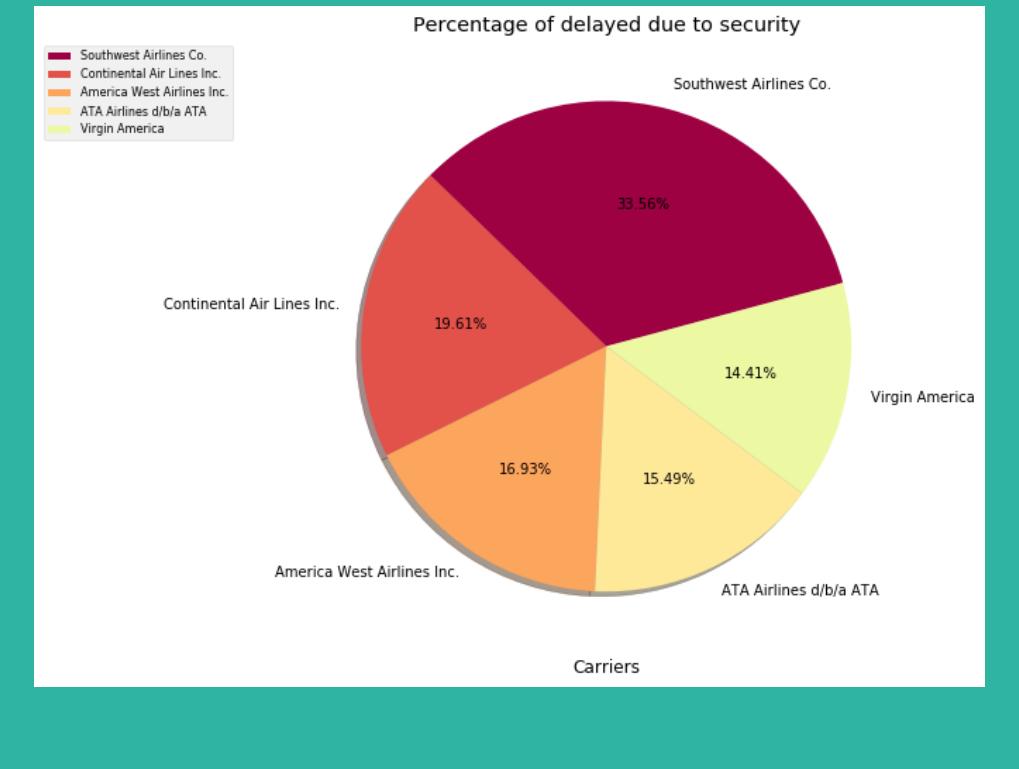
- Comair Inc.
- American Airlines Inc.
- Southwest Airlines Co.
- Atlantic Southeast Airlines
- American Eagle Airlines Inc.



Percentage of delayed due to National Aviation System delayed

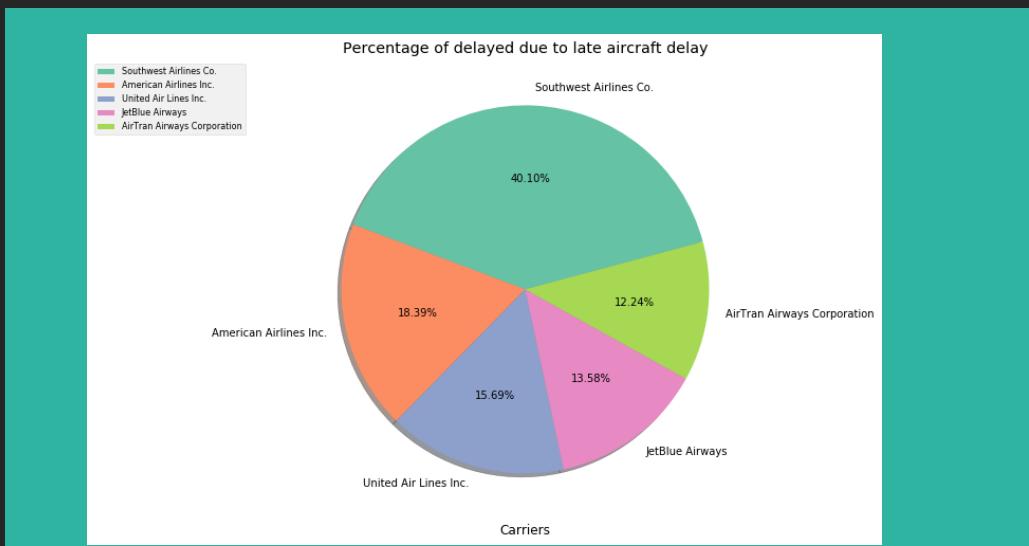
- Spirit Air Lines
- American Airlines Inc.
- Continental Air Lines Inc.
- United Air Lines Inc.
- Southwest Airlines Co.





Delayed Causes

Mainly there are 6 reasons for flights delayed those are listed and shown above in form of pie charts. In analysis I observed Southwest airline and American Airline are top 2 airlines with most number of delays.



Conclusion And Suggestions



Based on the data, Analysis and visualization I concluded that Southwest Airlines and American airlines are not reliable airlines because of most number of delays.



Whereas **Jet blue, Spirit airline, SkyWest airline** and other few airlines are reliable to choose for flying based on their arrival count, delayed count, cancelled count, diverted count and their delayed reasons.



In suggestion for the airlines are getting delayed like **Southwest airline, American airline** etc. should improve their performance based on delayed causes and increase quality of service and thereby gain customer satisfaction.

Achieved Result

Airline Company are using the latest technologies and data mining techniques to improve their organization and using the real-time data in order to provide their customers with individual offers, create better products and service, and increase customer satisfaction.

For instance Southwest airline reported \$22.4 billion revenue last year by using data mining and latest technology.



Thank you
