**Airlines and Airports  
Punctuality Performance Analysis**

Parag Siddamsettiwar Abhishek Upadhayay

Data Science Data Science

Indiana University Indiana University

[psiddam@iu.edu](mailto:psiddam@iu.edu) [abupad@iu.edu](mailto:abupad@iu.edu)

**Abstract**

Flight data visual analysis of US air traffic to compare carriers and airports based on punctuality, delays, cancelations and diversions.

**Terms**

IATA – International Air Transport Association

BTS – Bureau of Transportation Statistics

DOT – Department of Transportation

**1 Introduction**

In today’s busy world, time is a very precious commodity. Whether you are travelling for work or leisure, mode of transportation plays a very important role in your trip’s success. Flights are the most convenient and fastest way to travel. In fact, frequent travelers prefer living in a city well connected by air transport. Airlines and airports are becoming critical necessity for every individual, as we are moving towards a more collaborative global economy.

**1.1 Motivation**

More and more people are taking flights and according to a report by IATA, around 8.2 billion people will be travelling by flights by 2037 (IATA, 2018). This trend will continue to grow due to growing economy and a need to save time by travelling using the fastest mode of Transportation.

Flight delays and cancellations are a major cost escalator for businesses, passengers and Airlines. In 2010, passengers lost as much as $16.7 billion due to schedule buffer, delayed flights, flight cancellations and missed connections (Airlines for America, 2010).

Whether you are going on a business trip or going on a vacation, planning and booking a flight that will make the entire trip as planned and pleasant is very important. The hassle to adjust the entire trip or adjust the bookings takes a toll on everyone including passengers, airlines and businesses. According to a research, an average Briton wastes 16 days of their holiday time waiting for their flight to take off (Andrew Hough, 2010).

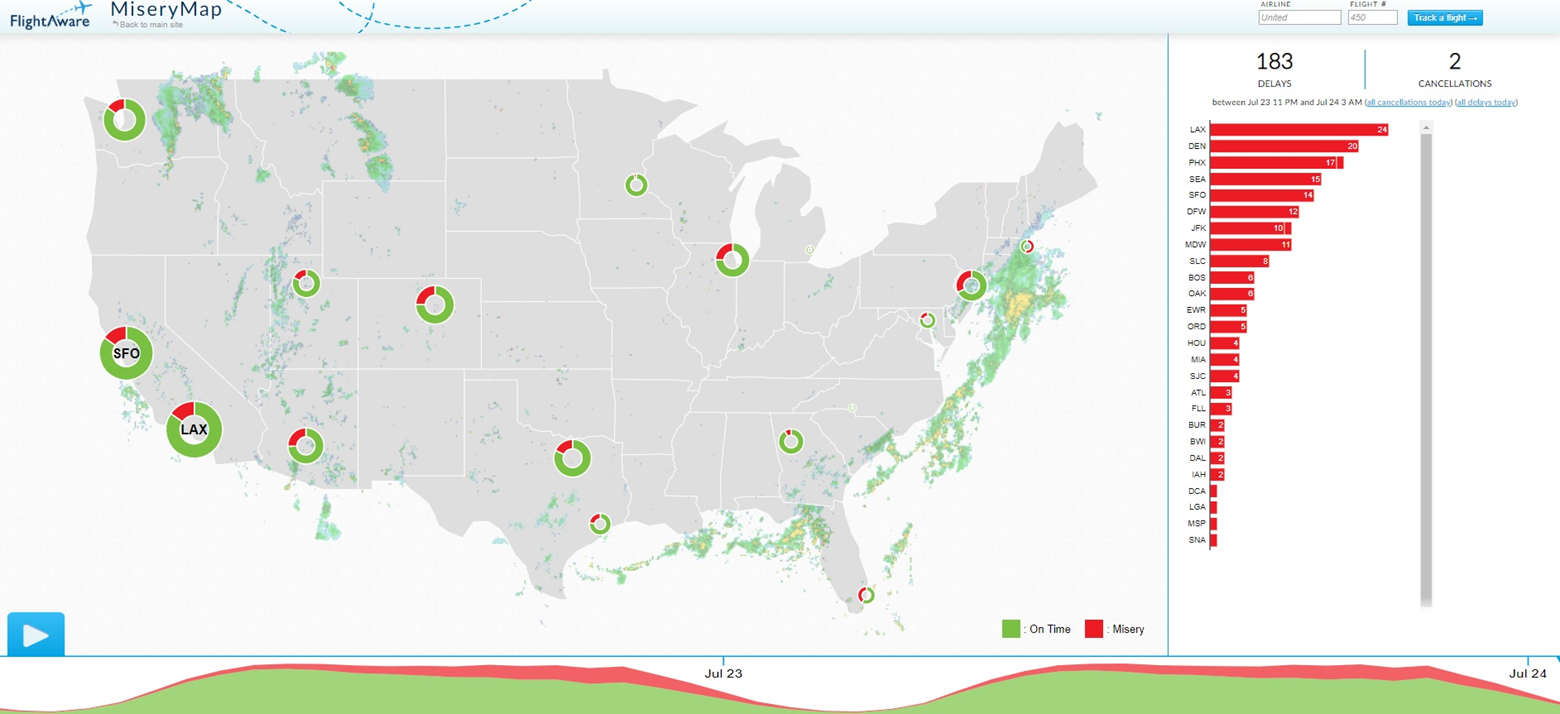
It will really help if we can analyze and visualize the flight data to get a better understanding of Airlines and Airports to make an informed choice before finalizing our travel plans

**1.2 Existing work**

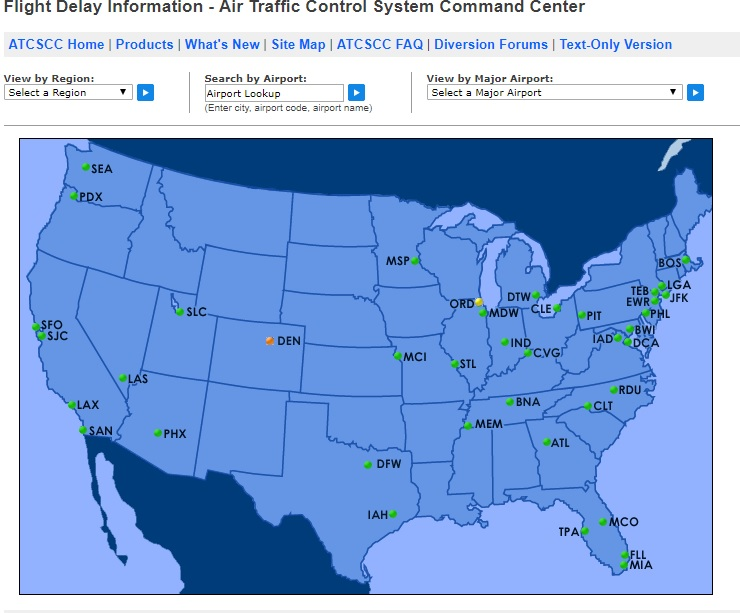
Even though there are tons of travel websites to book flights, there are only a handful websites that provide an analysis of punctuality of airlines and airports.

Even the ones that are available are either on current data (real time or past few hours) and not an analysis on historical data for a proper analysis of trends or behavior.

*FlightAware’s flight delay visualization (FlightAware, 2019)*  
This visualization shows some big Airports only, no information is available for Airlines and no option to choose other airports e.g. Indianapolis, Cincinnati etc. More importantly, this is based on current data (last few hours) and not an analysis on a historical dataset.



*Flight Delay Information - Air Traffic Control System Command Center (FAA, 2019)*  
This visualization also shows airports and it’s also based on current data (last few hours) and not an analysis on a historical dataset.  
Its interactive and lets user select airport or flights and provides current situation for that.



There is not enough detail available that can help us in making decision for future travels from airports and airlines based on past trends and airlines behavior.

Even though a flight may show on time on these current data visualizations, there is a possibility that the flights of an airline are regularly late due to unavailability of aircrafts or pilots. We may want to avoid taking that Airline

**1.3 Contribution (To be done)**

**2 Process**

Different countries have different airline authorities that control airline traffic for their respective countries. Usually these authorities hosts the data of these airlines including delay information.

For United States, U.S. Department of Transportation, a government agency, maintains and make this data public. Thanks to Bureau of Transportation Statistics, which is part of DOT, we have found the source for reliable, robust and continuous data of flights including delays, cancellation and diversion information.

These government agencies also regularly publish data via their open data Data.gov platform

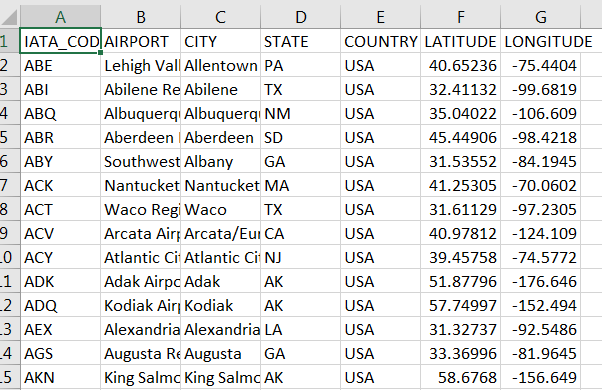
We will be pulling on time performance dataset of all the flights of United States. Such airline dataset is generally associated to IATA\_CODE which is nothing but a unique code given to every airline and airport. It is used to across databases for easy cross-reference.

Let us look at couple of basic datasets:

**Airlines:** This is needed to map IATA\_CODE with actual airline names



Airports: This is needed to map IATA\_CODE to actual names of airports.



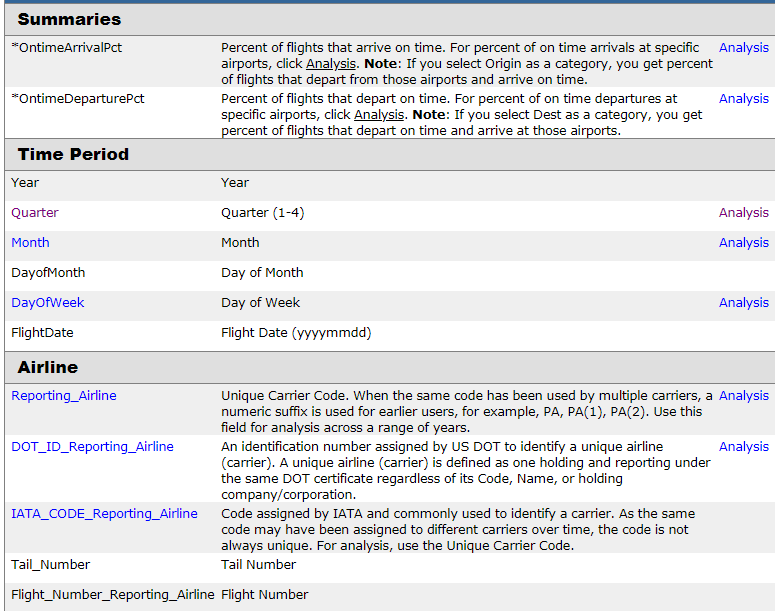
On-time performance dataset of USA domestic flights. This dataset is regularly updated and published which can be downloaded from following URL: <https://www.transtats.bts.gov/Tables.asp?DB_ID=120&DB_Name=Airline%20On-Time%20Performance%20Data&DB_Short_Name=On-Time>

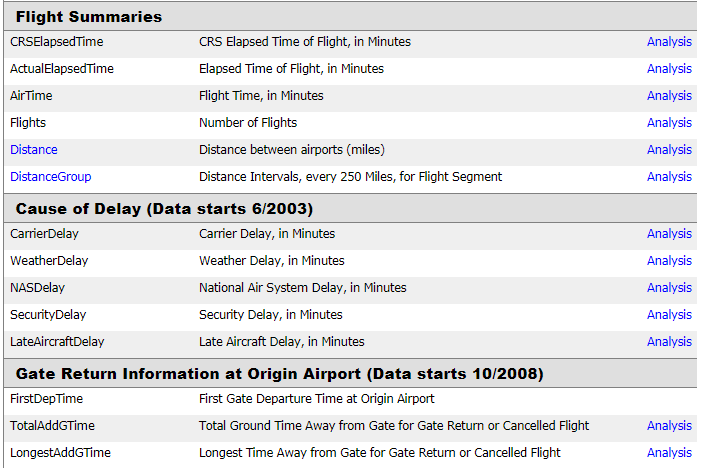
This is a large dataset with millions of records depending on how much historical data we download.

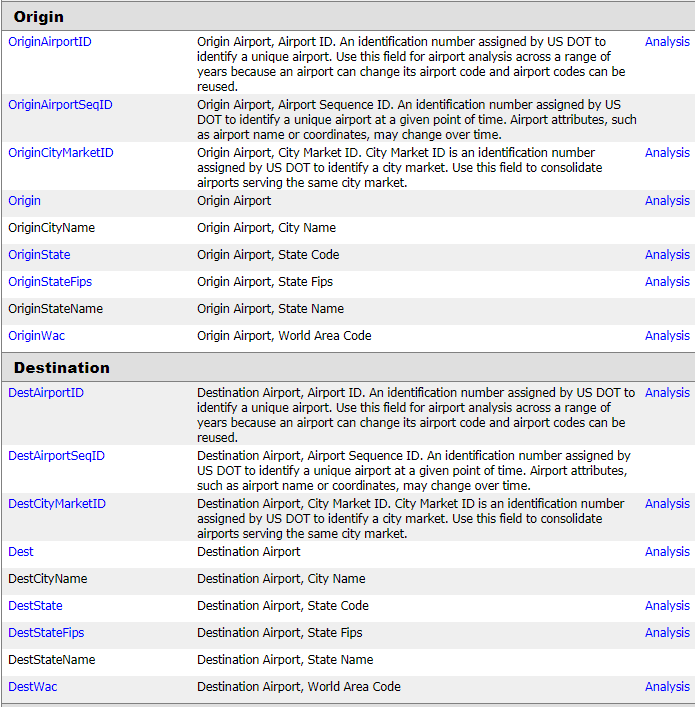
For showing high-level information, header is shown below to list all the dataset column names:

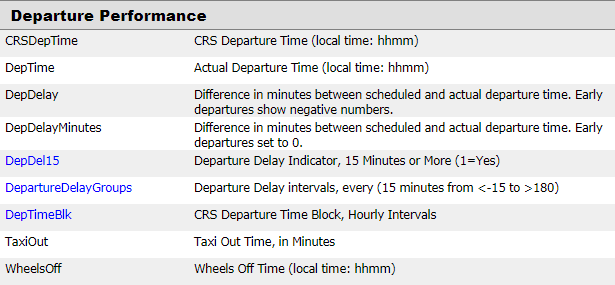
Index(['YEAR', 'MONTH', 'DAY', 'DAY\_OF\_WEEK', 'AIRLINE\_CODE', 'FLIGHT\_NUMBER','TAIL\_NUMBER', 'ORIGIN\_AIRPORT', 'DESTINATION\_AIRPORT','SCHEDULED\_DEPARTURE', 'DEPARTURE\_TIME', 'DEPARTURE\_DELAY', 'TAXI\_OUT','WHEELS\_OFF', 'SCHEDULED\_TIME', 'ELAPSED\_TIME', 'AIR\_TIME', 'DISTANCE','WHEELS\_ON', 'TAXI\_IN', 'SCHEDULED\_ARRIVAL', 'ARRIVAL\_TIME','ARRIVAL\_DELAY', 'DIVERTED', 'CANCELLED', 'CANCELLATION\_REASON',’AIR\_SYSTEM\_DELAY', 'SECURITY\_DELAY', 'AIRLINE\_DELAY','LATE\_AIRCRAFT\_DELAY', 'WEATHER\_DELAY', 'DATE', 'IATA\_CODE', 'AIRLINE','AIRPORT', 'CITY', 'STATE', 'COUNTRY', 'LATITUDE', 'LONGITUDE'],dtype='object')

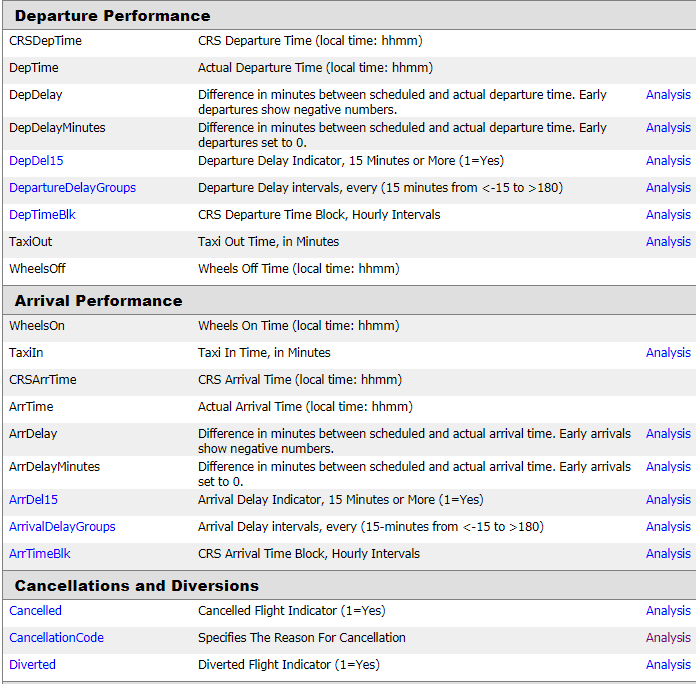
Additional details of these can be found on the transtats website which is as follows:

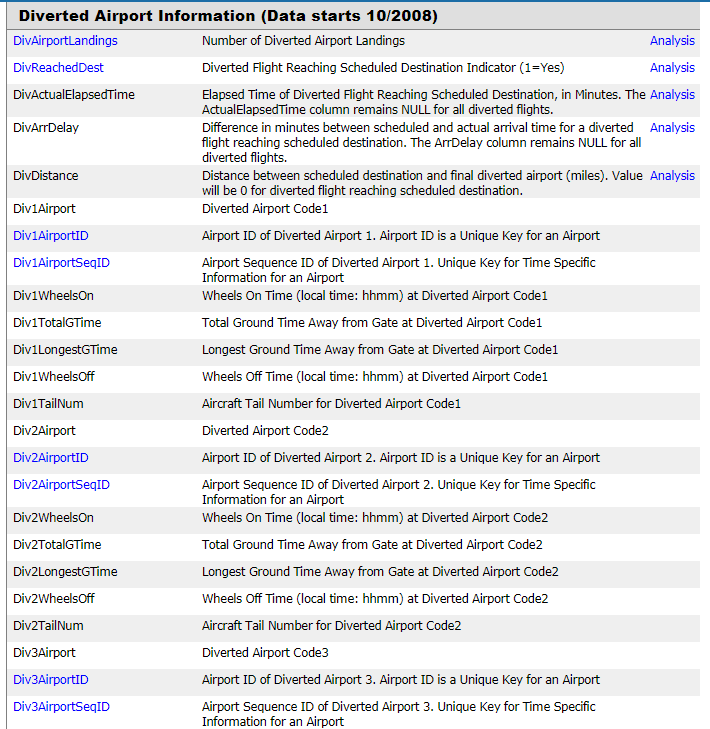














**3 Results**

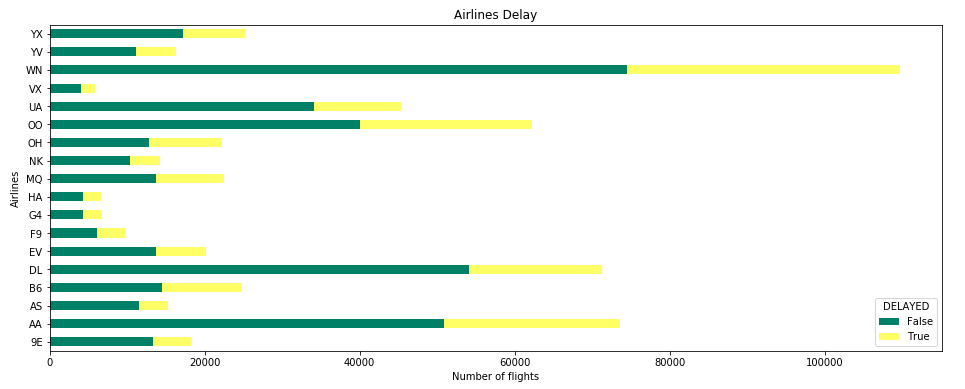
After carefully studying the data columns available to us, we have decided on downloading few of these that makes more sense in current analysis. We downloaded 10 years of data and tried to load it into python for doing exploratory analysis. As this data se very large and we have limited memory available on our local computers, we ended up noticing extremely slow responsive behavior of python program. It was taking way too more time to load data, combine it. We wanted to focus our work towards creating visualization and not on how much data we can bring in so we decided to restrict our dataset to 1 year of flights data. After using only 1 year of data, we were able to load it comparatively quickly.

Top 20 delayed flights

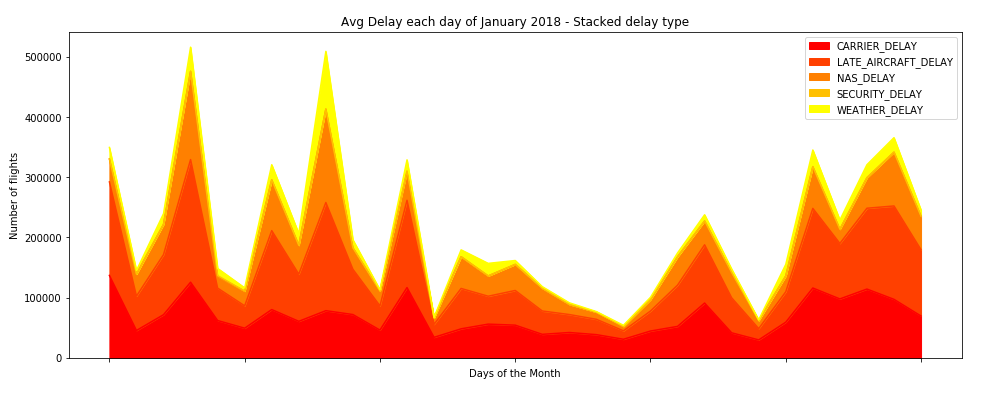


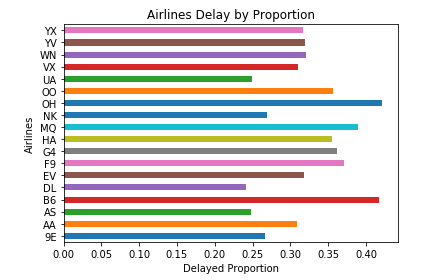
The most delayed flight was almost 2000 minutes delayed, almost a day

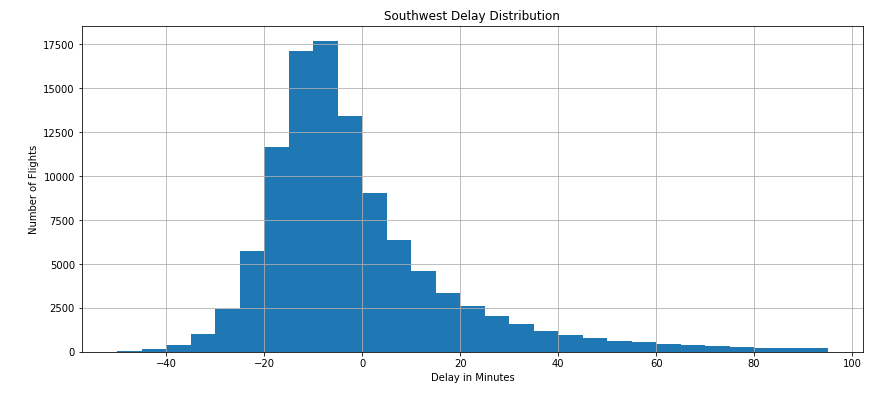
On time vs delayed ration of flights per airline

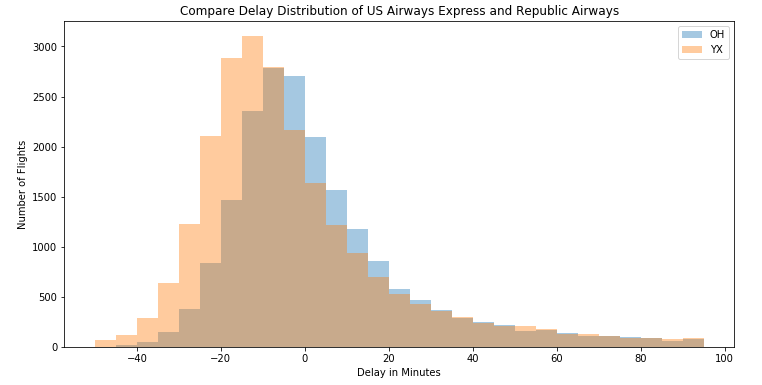


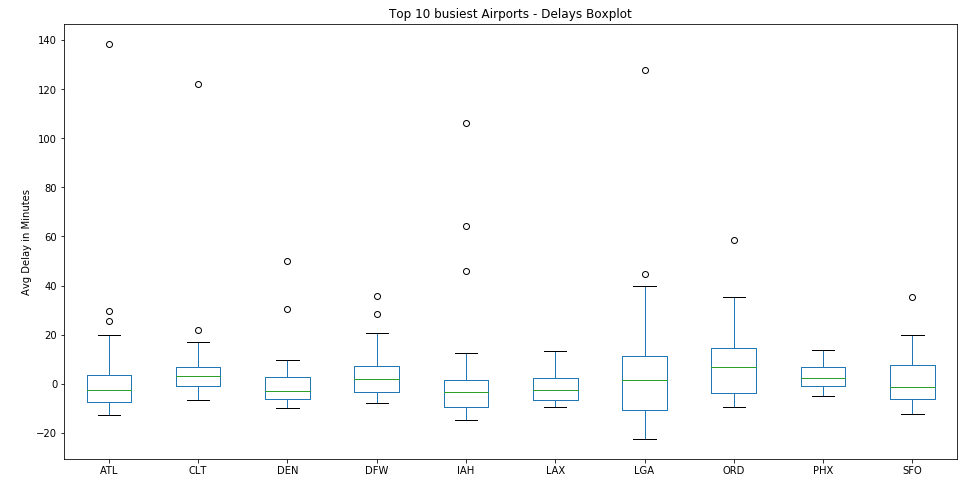
Flight delay reason heatmap

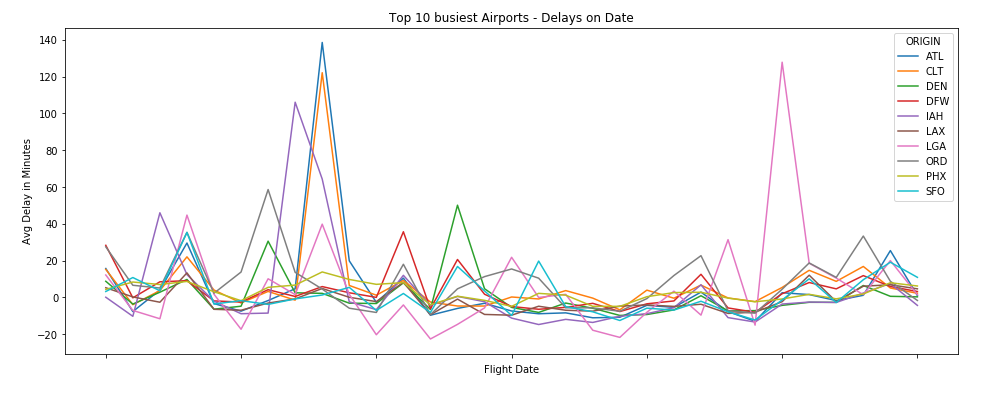


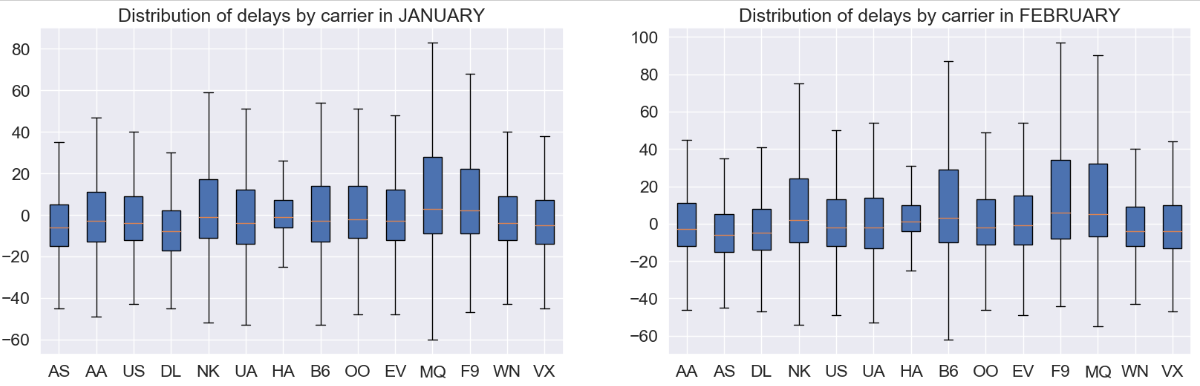


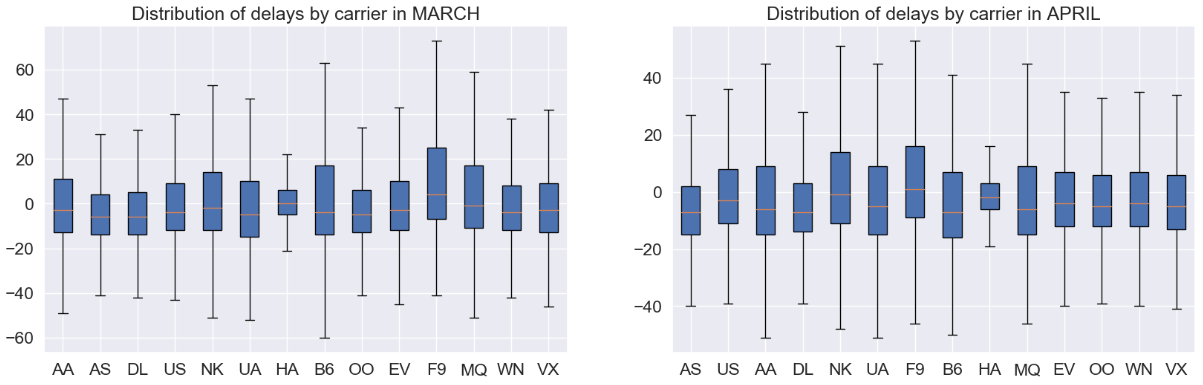


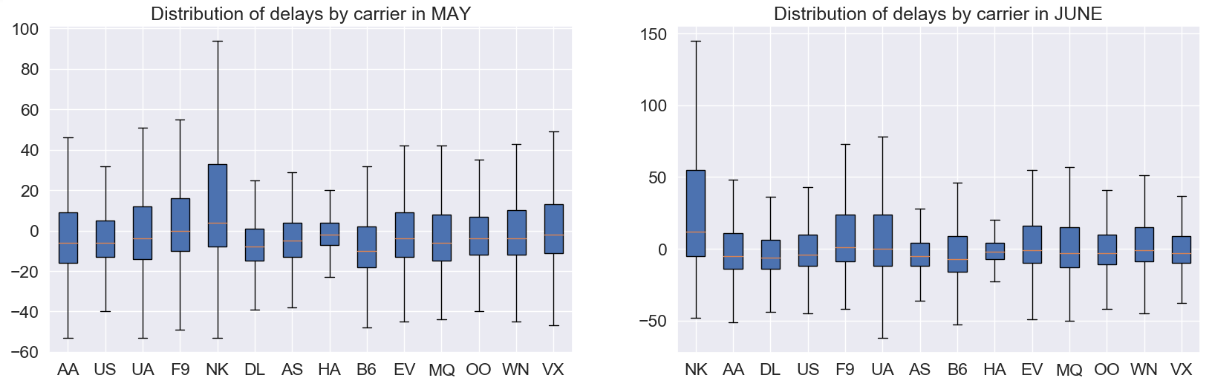


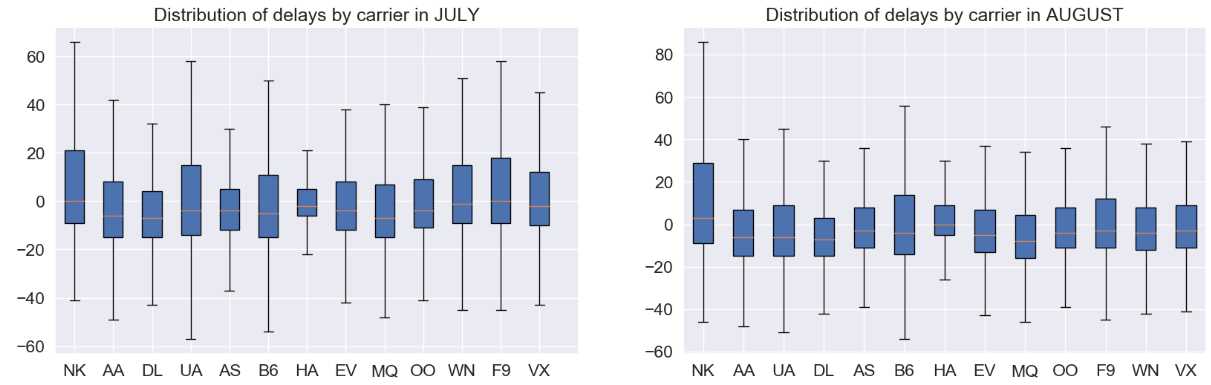


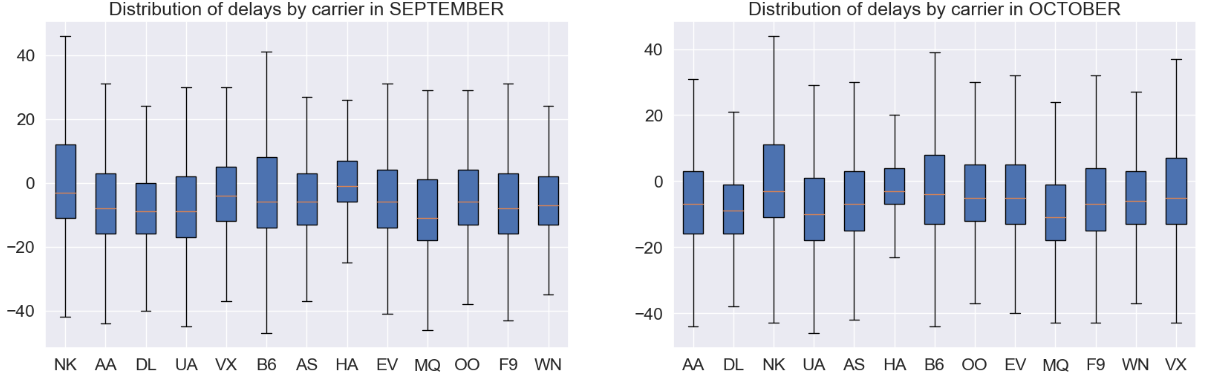


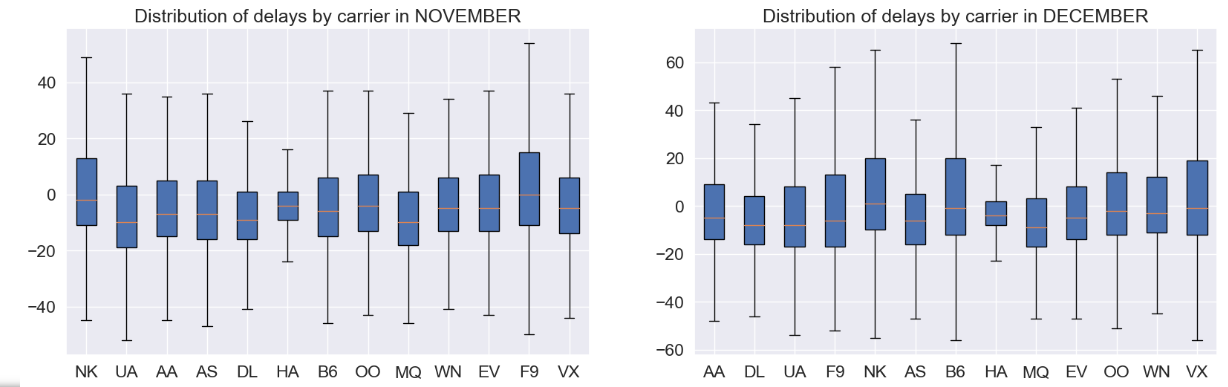




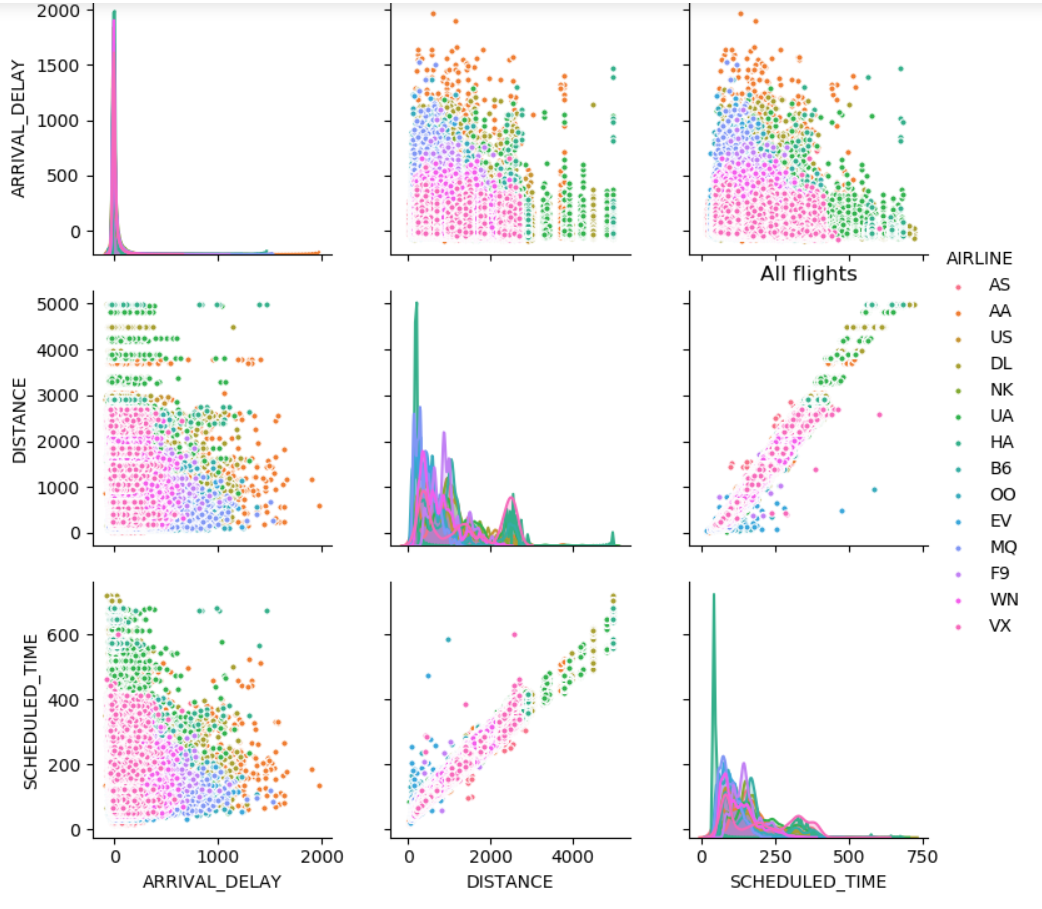


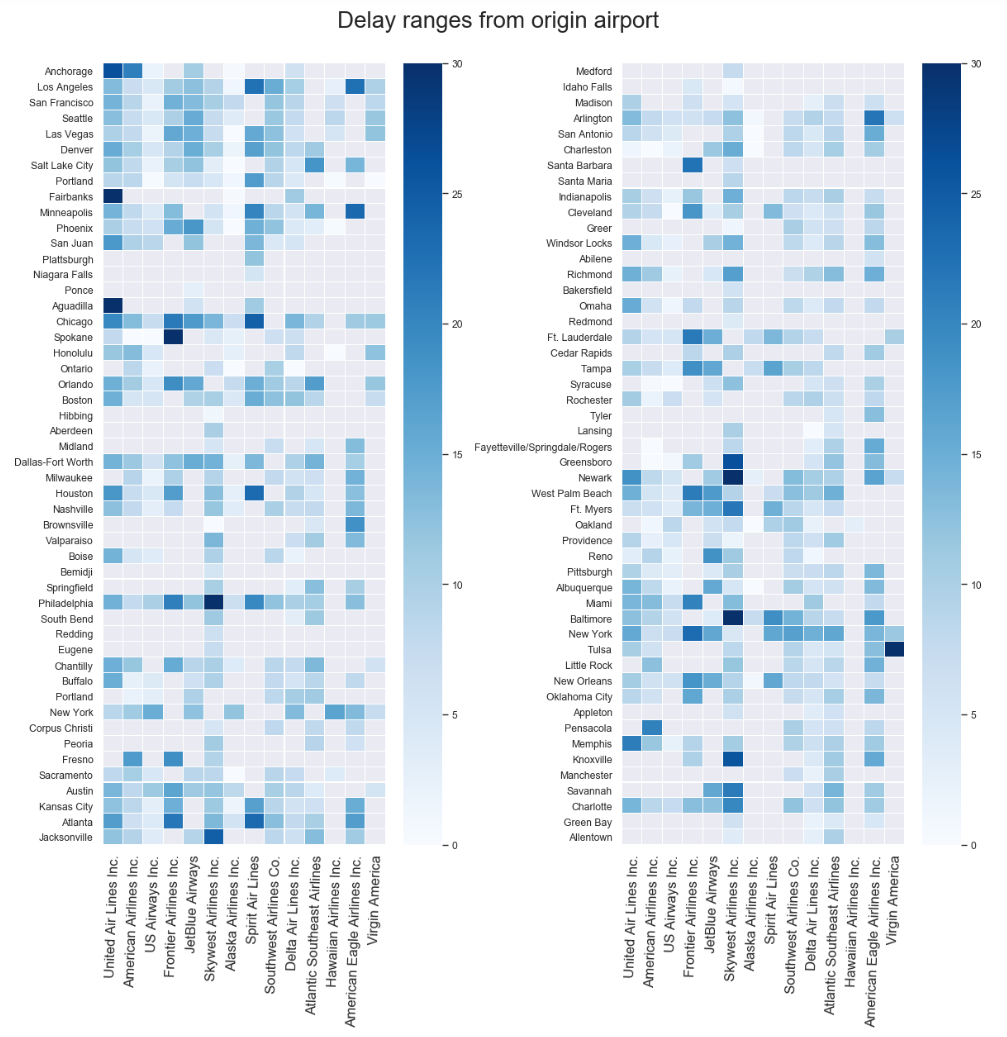






Compairing airline, arrival delay, distance and scheduled time using pairplot

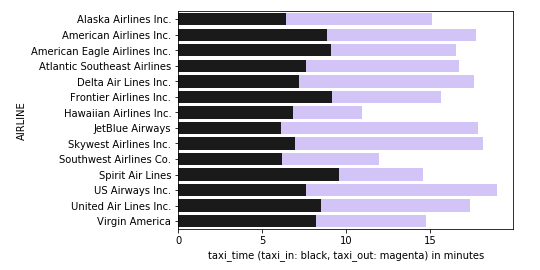




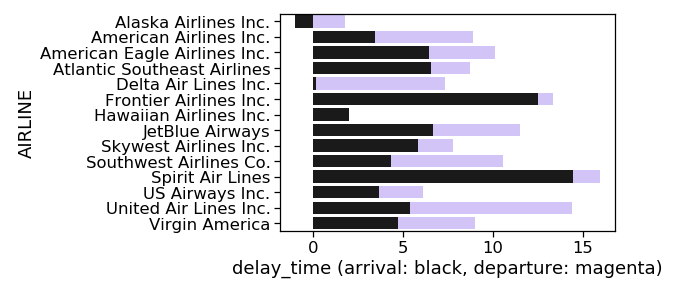
Calculating cancellation rate, diversion rate and flight percentage for each airline carrier.



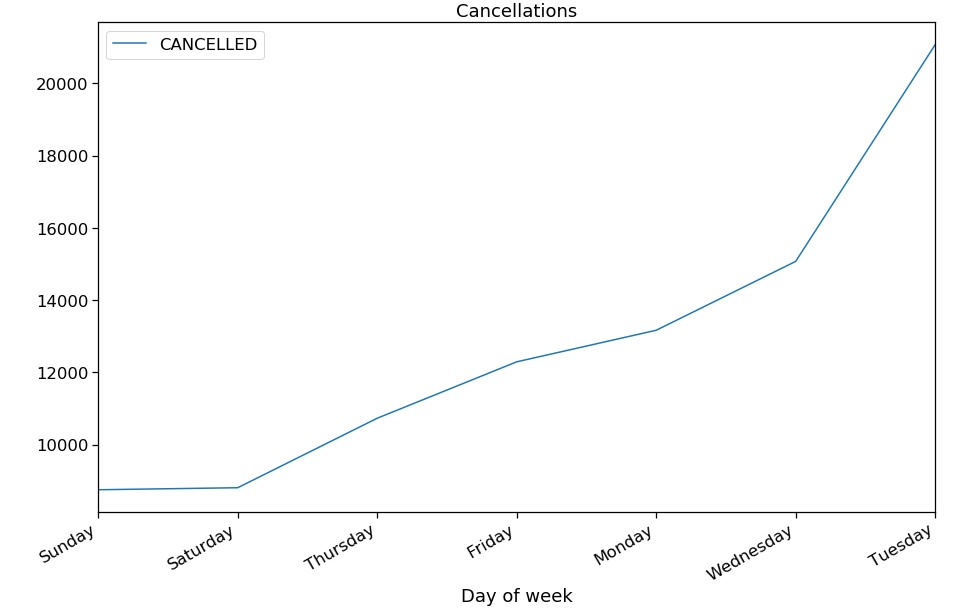
Taxi-in vs taxi-out in stacked bar chart for every airline



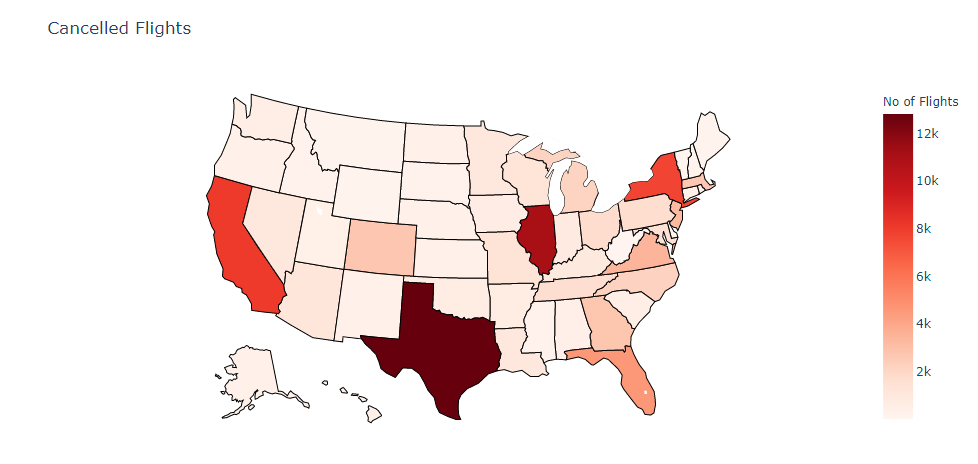
Arrival vs departure delay for every airline



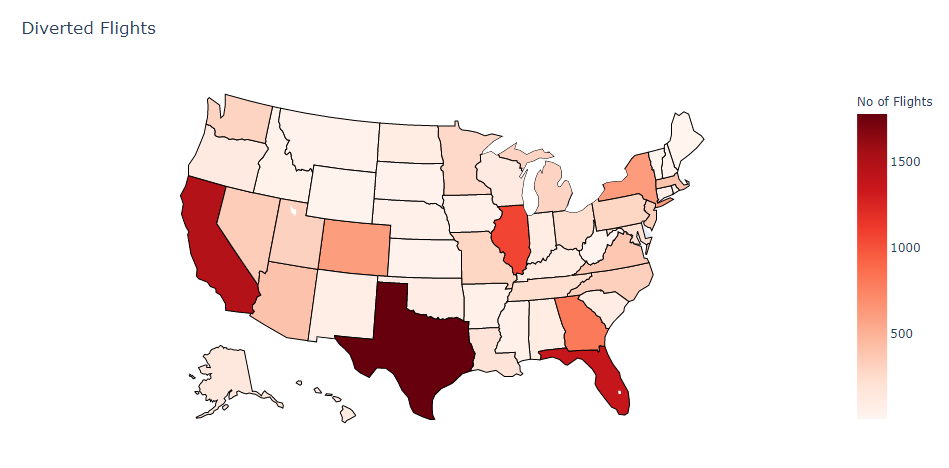
Week by cancellation



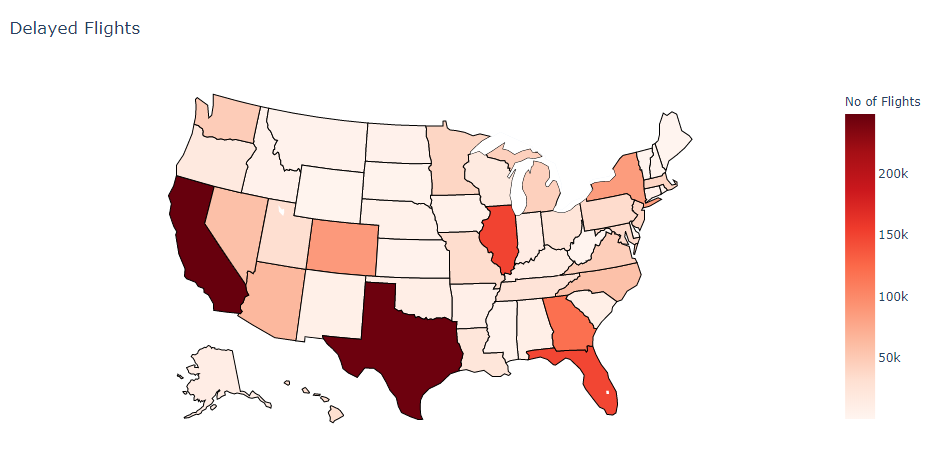
Choropleth state wise map visualization for cancelled flights



Choropleth state wise map visualization for diverted flights



Choropleth state wise map visualization for delayed flights



**4 Conclusion**

**5 Future Work**

# **6 References**

Airlines for America. (2010). *Annual U.S. Impact of Flight Delays (NEXTOR report)*. Retrieved from airlines.org: http://airlines.org/data/annual-u-s-impact-of-flight-delays-nextor-report/

Andrew Hough. (2010). *Holidaymakers waste 16 days waiting in airports over lifetime*. Retrieved from telegraph.co.uk: https://www.telegraph.co.uk/travel/travelnews/7546209/Holidaymakers-waste-16-days-waiting-in-airports-over-lifetime.html

FAA. (2019). *Flight Delay Information*. Retrieved from fly.faa.gov: https://www.fly.faa.gov/flyfaa/usmap.jsp

FlightAware. (2019). *MiseryMap*. Retrieved from FlightAware: https://flightaware.com/miserymap/

IATA. (2018). *IATA Forecast Predicts 8.2 billion Air Travelers in 2037*. Retrieved from IATA: https://www.iata.org/pressroom/pr/Pages/2018-10-24-02.aspx