US Flights Delay Analysis

Arth A. Talati

GAFL 531 Research Brief

Professor Sangenito

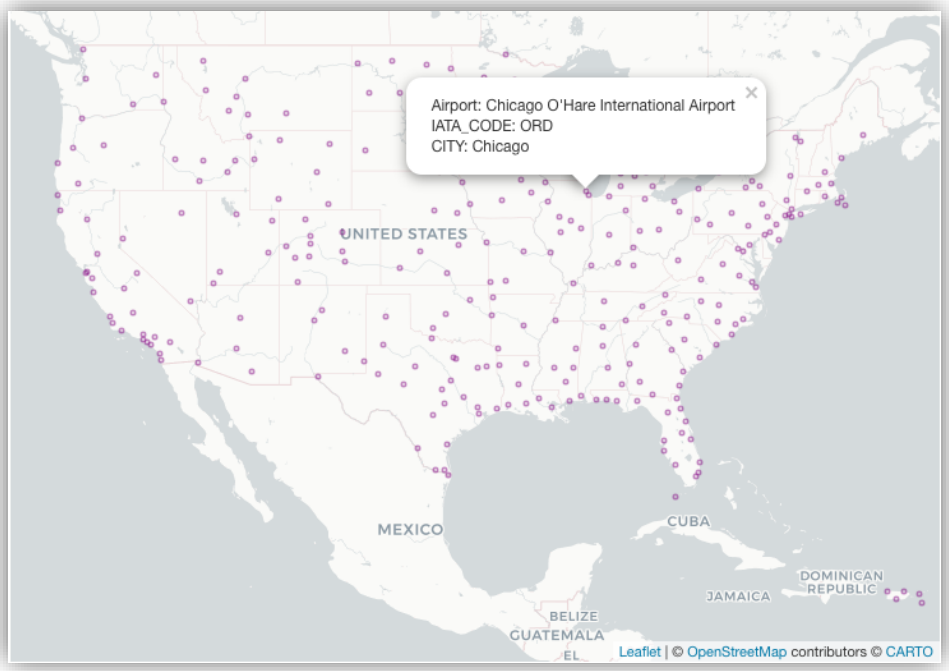
Abstract

In this project, I analyse the dataset of 2015 flight delays and cancellations by US Department of Transportation to find the determine the effects of delays on flight operation. The main reasons for flight delays are weather-related but in some cases there are also airline or airport related flight delays. This analysis tries to find patterns in this regard and build a tool for researchers and individuals to dig in and get more personalised information for decision making and further research.

US Flights Delay Analysis

Introduction:

The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large airline companies. Summary information on the number of on-time, delayed, cancelled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations. In 2015 there are lots of flight delays in United States for some reason. Nearly one third of all flights in this data set have delays. Also one fourth of all flights in this data depart early for their destination. There are number of factors which affect these complex flight operation and in this analysis we try and find these relations. This Kaggle dataset contained information about each flight operation in detail, which we would discuss in the next section.



Map of all Airports in Contiguous United States

Overview of Dataset:

As for the rough overview of the dataset, this analysis covers about 5000 public airports in 50 states and about 14000 private airports. All these airports together handle about 44,000 flights a day. The dataset has about 5.8 million data-points and about 30 odd features which in some or other contain information about the flight details and various types of delays.

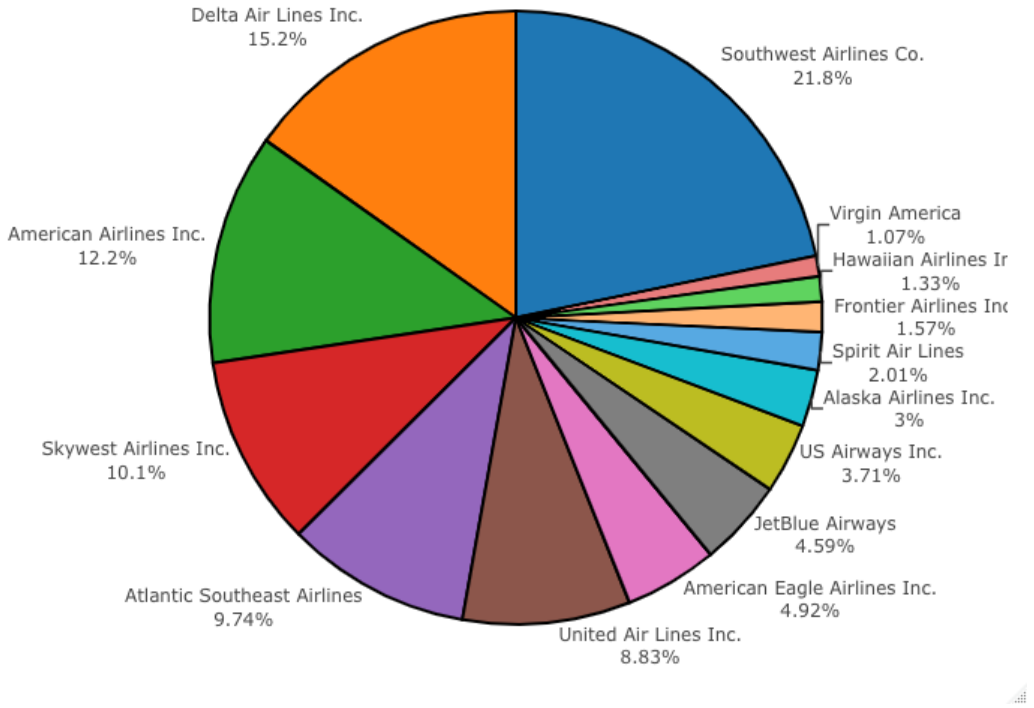
Each entry of the flights.csv file corresponds to a flight have been recorded in 2015. A description of these variables is as follows:

- **YEAR, MONTH, DAY, DAY_OF_WEEK**: dates of the flight
- **AIRLINE**: An identification number assigned by US DOT to identify a unique airline
- **ORIGIN_AIRPORT** and **DESTINATION_AIRPORT**: code attributed by IATA to identify the airports
- **SCHEDULED_DEPARTURE** and **SCHEDULED_ARRIVAL** : scheduled times of take-off and landing
- **DEPARTURE_TIME** and **ARRIVAL_TIME**: real times at which take-off and landing took place
- **DEPARTURE_DELAY** and **ARRIVAL_DELAY**: difference (in minutes) between planned and real times
- **DISTANCE**: distance (in miles)

An additional file of this dataset, the airports.csv file, gives a more exhaustive description of the airports and airlines.csv gives specific information (IATA Codes) for different airlines. These csv files had a couple of missing values for latitude and longitude, which I imputed manually to get a map of all airports.

Data Insights:

Before diving deep into my analysis I wanted to get some overview of the data at hand. I first tried and wanted to look for the number of airlines and their market share in the civil aviation business.

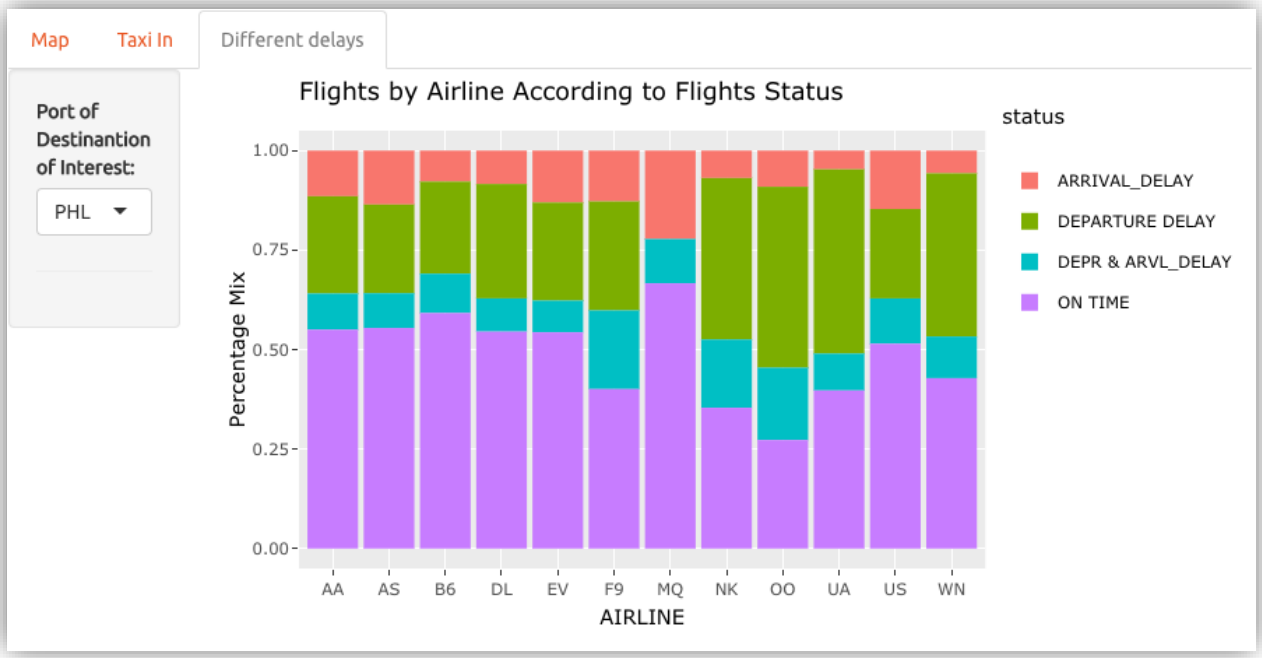


Pie chart showing airline share in civil aviation

The percentages are the number of flights operated by the carrier by the total number of flights in 2015. We can see the big fishes here as the Southwest, Delta, American. Many smaller carriers also operate such as Virgin America, Hawaiian, Frontier, Spirit, etc. It would be interesting to compare the big and the small player when it comes to on time performance statistics.

This plot from my shiny app for Philadelphia International Airport the comparison of Arrival and Departure delays for all market players. It is interesting to see that for Philadelphia Airport, airlines F9, NK and OO which are Frontier, Spirit and SkyWest respectively have a pretty bad

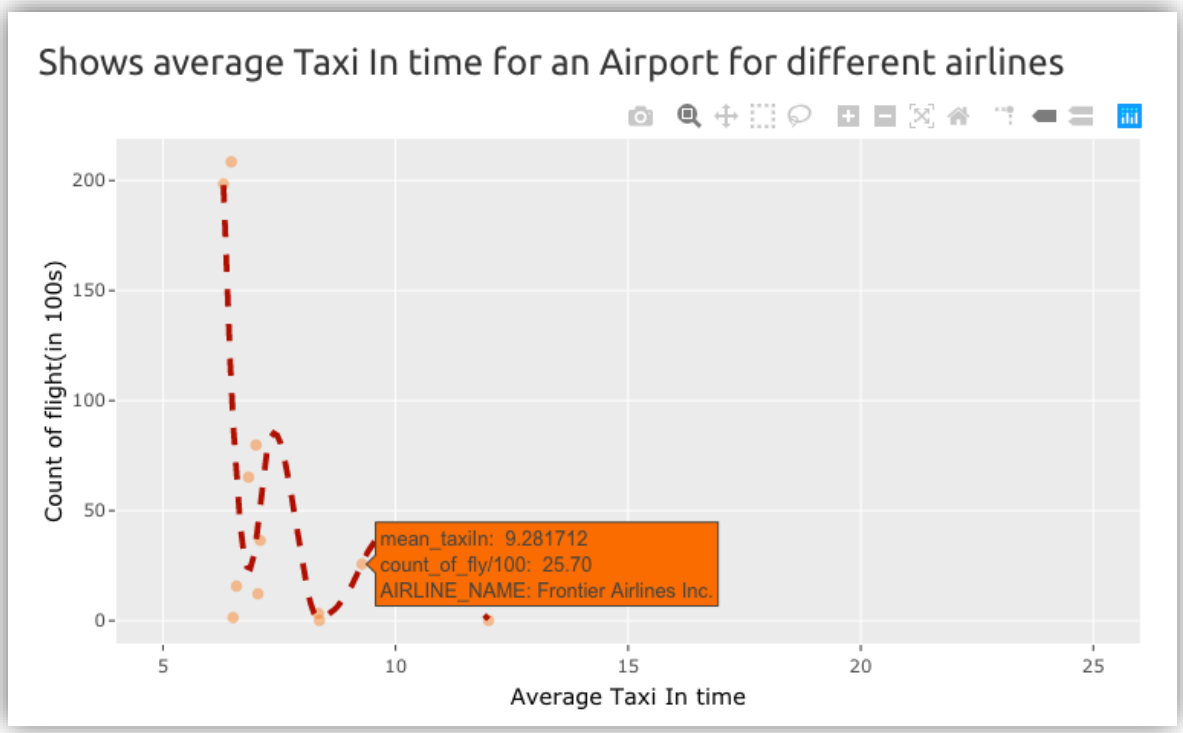
'ON TIME' performance. Whereas AA, DL and MQ airline codes which correspond to American, Delta and Jetblue perform much better. It could be said that as Philadelphia is not a major hub for any of these airlines, implying that it does not have any bias towards any airline carrier, it could be said that you are less likely to experience any kind of delay for major carriers whereas smaller carriers have more chances to be delayed.



Airlines and their performance

Next, I wanted to explore the reasons for the above different on time performance across various airlines and so I thought it would be a good idea to compare the Taxi In and Taxi Out timing for different carriers at airports. Apart from maintenance and weather related delays, an aircraft mostly gets delayed at the Tarmac and finding these patterns across airports could provide critical insight to optimize operations for airlines as well as airports. In this plot, for a given airport, we have the mean taxi times for all airlines.

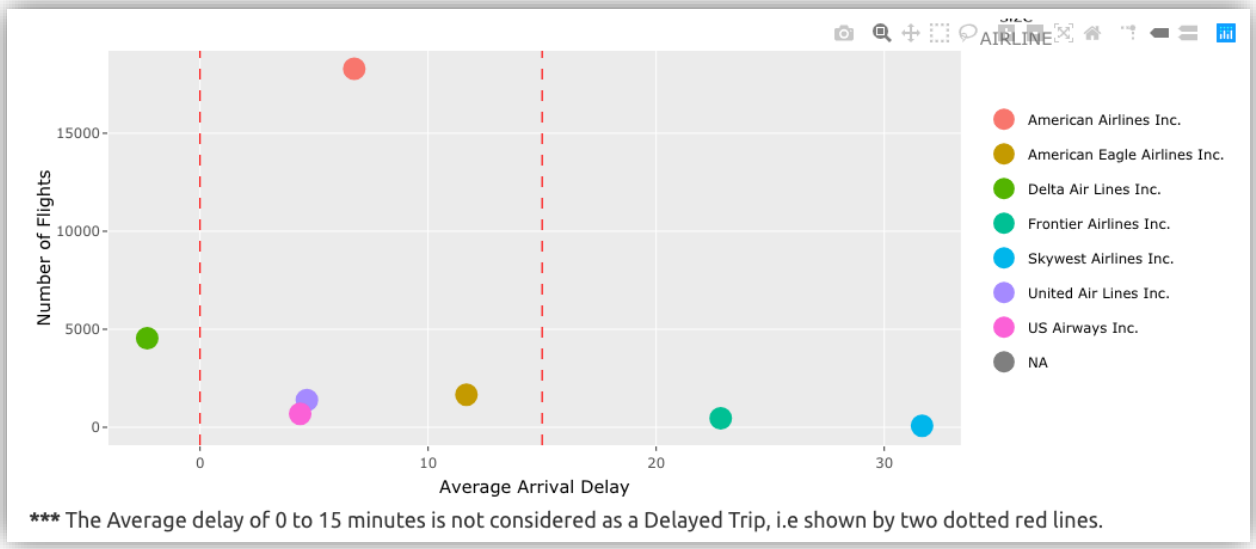
Frontier airlines has 9.24 minutes and SkyWest has 8.36 minutes as mean taxi time whereas Delta airlines has 6.83 minutes and Jetblue has 6.53 minutes as mean taxi time. This analysis does provide some support to the assumption that an airline facing greater arrival delays, in fact tends to have higher Taxi In time.



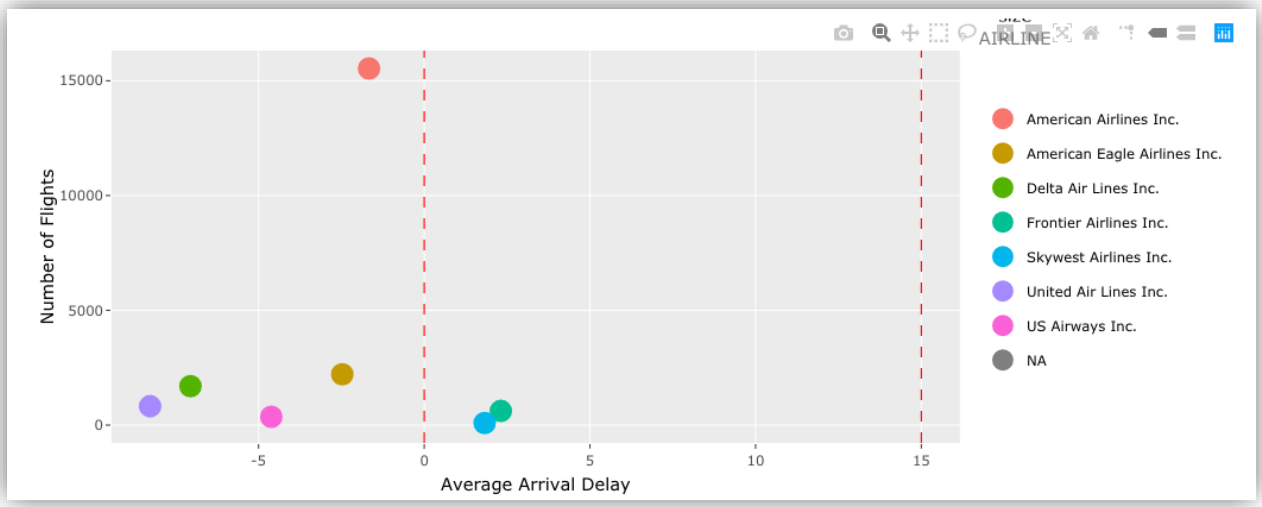
A comparison of Taxi In times for Philadelphia International Airport

Finally I wanted to see the effect of a bunch of factors on the arrival delay patterns for various airports throughout the nation. In the airline industry, the arrival and departure timings are well now to a good predictor of the possible delay the flight might experience. Its quite well-known that midnight and early morning flight usually don't get delayed as there is less traffic during these hours, whereas there is a peak rush around noon and at evening. A person planning a important travel almost always takes into account the time of the day for when the travel is to be planned. Another important factor one always considers is the day of the week of travel. It can be

a good predictor for large amount of business related travel and also weekend rush. I was interested in studying the effects of these parameters on mean arrival delay and see if they confine the norm.



Arrival delays for Miami Intl. Airport for all days and busy afternoon hours.



Arrival delays for Miami Intl. Airport for all days and early morning hours.

As you could see in the images above, for the morning hours, none of the carriers cross the 15 min mark and all are On-Time, majority of them even arriving before the scheduled time. But as we compare it with the busy afternoon hours, we can see that the graph shifts to the right now and the budet airlines/ small players are now showing delayed behaviour. This confirmns our earlier assumption that time of the day is an important factor while modelling the delays for an airport.

Public Policy Matters:

I hope this analysis is found useful by anyone who aims to improve the aviation sectors and reduce passenger woes. My intent was to provide better insight in modelling the different reasons that cause delays in airline operation. I hope that it is particularly useful for government agencies which try and allot slots and gates to the airports. This analysis could also be used in future to model the designs to optimise Taxi In/Out times to reduce delays in the pipeline.