**Executive Summary**

Today's the need for punctuality and saving time is becoming increasingly competitive for both carriers and travelers. Carriers must learn travelers’ time is important. It is the travelers job to give information to carriers in the most effective way possible. As the years go by, the carriers-airports system will become even more competitive, and the need for effective control on time and efficiency is a must. The key to effective carriers and airports relation is the communication, mutual support, and modernizing.

Few carriers provide good timing. In fact, some carriers find themselves cramping for time. We have made very simple analysis to effectively learn on domestic flights within USA. The result is encouraging. Generally, 66 percent of flights are on time.  
  
As carriers-airports system become more entangled, and are required to increase their standards. We believe punctuality improvement for regional flights will prove to be a very worthwhile and profitable investment in the future.

**Introduction**

We obtained our data from the, [United States Department of Transportation](http://www.transportation.gov), United States Bureau of Transportation Statistics ([BTS](https://www.transtats.bts.gov/tables.asp?DB_ID=120)). A large dataset of csv files contain on-time arrival data for non-stop domestic flights by major air carriers, monthly from 1987 to 2018, and include information such as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times, information about canceled or diverted flights, flight distance and more. Large data sets are often complicated to analyze as a single unit. In case of large dataset, it is recommended to stratify ([Subsample](https://en.wikipedia.org/wiki/Stratified_sampling)) a large dataset to manageable size and to reduce heterogeneity. In our case, the flight data, we stratify by month. We targeted the variation between given attributes as it is more informative than single values (means, medians and mode). Python computer language and its pandas library were heavily used to show variation. The sheer size and complexity of the data demystified by using scatter matrix and views in terms of origin/destination pairs.

**Objectives**

We appraised the data for practical benefits, hoping to glean trends that can be generalized to provide gains to efficient relation between carriers and airports including customer satisfaction, and showing limitation of the current relations.

**Data Analysis**

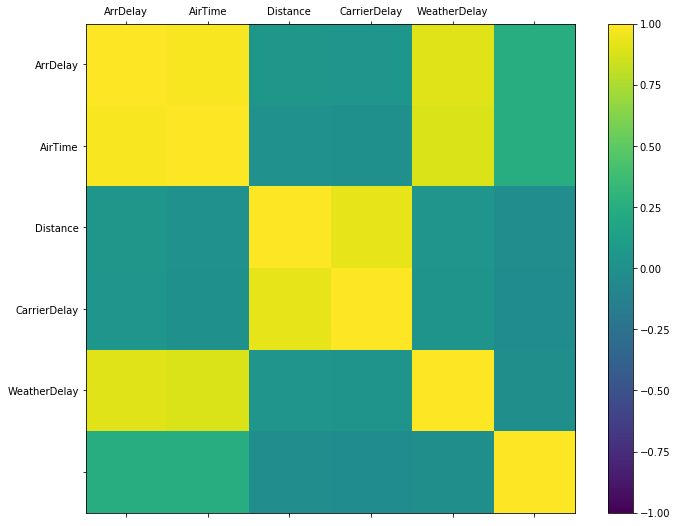
Pandas data frames are constructed with assistance of Python computer language with application of basic statistics. We analyzed flight delay variable. Delays can be driven by many factors, some are explainable other are dogmatic. There are about six millions flight per year, with data for origin airport, destination airport, and delay. We hypothesized that weather will be the ultimate cause of most of the delays.

**Conclusions**

**Appendices**

* Over all. there are 66 percent of flights that ar*rive on time.*
* Departure Times: As the day progress, there seem to be an increase in departure times.
* Departure Delay: Shorter departure delays for longer distance flights.
* Arrival Delay: Arrival Delay correlate with departure delay, with exception of longer flight, which seems to catch up and still arrive on time. Other delays (Weather Delays, etc) seem to be uncorrelated within with general trend of delays.
* MD, NV and TT are the 3 origin States/Territories that have the most number of delayed flights while MT and AK has the least delays.

Below, we show a figure summarizes the correlation of few kinds of the delays. Let's say it is of interest to see what type of delays is intuitive and explainable. For example, the figure shows correlation between Carrier Delays and distance while no correlation between weather delays and distance.



There are about six millions flights with flight duration (AirTime) and flight distance (Distance). Flight duration is measured in minutes and flight distance (distance of shortest path between origin and destination, not actual distance). Note that our dataset only includes non-stop flights that are domestic to the United States. The correlation coefficient between the two variables is 0.925. The fitted regression of AirTime on Distance is (see figure below):

AirTime=f(distance)~ m\*(distance)+intercept,

where m is the slope.

**AirTime ~ 16.47 + 0.11 \* Distance**

Based on this formula Aircraft requires 0.11 minutes to travel one mile, and 60/0.11=522 miles per hour.

