# Hypotheses Testing on Real-Time Flight Data


Neelaabh Gupta, Sharad Gupta, Sahil Sobti, Atmika Sharma

**CSE 544: Probability & Statistics for Data Scientists**

Stony Brook University

GitHub Link: https://github.com/sahilsobti/Flight_Delay_Analysis



**Abstract**

*Thousands of flights operate everyday in the United States of America. However, there are some flights that gets delayed and a few that get cancelled. Flight delays causes a lot of inconvenience to passengers. It can make them late to their scheduled events, thus prompting outrage and frustration. Additionally, travelers may not generally be qualified for compensation when a delay happens. Most airlines claim that a few reasons that causes delays are airline outages, weather conditions, air traffic congestion, aircraft maintenance issues, late arrival of connecting flights and security issues. Flight delays impact airlines, airports, and passengers. Their prediction is pivotal for all players of commercial aviation. This paper poses some hypotheses and shows how statistical methods like Regression, Time Series Analysis, Parametric Inference and Correlation Analysis can be used to prove/disprove those hypotheses to gain valuable insights.*


## I. Introduction

Flight delay is one of the most important indicator of a commercial airline's performance. A small disruption in the airline's operations can lead to massive financial losses. For example, a small system outage in the world's busiest airport (Atlanta's Hartsfield-Jackson) in 2016, cost Delta Airlines $50 million USD. Therefore, it is of utmost importance for an airline carrier to regulate & minimize flight delays. Commercial aviation players define flight delay as the period represented by the difference between scheduled and real-time departure/arrival of flights. In 2015, 22% of the flights got delayed which shows the magnitude of losses incurred by commercial airlines. The major reasons of delays were Carrier & Weather Delays.

Flight delays have mostly negative economic impacts for airlines, environment and most importantly passengers. Airlines suffer penalties, fines and additional operation costs at the hands of Federal Aviation Administration. Such fines include crew retention in airports as well as holding planes on the tarmac prolonged times. Additionally, delays also cause environmental damage due to an increase in fuel consumption and gas emissions. Last but not the least, given the uncertainty of flight delays, passengers usually plan to travel many hours earlier to the airport, increasing their trip costs, to ensure their timely arrival.

Federal Aviation Administration classifies flight delays in the following categories:

- **Carrier Delay**: Within the control of the air carrier. Includes aircraft damage, arrival of crew, baggage, engineering inspection, fueling, maintenance, slow boarding.
- **Weather Delay**: Caused by extreme or hazardous weather conditions.
- **NAS Delay**: Within the control of the National Airspace System (NAS) may include: non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control.
- **Security Delay**: Evacuation of a terminal or concourse, re-boarding of aircraft because of security breach.

To better understand and analyze the flight delay scenario in the US, flight data was taken from the Bureau of Transportation Statistics. The dataset contains details of every flight in the US along with information on the types of delays/cancellations and amount of delays in minutes respectively. Four broad topics of hypotheses were created with a few sub-hypotheses in each group. The 4 groups are:

- Flight Delays/Cancellations
- Flight Causality

- Flight Distributions
- Flight Delays Prediction/Correlation

## II. Dataset

The dataset was taken from Bureau of Transportation Statistics, USA. The name of the dataset is Airline Performance Data. The table contains on-time arrival data for non-stop domestic flights by major air carriers, and provides such additional items as departure and arrival delays, origin and destination airports, flight numbers, scheduled and actual departure and arrival times and cancelled flights. We've done most of our analysis on 2016 Flight Data, and used previous years (2011-2016) data for our Predictive Analysis.

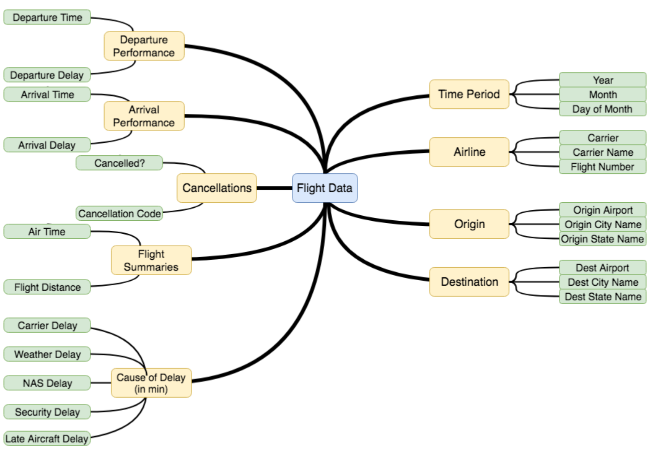


**Figure 1:** *DataTable Schema*

The dataset we use for most of our analysis, i.e 2016 Flight Delay/Cancellation data, had 1,032,000 records. There were 25 columns in the dataset, whose schema is shown in Fig1. However, the dataset was quite raw. Thus, a lot of preprocessing needed to be done. The major preprocessing phases can be seen in Fig2.

## III. Hypotheses

### i. Topic I: Flight Delay/Cancellations

#### i.1 The arrival delay for the shorter distance flights is same as compared to longer distance flights

Through this hypothesis we are trying to find whether distance between airports has any impact on flight arrival delays. This will help us with insights on how airlines can schedule flights in a better manner based on distance between different cities across USA.

This analysis is done with the flight arrival delay data across USA in year 2016.To test the hypothesis,

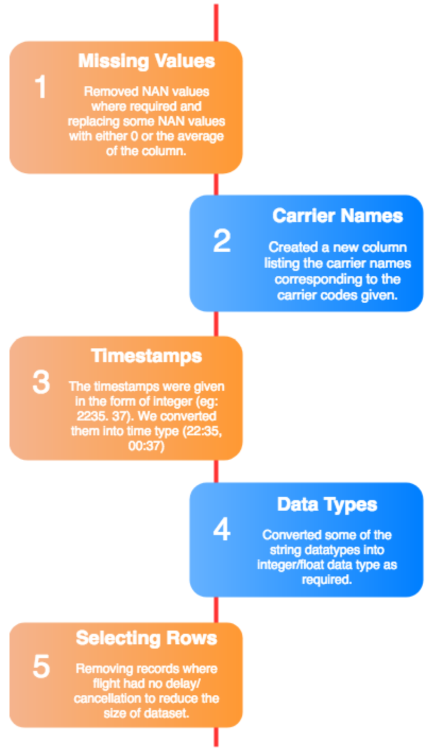


**Figure 2:** *Pre-processing Data*

shorter distances are assumed to be lesser than 1250 miles and longer distances above 1250 miles. We are also assuming the the shorter and longer duration flight to have **asymptotic normal** and independent delays.

Initially, we have divided the entire flights data into 11 buckets based on the distance between the departure and arrival airport distance. Each bucket includes distances increasing every 249 miles while the last bucket includes distances of 2500 miles or greater. We have further cleaned the data using Tukey's rule by dividing the total data into 3 different quartiles i.e. 25%(Q1), 50% (Q2) and 75%(Q3) . The outliers are then removed based on the equation shown below

$$Outlier = Q3 + 1.5 \times IQR \text{ or } Q1 - 1.5 \times IQR \quad (1)$$

Here IQR = Inter Quartile range. This data is further processed to find the mean of arrival delay for each bucket mentioned above. From the graph, it is quite clear that shorter distance does have higher arrival delays across entire US. Now, we have proved the same using one tailed T-test and one tailed Wald test where we have compared the arrival delays for different flight distances.

- Flight arrival delays for distances below 1250 miles in 2016

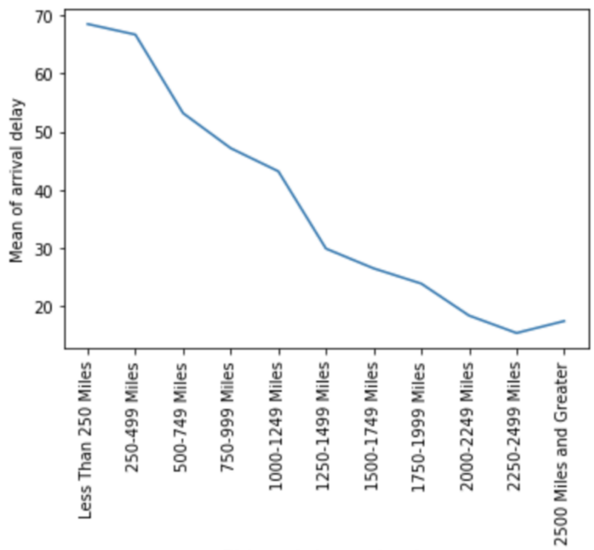


**Figure 3:** *Mean Arrival delay vs flight duration*

- Flight arrival delays for distances above 1250 miles in 2016

| Test (2 population) | Critical value | Test value | P value | Result |
|---|---|---|---|---|
| **Upper tailed T test** | 1.6495 | 6.02 | 0.001 | Reject Null hypothesis |
| **Upper tailed Wald's test** | 1.6495 | 6.717 | NA | Reject Null Hypothesis |

**Figure 4:** *Test results between longer and shorter duration flights*

We have therefore proved that shorter distance flights have shown to cause longer arrival delays compared to longer distance flights. It is therefore of utmost importance for the airlines to investigate further on these shorter routes in order to prevent loses in future.

### i.2 Small scale commercial airlines perform no better than the large scale commercial airline with respect to arrival delays irrespective of flight distance

The objective of this hypothesis is to know which kind of airlines have shown the least arrival delays in the past.It provides travelers with an insight to pick the right airlines for their journey in situations where they cannot afford to have flight delays

This analysis is done with the flight arrival delay data across USA in the year 2016. We have assumed that arrival delay of each airline is independent & **asymptotic normal**. To test this hypothesis, we initially divided the past flights arrival delay data based on whether the journey was within 700 miles or above. We have processed this further to find the average arrival delay for each airline in 2016 in USA. In order to test our hypothesis, we have used Waldś test for 2 independent populations on these aforementioned airlines.

Flights with distance of 700 miles or less

| CARRIER_NAME | ARR_DELAY |
|---|---|
| Hawaiian Airlines | 29.414448 |
| SouthWest Airlines | 45.672837 |

Flights with distance of more than 700 miles

| CARRIER_NAME | ARR_DELAY |
|---|---|
| Alaska Airlines | 43.223789 |
| SouthWest Airlines | 48.499967 |

**Figure 5:** *List of top 2 airlines with minimum arrival delays under 2 different scenarios*

| Distance (Miles) | Between Airlines | Wald test | Critical value for lower tailed test | Result |
|---|---|---|---|---|
| **<= 700** | Hawaiian vs Southwest | 19.9644 | -1.645 | Reject Null |
| **> 700** | Alaska vs Southwest | 5.4 | -1.645 | Reject Null |

**Figure 6:** *Results of Waldś test for routes under 700 miles and above*

The above can also be seen from the year-round arrival delays in both the airlines in 2016

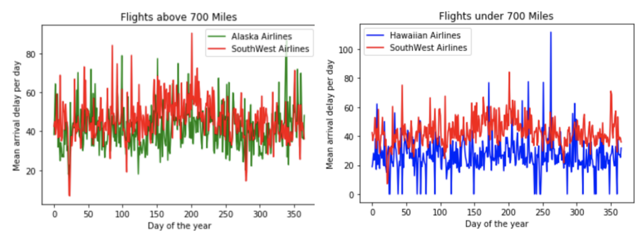


**Figure 7:** *Arrival delays of airlines in 2016*

Hence, we have confirmed that smaller airlines like Hawaiian and Alaska airlines have comparatively lesser arrival delays than bigger commercial airlines like Southwest airline. This hypothesis is useful for customers to book the right airline for their journey when arrival time is a major constraint for them.

### i.3 Rate of cancellation of flights is uniform across the day

Flight cancellations cause inconvenience to both airlines and passengers all across the year. Avoiding these situations is very important to prevent loses. Therefore, through this hypothesis, we want to figure out what is the trend in the cancellations every day at JFK airport in 2016.

This analysis is done with the flight cancellation data of John F. Kennedy airport, New York for the spring months of March, April and May in the year 2016.

We have initially extracted the flight cancellation data of JFK airport and divided the cancellations based on different time periods of the day as follows: Late Night : 12 - 4 am | Early Morning : 4 - 8 am | Morning

: 8 - 12 am | Afternoon : 12 - 4 pm | Evening : 4 - 8 pm | Night : 8 - 12 am. Based on these bins, we found the cancellations for each of these bins in the spring months and observed that late night to have the maximum cancellations on average.We therefore used 2 population Waldś tests to check if late night has the same cancellation distribution as that of other time intervals.

| Time Slot 1 | Time Slot 2 | Wald's Test | Critical value | Result |
|---|---|---|---|---|
| Late Night | Early Morning | 5.63 | 1.96 | Reject NULL |
| Late Night | Morning | 5.63 | 1.96 | Reject NULL |
| Late Night | Afternoon | 5.632 | 1.96 | Reject NULL |
| Late Night | Evening | 5.38 | 1.96 | Reject NULL |
| Late Night | Night | 5.605 | 1.96 | Reject NULL |

**Figure 8:** *Waldś test under different time slots*

Based on these results, we can now say that the flight cancellations do not have a uniform distribution across the day and hence it is advisable not to travel at a time slot which causes the maximum number of cancellations on average every day when the time is a priority for you.

## ii. Topic II: Causality

### ii.1 Arrival delays at an airport affect departure delays at the same airport

This hypothesis helps us investigate whether arrival delays will have an impact on departure delays incurred by flights of the same carrier. We look into whether a ripple effect is seen at any particular airport for connecting flights of an airline carrier.

This analysis is done with the flight arrival and departure delay data at LAX, Los Angeles in the year 2017. We have taken into consideration JetBlue flights that have incurred arrival delays of more than 50 minutes. For departure delays, we have considered JetBlue flights departing aon the same day from LAX. Since we apply KS-Test we do not need any assumptions for the two distributions being compared.

To test this hypothesis, we consider average arrival delays of incoming flights and average departure delays of outgoing flights. We apply KS Test to average arrival delay distribution for JetBlue airlines at LAX in 2017 with distribution for average departure delay incurred by incoming LAX flights with some arrival delay in the same year. The condition for accepting

| KS-Test | Critical value | Test value | Result |
|---|---|---|---|
| Arrival Delays and Departure Delays | 0.14 | 0.092 | Accept Null hypothesis |

**Figure 9:** *KS-Test on Arrival and Departure Delay Distributions*

null hypothesis is for the KS-Test statistic to be less than critical value. We accept the null hypothesis: arrival delays at LAX are distributed similar to departure delays at LAX. This shows that arrival delays impact departures thus causing a chain of delays across an airline route.

### ii.2 A disruption in operations of a certain airline does not cause any carrier delays over certain period of time

The hypothesis is basically to find if there is an impact of operational difficulties on the carrier delays of an airline.Through this hypothesis, we will get to know how severe is the impact of an operational glitch on the overall carrier delay of an airline.

We have used the carrier delay data for United States in the year 2016 and the populations are assumed to be **asymptotic normal** and independent for the Wald test.

**Carrier delay** is the delay which is in the control of air carriers. It is caused by aircraft cleaning, damage, engineering inspection, cargo loading etc.

We started our hypothesis with an article "Delta's flight operations return to normal" where they mentioned how delta airline came back to normalcy after an outage in august 2016. We used this piece of evidence to see if the same can be reflected in our data.

The peculiar peak in the august month only for delta airline is a valid proof of the fact that the outrage did impact the carrier delay. Now, in order to proof this through statistical methods, we have compared the two distributions of delta airlines i.e. mean carrier delay per day in august vs the mean carrier delay per day for the entire year. When we put these two populations under upper tailed- Wald's test for 2 populations, we got the result of Wald statistics as 1.889 where the critical value for upper tailed test is 1.645. So, we accept the alternate hypothesis that disruption in delta airlines did cause more carrier delay.

Based on the results, we have found that the operational difficulties can cause a carrier delays for the airline and therefore it is wise for the travelers to choose the airline appropriately in such scenarios.

### ii.3 Nearby airports will be affected due to delays at a certain airport

The major reason for doing this kind of hypothesis is to find if the airports which are closer to each other have the delays at the same time. After testing this hypothesis, we want to know if delays affect the airports

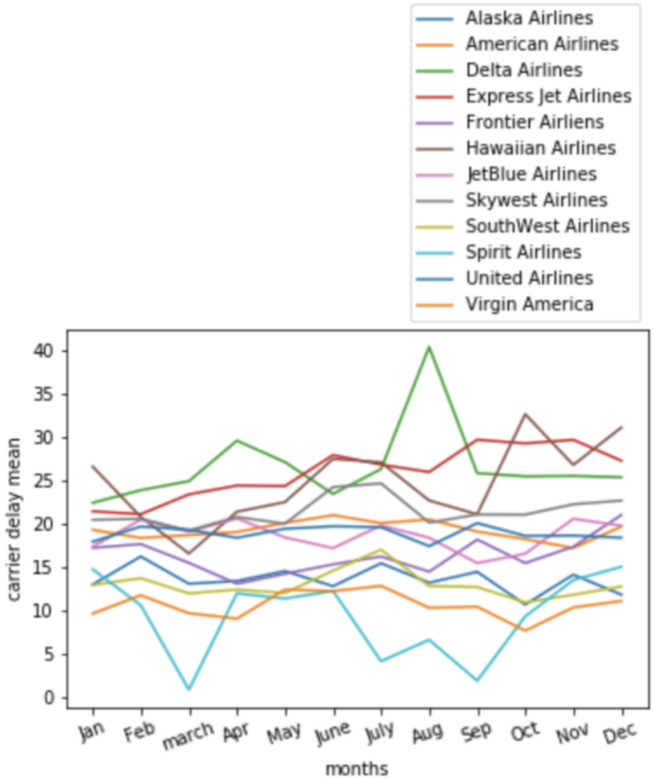


**Figure 10:** *Every month carrier delays of different airlines in 2016*

the same way and how travelers can use these insights to make a sound decision.

We have taken departure delay to test this hypothesis. We have also assumed all airports that are located in the same state, are very close to each other.

To test this hypothesis, we have considered the state of California as our subject. We first selected four random big and smaller airports such as Los Angeles International Airport(LAX), San Jose International Airport(SJC), Palm Springs International Airport(PSP) and John Wayne Airport (SNA). Then we extracted all the flights schedules in which the origin was one of the mentioned airports. This flight data was further processed to find the mean departure delay each day from all these airports. The distributions obtained was put through KS test to find if they are similar or not.

| Airport1 | Airport2 | Ks test value | Critical value | result |
|---|---|---|---|---|
| **LAX** | SJC | 0.33 | 0.07 | Reject NULL |
| **LAX** | PSP | 0.37 | 0.07 | Reject NULL |
| **LAX** | SNA | 0.26 | 0.07 | Reject NULL |
| **SJC** | SNA | 0.097 | 0.07 | Reject NULL |
| **PSP** | SNA | 0.32 | 0.07 | Reject NULL |

**Figure 11:** *KS test results between different nearby airports*

Therefore, based on the results above, we have rejected the null hypothesis that nearby airports have similar departure delay distribution. We also conducted the same experiment on others states like Colorado and New York and the results were quite similar.

Based on these results, it is quite clear that airports nearby do not have delays at the same time and therefore, it is beneficial for the travelers who are stranded at the airport due to long departure delay to look out for flights at the nearby airports.

## iii. Topic III: Distribution

Flight delay distributions and their comparisons provide valuable insights for both customers and carriers. It aids in better decision making for the customer during bookings and allows airlines to cut down on costs and minimize losses.

### iii.1 In winter months, different airlines have similar weather delay distributions [KS-Test]

Weather delays include any disruption in flight schedules caused due to extreme weather conditions such as thunderstorm, rain, clouds and snow. Through this hypothesis we are trying to find if extreme weather conditions during winter months impact all airlines at a particular airport similarly.

We consider flight weather delay data at JFK, New York in the year 2017. The airlines we have considered include American Airlines, Delta Airlines and JetBlue Airlines. We compare distributions for the aforementioned airlines in the months January, February, November and December. Since we apply KS-Test we do not need any assumptions for the two distributions being compared.

In order to test this hypothesis, for each winter month, we have considered average weather delay for each day in the month. The statistical method used in this case is the KS-Test. We compare the distribution for average weather delay for Airline A in Month X at JFK with the distribution of average weather delay for Airline B in Month X at JFK.

The condition for accepting null hypothesis is for the KS-Test statistic to be smaller than the critical value. Thus, the above results allow us to accept the null hypothesis that for each winter month, different airlines at a particular airport have similar weather delay distributions.

Our acceptance of the null hypothesis helps us provide insights to a customer such as booking a different airline during a period of extreme weather is not beneficial in any way.

| KS-Test | Airline Weather Delays (New York) | Critical value | Test value | Result |
|---|---|---|---|---|
| January | United and American | 0.242 | 0.0645 | Accept Null hypothesis |
| | United and JetBlue | 0.242 | 0.1613 | Accept Null hypothesis |
| | JetBlue and American | 0.242 | 0.1612 | Accept Null hypothesis |
| February | United and American | 0.242 | 0.107 | Accept Null hypothesis |
| | United and JetBlue | 0.242 | 0.142 | Accept Null hypothesis |
| | JetBlue and American | 0.242 | 0.178 | Accept Null hypothesis |
| November | United and American | 0.242 | 0.068 | Accept Null hypothesis |
| | United and JetBlue | 0.242 | 0.068 | Accept Null hypothesis |
| | JetBlue and American | 0.242 | 0.103 | Accept Null hypothesis |
| December | United and American | 0.242 | 0.178 | Accept Null hypothesis |
| | United and JetBlue | 0.242 | 0.142 | Accept Null hypothesis |
| | JetBlue and American | 0.242 | 0.107 | Accept Null hypothesis |

**Figure 12:** *KS-Test on pairs of airline carriers for each winter month*

### iii.2 In winter months, different airlines have similar weather delay distributions [Permutation-Test]

We now apply the Permutation-Test to compare the distribution for average weather delay for Airline A in Month X at JFK with the distribution of average weather delay for Airline B in Month X at JFK.

| Permutation-Test | Airline Weather Delays (New York) | Critical value | Test value | Result |
|---|---|---|---|---|
| January | United and American | 0.05 | 0.339 | Accept Null hypothesis |
| | United and JetBlue | 0.05 | 0.429 | Accept Null hypothesis |
| | JetBlue and American | 0.05 | 0.772 | Accept Null hypothesis |
| February | United and American | 0.05 | 0.955 | Accept Null hypothesis |
| | United and JetBlue | 0.05 | 0.284 | Accept Null hypothesis |
| | JetBlue and American | 0.05 | 0.888 | Accept Null hypothesis |
| November | United and American | 0.05 | 0.505 | Accept Null hypothesis |
| | United and JetBlue | 0.05 | 0.876 | Accept Null hypothesis |
| | JetBlue and American | 0.05 | 0.423 | Accept Null hypothesis |
| December | United and American | 0.05 | 0.545 | Accept Null hypothesis |
| | United and JetBlue | 0.05 | 0.689 | Accept Null hypothesis |
| | JetBlue and American | 0.05 | 0.410 | Accept Null hypothesis |

**Figure 13:** *Permutation-Test on pairs of airline carriers for each winter month*

The condition for accepting null hypothesis is for the Permutation-Test statistic to be greater than the critical value. We thus accept the null hypothesis that for each winter month, different airlines at a particular airport have similar weather delay distributions.

### iii.3 Departure delays at airports are normally distributed

We check departure delays at different airports for normality through this hypothesis. This analysis is done with departure delay data for all airline carriers with flights originating from SFO (San Francisco), JFK (New York) and LAX (Los Angeles) Airports in the year 2017. Since we apply KS-Test we do not need any assumptions for the two distributions being compared, one of which is a general standard normal distribution.

To test this hypothesis, we consider average departure delay of flights originating from a particular airport. We first transform our departure delay into a standard normal distribution and apply KS Test to compare the transformed average departure delays for all airlines originating from SFO, JFK and LAX in 2017 with a general standard normal distribution.

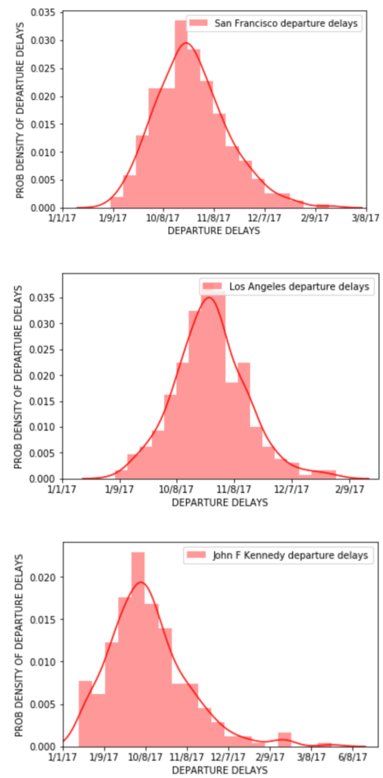


**Figure 14:** *Probability density distribution at SFO, JFK, LAX*

| KS-Test (SFO) | Critical value | Test value | Result |
|---|---|---|---|
| Departure delays | 0.07 | 0.053 | Accept Null hypothesis |

| KS-Test (JFK) | Critical value | Test value | Result |
|---|---|---|---|
| Departure delays | 0.07 | 0.082 | Reject Null hypothesis |

| KS-Test (LAX) | Critical value | Test value | Result |
|---|---|---|---|
| Departure delays | 0.07 | 0.051 | Accept Null hypothesis |

**Figure 15:** *KS-Test for Normality at SFO, JFK, LAX*

Since we have proved that SFO and LAX departure delay distributions are normally distributed, we've found their mean and variance using **MME and MLE**

$$\hat{\mu_{MME}} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\hat{\mu_{MLE}} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\hat{\sigma}^2_{MME} = \frac{\sum_{i=1}^{n} X_i^2}{n} - (\frac{\sum_{i=1}^{n} X_i}{n})^2$$

$$\hat{\sigma}^2_{MLE} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

Using the above formulae we calculated

$$\hat{\mu}_{MLE} = \hat{\mu}_{MME} = 51.24$$

and

$$\hat{\sigma}^2_{MLE} = \hat{\sigma}^2_{MME} = 201.25$$

for SFO. Similarly we can find these values for normally distributed LAX. We have shown the probability of our estimator is falling within the range [23.43, 79.04] for SFO with 95% **confidence interval (CI)**. Similarly for LAX is [27.18,76.25]

The condition for accepting null hypothesis is for the KS-Test statistic to be smaller than the critical value. We accept the null hypothesis that departure delay of flights originating from SFO and LAX are normally distributed while those originating from JFK are not normally distributed.

## iv. Topic IV: Prediction & Correlation

Flight delay alerts are helpful, but they usually depend on airlines notifying you in a timely fashion... and they are not always quick on the draw. We're happy that we have taken the first step in filling this gap.

Imagine getting to know the delay a flight will incur during inclement weather conditions in advance, this could give us a hint of trouble before we set foot in the airport. This subsection deals with predicting the future delays of flights, followed by an extensive study of the impact of disasters on the departure delays of flights.

### iv.1 Predicting departure delay for 4 quarters based on historic data.

We're tasked to predict the flight departure delays for 2017 across 4 quarters using the previous yearâĂŹs data. We've taken the flight delay data for 2013-2016 as the training data. Assuming that the delay is linearly dependent on features taken into consideration for learning, we've predicted delays for Jan-March, Apr-June, July-Sept, Oct-Dec 2017 quarterly.

In order to predict the flight departure delays we have taken many features into consideration which are divided into two broad categories:

1. **Numerical Features**: Features whose domain are real no's greater than 0.
   - Distance: Distance travelled by the flight from origin to destination
   - Weather Delay: Weather Delay for the flight
2. **Categorical Features**: Features that take some categorical values only.
   - Month: The month in which the flight ran
   - Day of Month: The day of the month in which the flight ran
   - Day of Week: The day of the week in which the flight ran.
   - Origin: The origin airport from flight started.
   - Destination: The destination airport at which the flight ended
   - Scheduled Departure: The time at which the flight is scheduled for departure.

Before learning delays using above features, we first scaled the numerical features so that our learning is smooth, and for the categorical features we used one hot vector which gives the numerical label for unique categories. So, we took the above processed features for every flight in USA from 2013-2016, and then ran Linear regression algorithm using above features to predict the departure delay for every flight in 2017.

To show the results we divided our prediction into 4 quarters namely Jan-March, Apr-June, July-Sept, Oct-Dec and plotted the mean predicted delays and actual delays in those quarters. Below are the four figures corresponding to each quarter:

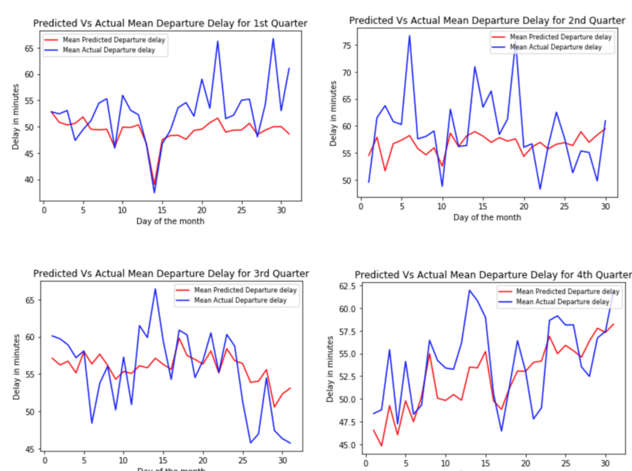


**Figure 16:** *Linear Regression Predictions*

As, we can see that the predicted values seem to fit close to the actual delays for 1st and 3rd quarter but for 2nd and 4th quarter there are some anomalies in the data, and the reason for that is the natural disaster which caused huge delays in the data. Thus, based on our statistical results, we have found that past flight delays data can be helpful for modeling the future departure delays and therefore can help people in planning their journeys effectively in future. Also, it can give more insights to airlines as to which months facing more delays, so that they can be prepared in advance to handle the situation in a better way.

### iv.2 Predicting delays for 2018, based on data for 2011-2017

Another very interesting approach to predict the future delays of an airline is using Time Series Analysis. We've predicted the delays of flights departing from New York in the month of Jan 2018. We've used departure delay data of the similar setting on the years 2011-2017 as training data.

The data thus obtained contained 216 values. Each value represented the average departure delay per day for a particular flight departing from New York. Thus, the data spans 8 years (2011-2018). We used the first 7 years as training, and the $8^{\text{th}}$ for testing. We further noted the average error over the $8^{\text{th}}$ year. Finally, we reported the average error across the last 31 days of the $8^{\text{th}}$ year. The average precision error thus obtained came out to be 14.4%.

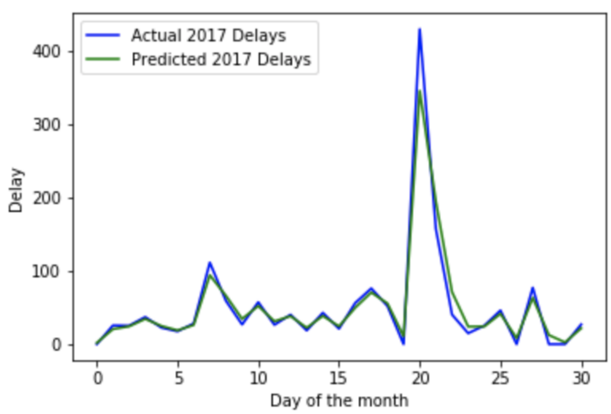


**Figure 17:** *Time Series Analysis Predictions*

### iv.3 Delays/Cancellations increases as the intensity of natural disaster increases.

United States of America is hit with a number of disasters throughout the year. The disasters range from being small scale like Heat, Wildfires to intense large scale disasters like Hurricanes and Blizzards that bring destruction in its wake. Disasters costs the affected region in all manners. One such impact is the airline disruption that happen at the affected regions. This study takes 5 disasters of different intensities and tries to find their impact on the affected regions. The hypothesis is as follows:

- **Null Hypothesis:** Delays/ Cancellations does not increase as the intensity of natural disaster increases.
- **Alternate Hypothesis:** Delays/ Cancellations increases as the intensity of natural disaster increases.

To test the hypothesis, we've used Wald's 2 Population Test. The 2 populations being, Mean Departure Delay/Average Cancellation per day in the affected region during the disaster, vs Mean Delay/Average Cancellation per day in that region throughout the year. The results for the 5 disasters below.

**Disaster 1: 2016 American Northeast Heatwave (Aug 8-11, 2016, New York)**

- **Scenario1: Duration of Disaster**: Mean Departure Delay = 117 min
- **Scenario2: The Whole Year**: Mean Departure Delay = 64 min

The Wald Statistic came out to be 13.37,

**Disaster 2: 2017 California Wildfires (October 11-13, 2017, California)**

- **Scenario1: Duration of Disaster**: Mean Departure Delay = 101 min
- **Scenario2: The Whole Year**: Mean Departure Delay = 57 min

The Wald Statistic came out to be 32.7.

**Disaster 3: Jan 2016 Blizzard (January 22-25, 2016, North-Eastern USA)**

- **Scenario1: Duration of Disaster**: Avg Cancellation/Day = 982
- **Scenario2: The Whole Year**: Avg Cancellation/Day = 41

The Wald Statistic came out to be 9.24.

**Disaster 4: Louisiana Floods (August 12,20, 2016, Louisiana USA)**

- **Scenario1: Duration of Disaster**: Avg Cancellation/Day = 10

- **Scenario2: The Whole Year**: Avg Cancellation/Day = 2

The Wald Statistic came out to be 3.27.

**Disaster 5: Hurricane Matthew (October 1-9, 2016, Florida/North Carolina USA)**

- **Scenario1: Duration of Disaster**: Avg Cancellation/Day = 150
- **Scenario2: The Whole Year**: Avg Cancellation/Day = 20

The Wald Statistic came out to be 2.33.

The critical value for the was 1.96 for 95% confidence interval. Thus, we can reject the null hypothesis and say that Delays increases as the intensity of natural disaster increases. Furthermore, we can see that as the intensity of the disasters increase, the flight does not get delayed, but gets cancelled. Even further, as the intensity gets even more, the Wald statistic increases. Thus we can infer from the study that Delays/ Cancellations increases as the intensity of natural disaster increases.

## IV. Prior Work

The Airline Performance Dataset by Bureau of Transportation Statistics is very popular among data scientists to study the impact of the flight delays on airline transportation in the United States. A thorough background research was done, to get inspiration for this study. One major source of inspiration was the Total Delay Impact Study.

In [1] the author compares different approaches for predicting delays using models such as Markov Jump Linear System, Linear Regression, and the Neural network architectures. By contrast, we have just focused on using linear regression with more useful features. In [2] they analyzed the properties of flight delays, including the distribution of total delays, the dependence on the day of the week and the hour-by-hour. This paper extends the study by incorporating more analysis of flight delays and the distribution of those across airports. In [3] statistical models of airport delay and single flight arrival delay were developed. Concepts like Multivariate regression, ANOVA were used to detect the pattern of airport delay, aircraft arrival delay and schedule performance. But, we did study on the pattern of delay with respect to distance and airlines. One major source of inspiration was the Total Delay Impact Study.

## V. Future Work

Our model works based on the assumption that the delay is linearly dependent on the input features but because these assumptions do not hold in real life flight delay data, our delay prediction can further be improved using more sophisticated models such as Polynomial Regression or Gaussian Mixture Models. These models will allow us to incorporate the non-linear relationships between the delay and features.

Also, since we work with limited no of features for our implementation of Linear Regression, addition of more insightful features such as National/Cultural Holiday data across the US and hour of day can help us further improve our prediction models.

To improve the accuracy of our time-series prediction, we can use advanced machine learning techniques such as Hidden Markov Model, Recurrent Neural Networks or Gaussian Processes.

For hypothesis testing involving airline schedules of more than 2 flights we can make predictions for delays and losses that are caused across long routes of connecting flights. There is also scope for extracting deeper insights by correlating weather condition data with flight delay data.

## References

[1] Karthik Gopalakrishnan and Hamsa Balakrishnan, Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017); A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks

[2] Pablo Fleurquin, Jose J. Ramasco, Victor M. Eguiluz; Characterization of Delay Propogation in the US Air Transportation Network; Instituto de Fisica Interdisciplinaria y Sistemas Complejos IFISC (CSIC-UIB), Campus UIB, E07122 Palma de Mallorca, Spain; Innaxis Foundation & Research Institute, Jose Ortega y Gasset 20, E28006 Madrid, Spain

[3] Yuqiong Bai (2006); Analysis of Aircraft Delay and Airport On-Time Performance

[4] Alice Sternberg, Jorge Soares, Diego Carvalho, Eduardo Ogasawara (2017); A Review on Flight Delay Prediction, CEFET/RJ.

[5] `https://www.statlect.com/fundamentals-of-statistics/normal-distribution-maximum-likelihood`