Flight Delay Analysis

Group - 28

Ramakanth Vemula
Krishna Anantha Padmanabhan

Big Data Analytics and Text Mining

Topics of Discussion

- Why flight data?
- Data Information
- Visualization - Exploratory Data Analysis
- Algorithms
- Future Work

Why Flight Data?

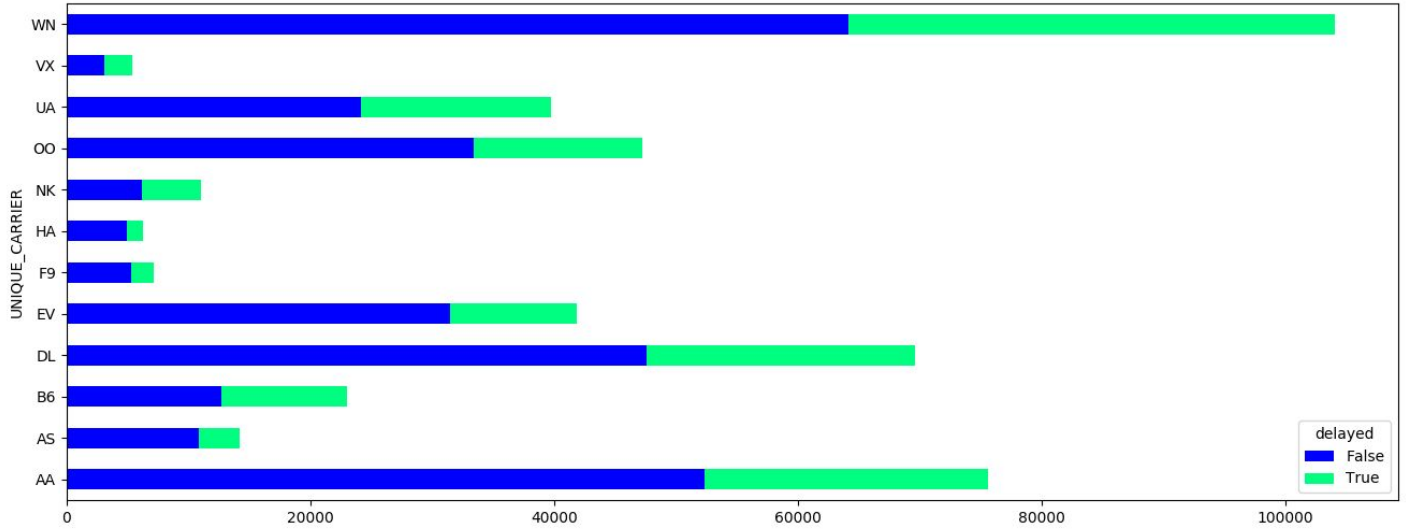
- Airport Authorities and carriers can use it to improve and streamline their services.
- Gain understanding into previous delays and factors affecting delays
- Can be used to appropriately move resources around to prevent delays
- Customers can use it to decide which flights to take or avoid

Data Information

- US Bureau of Transportation Statistics
- Years - Jan, 2014 to Dec, 2016
- Nearly 17 million records with comprehensive information
- Data contains - Timestamp, Carrier Info, Airport Info, City Info, Delay Info (Departure Delay, Arrival Delay, Weather Delay, Carrier Delay, etc)

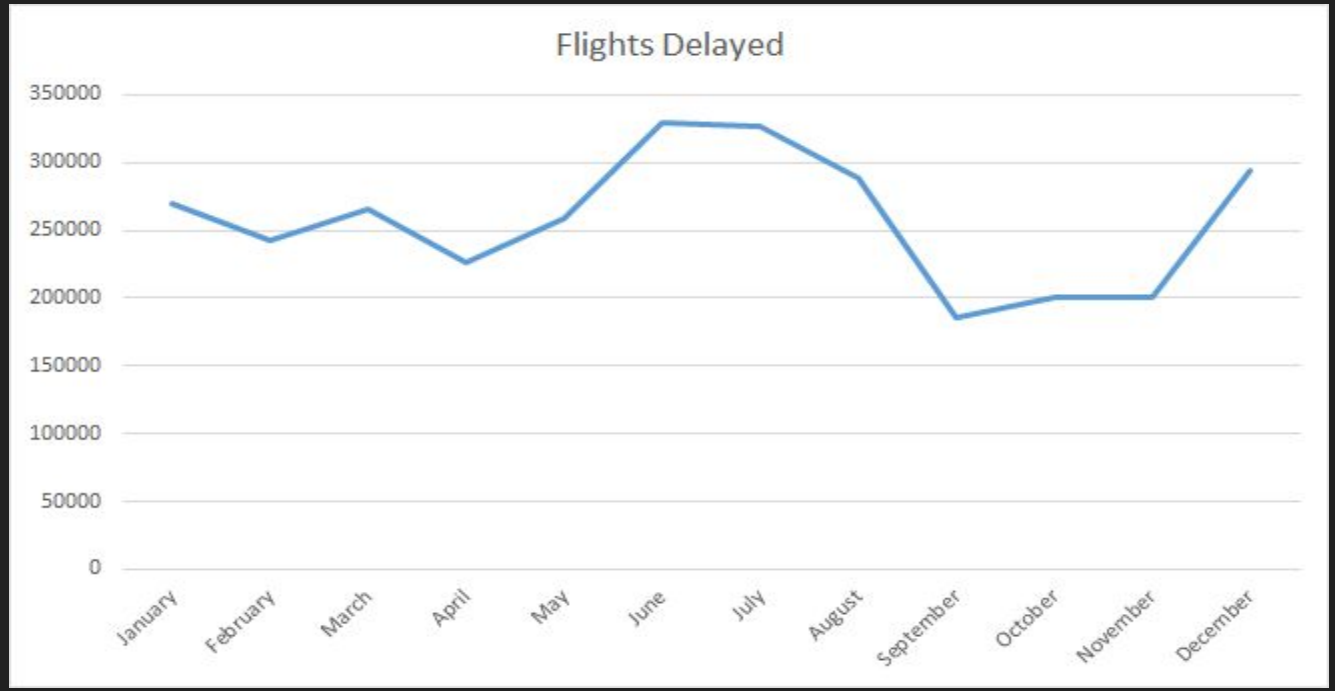
Visualization - Exploratory Data Analysis

Carrier Flights On Time/Delay vs. Total Flights



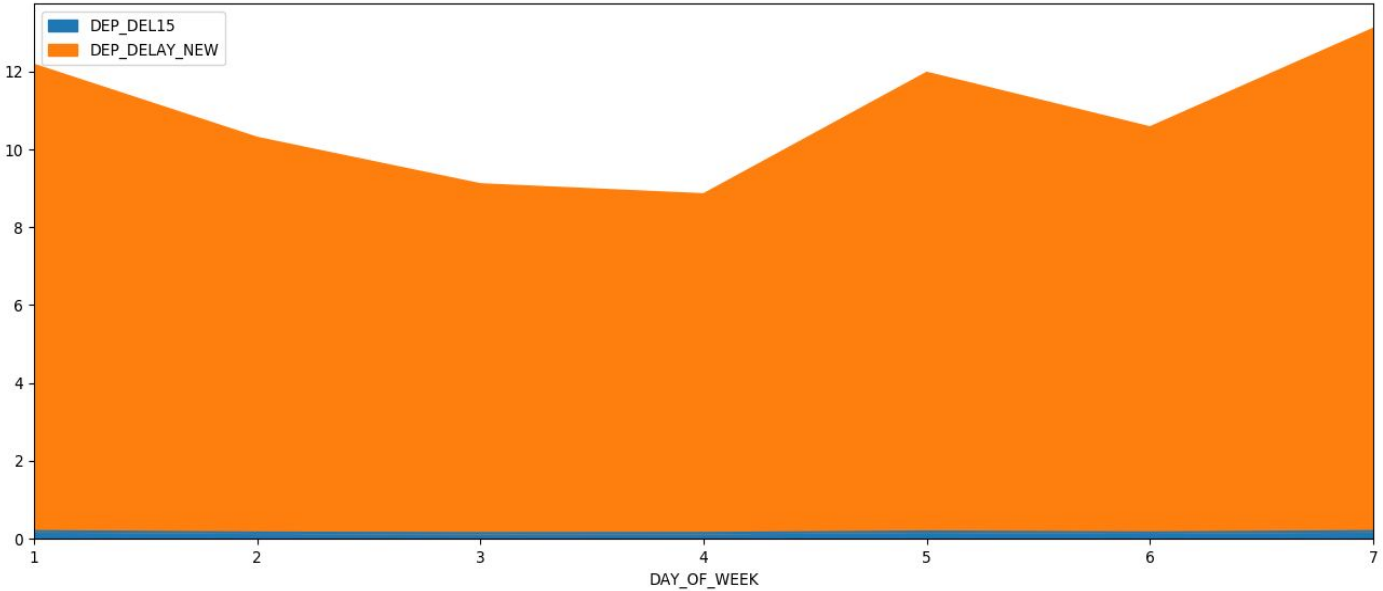
Visualization

Delays on months



Visualization

Delays on Days of the week



Algorithms

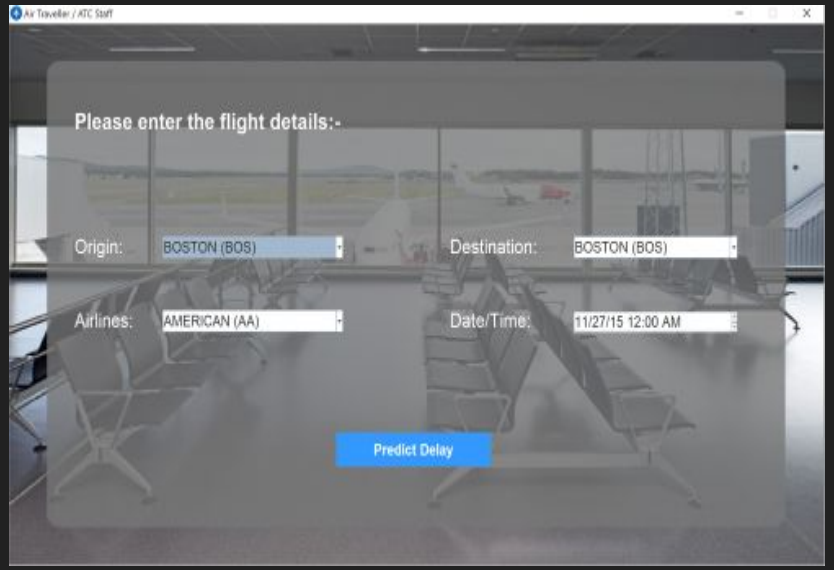
Implementations were done using Scala, Spark, ML libraries on Spark, and Python

Algorithm - Accuracy Percentage

- Logistic Regression - 83.145 %
- Naive Bayes - 81.119 %
- Decision Tree - 84.131 %
- Random Forest - 80.099 %

Future Work

- Front end interface that lets users select flights and displays delay - Yet to complete
- Analysing patterns in data that can help ATC



Thank You!