Deloitte.

Flight Delay Analysis

By: Reece Boyce

Business Problem

Increased Expense for Crew, Fuel, and Maintenance

Delayed flights cause more expenses for relocating staff
Fuel expenses increase due to idle time

Cost of Delays on the Industry

Estimated cost of $8 billion per year for the industry*
Estimated cost of $17 billion per year for passengers*

Decreased Customer Satisfaction

Airlines depend on repeat customers and word of mouth
Competition is fierce in the industry, airlines that experience delays frequently lose business.

Customers Do Not Like Uncertainty

Delayed flights cause stress and anxiety for travelling, especially for those customers traveling with connecting flights

*According to 2010 study commissioned by the Federal Aviation Administration

Data Gathering

Full Dataset:

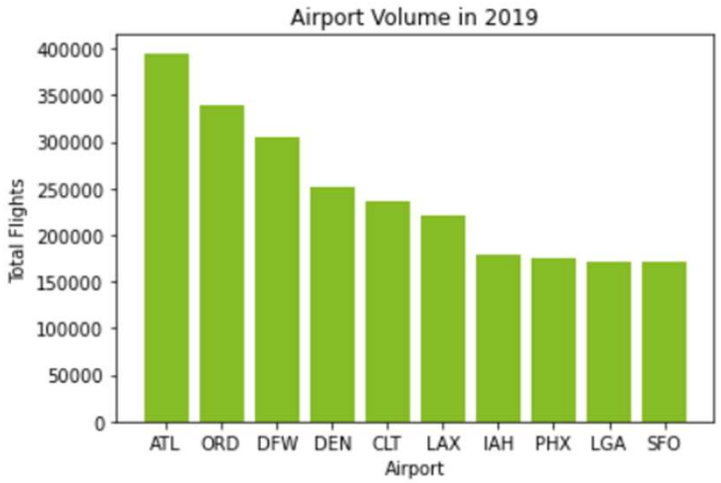
- Source: US DOT Bureau of Transportation Domestic Flight Data
- Includes Data from 2009 – 2019
- +4 million flights per year (40 million in total)

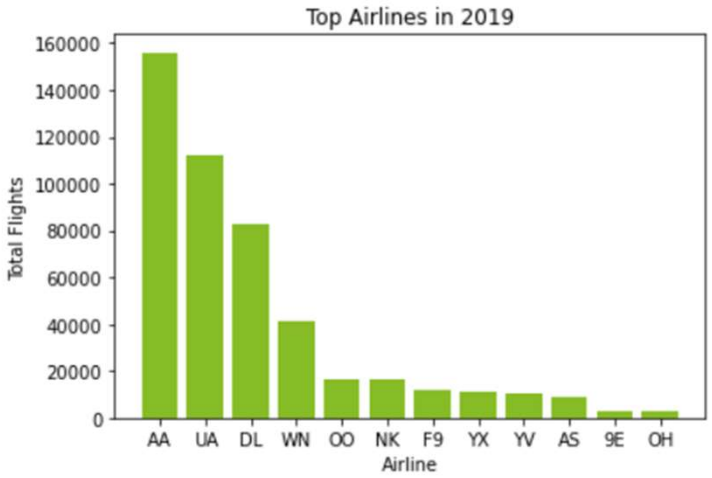
Selected Data:

- Top domestic carriers that travel nationwide
- Top 10 airports by volume
- 472,000+ data points used in analysis
- 2019 data used in analysis

Variables:

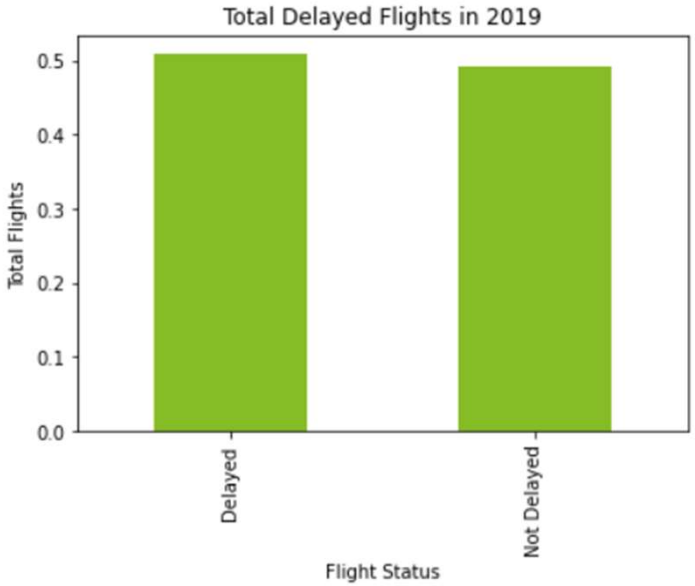
- Year
- Month
- Day of the Week
- Carrier
- Flight Number
- Departure/ Arrival times
- Origin
- Destination
- Departure time
- Time to Taxi out/in
- Wheels on/off
- Distance





Model Selection

Target Variable: Delay



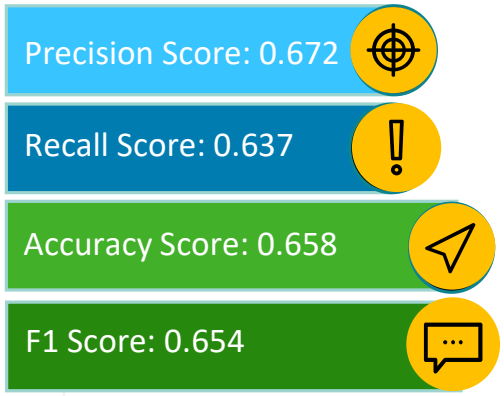
Problem: Binary Classification

	Computational Needs	Advantages
Logistic Regression	• Training time: 1.92 seconds • Prediction time: 0.01 seconds	• Less tuning needed, easier to perform on large datasets.
Random Forest	• Training time: 41.26 Seconds • Prediction time: 0.77 Seconds	• Can handle missing values. • Performs better with more explanatory variables.
XGboost	• Training Time: 17.05 Seconds • Prediction Time: 0.32 Seconds	• Less randomization than RF • More parameters than other two methods

Modeling

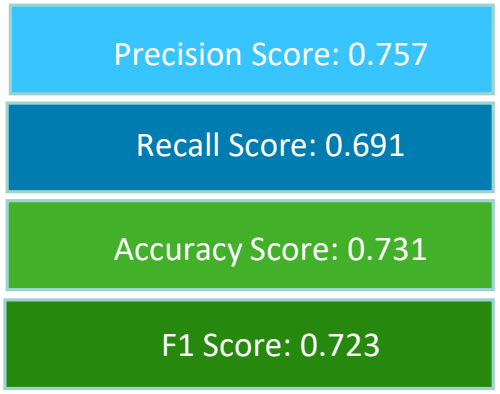
Logistic Regression

True Delays: 47,840
True Delays Predicted: 30,498



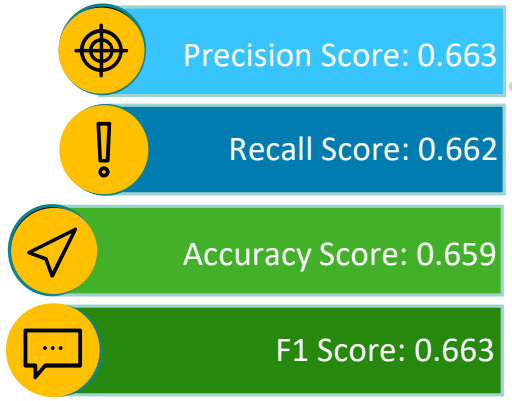
XGboost ★

True Delays: 47,840
True Delays Predicted: 33,064



Random Forest

True Delays:47,840
True Delays Predicted: 31,717



Recommendations and Way Forward