Avoiding Weather Delays at Miami International

It is estimated that a delayed flight cost an airline a minimum of $80.52 per minute. Airlines develop sophisticated maintenance schedules for aircraft not only to ensure safety for the passengers and crew, but also to avoid delays. However, weather has always been a travel x-factor. This purpose of this project is to develop a schedule to avoid weather delays when flying to Miami International Airport (MIA) during the summer travel months.

1. Data

Bureau of Transportation Statistics (BTS) is part of the Department of Transportation. BTS is considered one of the best sources for statistics for commercial aviation. Since 1987 the BTS has been collecting data of on-time performance. Since June of 2003, BTS has collected why a delay happened. I used data from the BTS starting with June of 2003 to present.

2. Method

The process planned for this project is to collect the data, prepare for analysis, then use several machine learning models to see which model returns the best results.

3. Data Cleaning

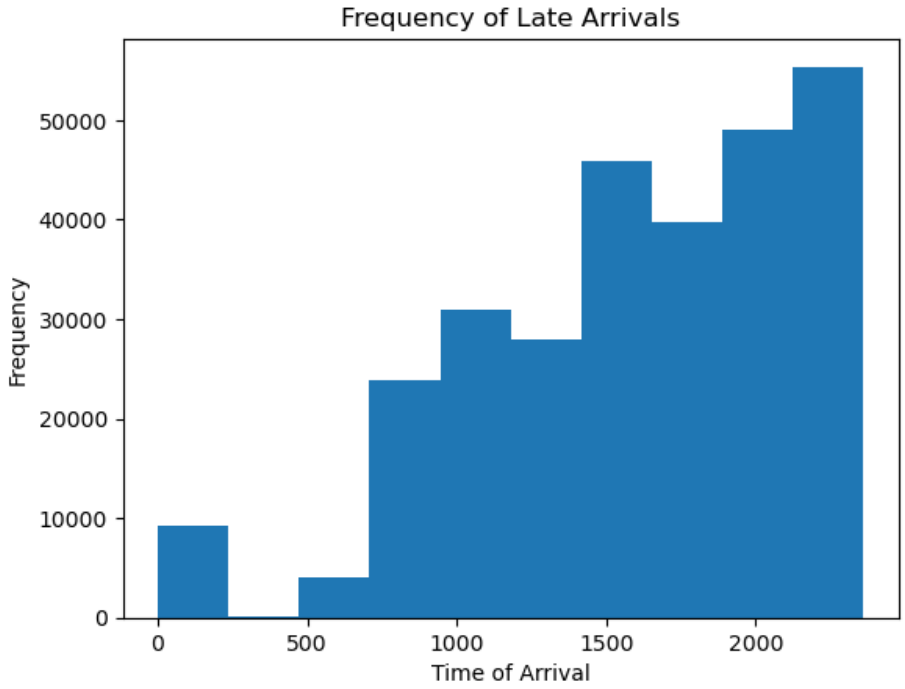
Data Cleaning

The BTS produces the on-time data monthly, the first step was to download all the files individually then combine them all to have one data set. Next since I was only dealing with flights in and out of MIA, I removed all flights that dealt with other airports. I then eliminated columns that dealt with flight cancellations, diverted flights and different airline identification data.

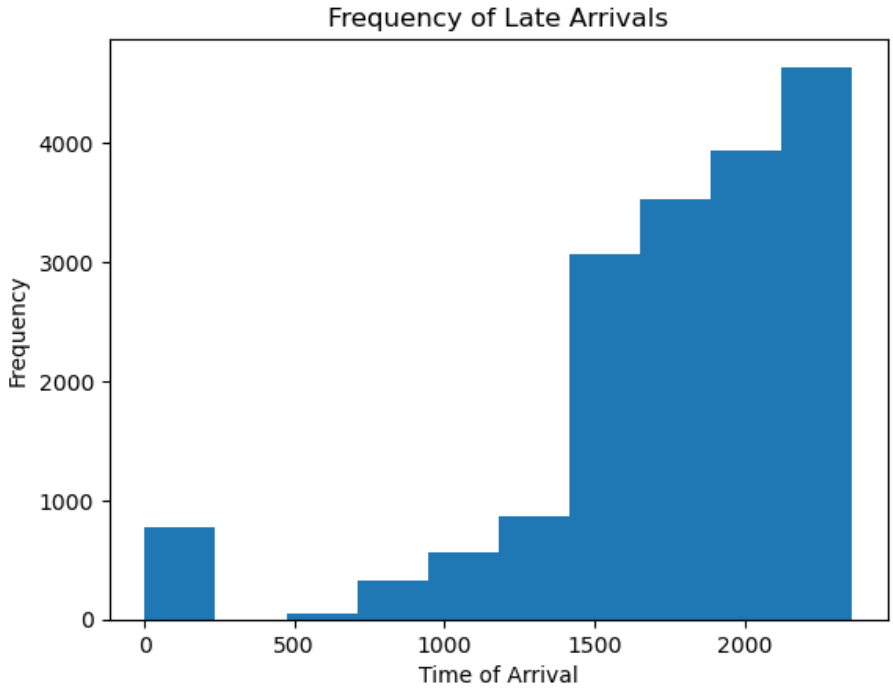
4. Exploratory Data Analysis

EDA

Here is two graphs comparing late arrivals. Below isva graph of all late arrivals.



This graph is late arrivals due to weather.



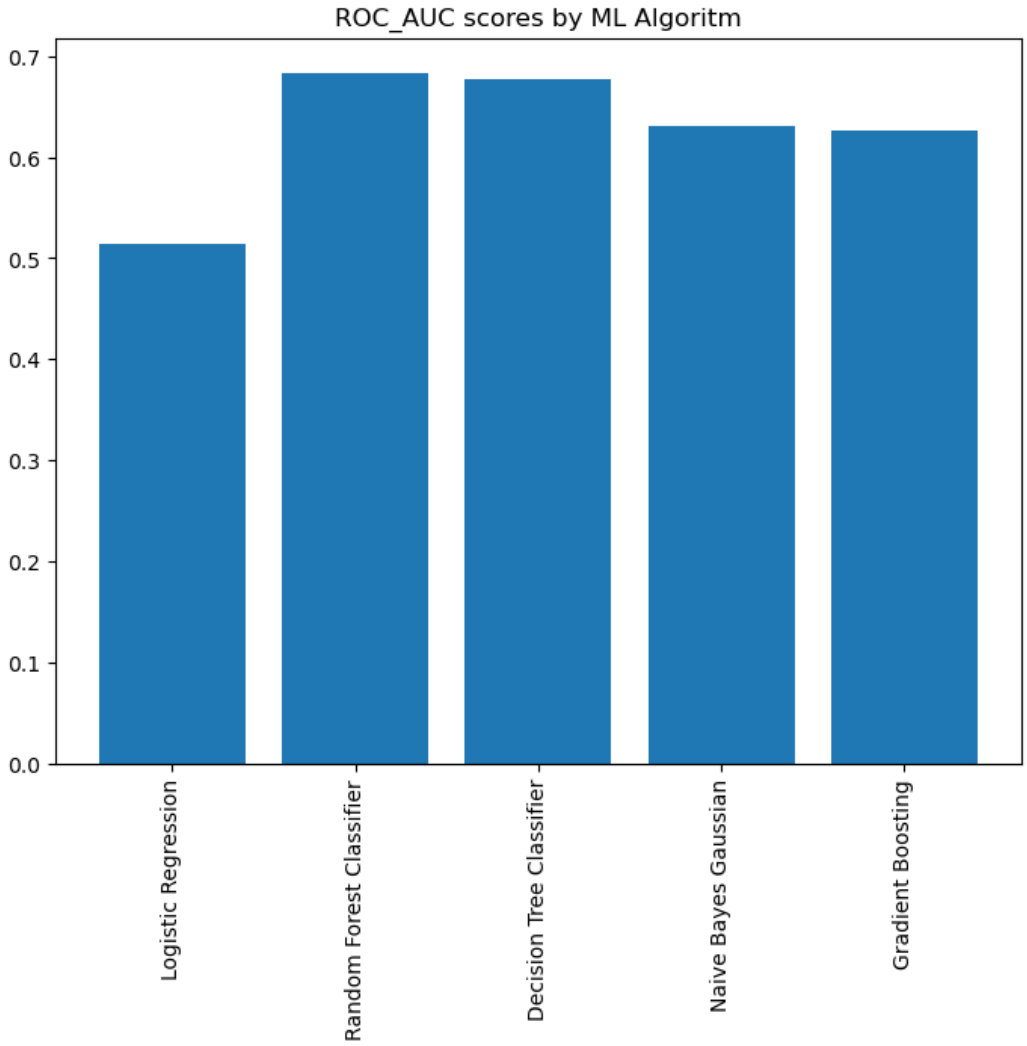
4. Algorithms & Machine Learning

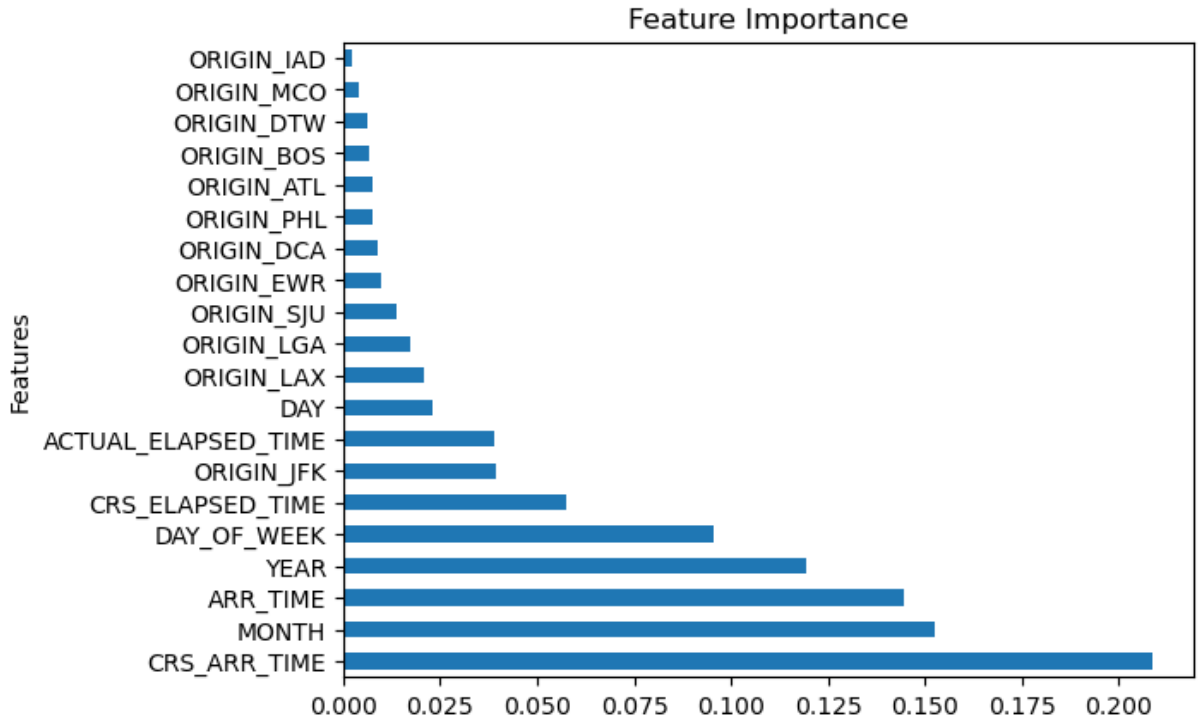
Machine Learning

Originally, I planned to use Logistic Regression, Random Forest Classifier, Decision Tree Classifier, Naïve Bayes Gaussian, Gradient Boosting and Singular Value Decomposition. However, Singular Value Decomposition was too computationally expensive, so I dropped it.

5. Initial Results

After processing the data through all the different machine learning algorithms,





The Random Forest Classifier and Decision Tree Classifier were the two highest performers with a score of .6837846369292927 and .6764139530069186 respectively. I tried to hyper tune both algorithms, but neither were able to increase performance with hyper tuning.

Further analysis determined that the scheduled arrival time was the most important feature followed by the month. I believe that using this model would help develop a schedule that avoids scheduling arrivals into MIA during times that a weather delay is likely.

6. Future Work

- Develop a model for departing flights.
- See if using time intervals opposed to exact times create more predictability.
- Apply the models to other airports to be able to create a scheduling network that meets customers demands as well as avoid weather conditions.

7. Contributions

- Raghunandan Patthar