IST 718:Big Data Analytics
Predicting Airtime Delays

Raghav Raheja
Pranay Lulla
Vaibhav Nigam
Palani Mutt

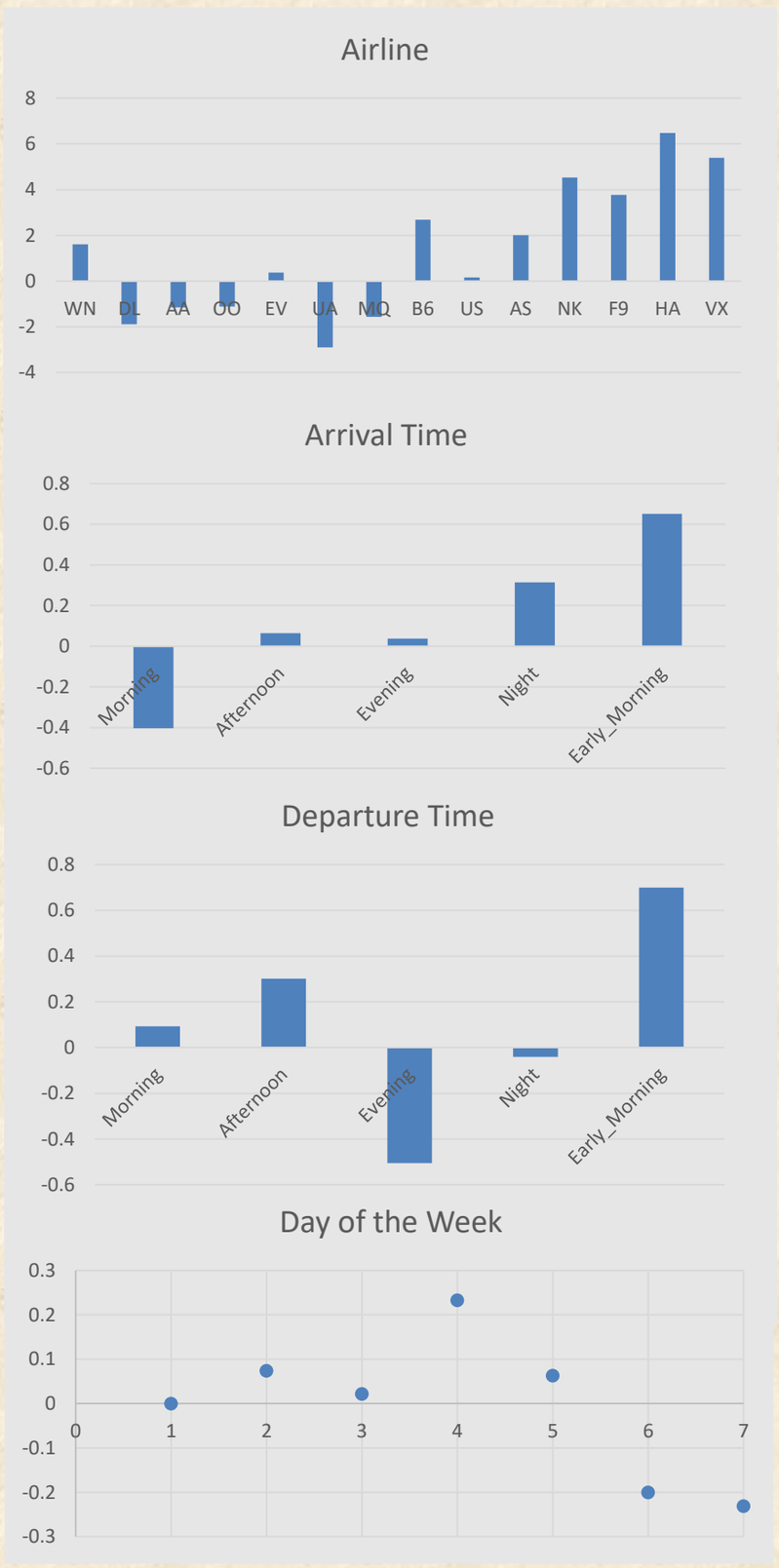
Problem Statement

- In 2015, around 21.72 % flights get delayed due to different reasons [1]
- Thus, we decided to predict what are the factors that effect the airtime delay, and identify important features causing delay

Data Description

- Our Dataset consist of 18 features and 5.7M records. It consist of data of year 2015 flight delays across United States. It contains a calculated columns Airtime_delay which is used as a predictor variable.

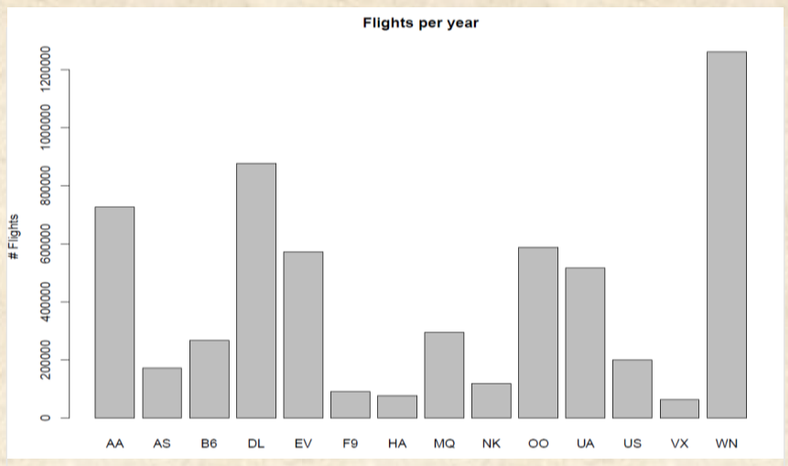
Feature Coefficients

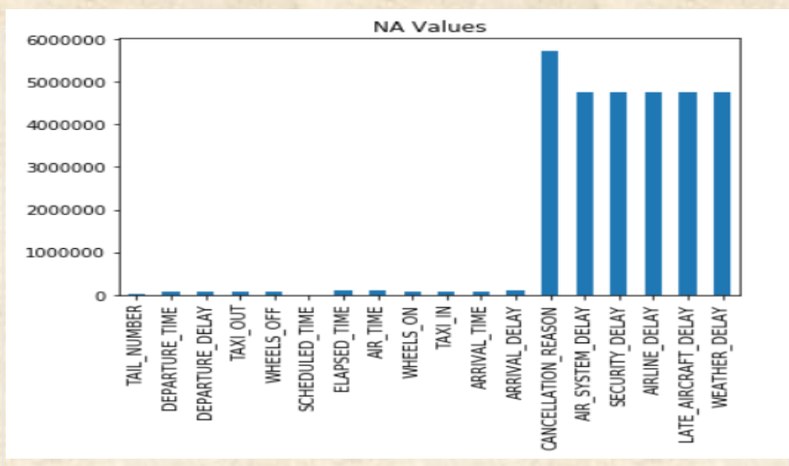


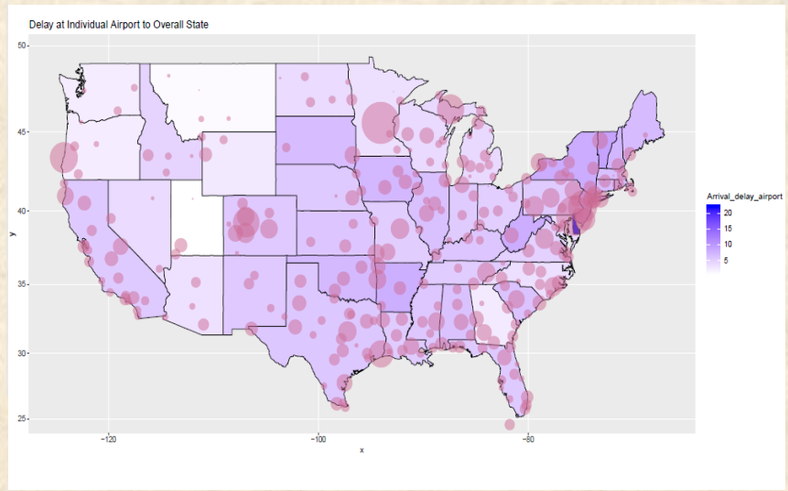
Conclusion

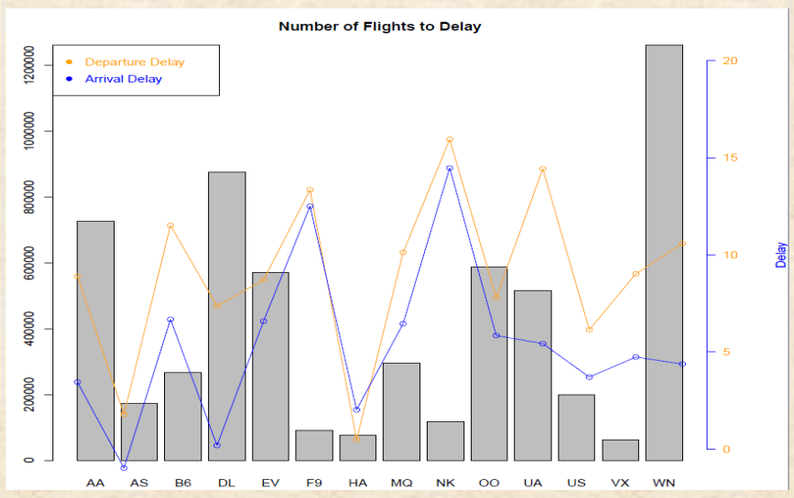
- We found that the best model in our case out of all the models that we ran was linear regression without the regularization parameters
- We have used twelve features to predict the airtime delay with the best root mean square error of 8.7100 on the testing data.

Data Exploration









Model Description

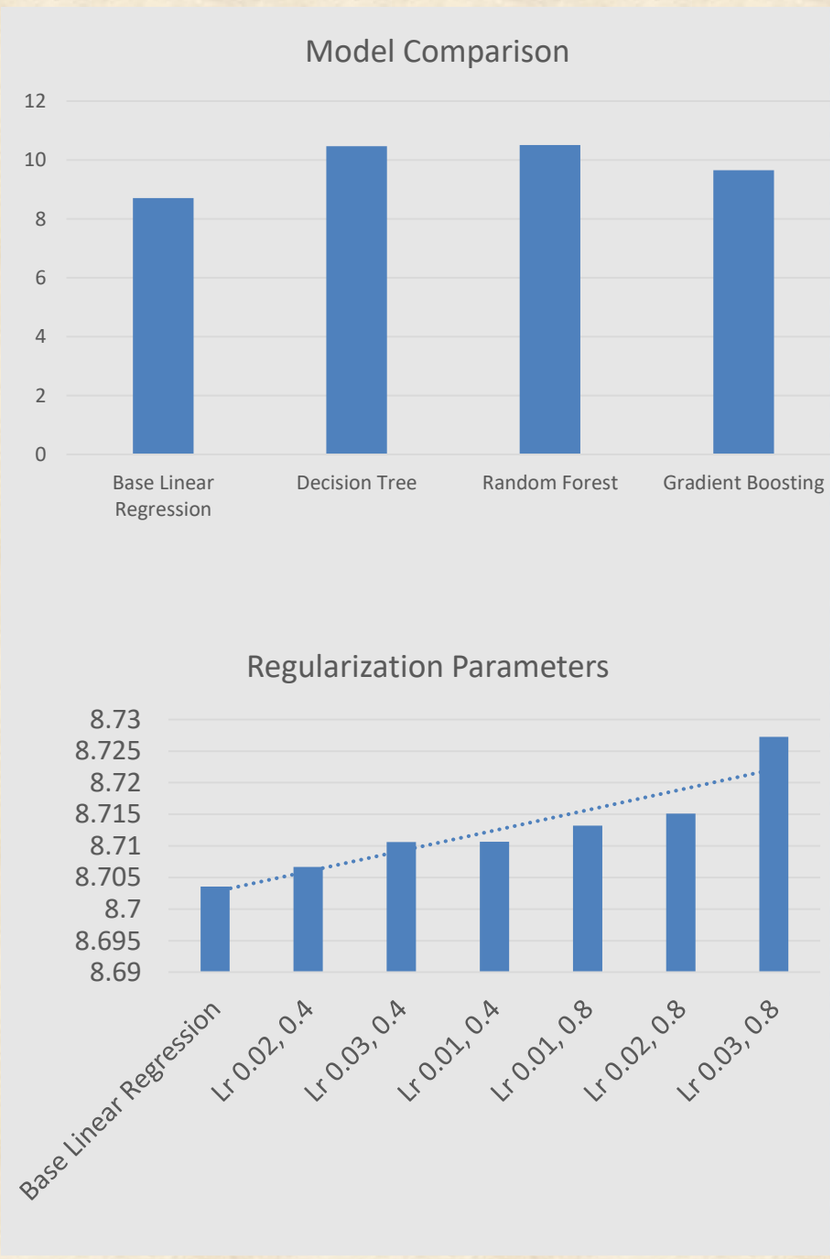
Model	Features	Techniques	Evaluation
Linear Regression	Month	String Indexer	12.8897
Linear Regression	Month	String Indexer, One Hot Encoder	12.8715
Linear Regression	Day	String Indexer	12.8901
Linear Regression	Day	String Indexer, One Hot Encoder	12.8841
Base Linear Regression	Taxi out, Distance, Taxi in, Departure delay, Month, Day, Day of week, Airline, Departure time, Arrival time, Origin airport, Destination airport	String Indexer, Vector Assembler, One Hot Encoder, Pipeline	8.7096
Random Forest			10.5067
Decision Tree			10.4701
Gradient Boosting			9.6540

Methods

- Data Cleaning: Handling NAs and derived columns
- Feature Engineering: Converting time into categories and derive columns
- Handling Categorical Variables: Testing the best model for conversion
- Models: Performed models to predict airtime delay of the flights

Objectives

- Exploratory data analysis to analyze the dataset for summarizing the main characteristics
- Calculating the Airtime Delay by the difference of Elapsed time and Scheduled time
- Applying Regression Modeling techniques



References
[1] https://www.transtats.bts.gov/HomeDrillChart.asp
[2] https://www.kaggle.com/usdot/flight-delays