

Business Intelligence and Data Mining



**Flight delays in US
JFK to BOS**

We have attempted to predict flight delays based on traffic, time, weather, carrier details

Problem Statement

Classify if a flight from JFK to BOS is going to be delayed by > 15 minutes

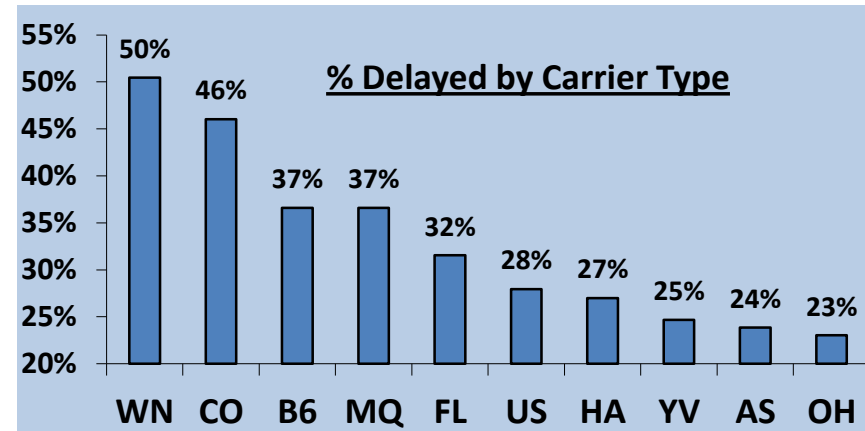
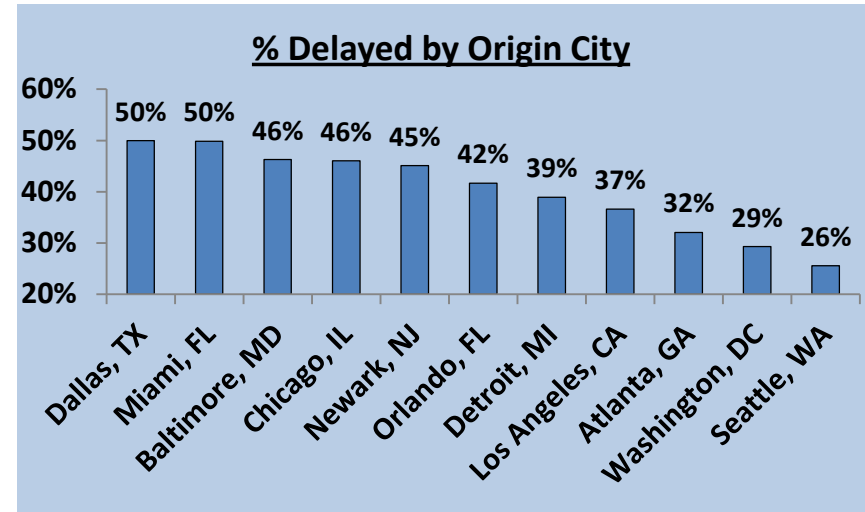
Applications of prediction

- Customers can choose flights based on sensitivity to delays
- Airlines can better optimize their operations

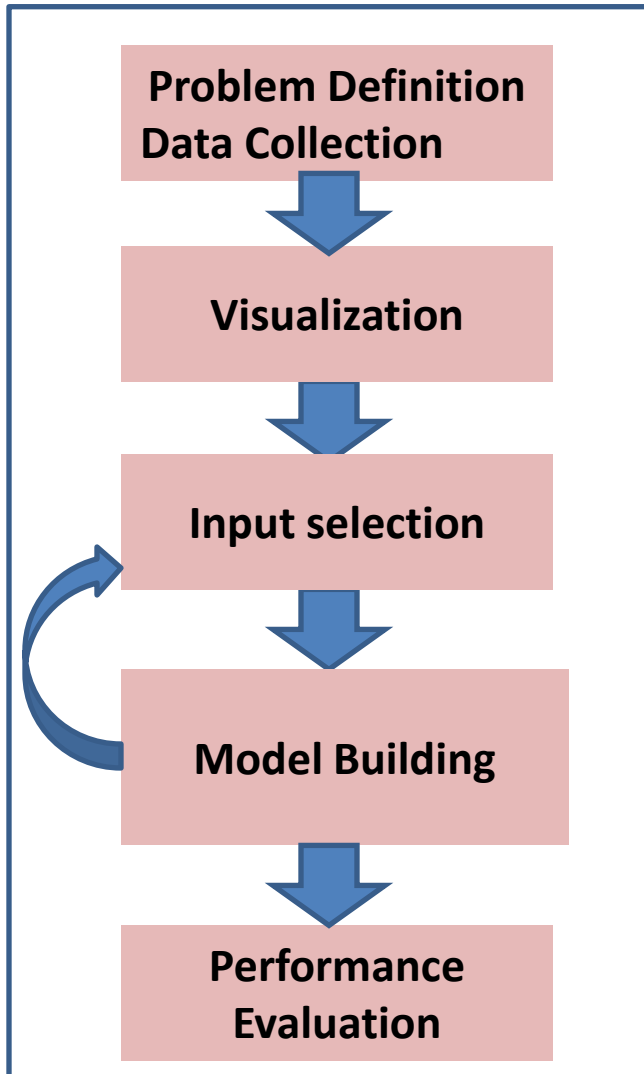
Baseline – Naïve rule

Out of ~8000 flights from JFK to BOS +2000 are delayed.

In the test data: 29%



Data and Method: We leveraged disparate data sources to build a comprehensive model

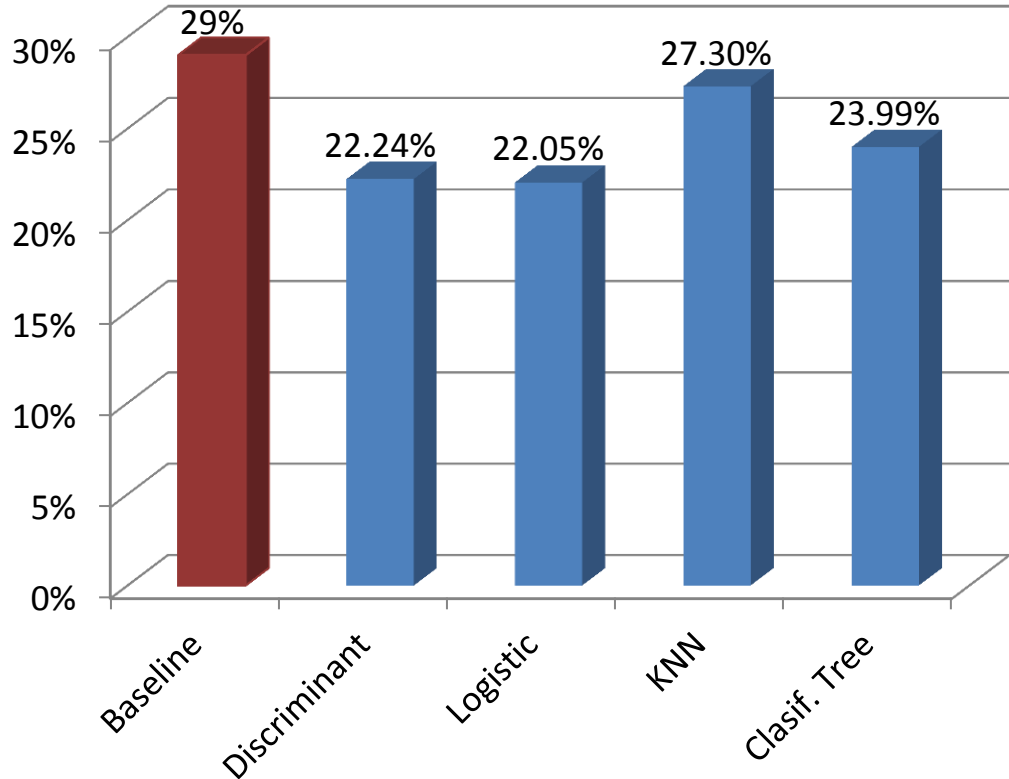


Category	Input Variable
Weather	Snow, Precipitation Temperature (Min, Max, Avg)
Time	Day Part, Day of Week
Flight	Carrier , Tail Number, Flight number
Airport Congestion	Daily # of flights arriving and departing at BOS and JFK
Travel Heavy Dates	Thanksgiving, Christmas Indicators



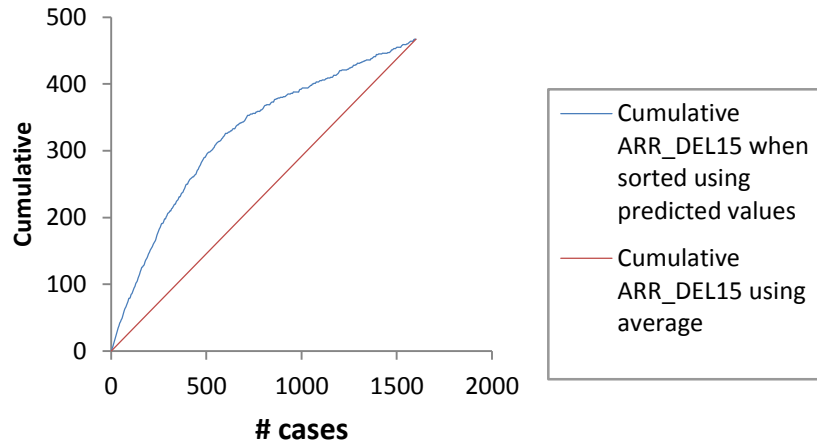
Results: The models are able to beat the naïve rule

Test Errors

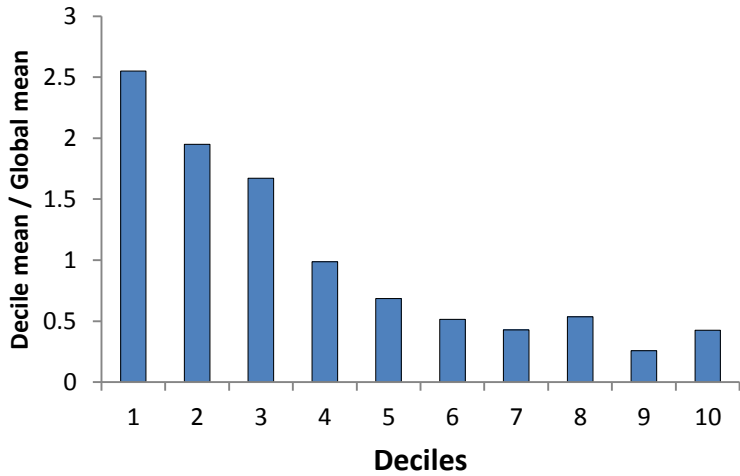


This model was built using ~8000 records of data for the JFK – BOS route and hence it is robust from a data perspective

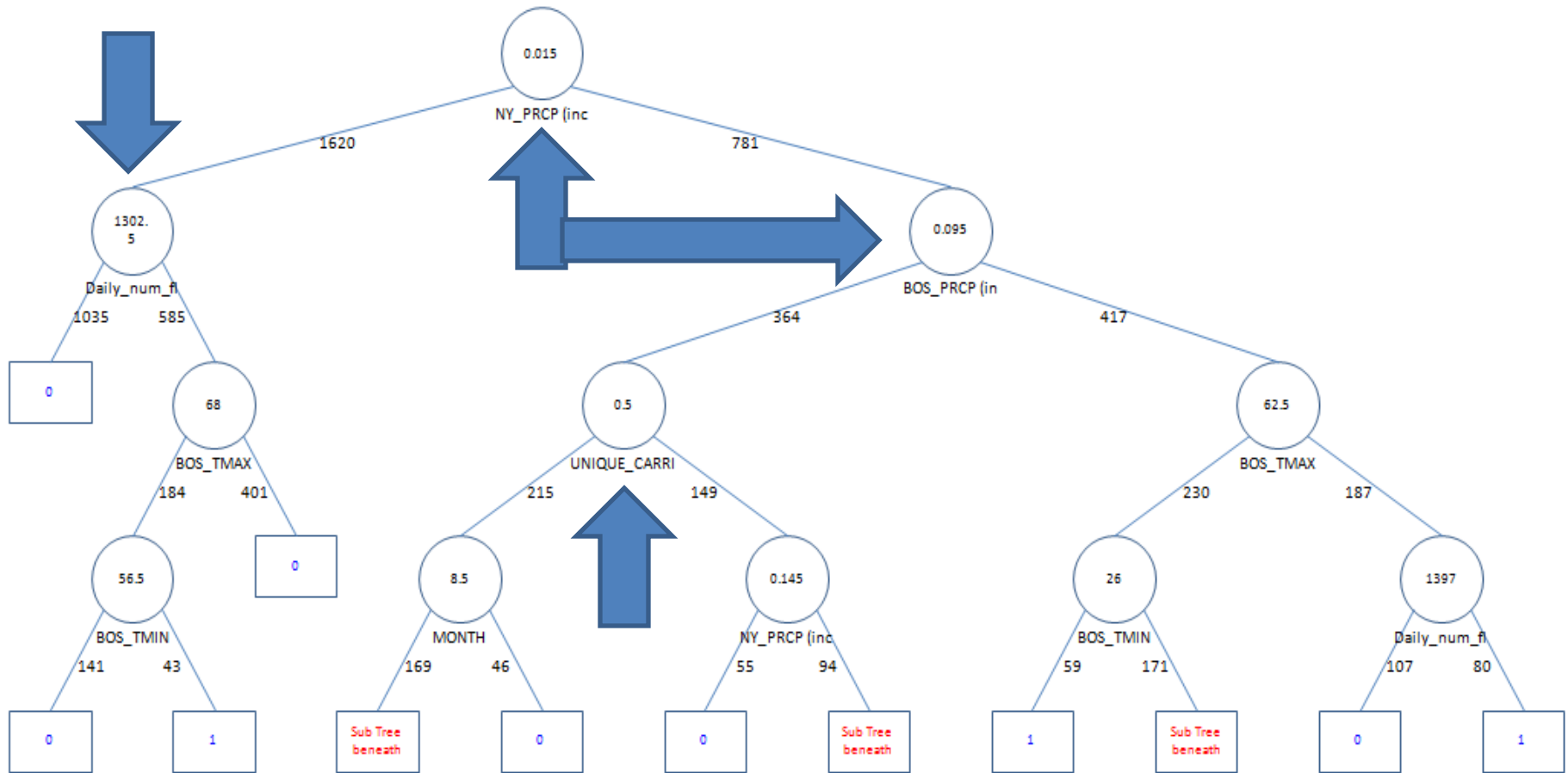
LR - Lift chart (test dataset)



DA - Decile-wise lift chart (test)



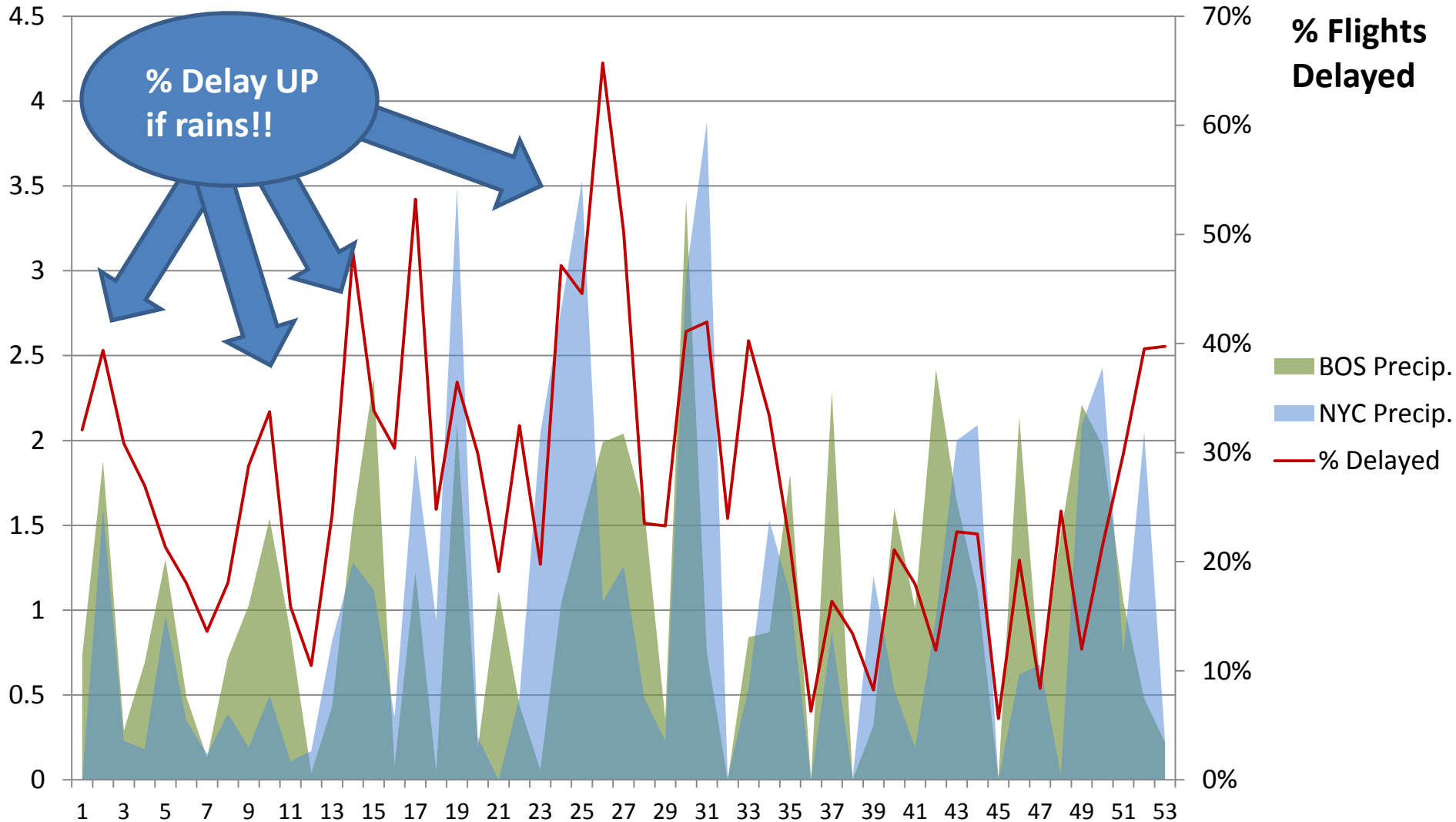
The classification tree model provides a good idea of the top predictive factors



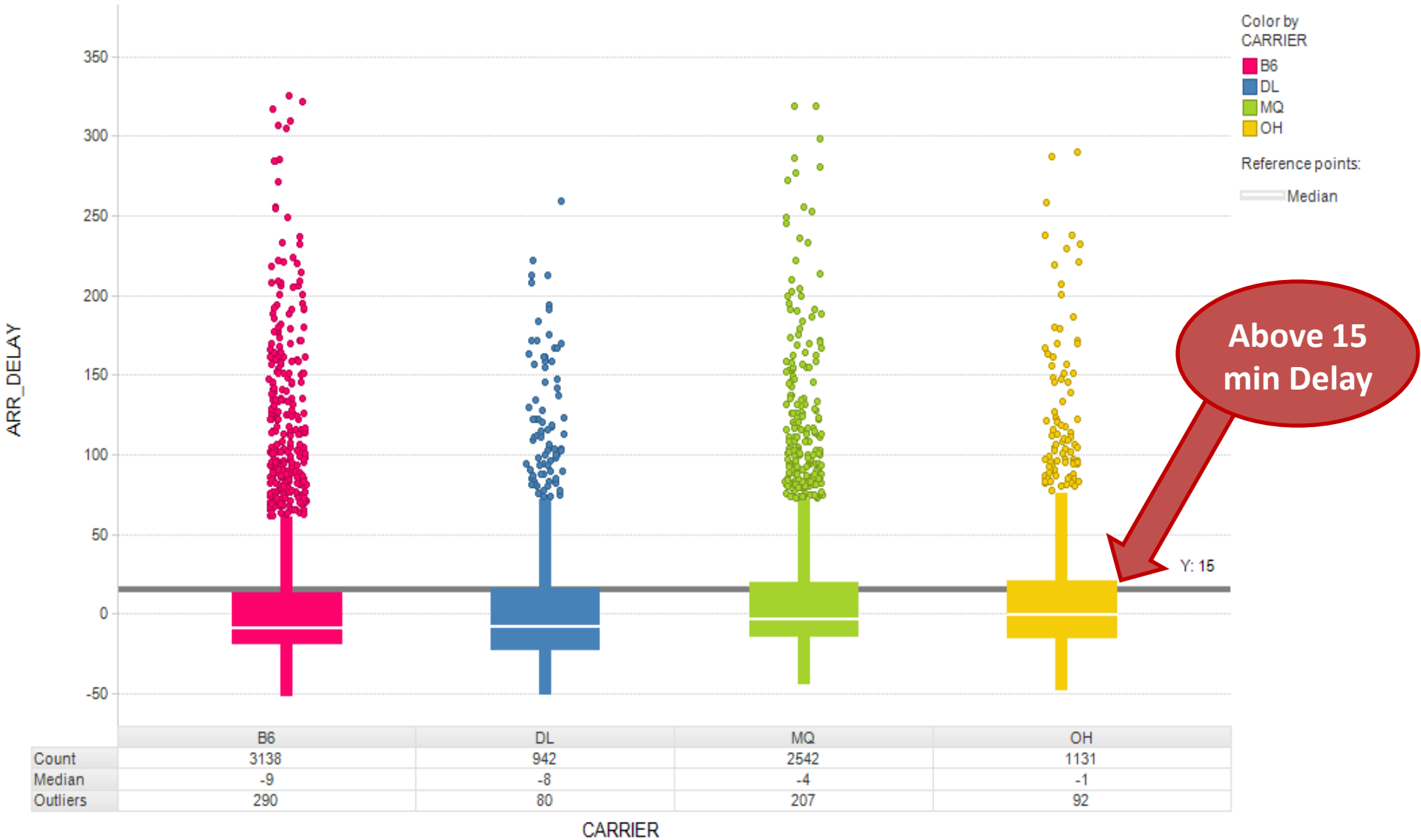
Precipitation in New York and Boston, airport congestion, max temperature and carrier are the major splitter variables.

% Delays vs precipitation. Weekly

Inches
Precipitation



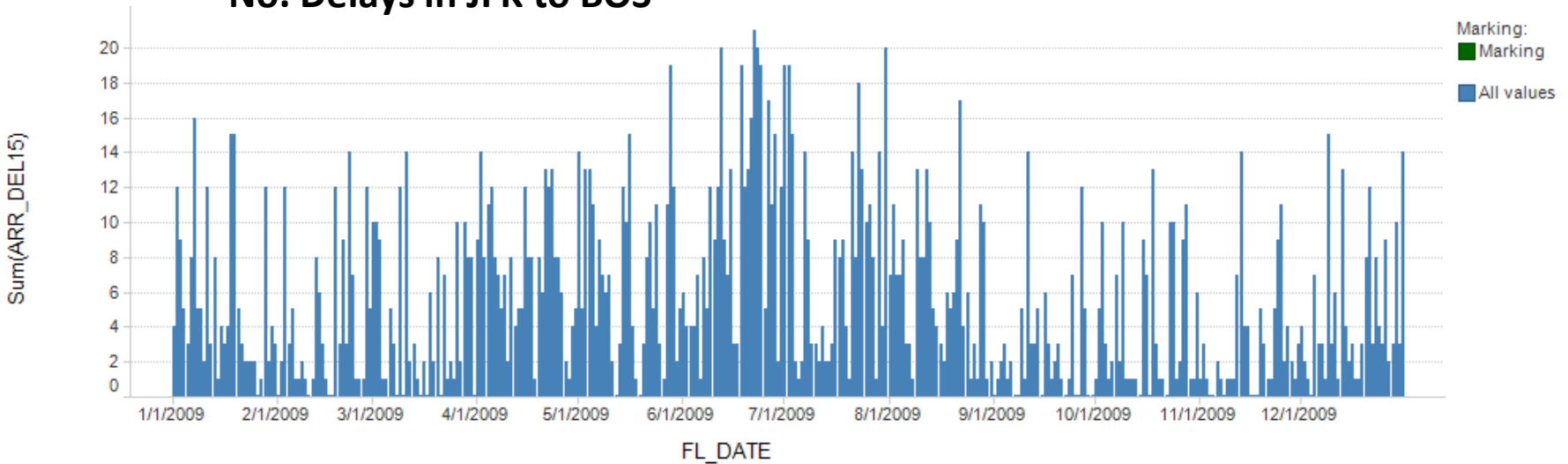
Boxplot distributuion of delays by carrier



Delays vs Congestion. Daily.

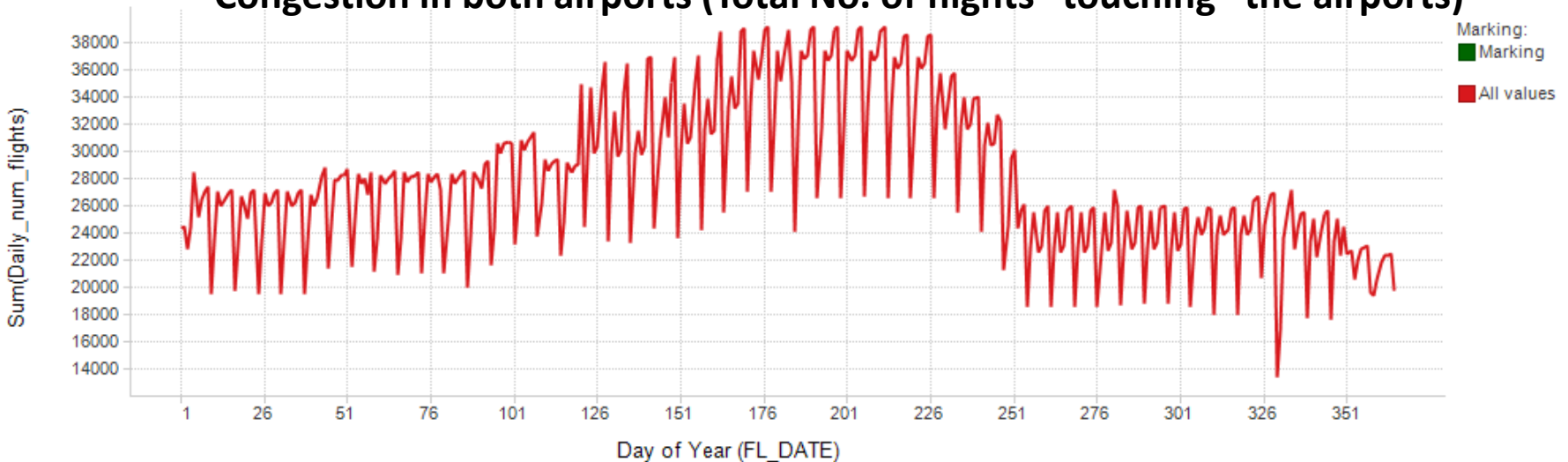
Bar Chart

No. Delays in JFK to BOS



Line Chart

Congestion in both airports (Total No. of flights "touching" the airports)



Conclusions

- The classification tree methodology was very useful to understand the important inputs
- Principal Components had hardly any effect on our dataset
- DA and LR worked the best.

Next Steps

Monetization

- Highlight the probability of delay information on travel booking sites – Yatra etc.
- Negotiate a higher margin on flights booked after checking delay information
- Mobile apps / Advertising

Model Improvement

- Include additional weather data
 - Visibility, Wind speed
- Disaggregate airport congestion data (different runways, etc...)
- Military maneuvers, exhibitions
- International flights