



Trends Contributing to Flight Delays

Gordon Duncan
Kenneth Casimiro
Sri Shatagopam
Haoliang Wei
Hao Yu
Yuqi Su

Motivation

- ❖ Flight delays are irritating
- ❖ Explore a flight delay dataset to discover correlated trends for delays.
- ❖ Visualize these trends and create a delay prediction model using these features.



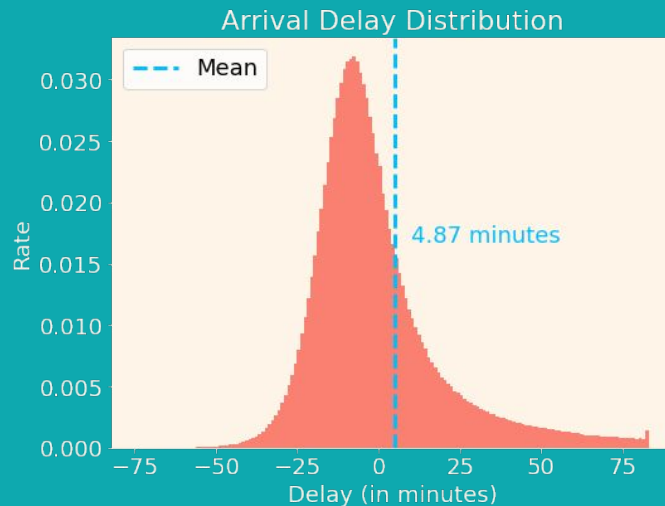
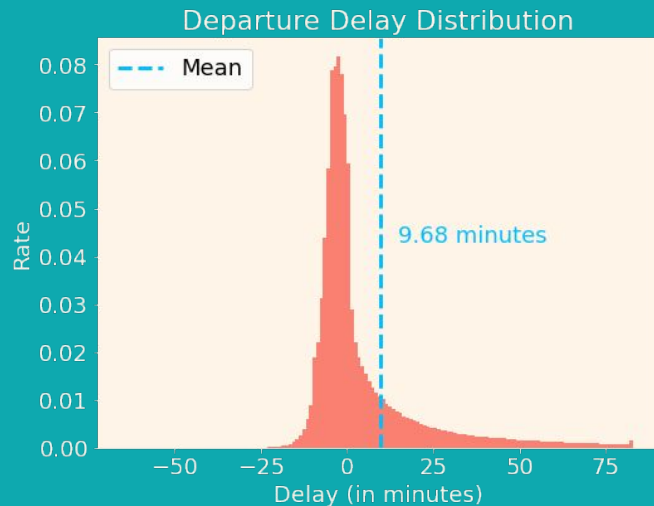
Methodology

- ❖ We use three connected datasets from Kaggle (flights, airline, airport).
- ❖ Limited to flight delays in the US for 2015.
- ❖ Visualize these trends and create a delay prediction model using these features.

kaggle™

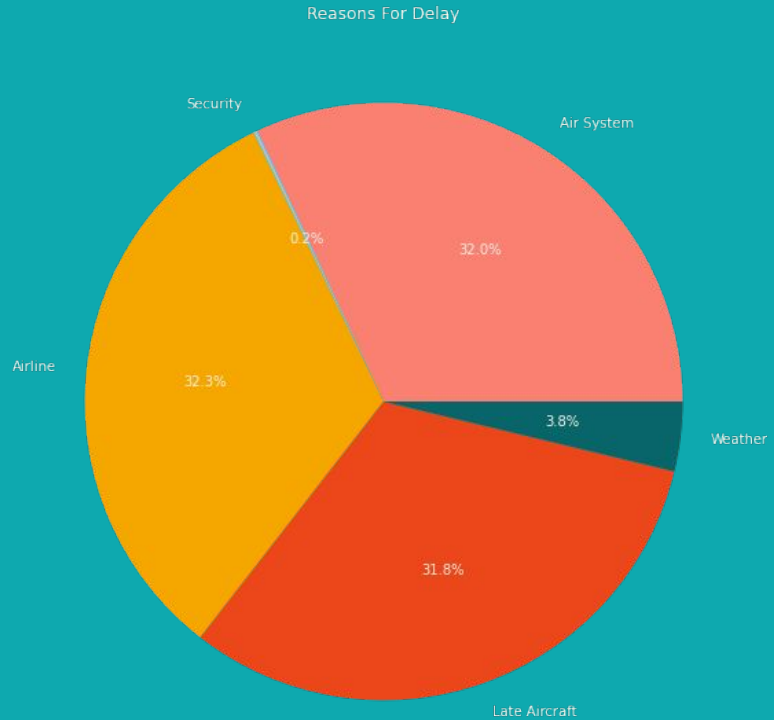


Exploratory Analysis



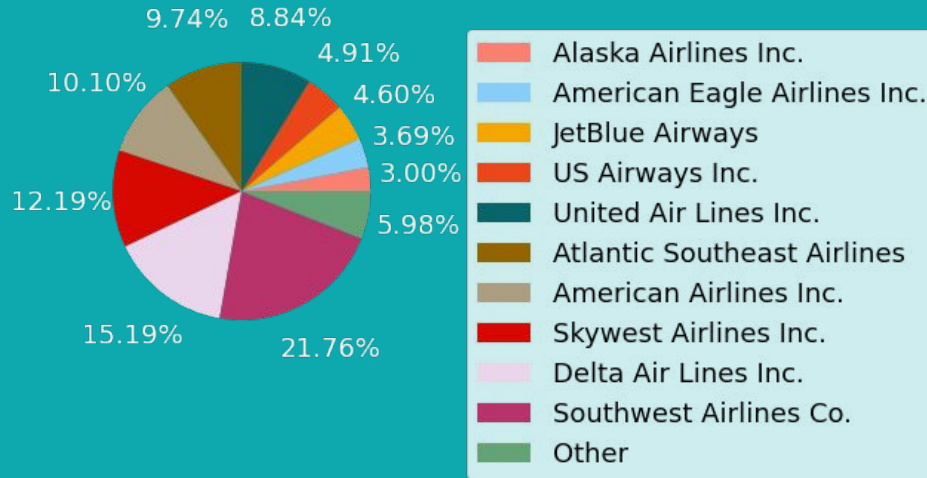
Delay Explanations

- ❖ Explanations are sparse but worth mentioning.
- ❖ Nearly evenly split between system failure, airline, and late arrival.



Limitations

Airline Representation



- ❖ Airlines are not evenly represented.
- ❖ Some airlines are still responsible for more delays.

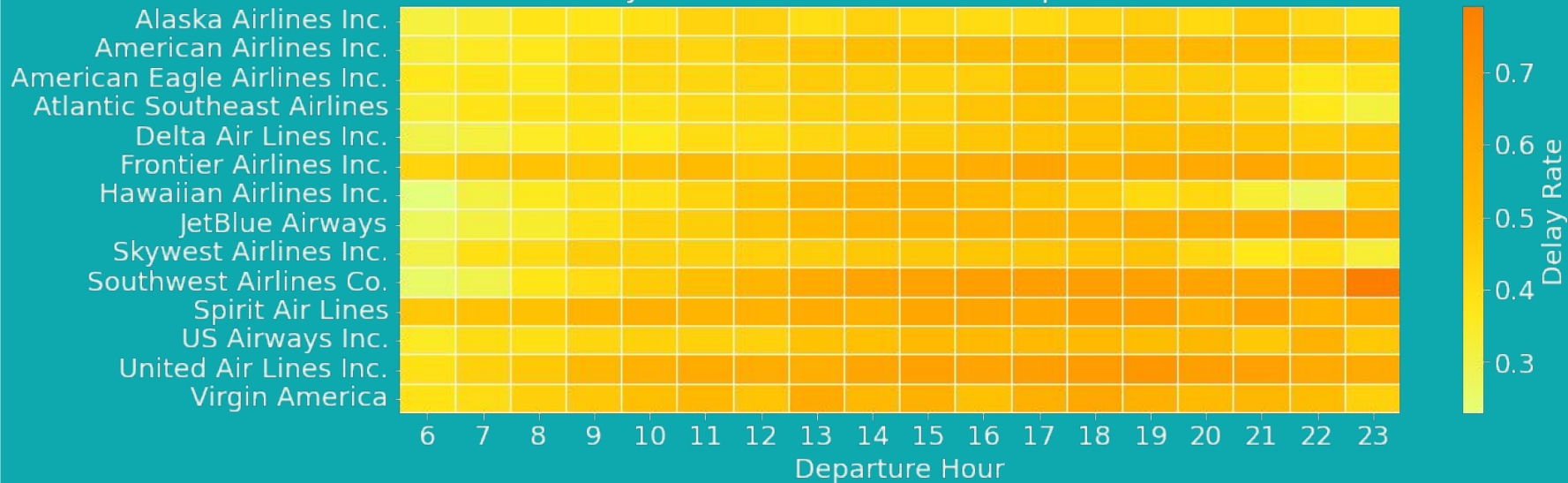
Delay per Airline

- ❖ E.g. Delta has 2nd highest share, but 2nd lowest delay proportion.
- ❖ E.g. Spirit has highest delay rate but very low representation.



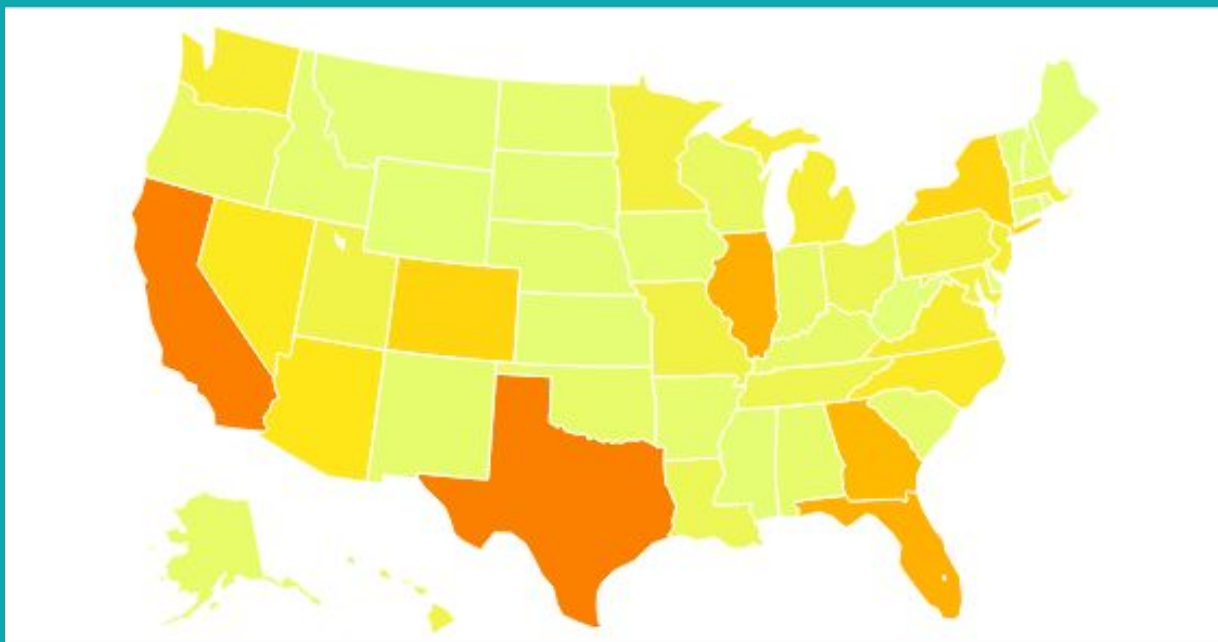
Airline Delays on Departure

Delays Rate vs. Airline and Departure Hour



Delay per State

Density of Flight Delays for U.S. States



Delay rate

300k

250k

200k

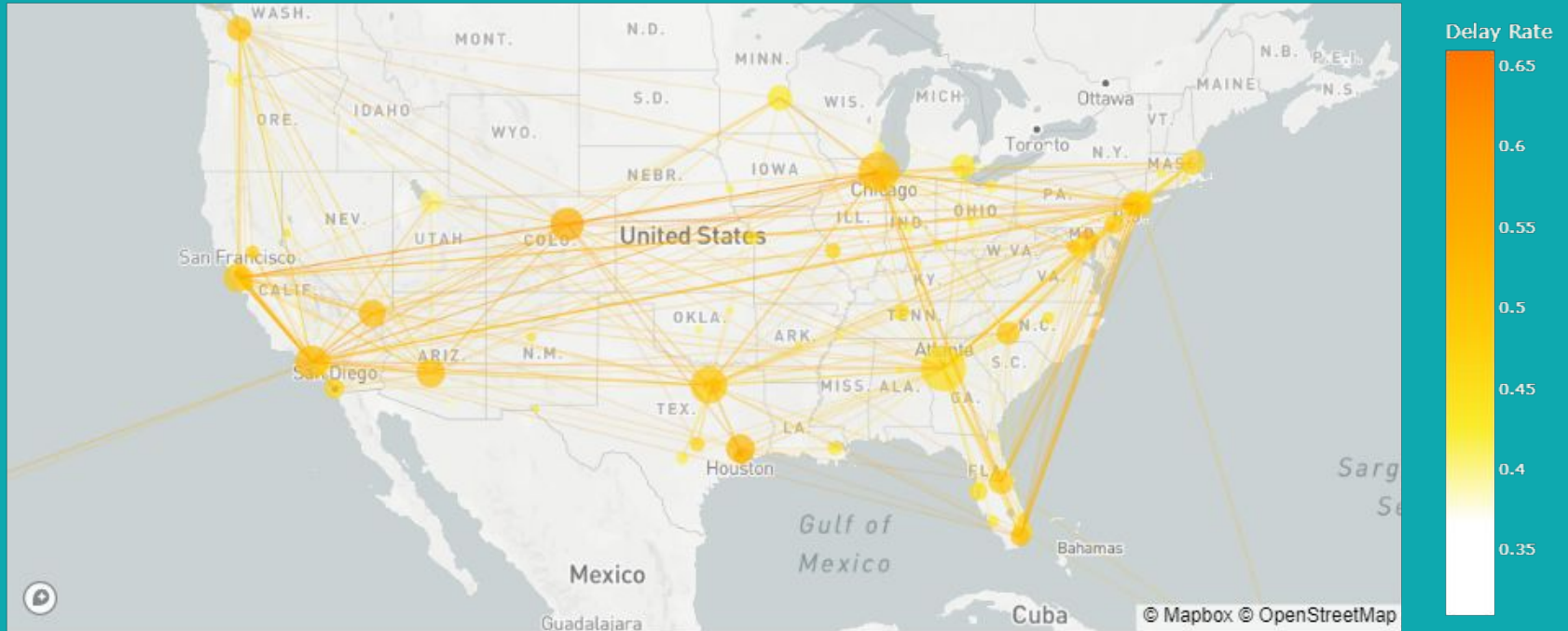
150k

100k

50k

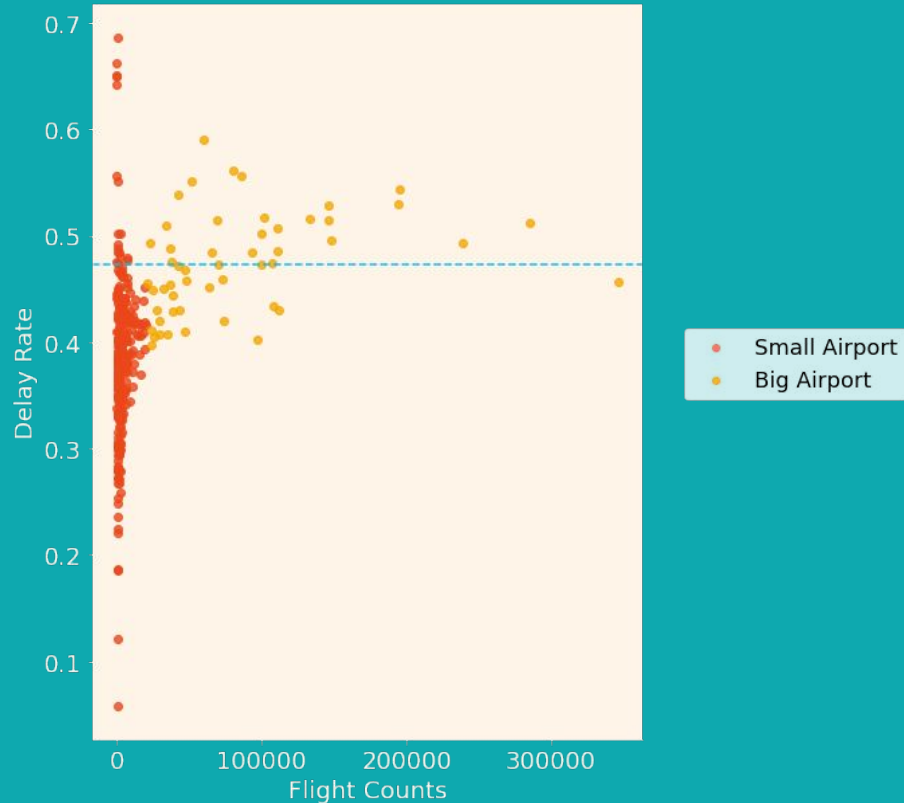
Delay per Airport

Delay Rate of Flights and Airports in US



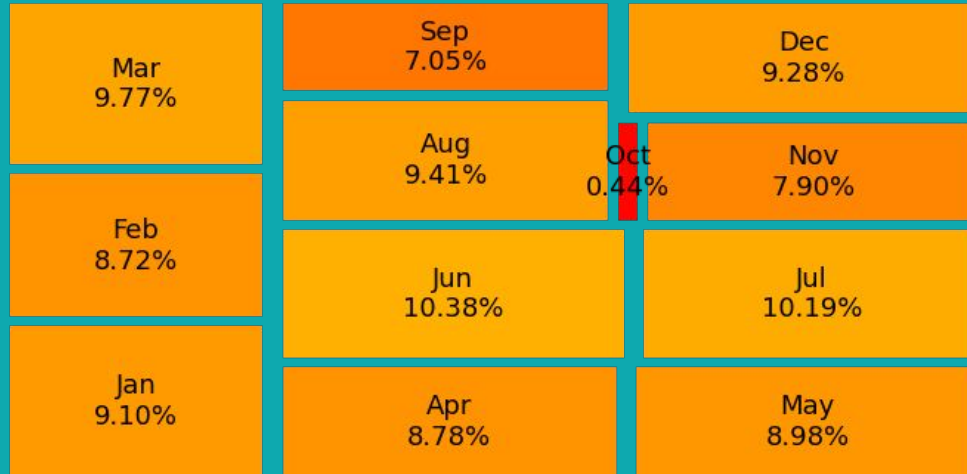
Delay per Airport

- ❖ Base delay rate of 0.47.
- ❖ Small airports have normal spread.
- ❖ Big airports have higher delay rates no matter what.



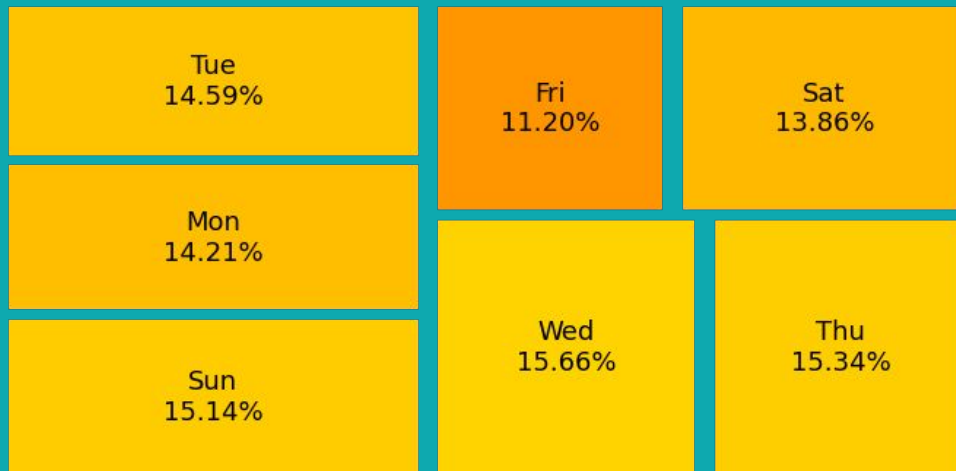
Delay per Month

Delay Percentages per Month in 2015

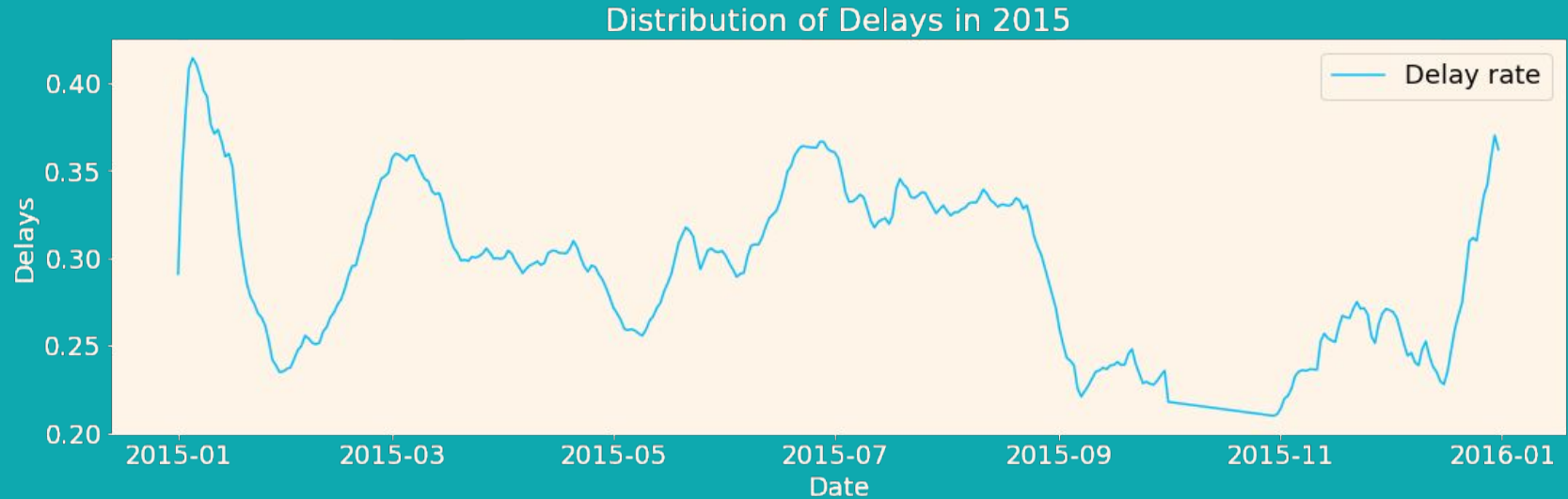


Delay per Weekday

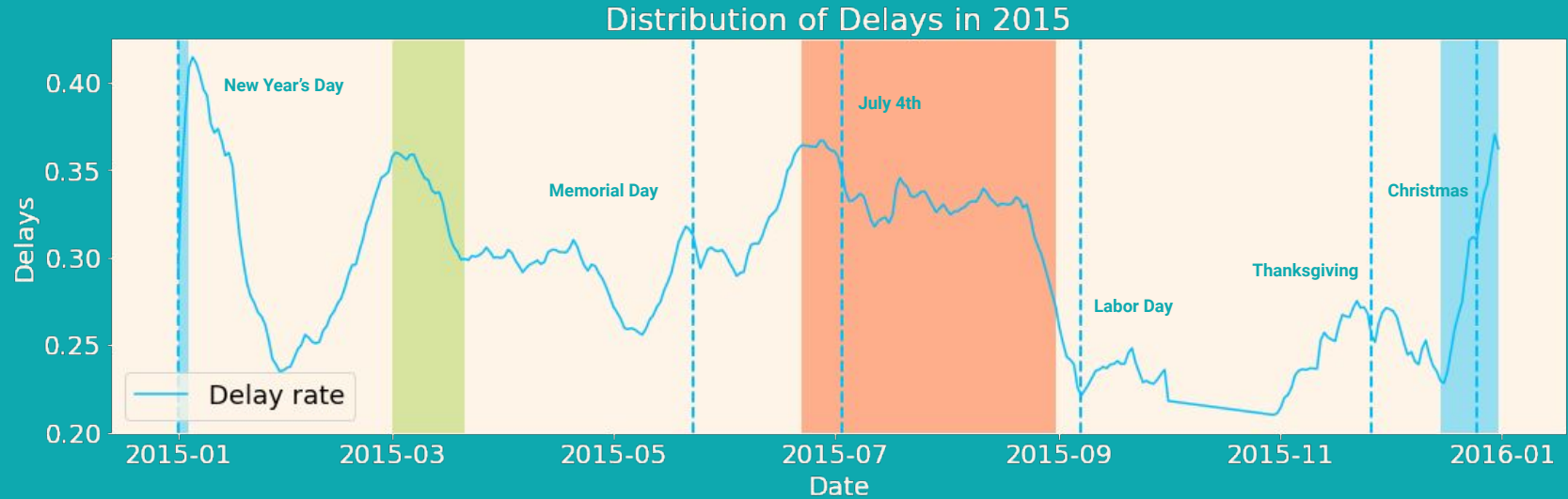
Delay Percentages per Weekday in 2015



Delay Over Time



Delay Over Time

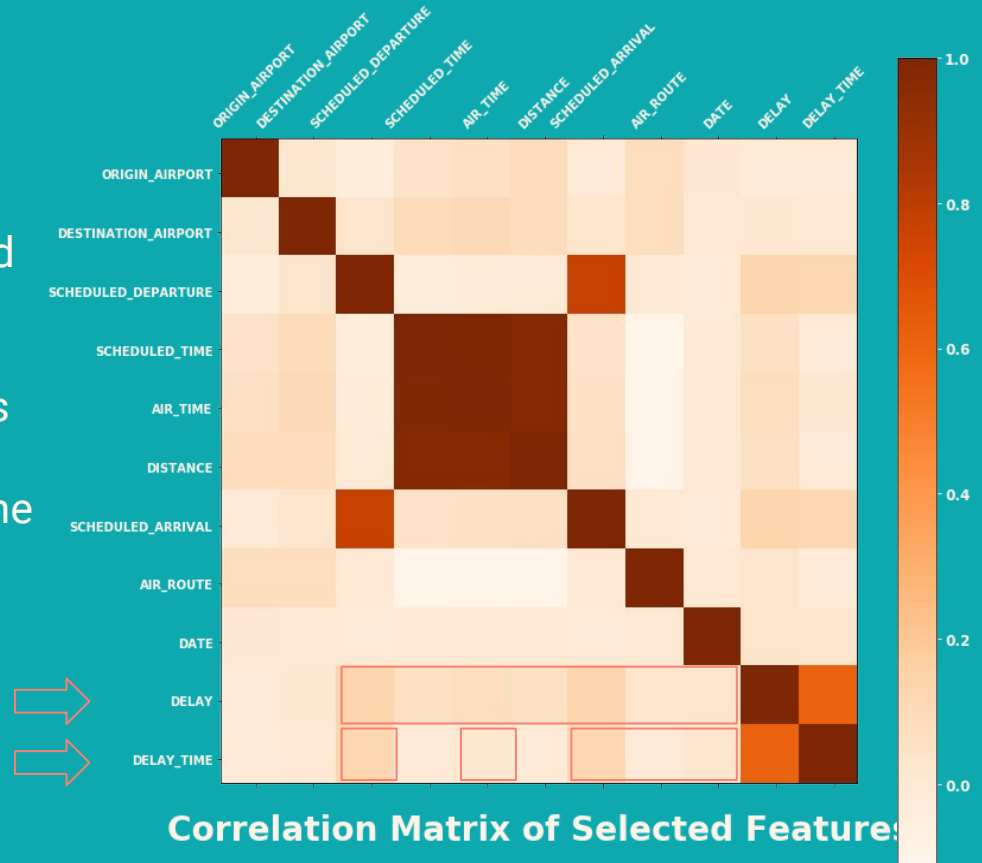


Trend Conclusions

- ❖ Noticeable trends can be found when comparing delay against:
 - Airline
 - Airport (and related geo. info)
 - Date of flight (extendable to month/weekday)
- ❖ These can be transformed into usable features when trying to predict flight delays.

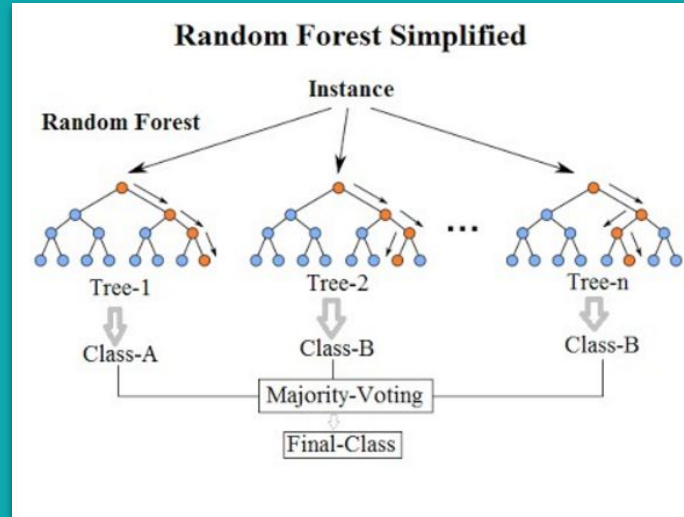
Delay Prediction

- ❖ Only consider features that would be known during booking
- ❖ Find the most correlated features
- ❖ Features then used in our machine learning models

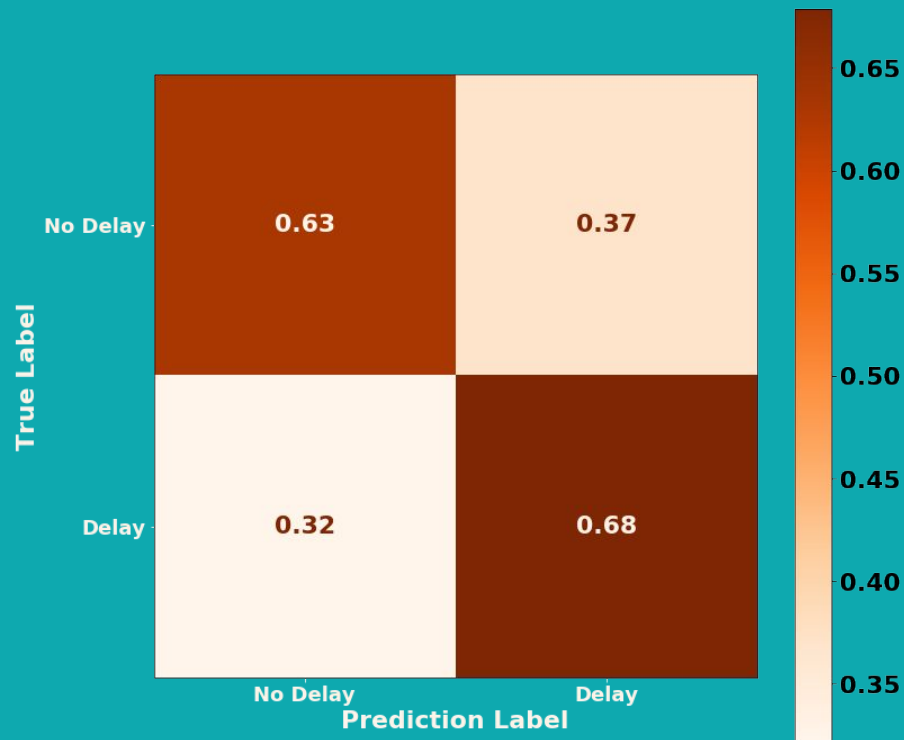


Delay Prediction

- ❖ Binary Classification
 - Delay vs No Delay.
 - Using Random Forest Classifier.
- ❖ Multi-Class Classification
 - Split delay into 15 minute gaps.

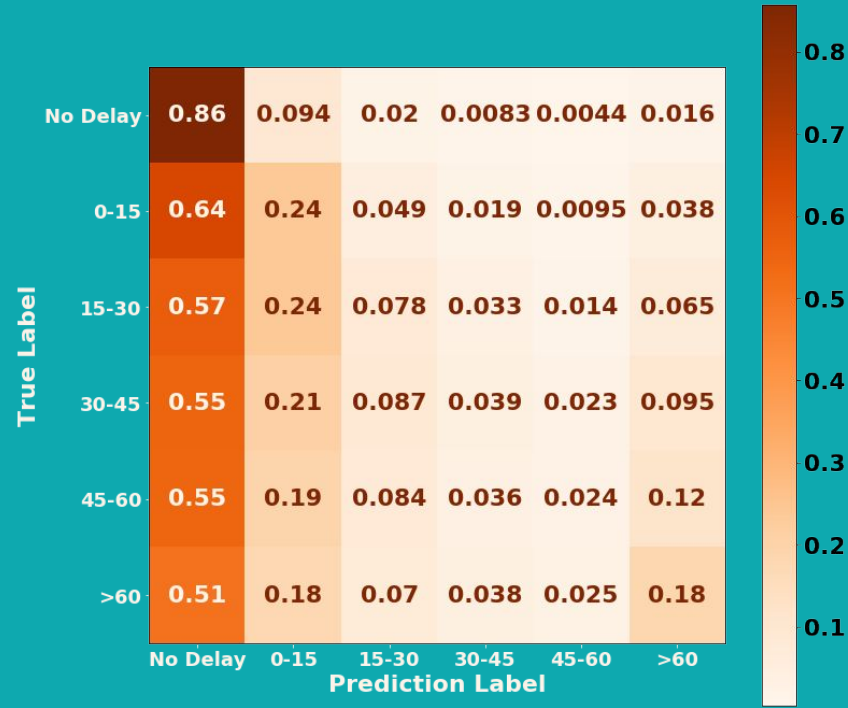


Random Forest



Confusion Matrix for Binary Classification (Random Forest)

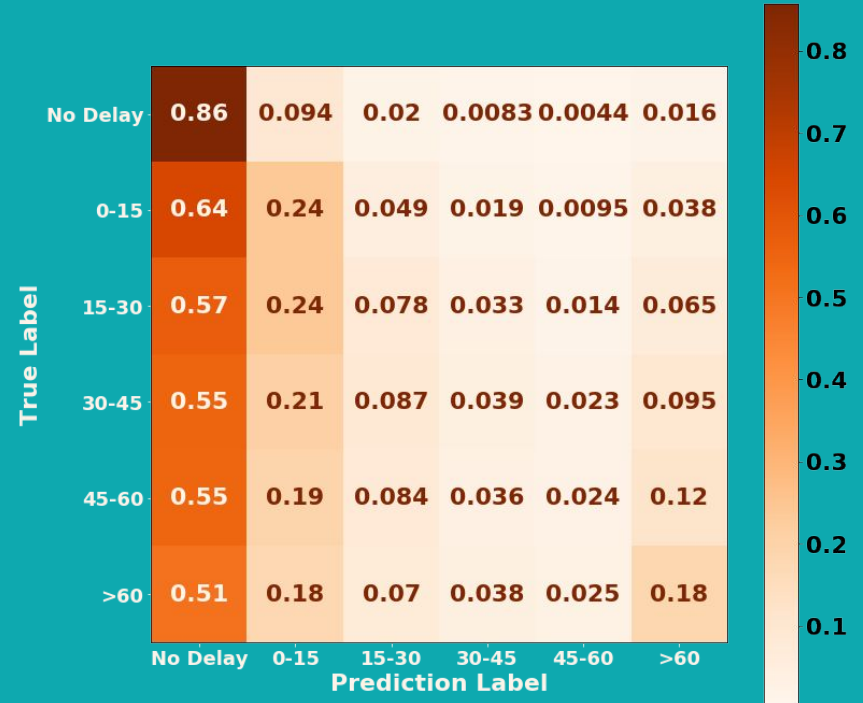
Multi-Class Random Forest



Confusion Matrix for Multi-Class Classification (Random Forest)

Conclusion from the model

- ❖ Final best accuracy of 67.58% vs delay rate of 47.28%.
- ❖ Confusion matrix show heavy bias for selecting No Delay.
- ❖ Predicting flight delay is hard.



Confusion Matrix for Multi-Class Classification (Random Forest)

Thank you!

Any Questions?

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik

