2015 U.S. Flight Delay Prediction

Using LightGBM, Random Forest, and Neural Networks

Azadeh Ansari

August 1, 2024

Introduction

- Overview of the Flight Delay data
- Importance of predicting delay amount

Data Description

- Description of the dataset
- Features: Day of the week, Month, Scheduled Arrival Time, Scheduled Departure Time, Taxi In, Taxi Out, Airline, etc.

Methodology

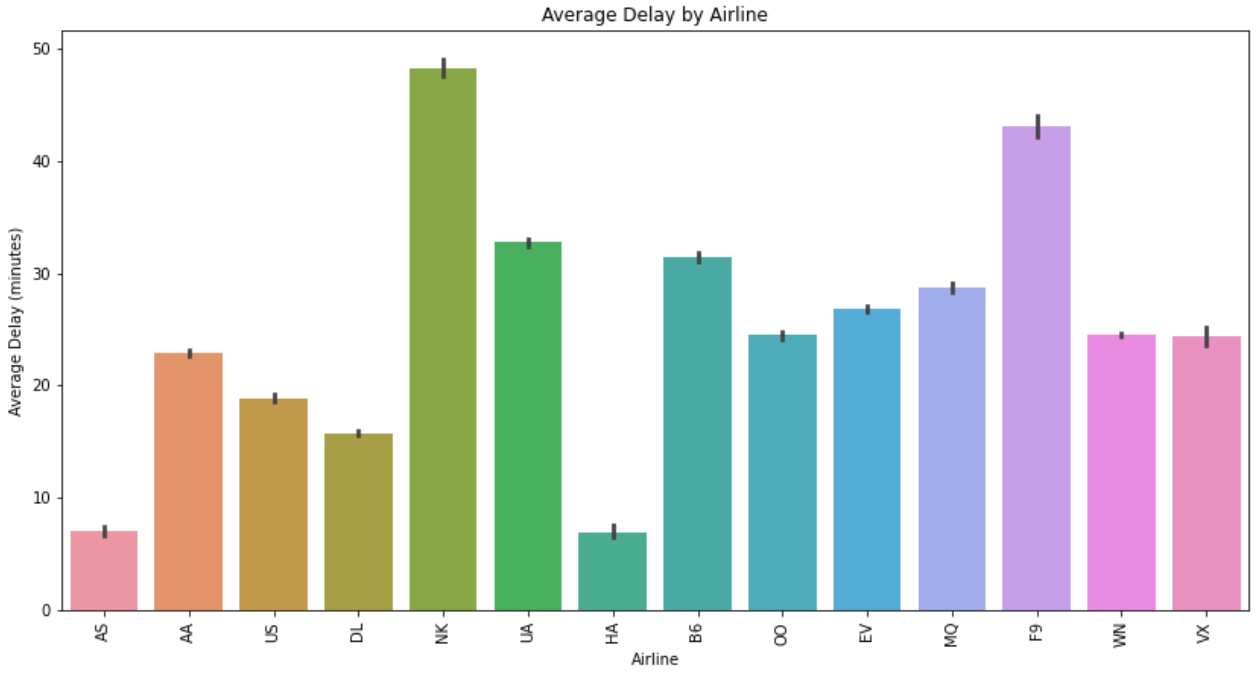
- Data Preprocessing: Cleaning, Feature Engineering, Normalization
- Models Used: LightGBM, Random Forest, Neural Networks

Evaluation Metrics

- RMSE
- Computational Time

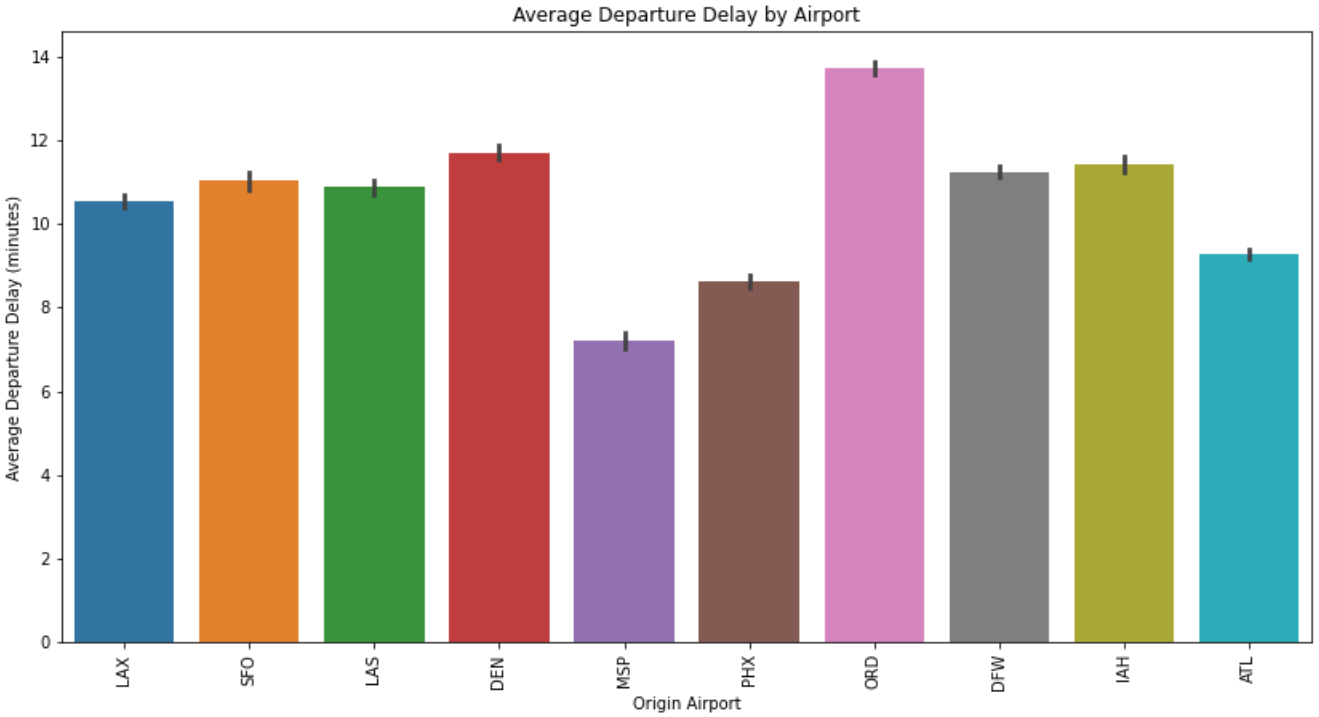
Explanatory Data Analysis

- Average Total Delay per Airline



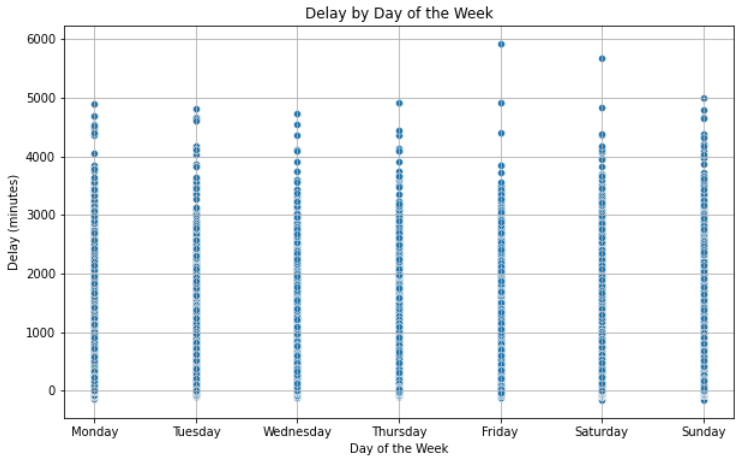
Explanatory Data Analysis

- Average Total Delay per Airline



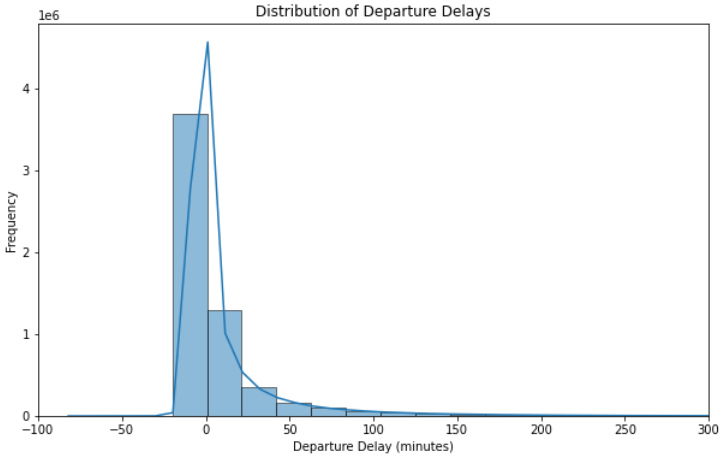
Explanatory Data Analysis

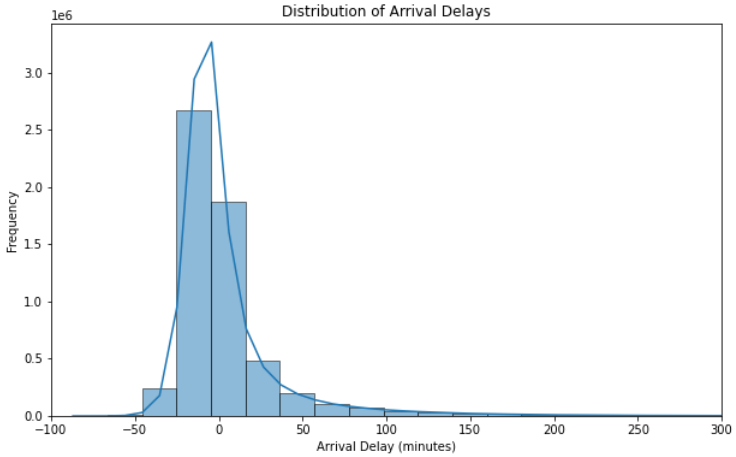
- Total Delay per Day of the Week



Explanatory Data Analysis

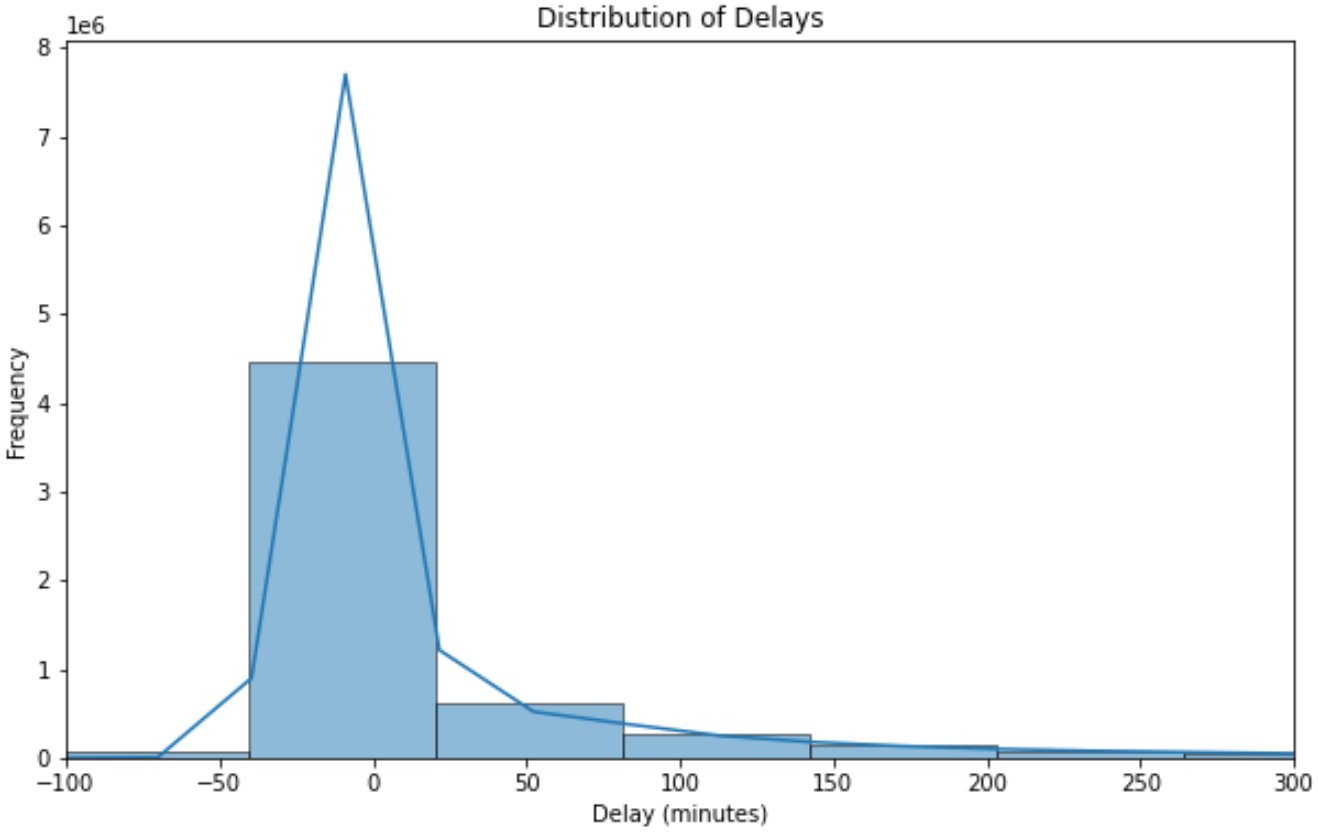
- Departure/Arrival Delay Distribution





Explanatory Data Analysis

- Total Delay Distribution



Comparison of Models

- RMSE on Test Data Set
- Computational Time

Model	RMSE	Computational Time (minutes)
LightGBM	102.44	28
Random Forest	104.58	263
Neural Networks	-	Did not run

Conclusion

- Summary of findings
- Best performing model
- LightGBM outperforms other models.