TEAM BETA – PROJECT 02

FLIGHT DELAY PREDICTION

CONTENT

INTRODUCTION
QUESTION

Is it possible to predict in advance the delay times of each flight??

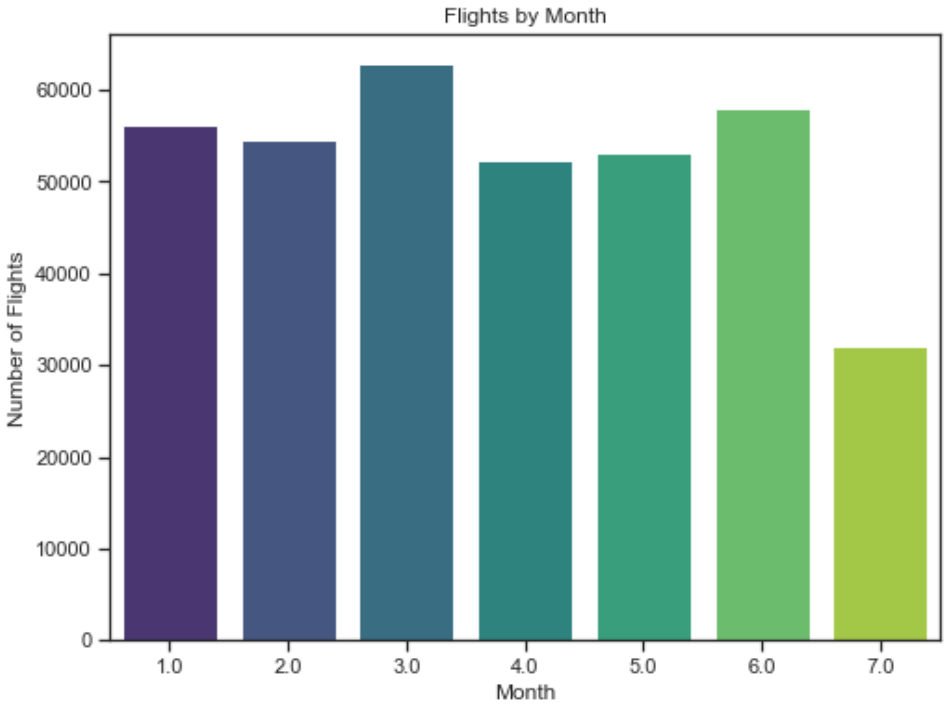
Data Collection, Cleaning & Pre-processing

Data was sourced from [specific source: https://www.kaggle.com/code/bobirino/predicting-flight-delay/notebook], including key attributes like

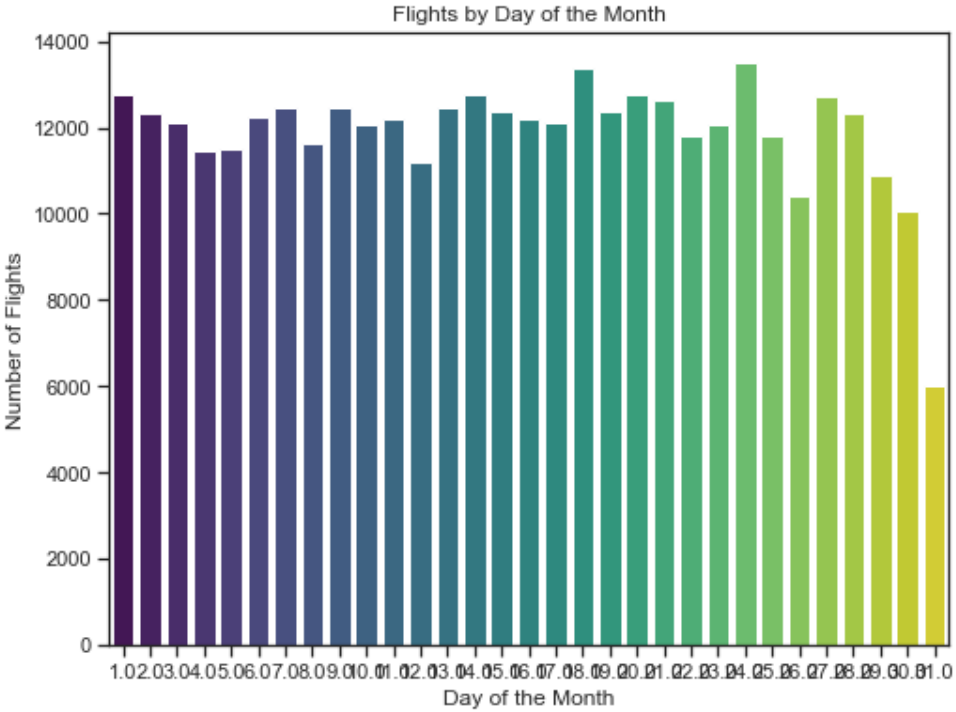
1.Year 2016
2.Month 1-12
3.DayofMonth 1-31
4.DayOfWeek 1 (Monday) - 7 (Sunday)
5.DepTime actual departure time (local, hhmm)
6.CRSDepTime scheduled departure time (local, hhmm)
7.ArrTime actual arrival time (local, hhmm)
8.CRSArrTime scheduled arrival time (local, hhmm)
9.UniqueCarrier unique carrier code
10.FlightNum flight number
11.TailNum plane tail number: aircraft registration, unique aircraft identifier
12.ActualElapsedTime in minutes
13.CRSElapsedTime in minutes
14.AirTime in minutes
15.ArrDelay arrival delay, in minutes:
16.DepDelay departure delay, in minutes
17.Origin origin IATA airport code
18.Dest destination IATA airport code
19.Distance in miles
20.TaxiIn taxi in time, in minutes
21.TaxiOut taxi out time in minutes
22.Cancelled *was the flight cancelled
23.CancellationCode reason for cancellation (A = carrier, B = weather, C = NAS, D = security)
24.Diverted 1 = yes, 0 = no
25.CarrierDelay in minutes: Carrier delay is within the control of the air carrie
26.WeatherDelay in minutes: Weather delay is caused by extreme or hazardous weather conditions
27.NASDelay in minutes: Delay that is within the control of the National Airspace System (NAS) may include: non-extreme weather conditions, airport operations, heavy traffic volume, air traffic control, etc.
28.SecurityDelay in minutes: Security delay is caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, inoperative screening equipment and/or long lines in excess of 29 minutes at screening areas.
29.LateAircraftDelay in minutes: Arrival delay at an airport due to the late arrival of the same aircraft at a previous airport. The ripple effect of an earlier delay at downstream airports is referred to as delay propagation

Visualization:

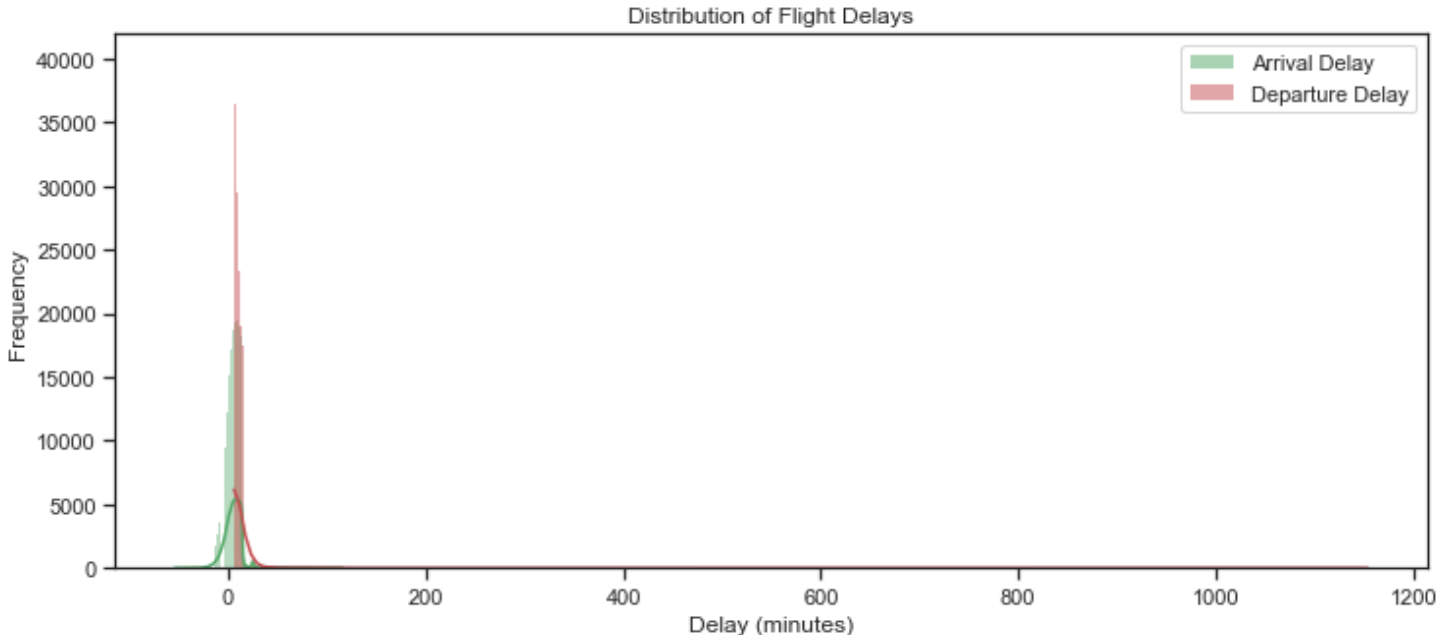
- **Day of the Week bar chart:**



- **Day of the Month bar chart:**

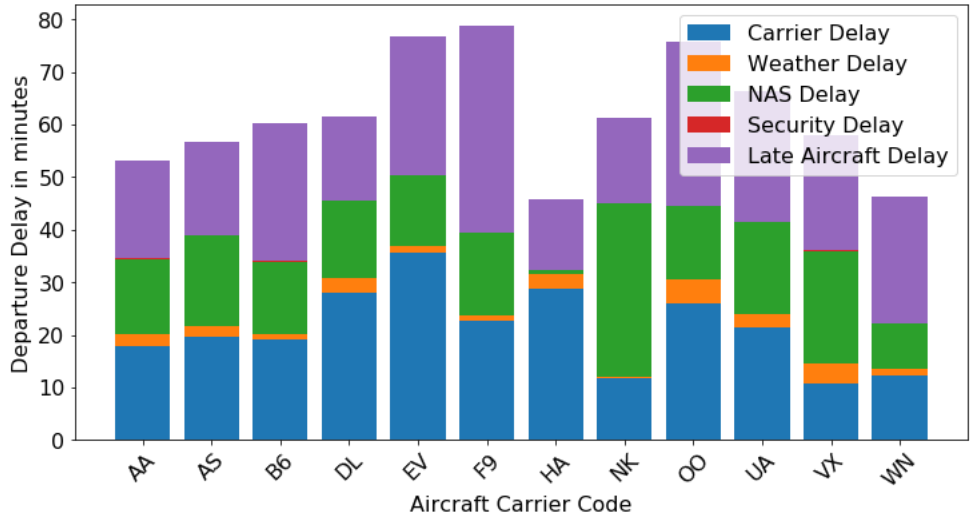


- **Distribution of Flight delay**

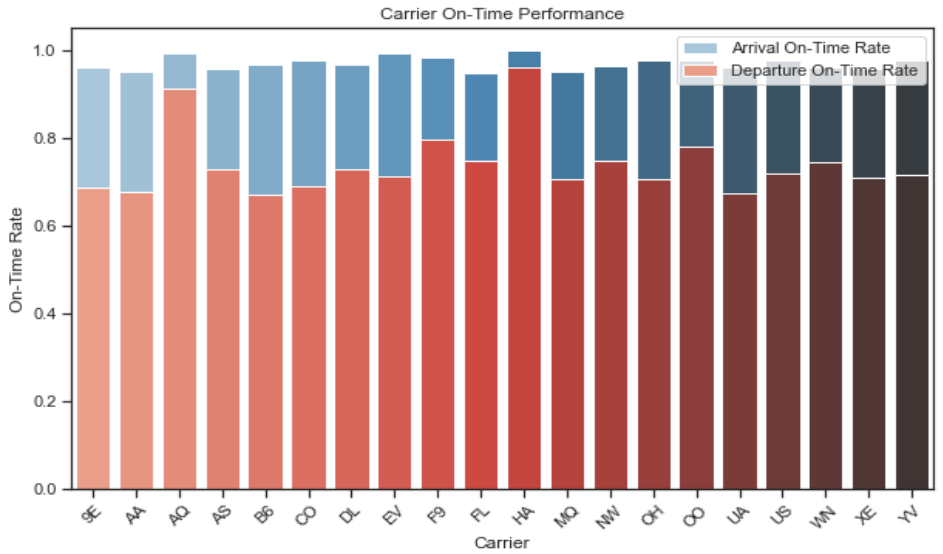


Visualization:

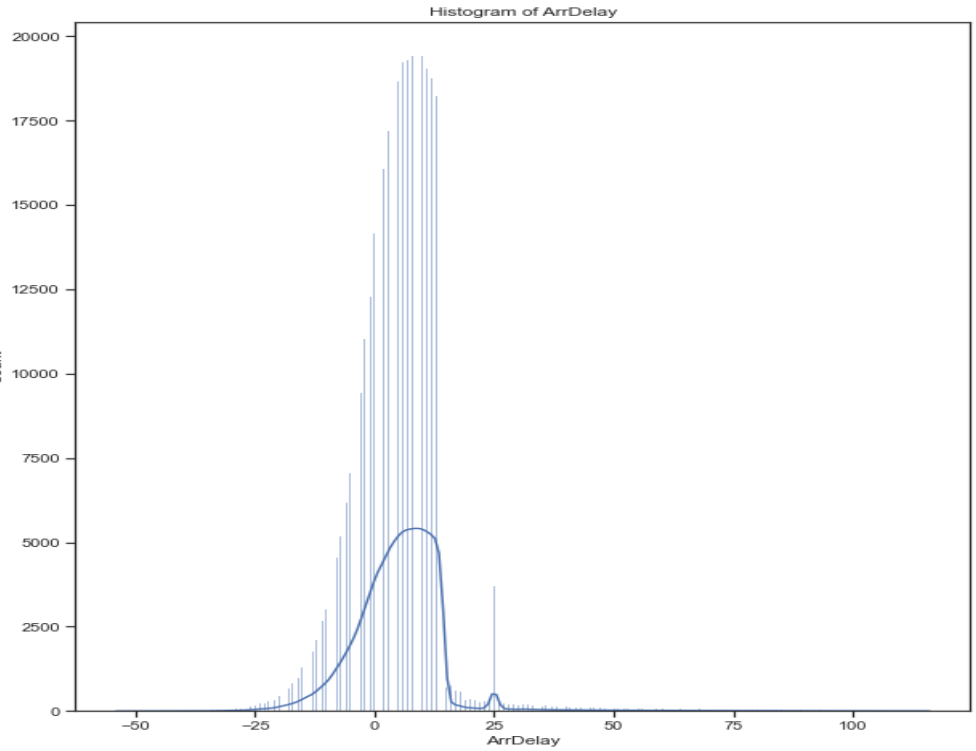
- **Carrier Delays and Reasons**



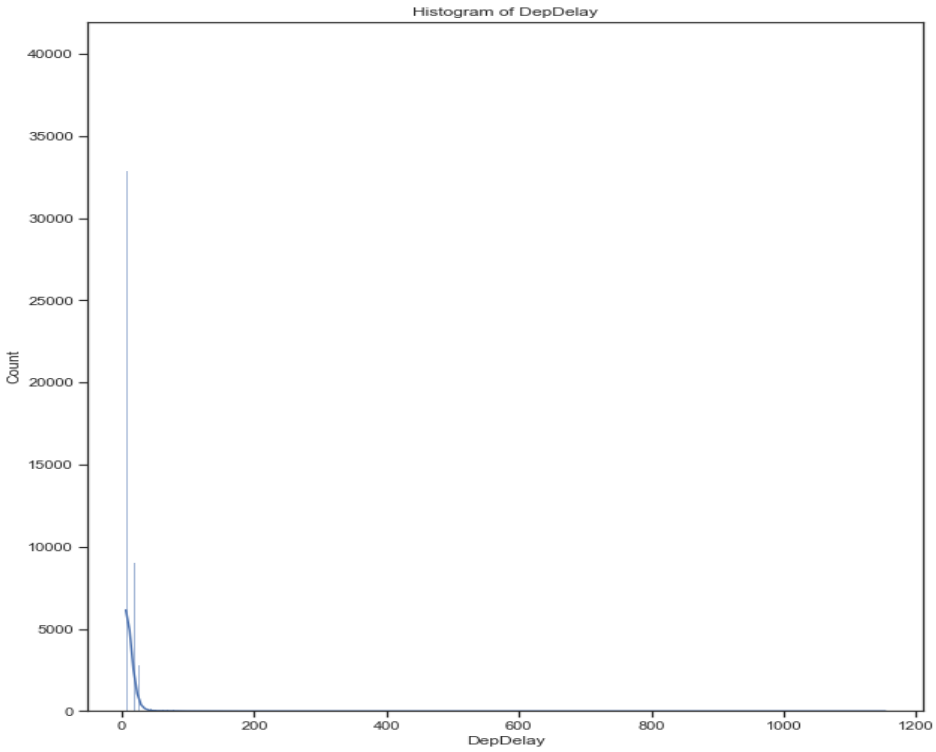
- **Carrier On-Time Performance**



- **Arrival Delay:**

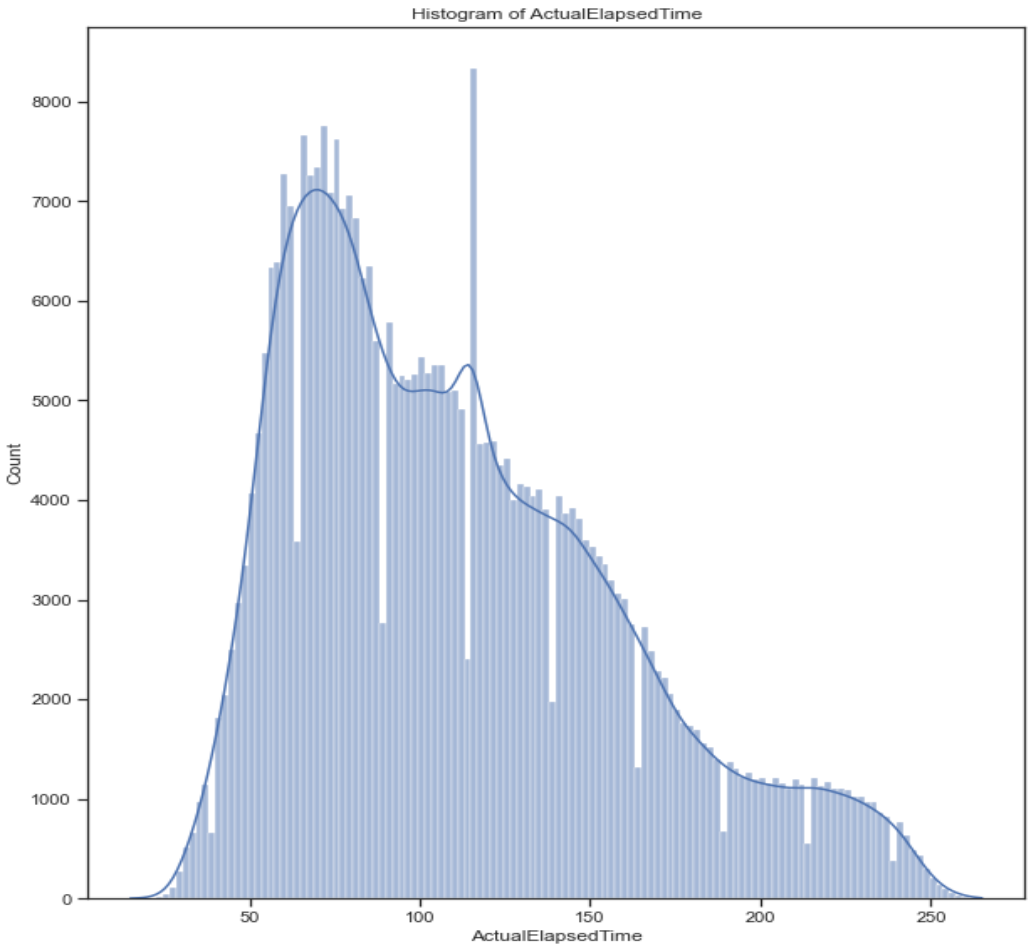


- **Departure Delay:**

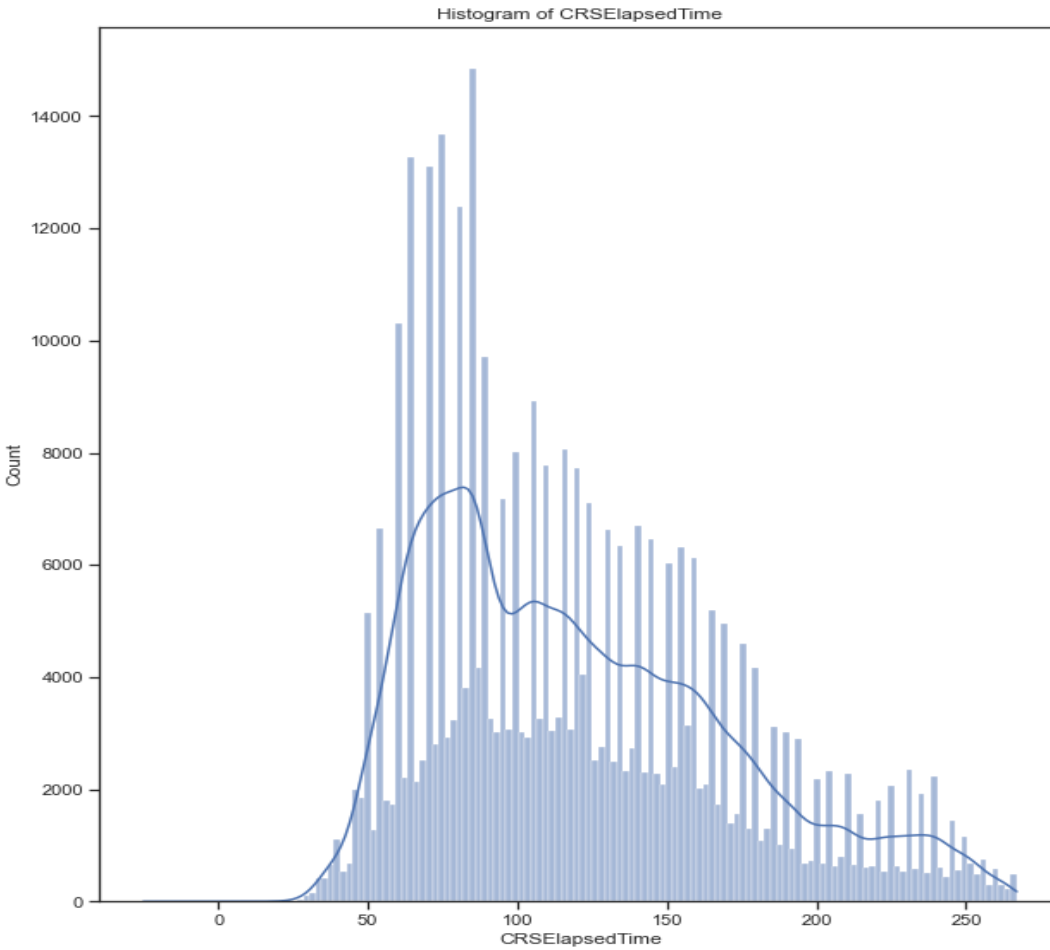


Visualization:

- **ACTUAL TIME OF FLIGHTS**



- **CALCULATED/ESIMATED TIME OF FLIGHTS**



FEATURE ENGINEERING

IMPORTANCE OF FEATURE ENGINEERING: ENHANCING PREDICTIVE POWER AND MODEL PERFORMANCE

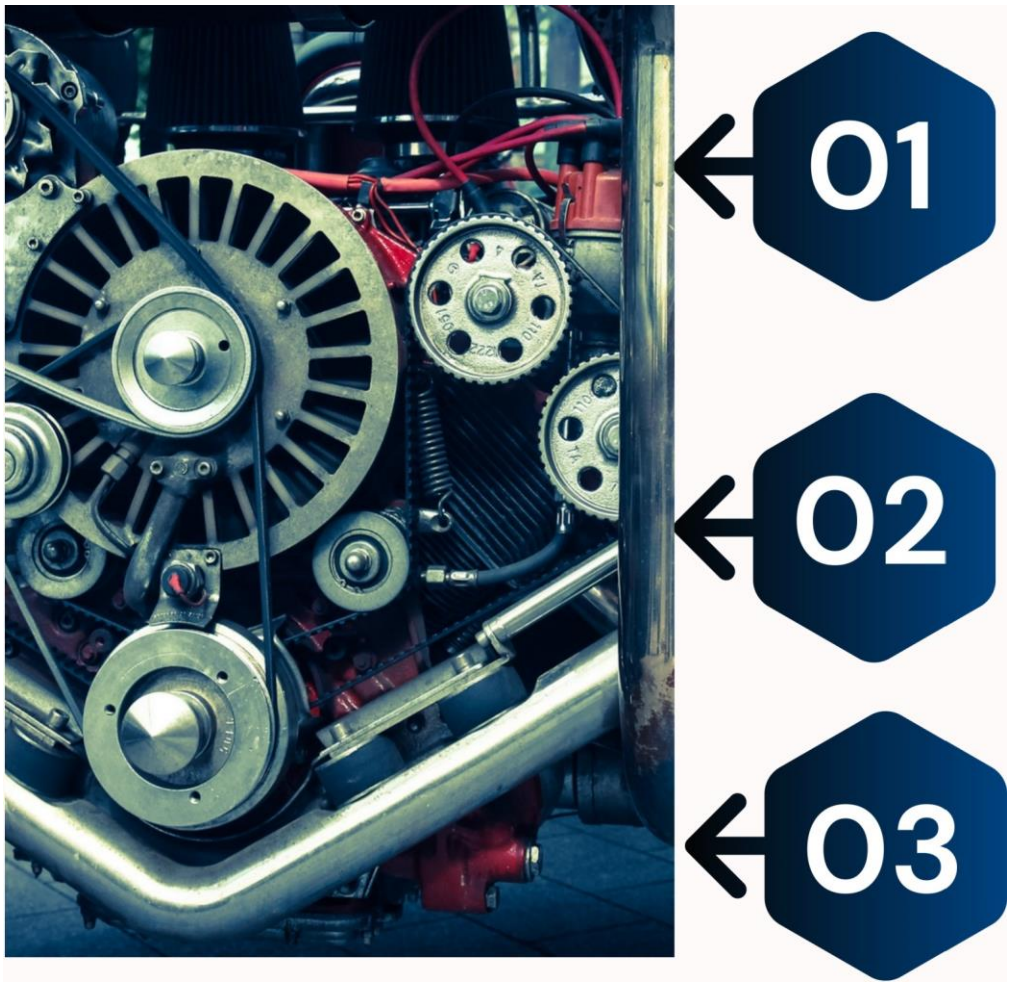
FEATURE TRANSFORMATION: APPLIED TRANSFORMATIONS (E.G., LOGARITHMIC, SCALING) TO IMPROVE DATA DISTRIBUTION AND REDUCE SKEWNESS

FEATURE EXTRACTION: EXTRACTED RELEVANT FEATURES FROM RAW DATA (E.G., FLIGHT DURATION, DEPARTURE TIME, AIRLINE, WEATHER)

FEATURE CREATION: GENERATED NEW FEATURES THROUGH COMBINATIONS, INTERACTIONS, OR DOMAIN KNOWLEDGE (E.G., FLIGHT DENSITY, DEPARTURE-ARRIVAL TIME DIFFERENCE)

Feature engineering enhanced the model's ability to capture patterns and relationship in the data. Key features like Elapsed time, Arrival Delay, Departure Delay etc. provided important context. Transformations ensured compatibility with the model, while new features offered additional insights. These eff orts played a pivotal role in achieving the desired outcomes.

MODEL DEVELOPMENT



METHODS

Random Forest Regression

- Boosting algorithm that combines weak learners to capture non-linear relationships and feature interaction effectively
- Handles higher-order dependencies and offers flexibility in controlling overfitting through regularization parameters
- Evaluation Metric: Mean Squared Error

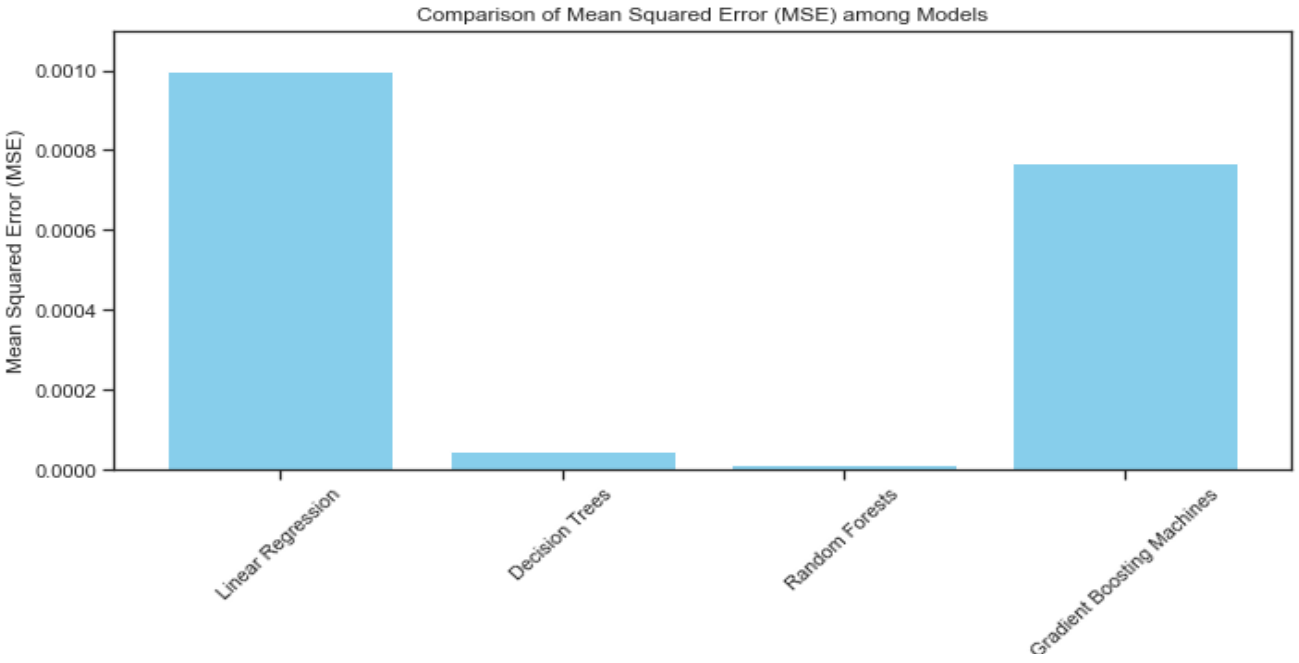
Gradient Boosting Regression

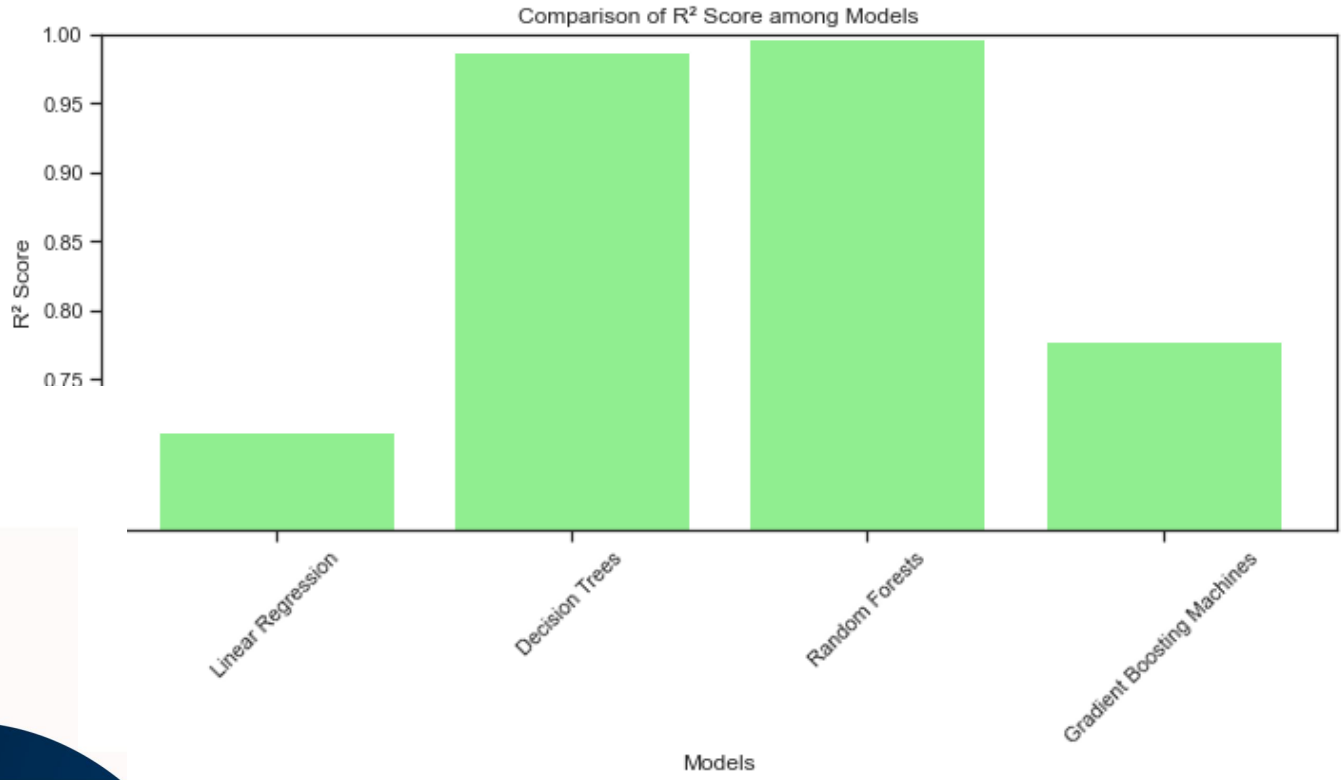
- Ensemble learning method that builds multiple decision trees and combines their predictions.
- Robust to outliers and noise, handles non-linear relationships and provides features importance.
- Evaluation Metric: Mean Squared Error

Linear Regression

- Linearity assumption: linear regression assumes a linear relationship between independent variables and dependent variable
- Linear regression coefficients; impact of variable changes

MODEL EVALUATION





FINAL CONCLUSIONS

In conclusion, our project successfully predicted flight delays with minute-level precision. We collected and processed data, performed feature engineering, visualized relationships, and built accurate predictive models. These models can improve operational efficiency and enhance the flight management experience for both airlines and passengers.