Final Project

Traffic Prediction

Presented By:

Vidusha Wilpita

Project Outline

- Problem Statement and Importance
- Dataset description
- EDA
- Model selection
- Model evaluation
- Findings/Conclusion
- Model in Use

Problem Statement and project importance

Problem:

- This project aims to use machine learning to predict optimal traffic light timings at four junctions. Inaccurate timings often lead to congestion due to rising populations and inefficient systems. By analyzing historical patterns, this project aims to create a model that optimizes timings dynamically, reducing delays and fuel wastage.

Why is it important?

- This project is important as it can transform urban transport. Traffic congestion causes economic losses and pollution. Creating an adaptable machine learning model for traffic light predictions can significantly improve traffic flow, leading to energy savings, lower emissions, and more sustainable cities.

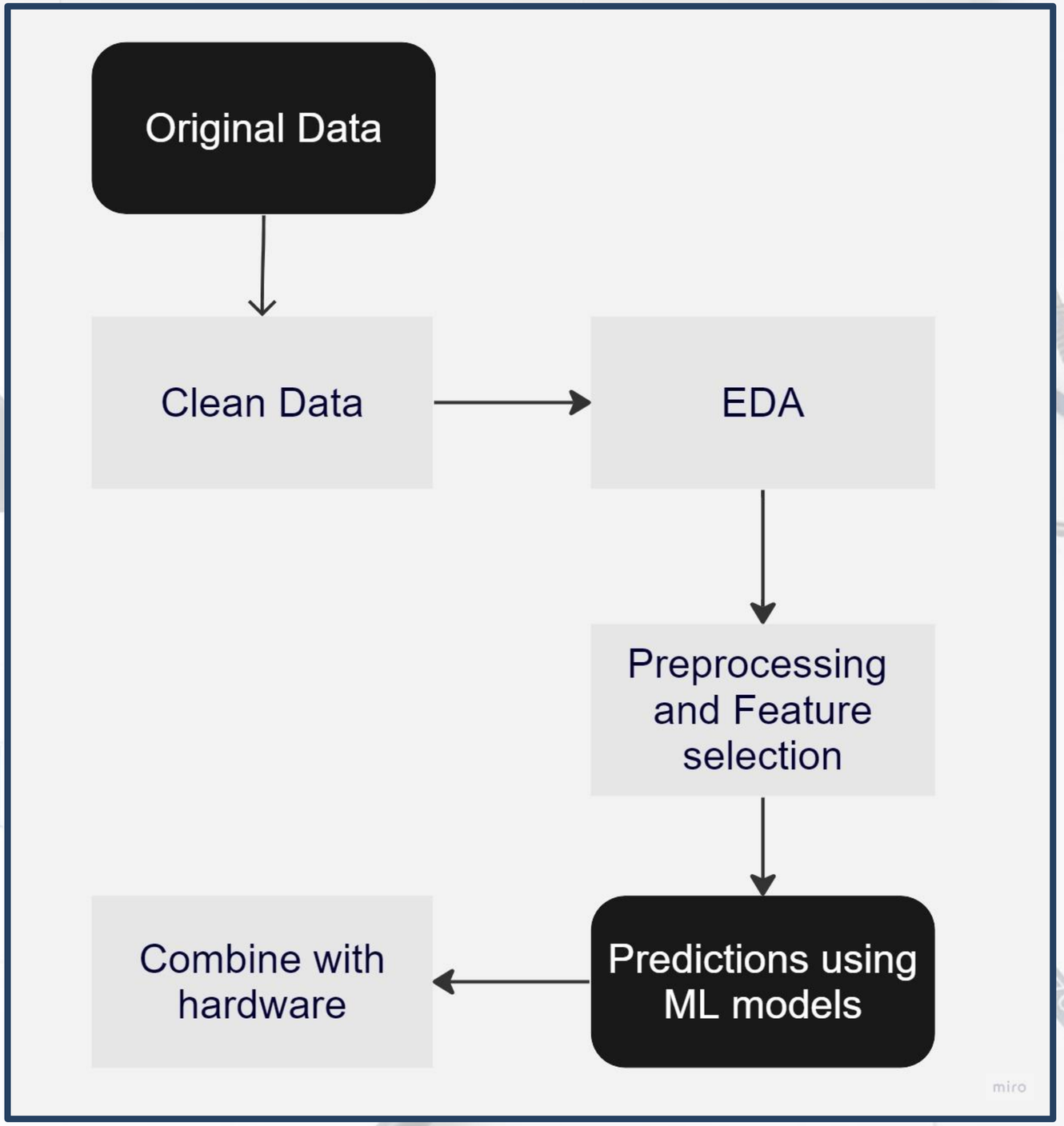
Dataset Description

- The dataset contains 48120 observations of the number of vehicles at four traffic junctions each hour, collected using sensors. It consists four columns and those are Datetime, Junction, Vehicles and ID.

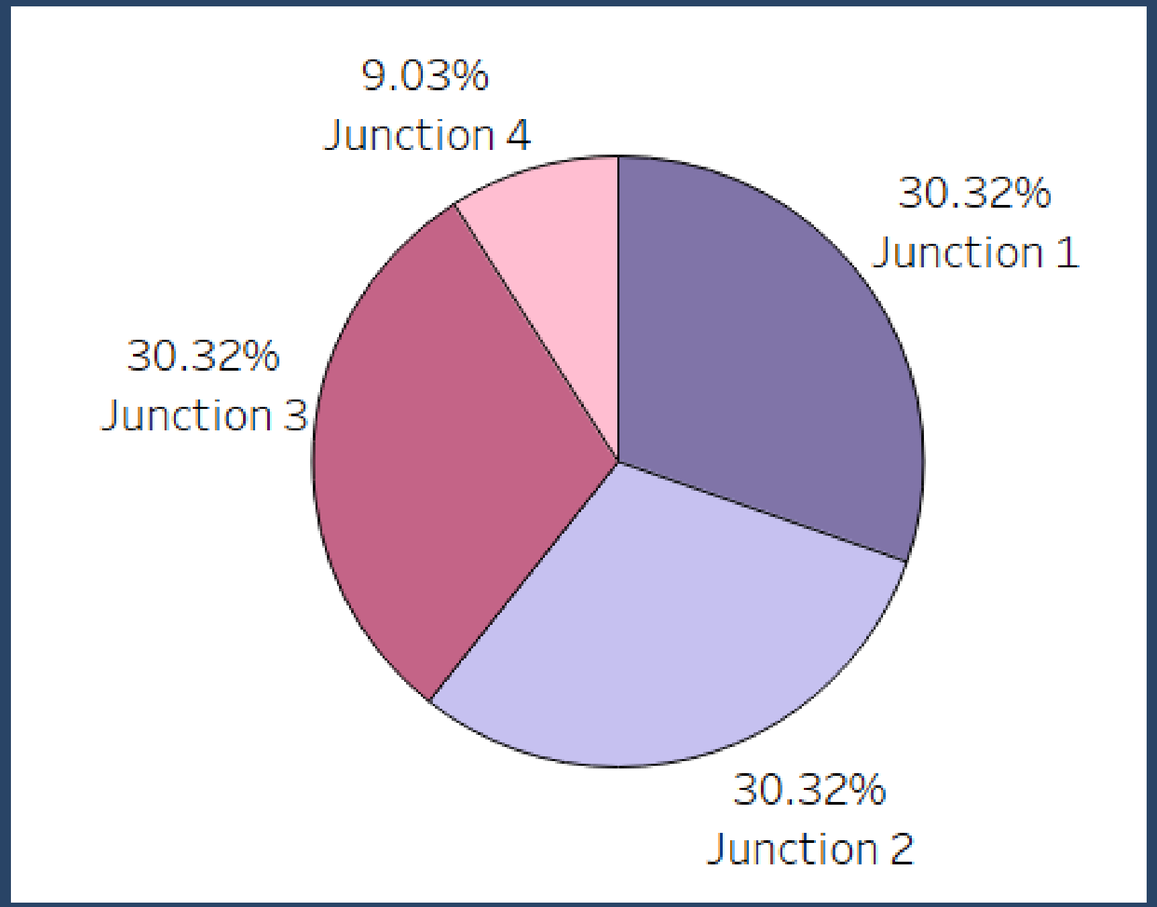
Project Objectives

1. To identify the most significant features for detecting the number of vehicles every hour.
2. Based on the model evaluation result, predict the number of vehicles at each junction on a given time on a given day.
3. Predict the time for traffic light at each junction and simulate it on traffic light controller unit.

Approach


Original Data
Clean Data
EDA
Preprocessing and Feature selection
Predictions using ML models
Combine with hardware
miro

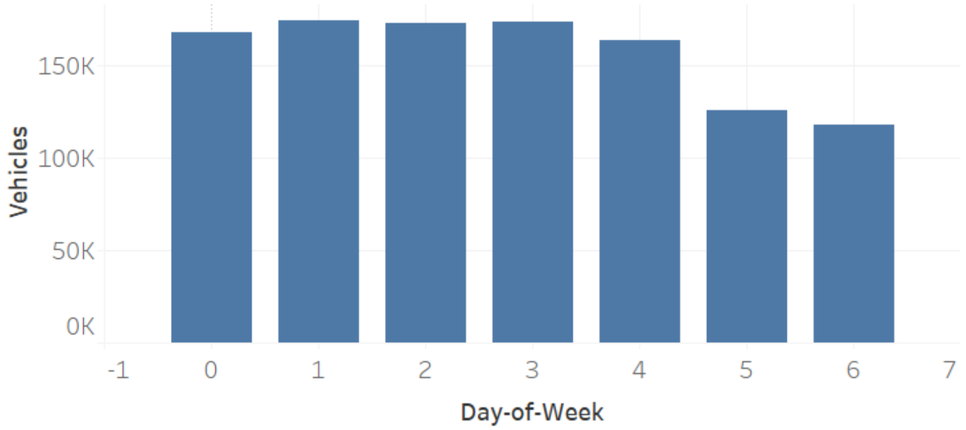
Dataset by Junction



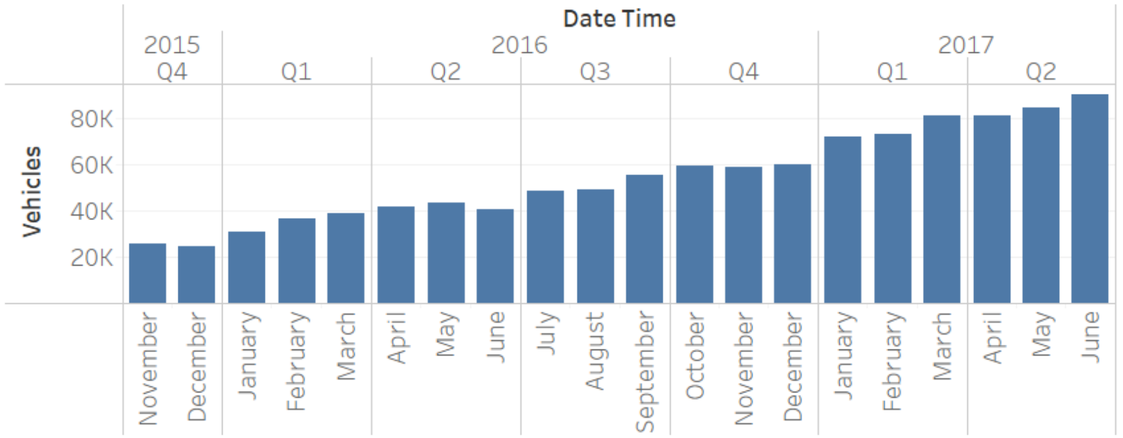
Feature Engineering

- ID column is removed as it has no association with the number of vehicles
- DateTime is separated to Year, Month, Day, Day of the week and Time
- Separate datasets created for each junction

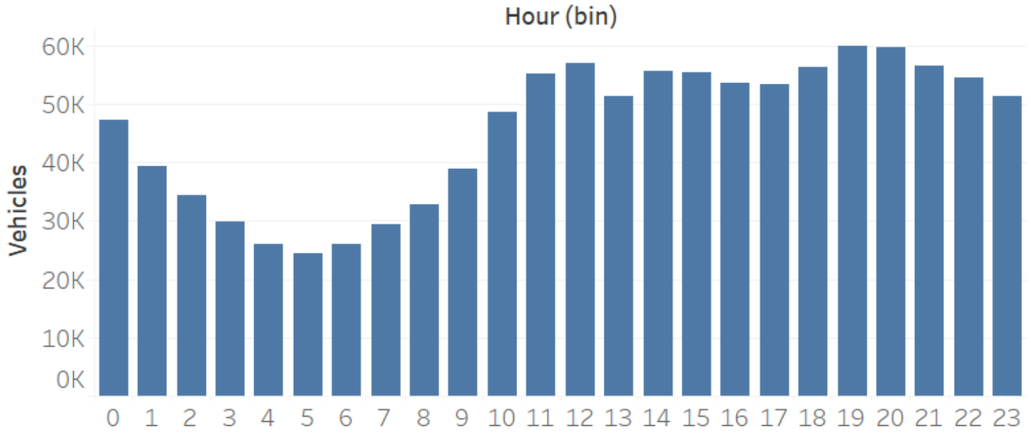
Day of the week versus Vehicle count



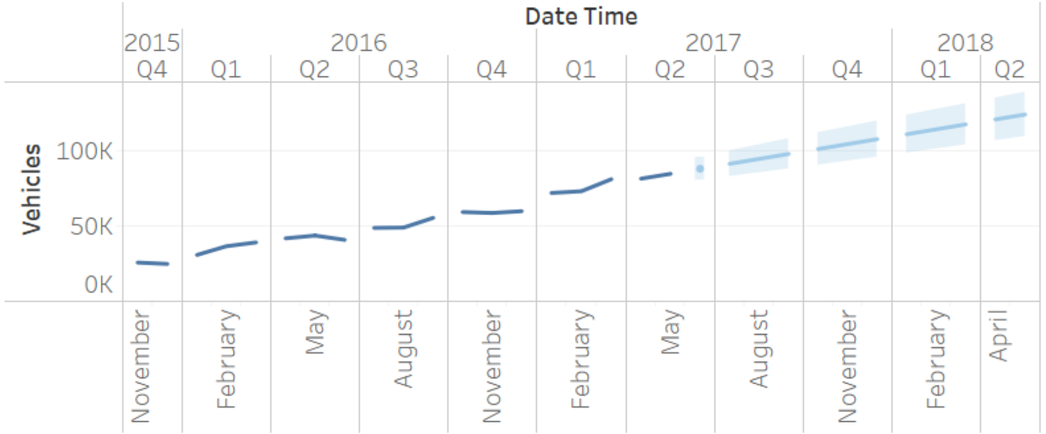
Vehicles by Month



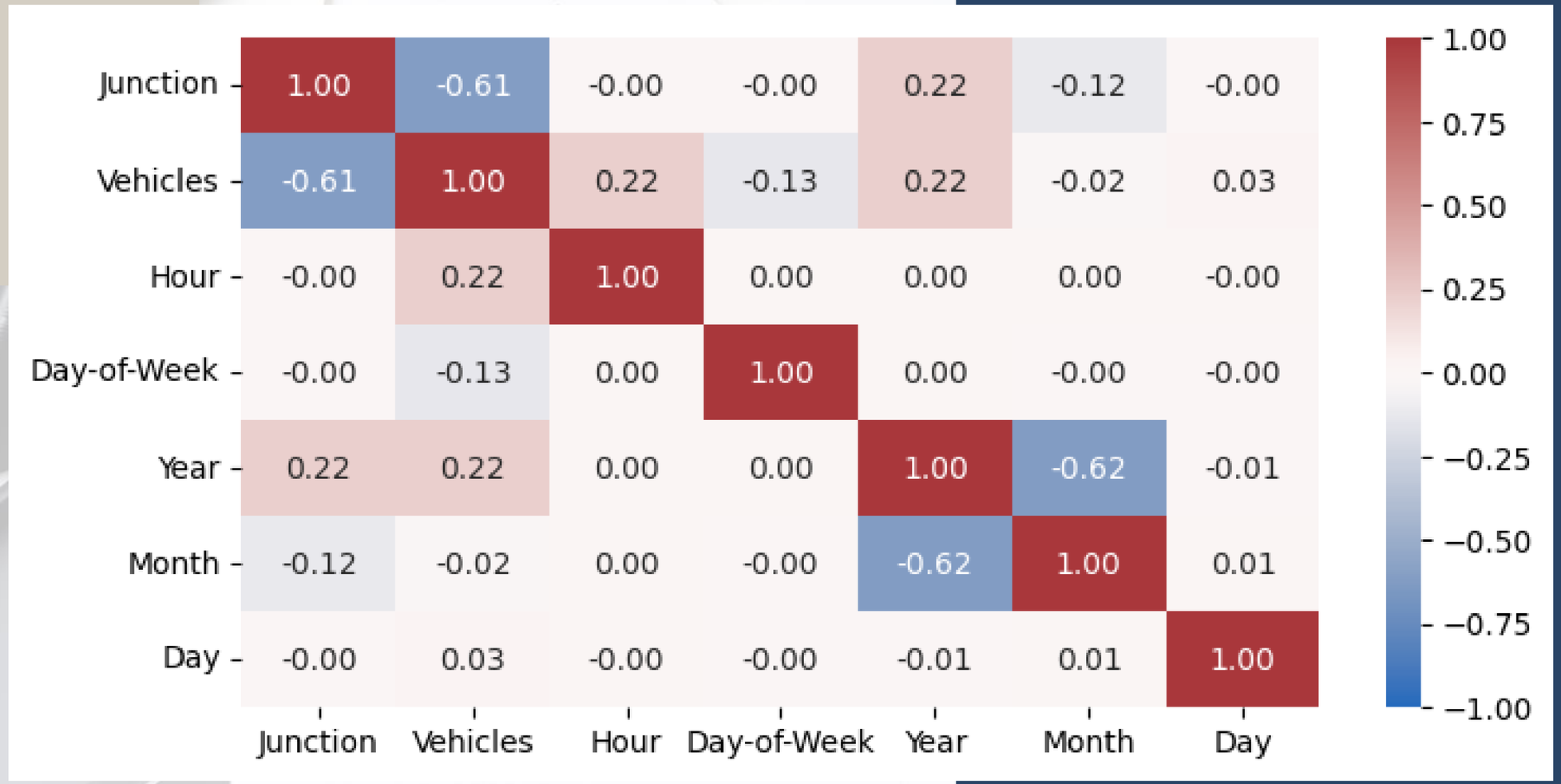
Vehicles by Time of the Day



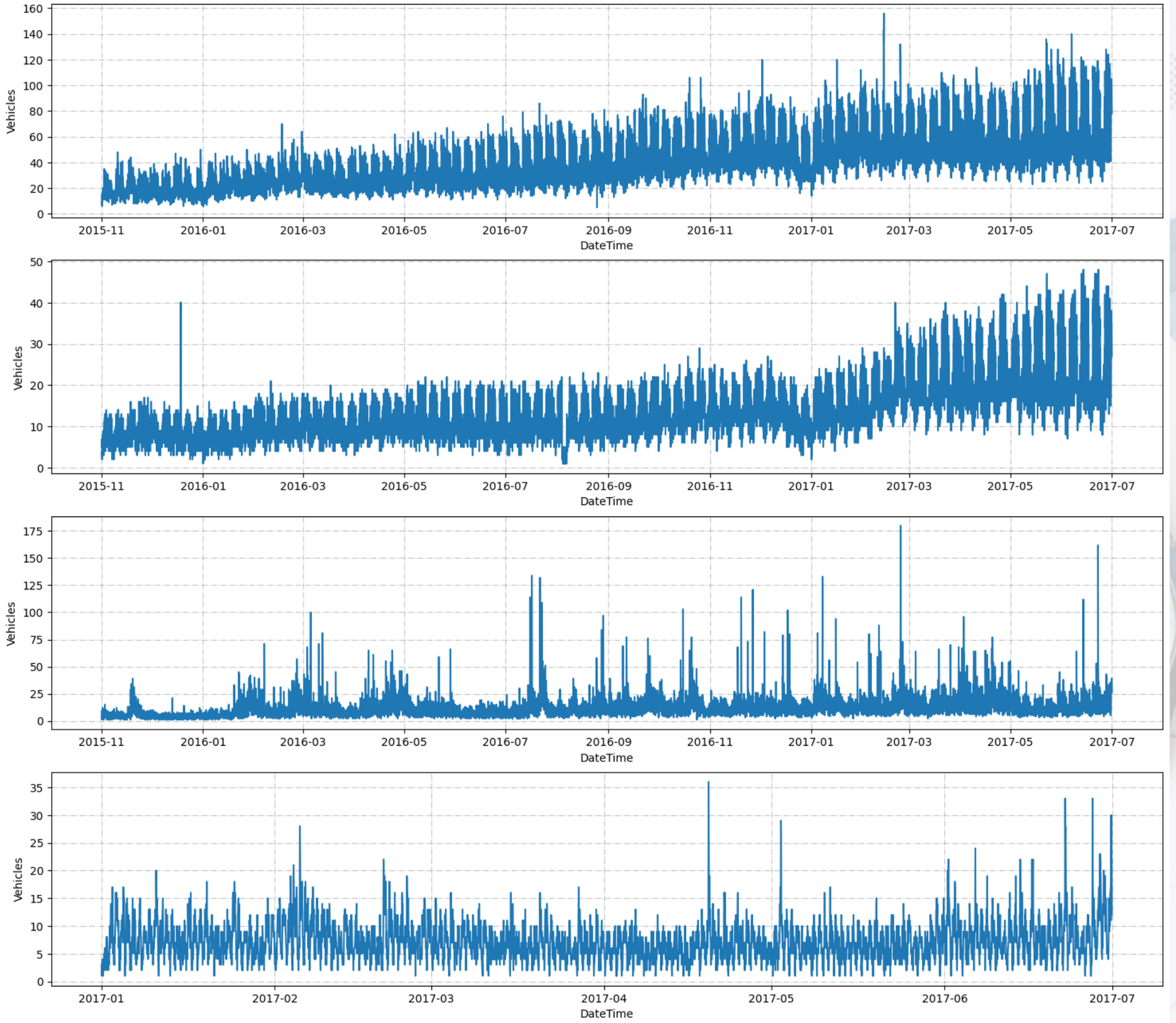
Predictions of Vehicles for Next year



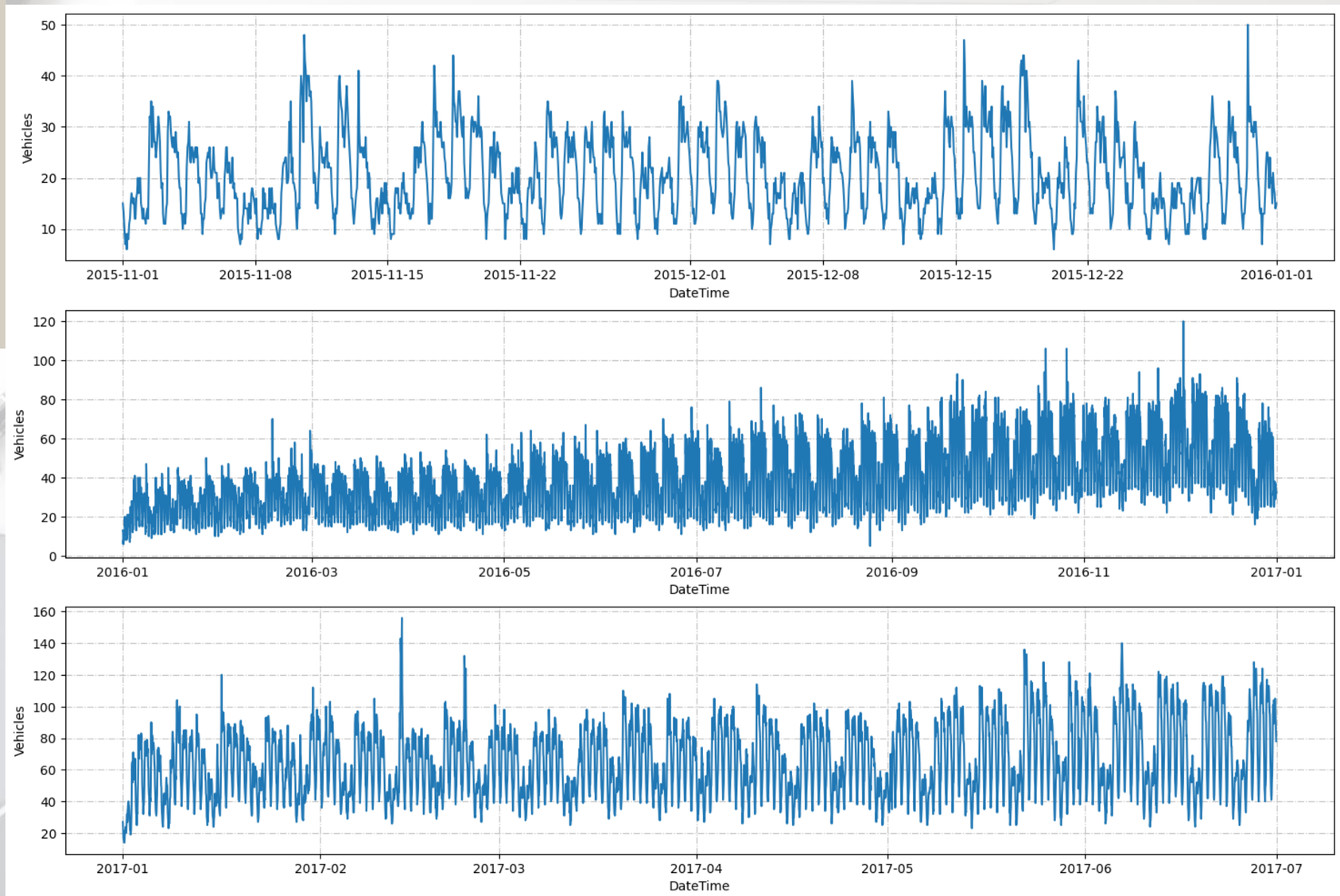
Heatmap of Correlation



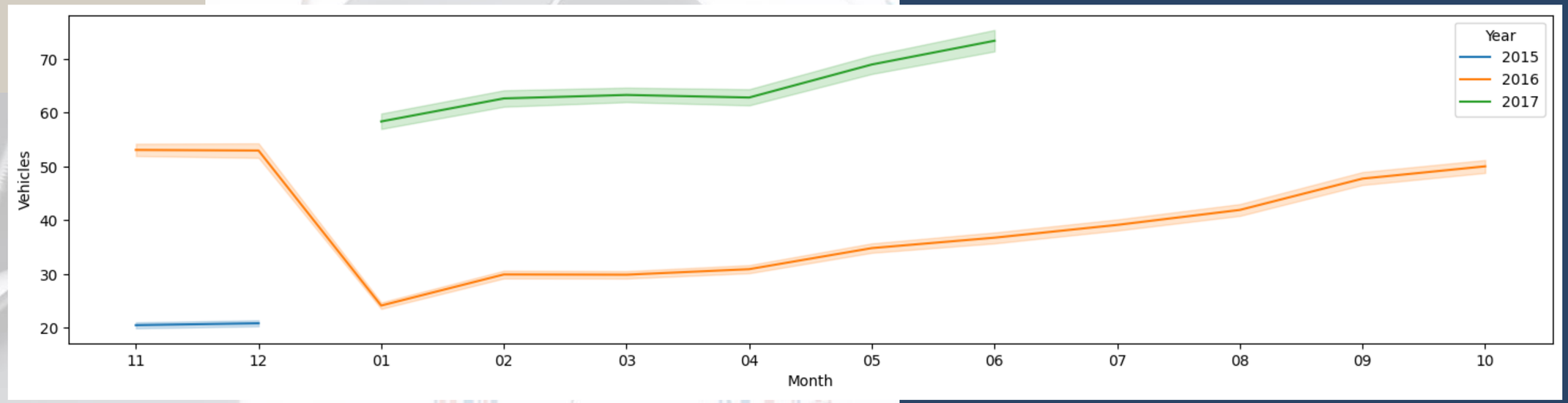
Vehicle count at each Junction from 2015-2017



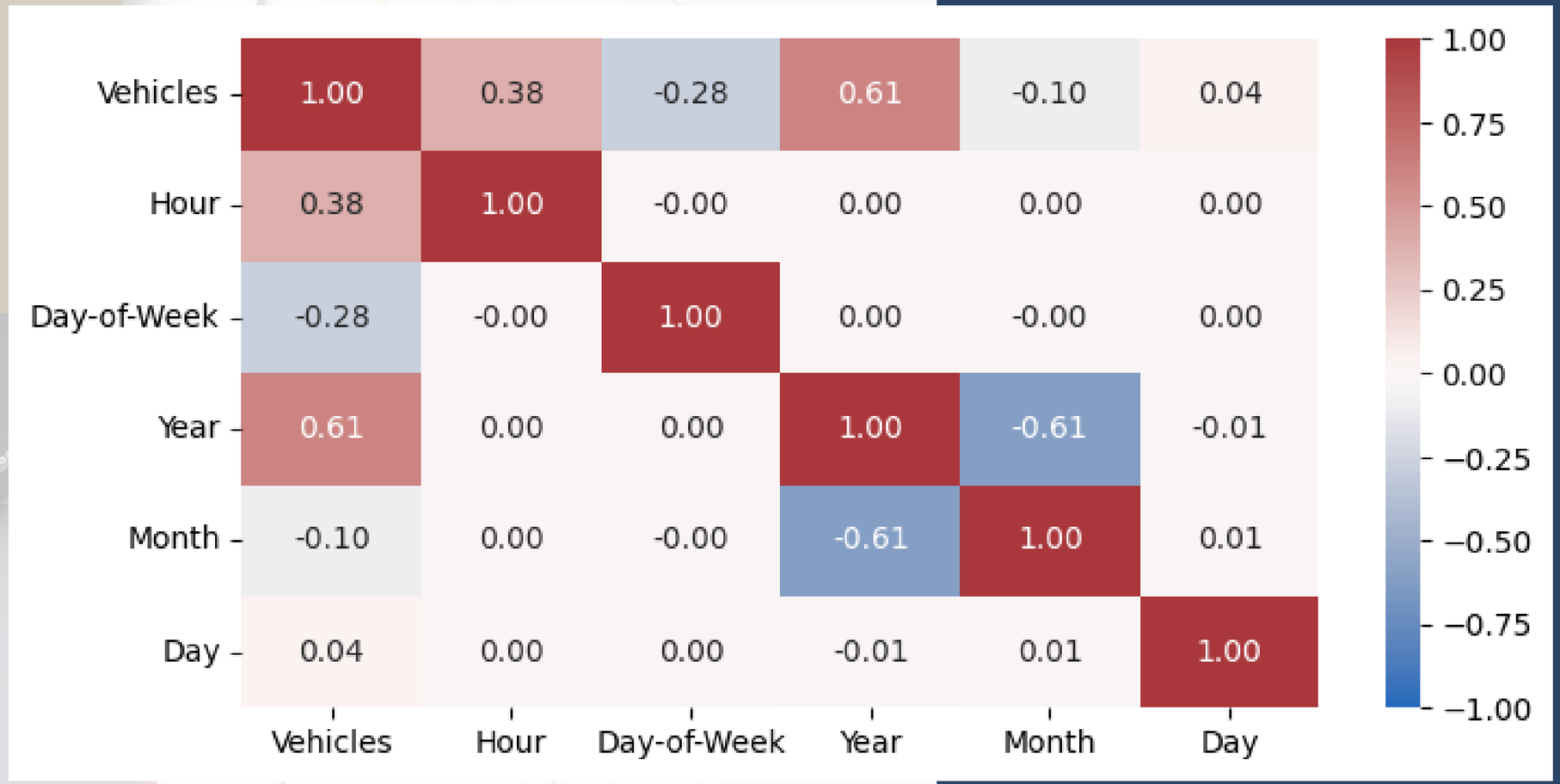
Vehicle count at by year



Vehicle count by Month



Heatmap of Correlation for Junction 1



ML Model evaluation

Random Forest Regression

```
Random Forest Regression Results
Full data --> mse: 12.639263663757273, r2 score: 0.9689869595286663
Junction 1 --> mse: 15.56239653991093, r2 score: 0.9701912190621951
Junction 2 --> mse: 5.102325522439192, r2 score: 0.9047621616969035
Junction 3 --> mse: 30.7244340870161, r2 score: 0.7209014031795145
Junction 4 --> mse: 6.813435673187572, r2 score: 0.5128956014780774
```

Polynomial Regression

```
Polynomial Regression Results
Full data --> mse: 41.796853009592304, r2 score: 0.8974427997987079
Junction 1 --> mse: 47.07896819936367, r2 score: 0.909823230231046
Junction 2 --> mse: 7.8427200427699395, r2 score: 0.8536111230055923
Junction 3 --> mse: 66.87979548789737, r2 score: 0.392468644875665 24
Junction 4 --> mse: 7.679502926329214, r2 score: 0.450978943178765 97
```

XGBoost Regression

```
XGBoost Regression Results
Full data --> mse: 27.08977949544147, r2 score: 0.9335296382604392
Junction 1 --> mse: 19.38231805774552, r2 score: 0.9628744023088939
Junction 2 --> mse: 5.214722032528236, r2 score: 0.9026642162391673
Junction 3 --> mse: 48.47958658496223, r2 score: 0.559614847519092
Junction 4 --> mse: 6.755071586727696, r2 score: 0.5170681518027536
```

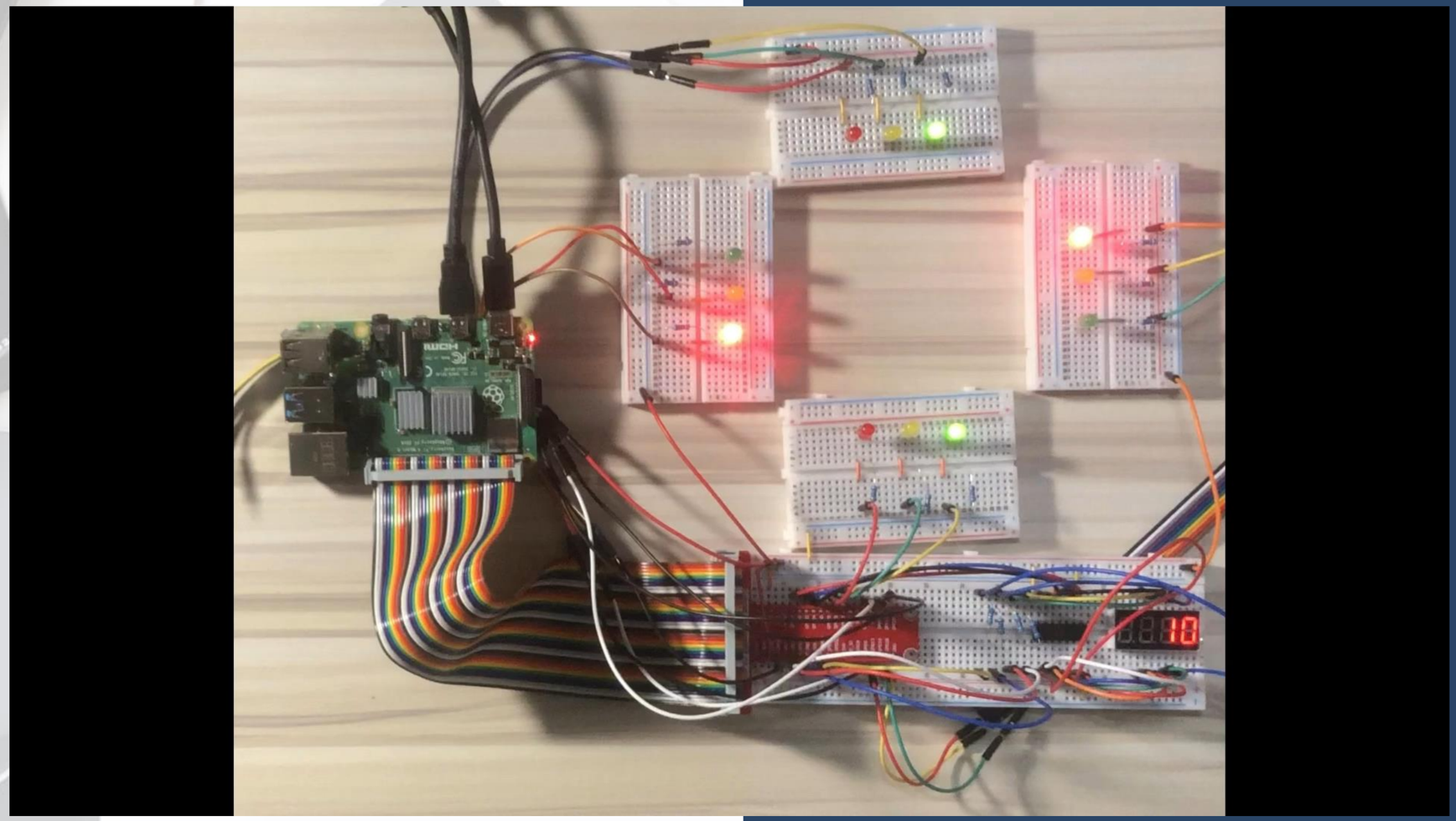
Model Selection

- Random Forest Regression

Results

- Prediction of number of vehicles for 2023 in each junction as output

Using ML predictions in timing traffic light controller



Challenges

- Limited time
- Learning to build the bridge between ML and embedded programming

Future Goals

- Deeper analysis
- Include image recognition to control left turn lights

Thank you!