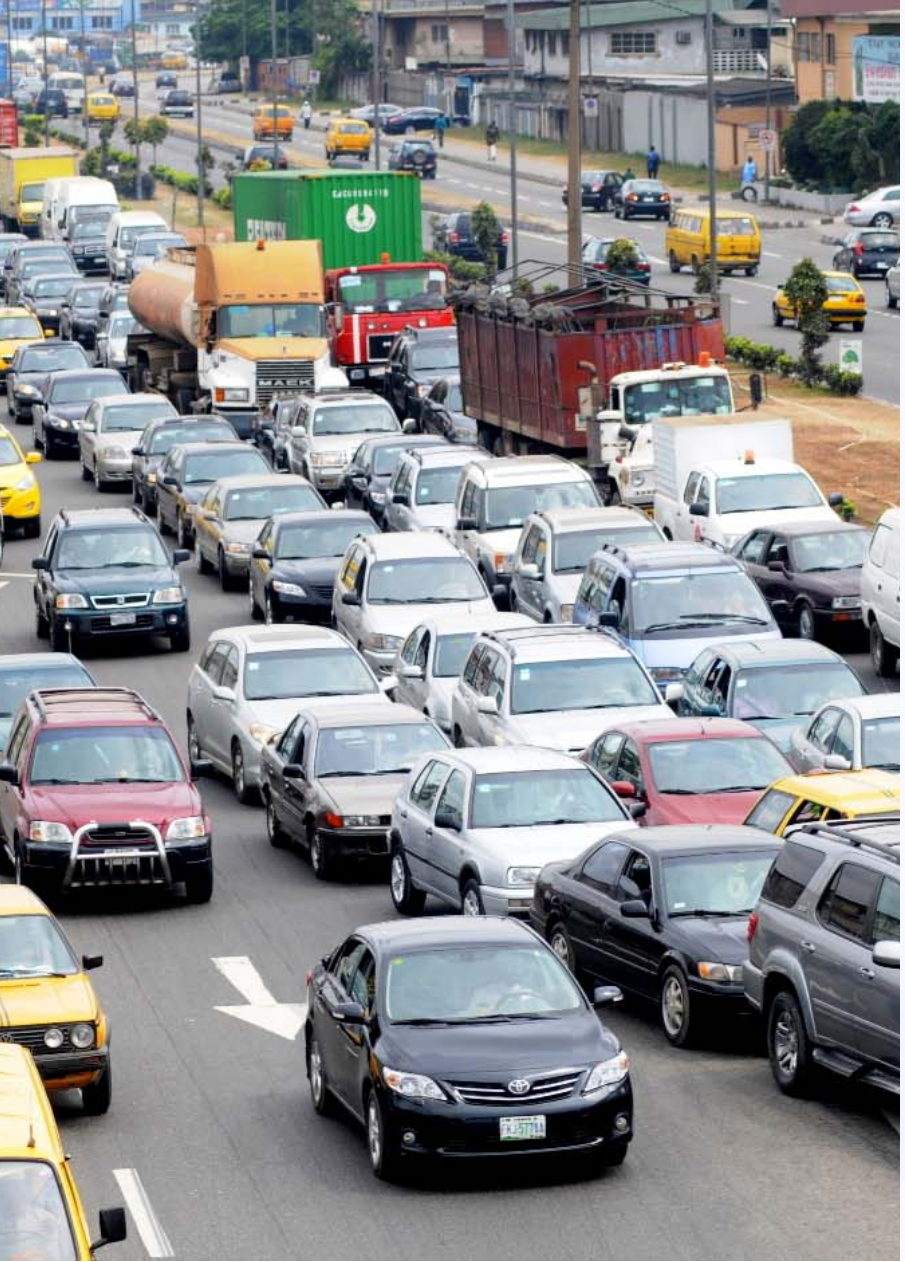


Predicting Traffic Volume using Machine Learning

Transportation plays a vital role in New Zealand's social and economic development. Developing cutting-edge solutions to aid the transportation industry is of utmost importance.

This project aims to predict traffic volume for a given date, time and location.

Navin Sanjay

19/05/2023

About me

- Navin Sanjay
  - Aspiring Junior Data Scientist
- Bachelor of Engineering (Honours) in Mechanical Engineering
  - University of Auckland (2020)
- Product Development Engineer at Fisher & Paykel Healthcare for Two Years

Agenda

1. Introduction
2. Project Context
3. Dataset
4. Data Insights
5. Modelling
6. Conclusions

Introduction

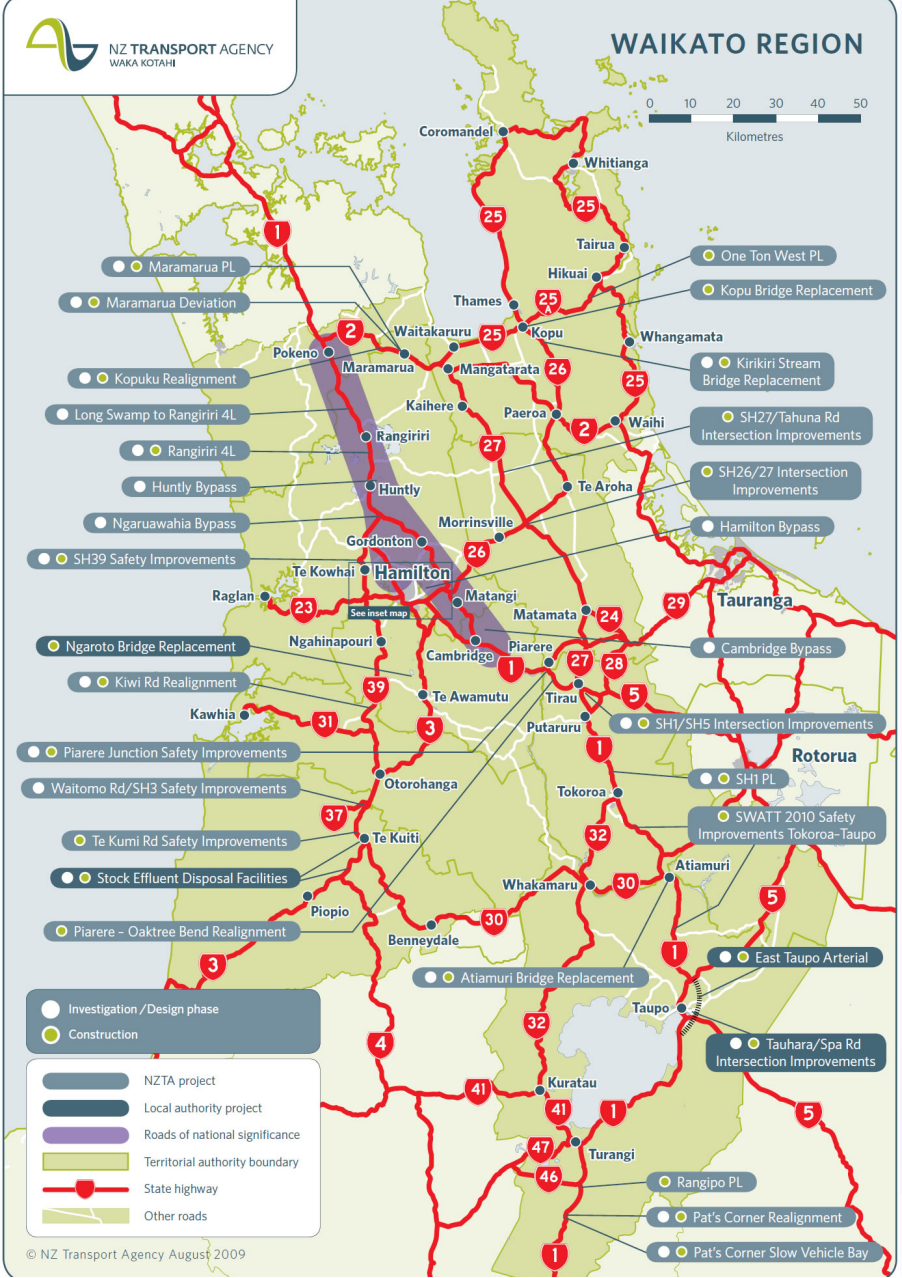
The transportation industry has a crucial role in the economic and social development of New Zealand. It provides connections between people, goods, and services, facilitating trade, tourism and overall movement. That's why its important to continue to develop solutions that will aid the transportation industry.

Project Context

Create a model which can **predict the volume of traffic for a specified point given a Date and Time.**

Stakeholders:

- **Transportation Agencies**
  - Helps agencies plan things such as road maintenance, traffic signal timing, and new infrastructure based on expected traffic demand
  - Help manage traffic flow, mitigate congestion etc
- **Urban Developers**
  - Give insights on how to design and develop transportation infrastructure to handle expected traffic volume
    - Road Networks, Highways, Bridges and Public Transportation systems
- Businesses and Retailers
- Communities & Residents



Dataset

The NZ Transport Agency dataset was used.

- Daily traffic volumes from state highway count sites in 15-minute intervals by vehicle class and direction.
  -

opendata-nzta.opendata.arcgis.com

TMS traffic quarter-hourly: Jan 2013...

Daily traffic volumes from state highway count sites in 15-minute intervals, by vehicle class a...

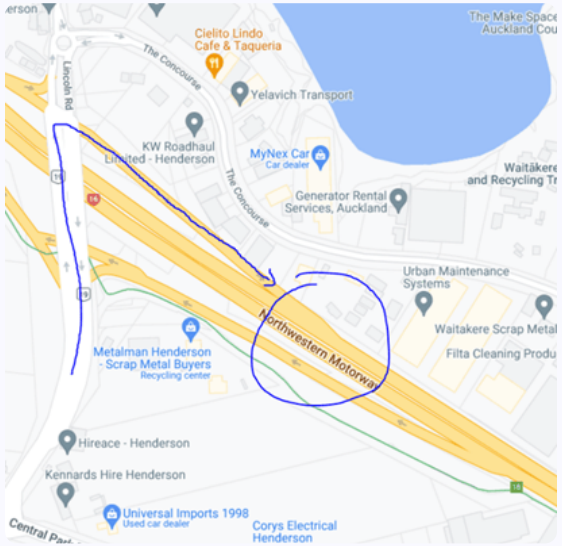
- The data covers the period of January 2013 to September 2020
  - This project looks at 2017 data

Dataset

Location

- Lincoln Rd SH16 Interchange East Bound, Henderson, Auckland, NZ, leaving the suburb going to the city

Location was chosen at a place that is near me. Looks at the amount of people **leaving** the suburb going on the motorway towards the city.



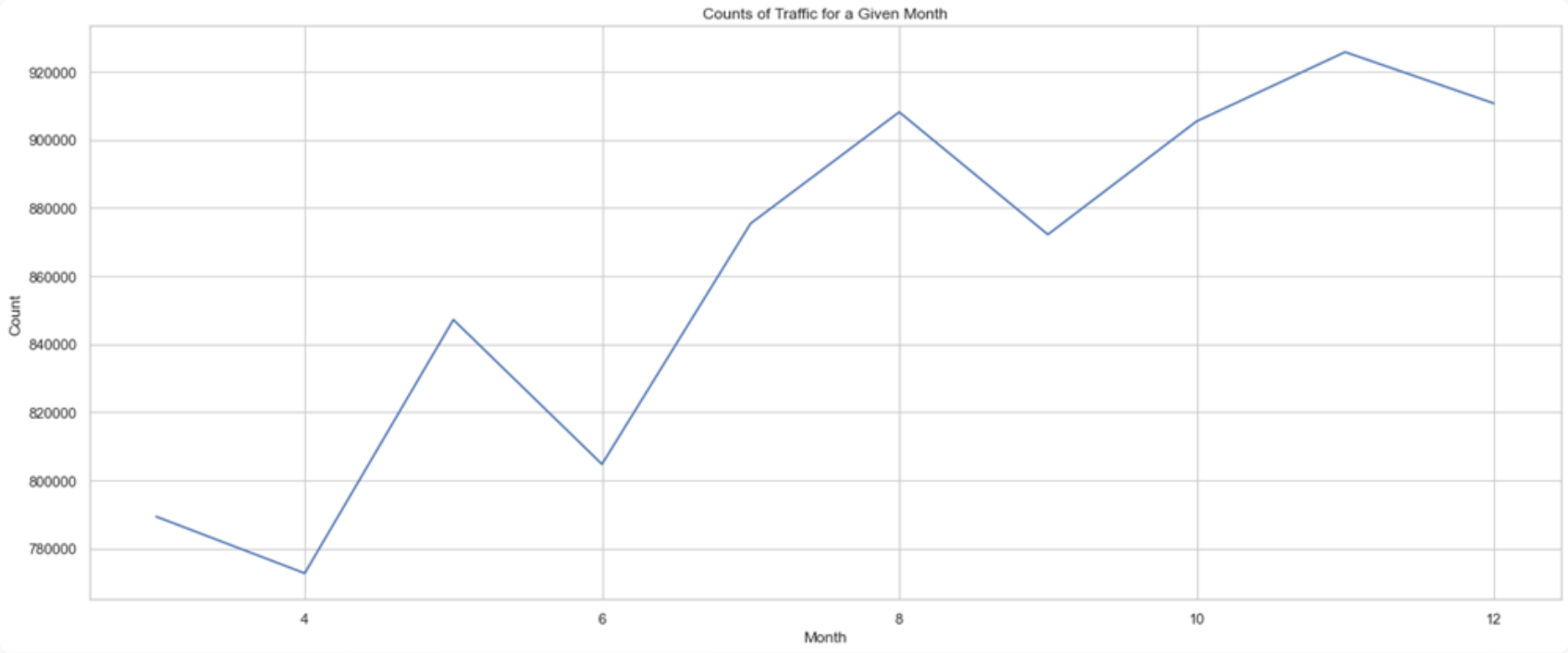
Dataset: Raw Dataset

	class	siteRef	startDatetime	endDatetime	direction	count
	L	01620013	22-FEB-2017 03:15	22-FEB-2017 03:30	2	23.0
	L	01620013	22-FEB-2017 21:15	22-FEB-2017 21:30	2	211.0
	L	01620013	22-FEB-2017 23:30	22-FEB-2017 23:45	2	46.0
	L	01620013	23-FEB-2017 18:30	23-FEB-2017 18:45	2	418.0
	L	01620013	24-FEB-2017 03:15	24-FEB-2017 03:30	2	19.0
	L	01620013	25-FEB-2017 10:45	25-FEB-2017 11:00	2	445.0
	L	01620013	26-FEB-2017 11:00	26-FEB-2017 11:15	2	355.0
	L	01620013	28-FEB-2017 15:15	28-FEB-2017 15:30	2	369.0
	L	01620013	04-MAR-2017 19:30	04-MAR-2017 19:45	2	238.0
	L	01620013	05-MAR-2017 08:30	05-MAR-2017 08:45	2	233.0

Shape: (29856, 6)

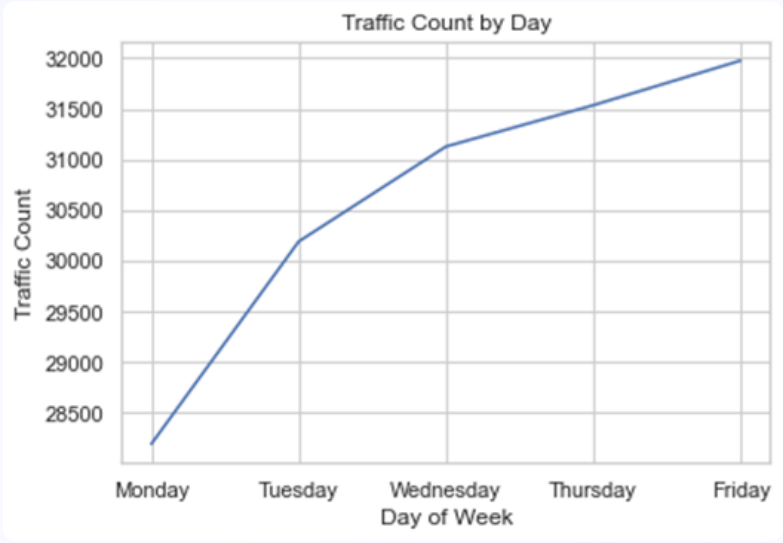
Uncovering Traffic Insights

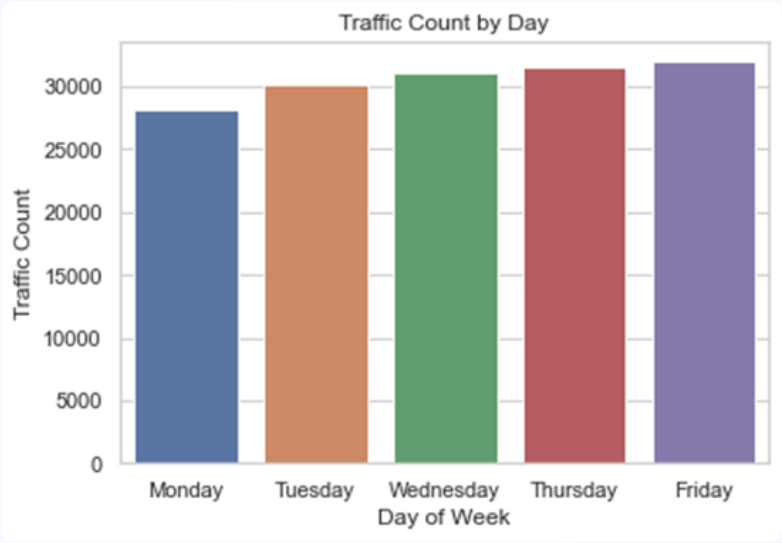
Insights: Yearly Trend of Traffic Count



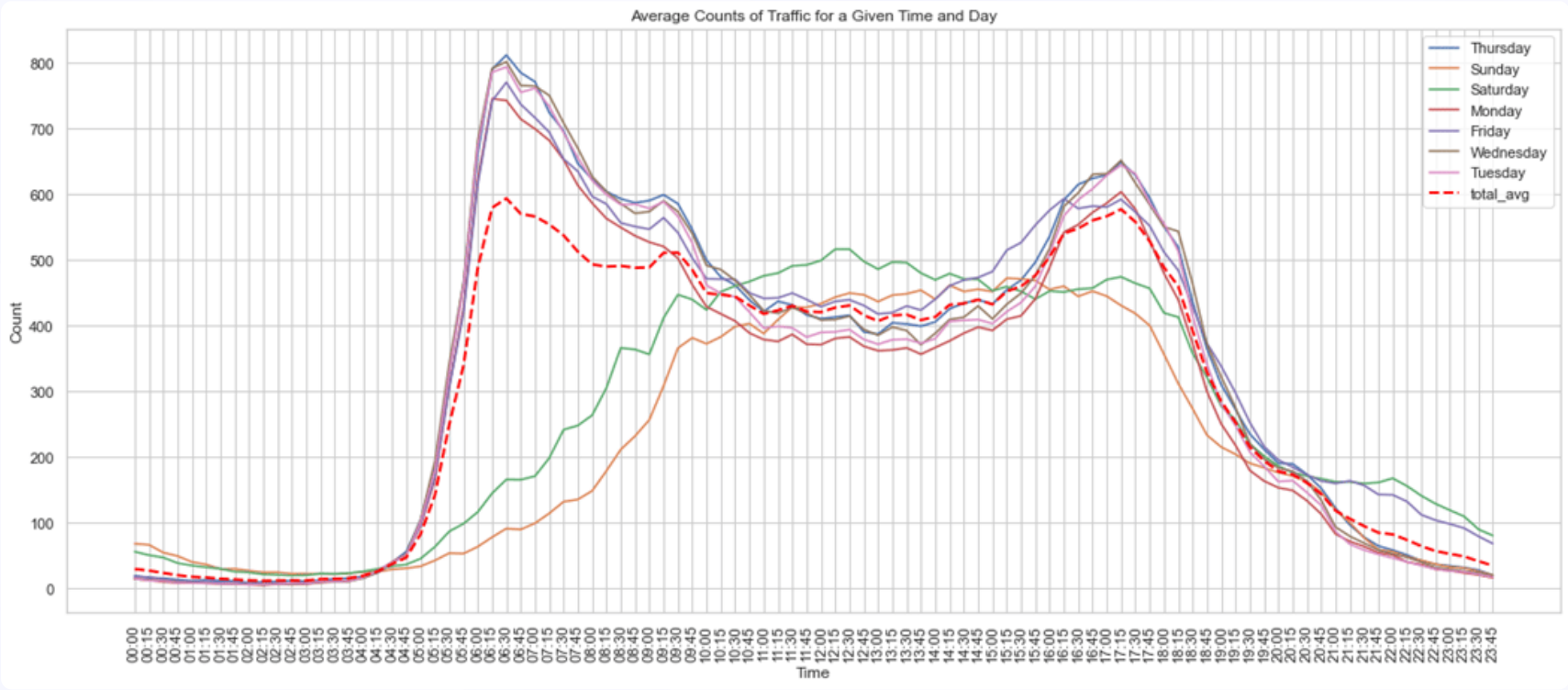
Total cars counted 2017: **8,840,597**

Insights: Traffic Count per Week Day

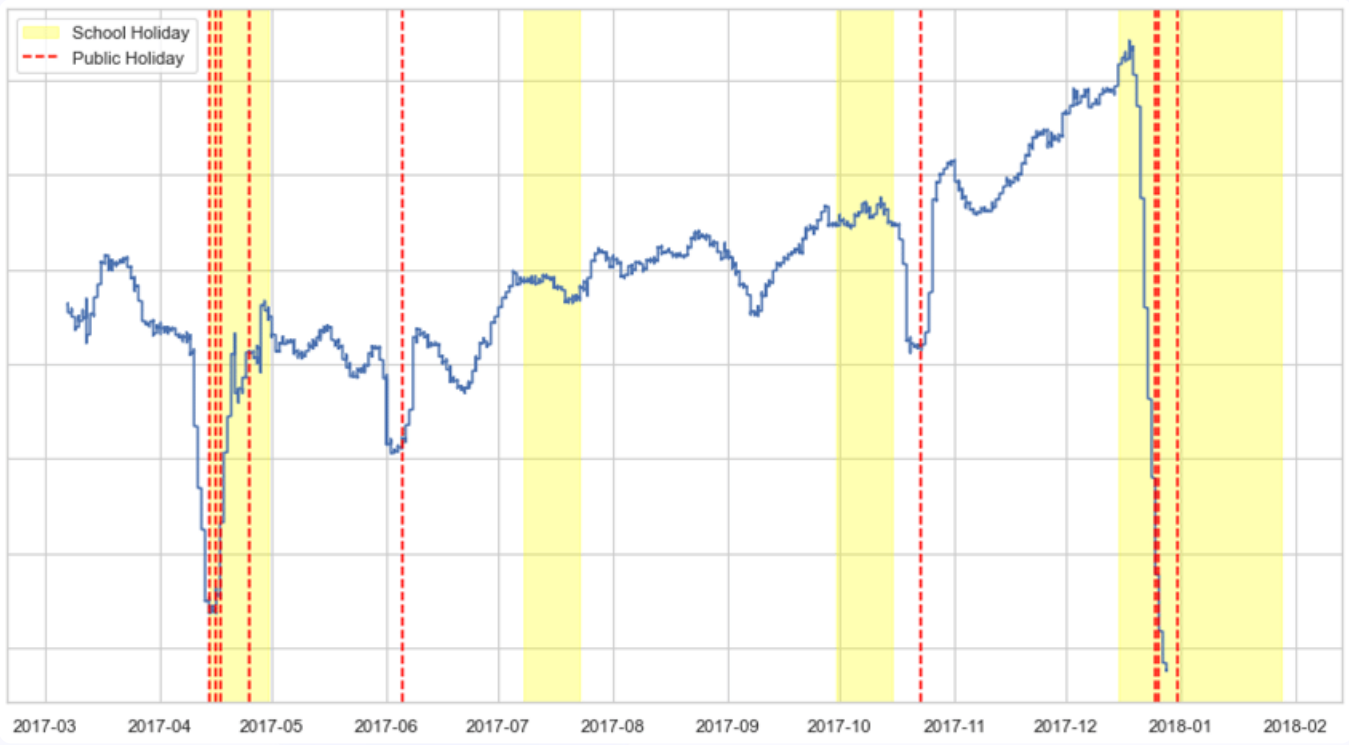




Insights: Traffic Count for a given Time and Day



Insights: Traffic Count Trend



Feature Engineering

- External Features were added:
  - **Public Holidays**
    - Can significantly influence traffic patterns. People aren't going to work/school.
  - **Average Monthly Temperature**
    - Temperatures can affect the activities people do on certain days
  - **Average Monthly Rainfall**
    - People may alter their travel plans or choose different routes due to adverse weather conditions
  - **Lagged Traffic**
    - Making forecasts based on previous traffic counts
    - 1 hour lag, 24 hour lag and 1 week lag

Feature Selection: Correlation Matrix

1

Data Pre-Processing

- Feature Engineering/Selection
- Define predictors
  - 1 week traffic lag
  - Average Monthly Rainfall
  - Average Monthly Temperature
  - Public Holidays
- Define Target
  - Traffic Count (y)

2

Choose Models

Supervised Learning in Time Series Forecasting

- Linear Regression
- Decision Tree Regression
- **Random Forest Regression**
- Prophet
- *Hyper-parameter Tuning*
- *Feature Importance*
- *Forward Feature Selection*

3

Modelling

Train models

Evaluate Performance

- Mean Absolute Error
- Root Mean Squared Error
- R^2

Test Train Split

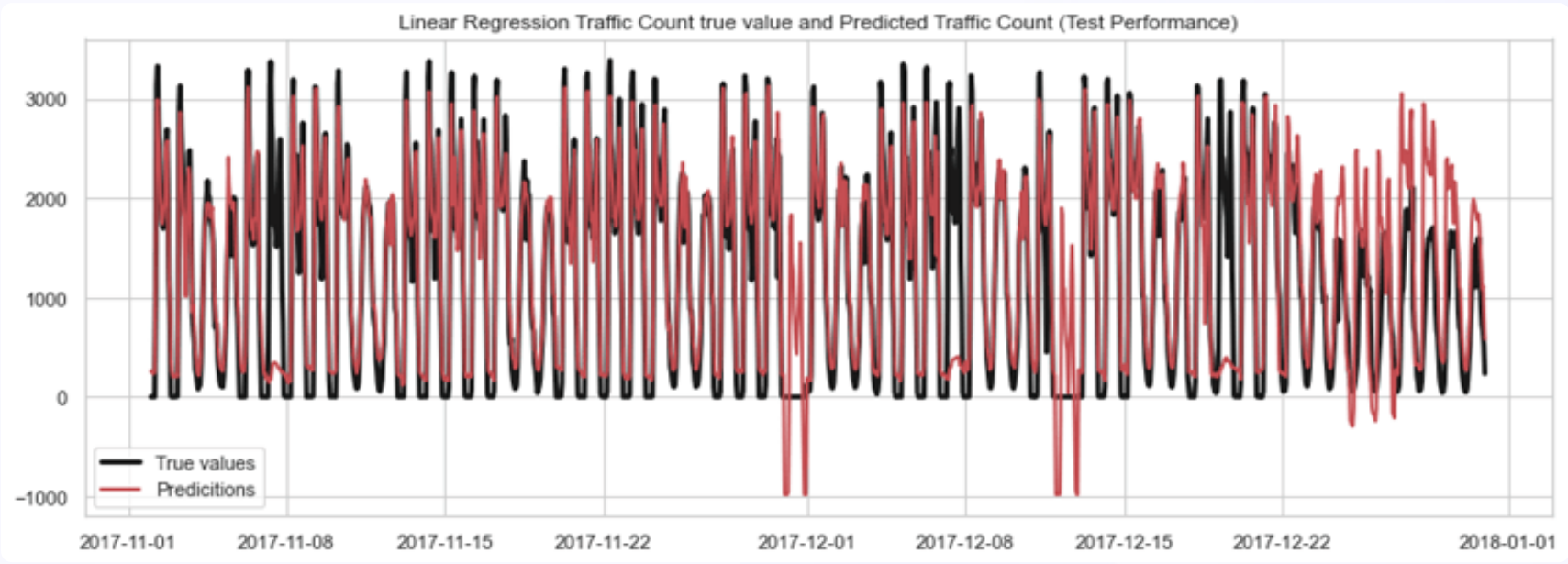
- 80% Train, 20% Split
  - Train: March-November
  - Test: November - December

Model Comparison

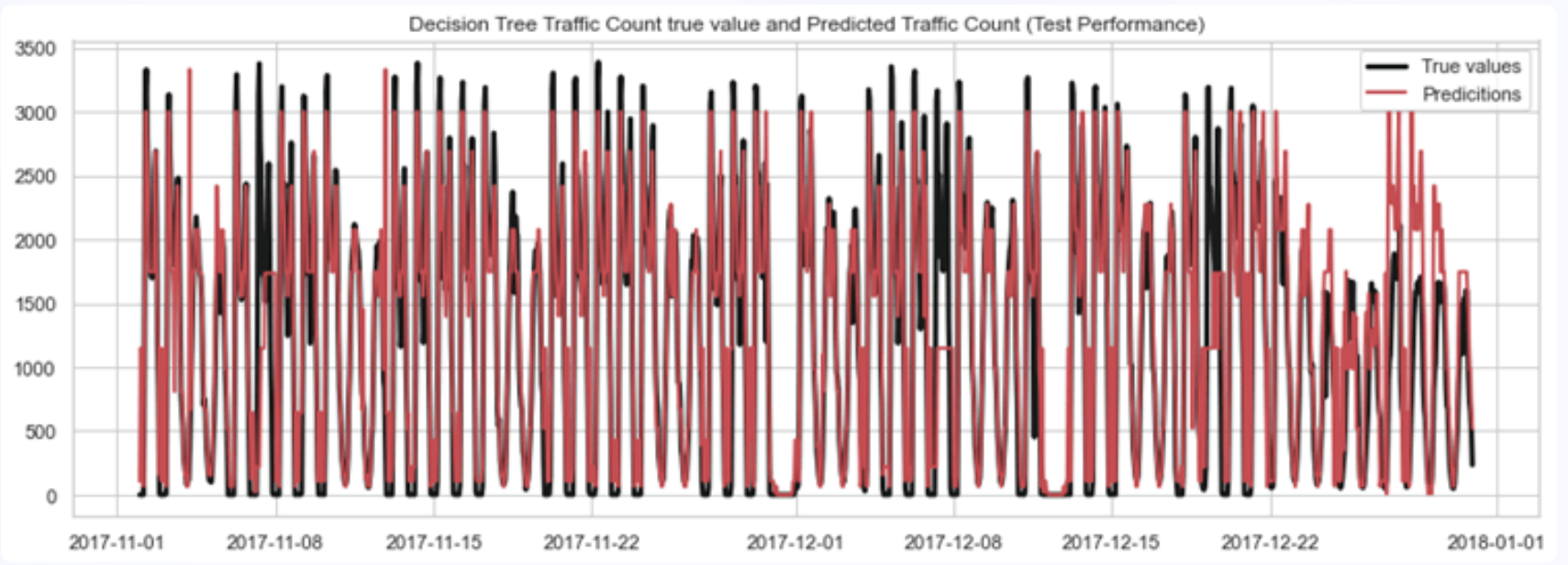
TEST				
Metric	Linear Regression	Decision Trees	Random Forest	Prophet
MAE	287.01	256.75	252.66	426.58
RMSE	524.54	461.95	447.64	584.18
R2	0.72	0.78	0.80	0.63

Model Comparison

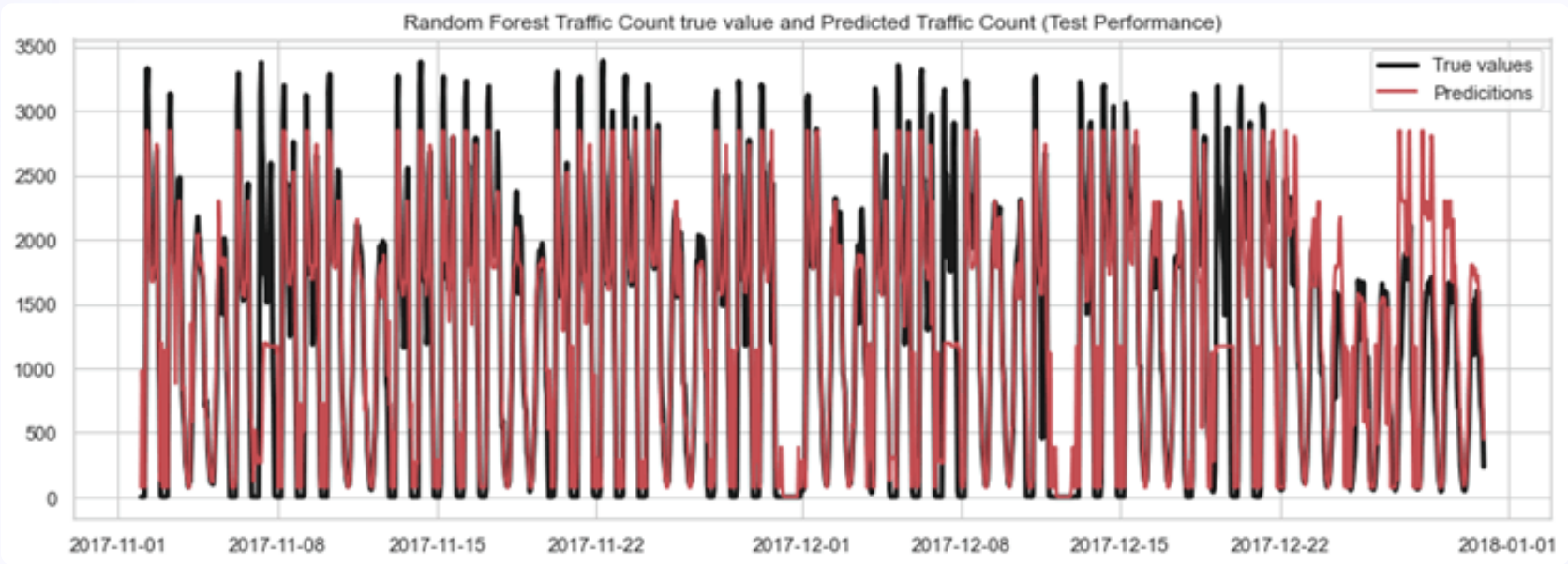
- Linear Regression:



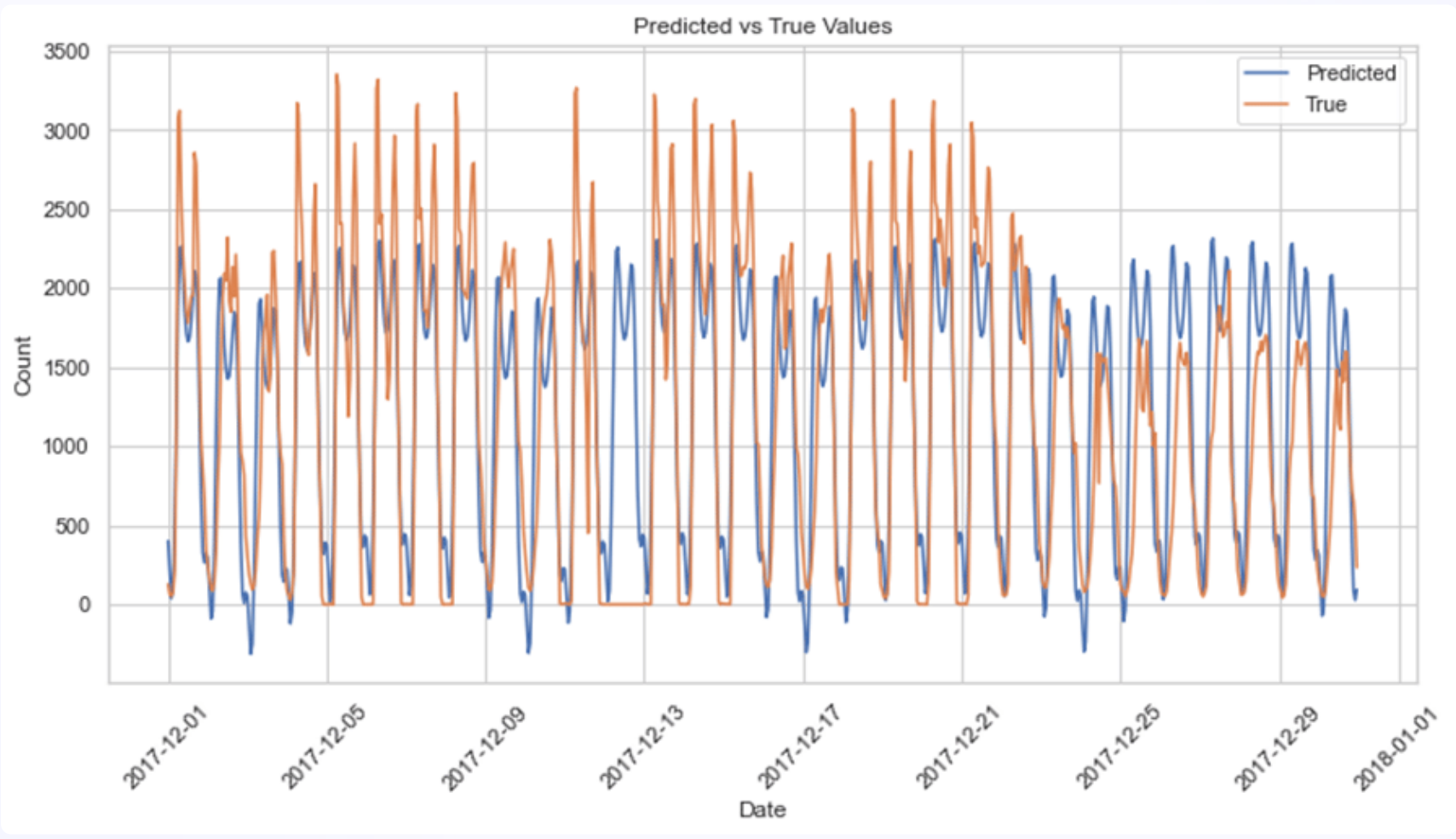
- Decision Tree:



- Random Forest Regression:



- Prophet Model:



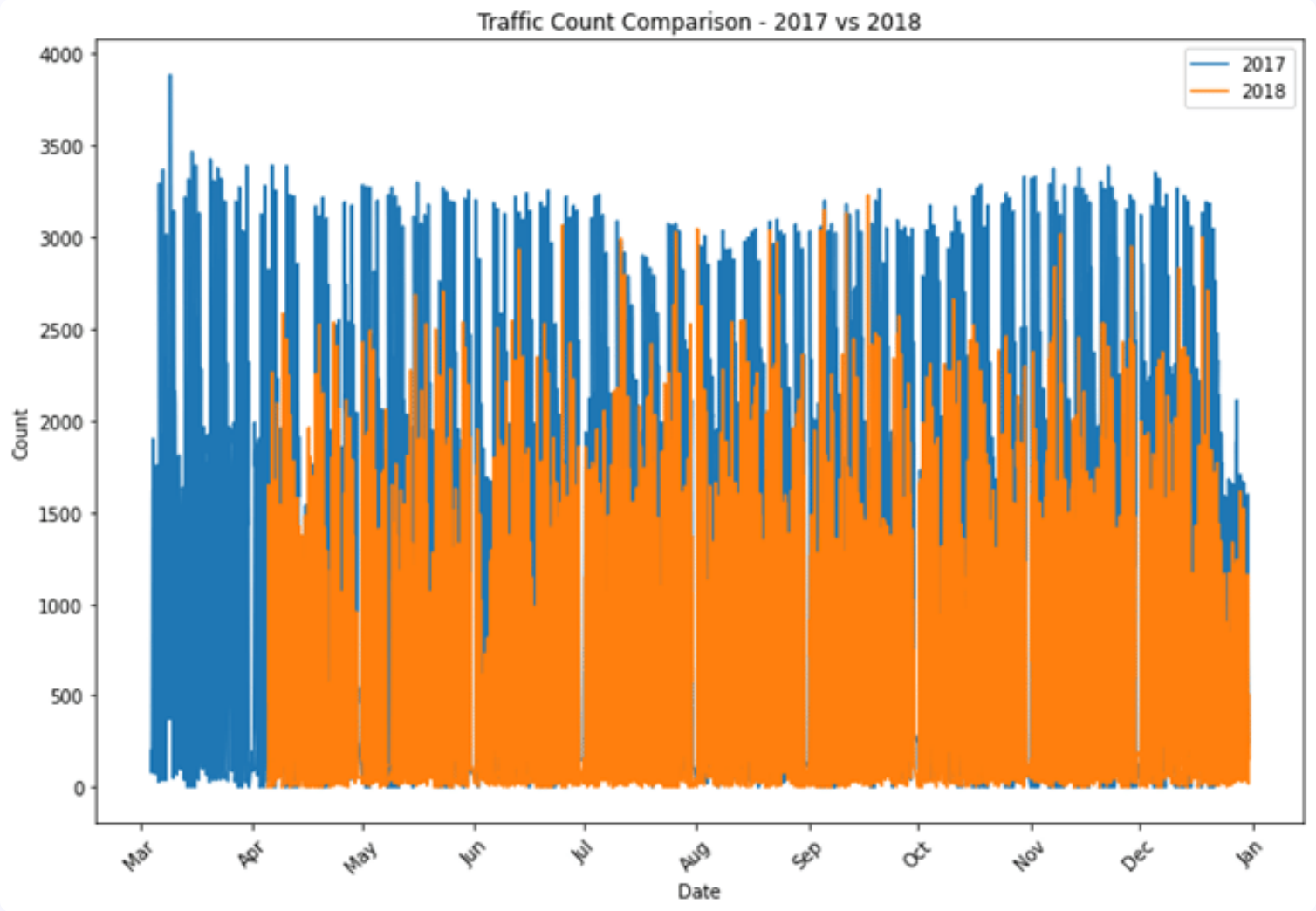
Model Comparison: 2018 Data

- The models were initially trained and tested on 2017 data
- Those models were then tested on the 2018 data to see how well they would perform

Model Comparison: 2018 Data

2018 Data				
Metric	Linear Regression	Decision Trees	Random Forest	Prophet
MAE	356.28	371.96	364.95	475.63
RMSE	506.37	548.80	535.96	606.54
R2	0.30	0.17	0.21	0.01

Model Performance: 2018 Data Limitations



Summary of Results

- In this presentation, we explored the problem of predicting traffic volume using machine learning. We discussed the dataset, data insights, feature engineering, model comparison, and model performance. We also highlighted the limitations of the model and suggested ways to improve it.
- Models were trained/tested on 2017 data
  - Trained on months: March - November
  - Tested: November - December
- Features:
  - 1 Week Lagged Traffic
  - Average Monthly Temperature
  - Average Monthly Rainfall
- Final Model:
  - Random Forest Regression
  - 80% Accuracy, Error of ~252 cars

Going Forward

To improve the model, more historical data needs to be obtained and cleaned up, and more robust data is also required. The model will need to be retrained on this data and additional models such as LSTM, ARIMA, and spatial temporal modelling can be used.

To deploy the model, a web-based app can be developed.

Appendix

- https://opendata-nzta.opendata.arcgis.com/datasets/NZTA::tms-traffic-quarter-hourly-jan-2013-to-sept-2020/about
- https://www.holidays-info.com/new-zealand/calendar/auckland/2017/
- https://facebook.github.io/prophet/
- https://www.transport.govt.nz/assets/Uploads/Report/TransportOutlookFutureOverview.pdf

End of Presentation!

Thank you for your attention.

Questions?

Thank you for joining today. I hope you found this presentation informative and insightful. If you have any questions or comments, please don't hesitate to ask.