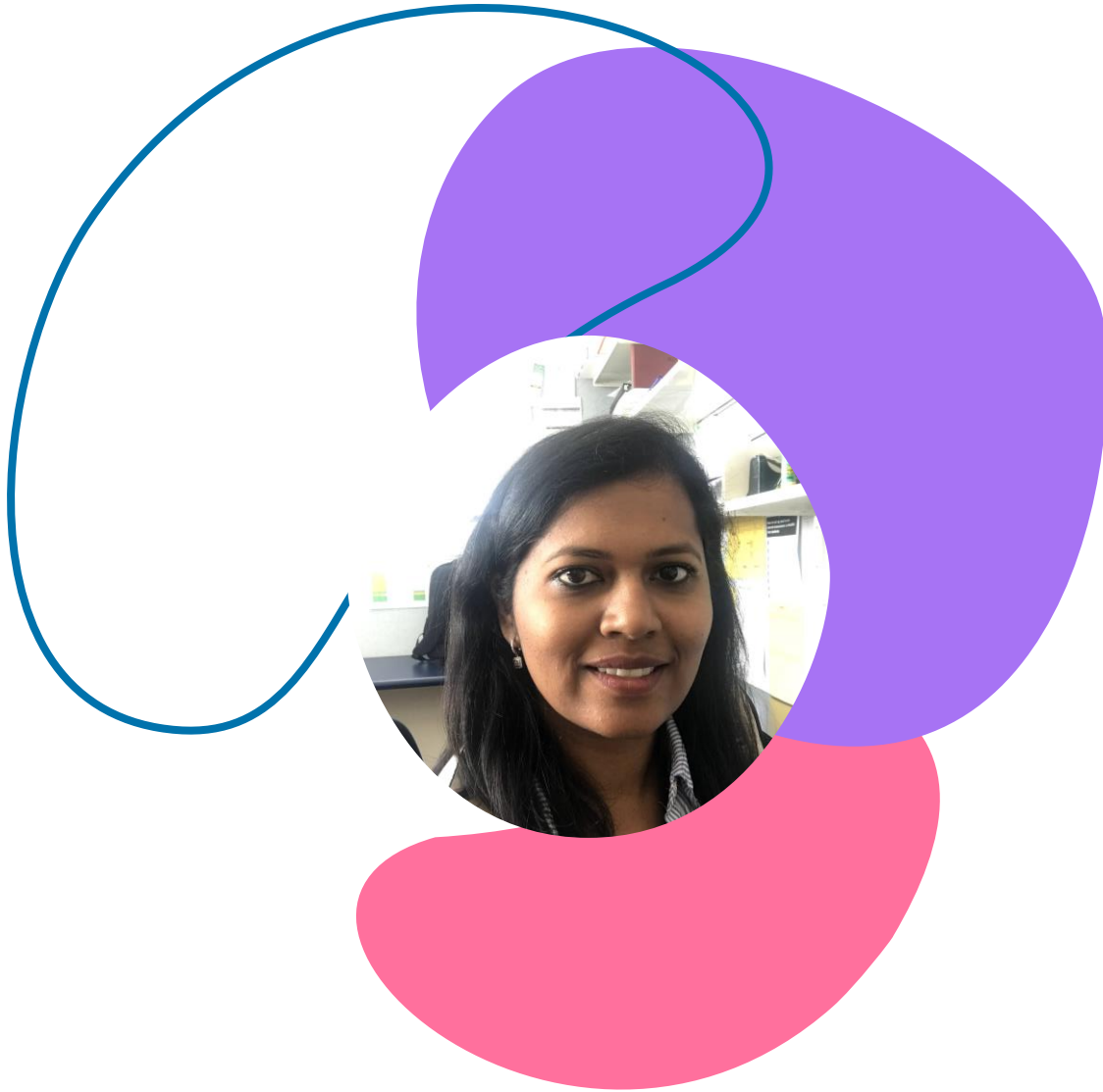


TRAFFIC FLOW PREDICTION

CAPSTONE PROJECT

Esther Dantra





About Me

Esther Dantra

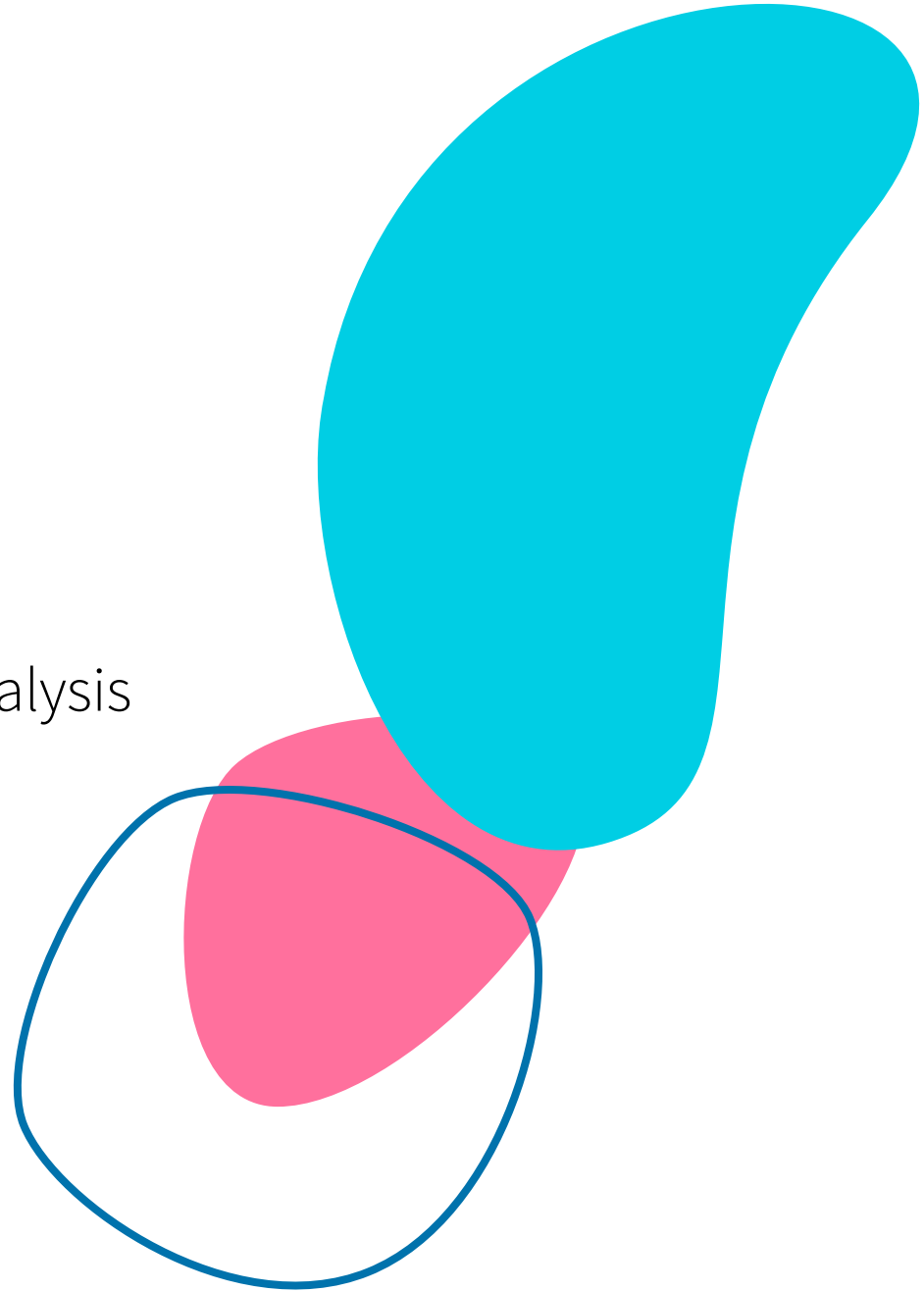
BSc, PGDipSci, GradDipTchg



<http://nz..linkedin.com/in/esther-dantra>
esther.mayari@gmail.com

AGENDA

- Business Problem
- The process
- About the dataset
- Initial Observations – Exploratory data analysis
- Models used and their metrics
- Chosen model efficiency
- Limitations and Summary



Business Problem

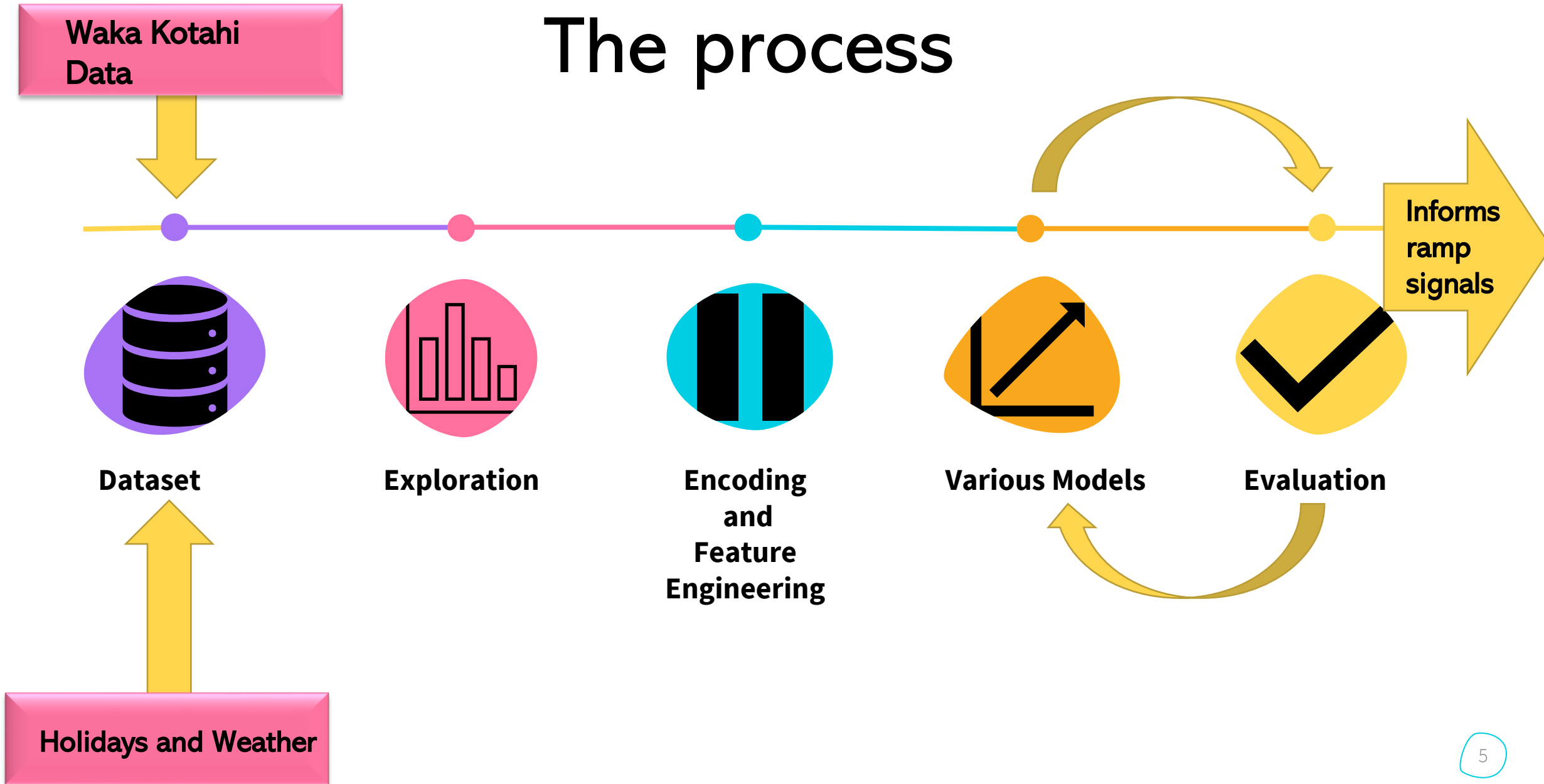
GOAL:

- To explore the factors that impact Traffic flow on motorways, including weather conditions and categorization of the day.
- To build a model which can predict traffic flow on Auckland Motorways and help inform traffic management systems.

"...costs of congestion in Auckland are approximately \$1,250 million per year...."



The process



About the dataset



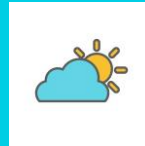
Traffic Count

Traffic counts at various sites across highways in New Zealand from 2018-2019



Holidays

Public Holiday and School Holiday Info from Auckland



Weather

Hourly Weather data from Auckland.
Includes: Cloud cover, visibility, Temperature etc



Monitoring Sites

Approx. 2000 sites. Has info on site type, location, annual average daily traffic count for previous years.

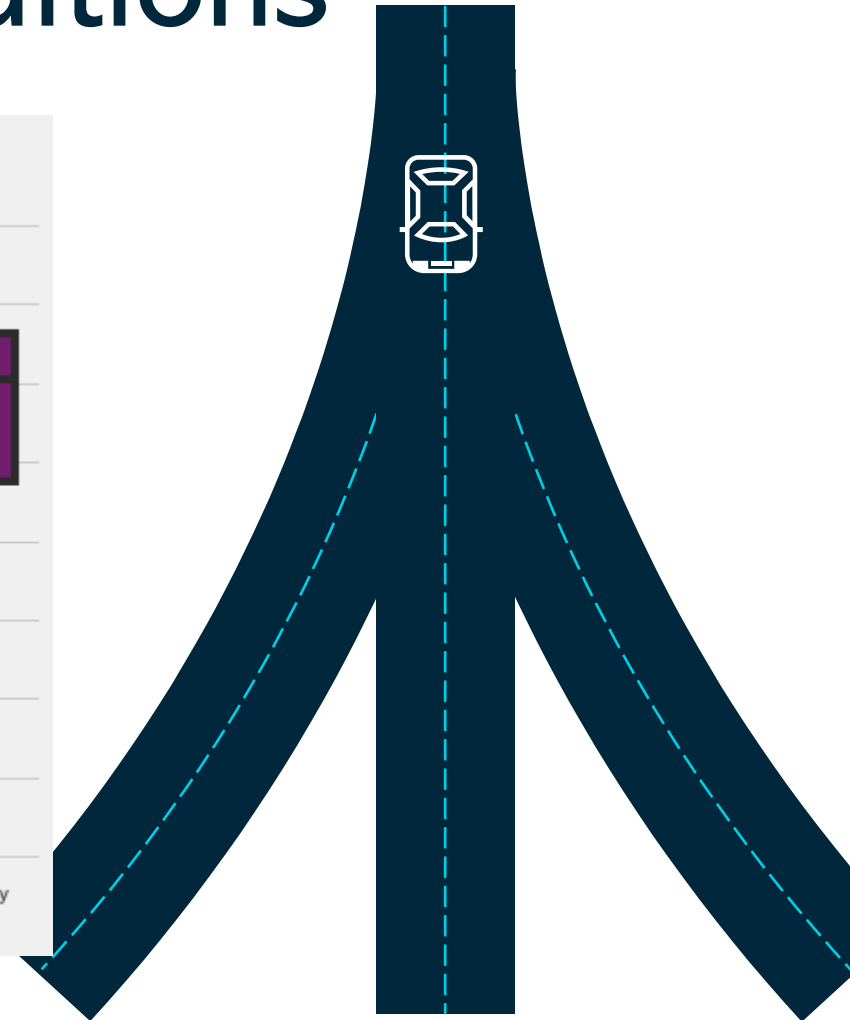
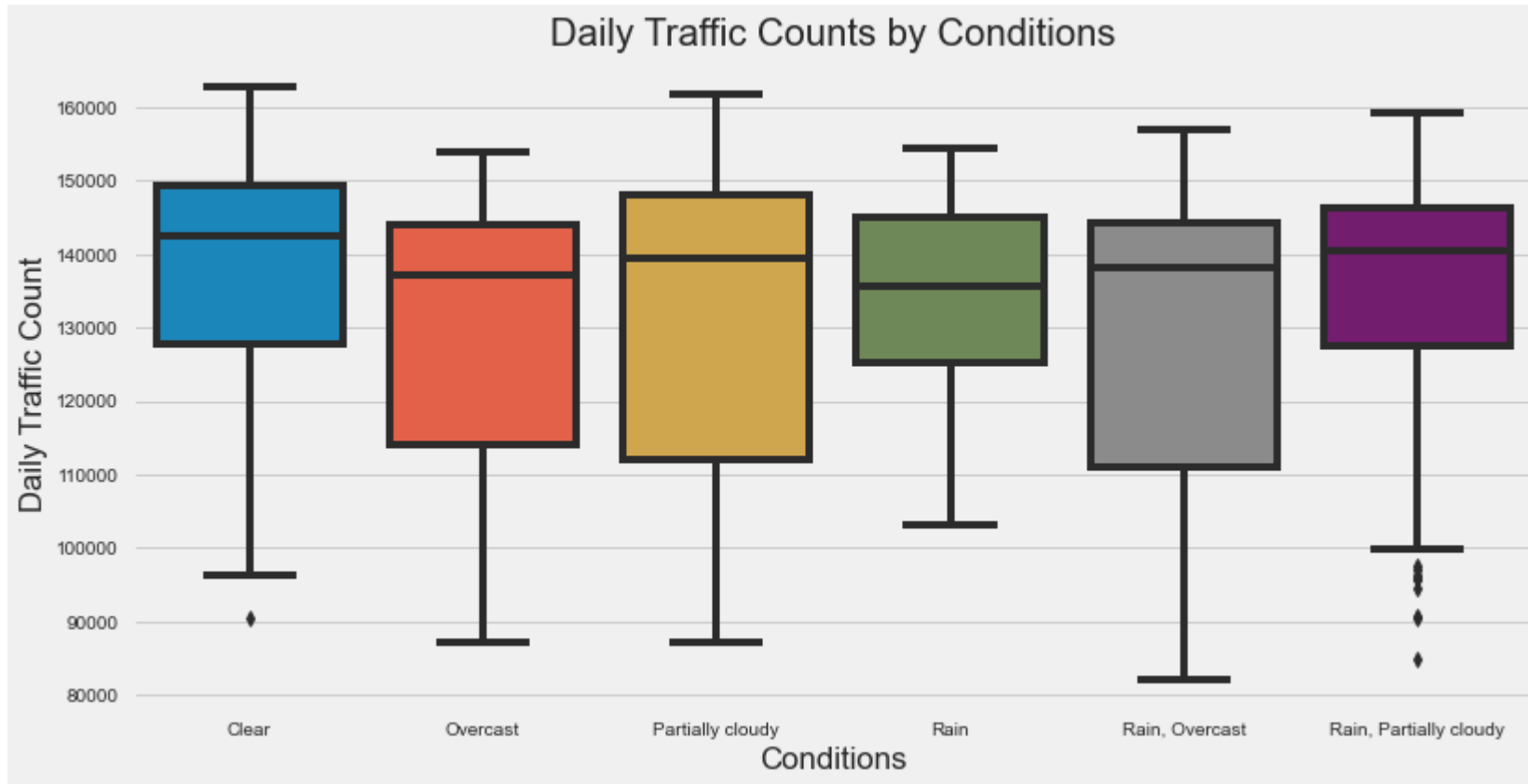


Final Dataset

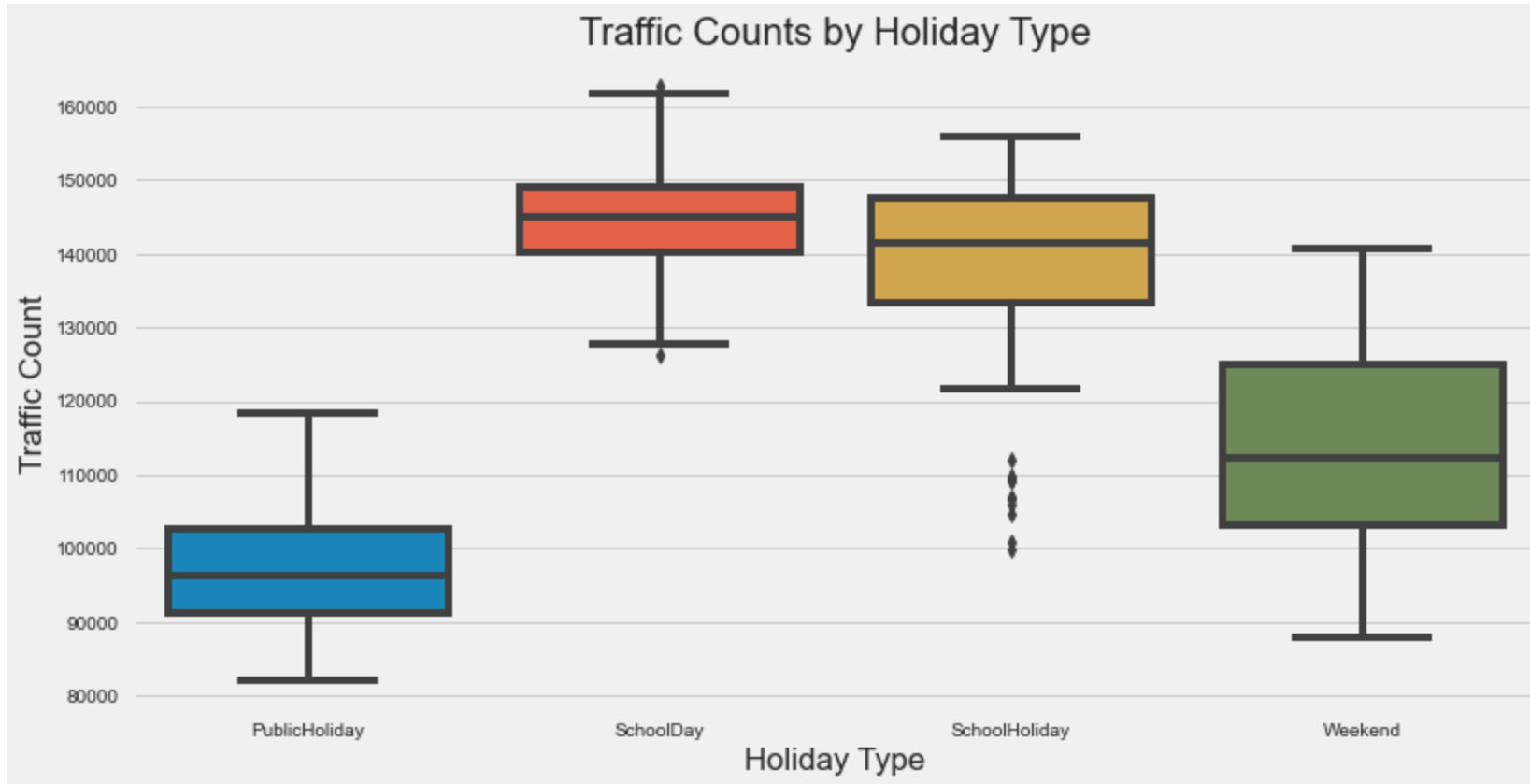
Number of features = 14
About 15000 rows per site.
Lincoln Road and Te Atatu Junction

	siteRef	X	Y	year	lane	percentHeavy	AADT5yearsAgo	AADT4yearsAgo	AADT3yearsAgo	AADT2yearsAgo	AADT1yearAgo	Minimum Temperature	Maximum Temperature
startDatetime													
2018-01-01 00:00:00	1610011	174.654029	-36.858365	2018	Inc	8.4	37636.0	39823.0	40574.0	37891.0	39148.0	18.0	
2018-01-01 01:00:00	1610011	174.654029	-36.858365	2018	Inc	8.4	37636.0	39823.0	40574.0	37891.0	39148.0	18.0	

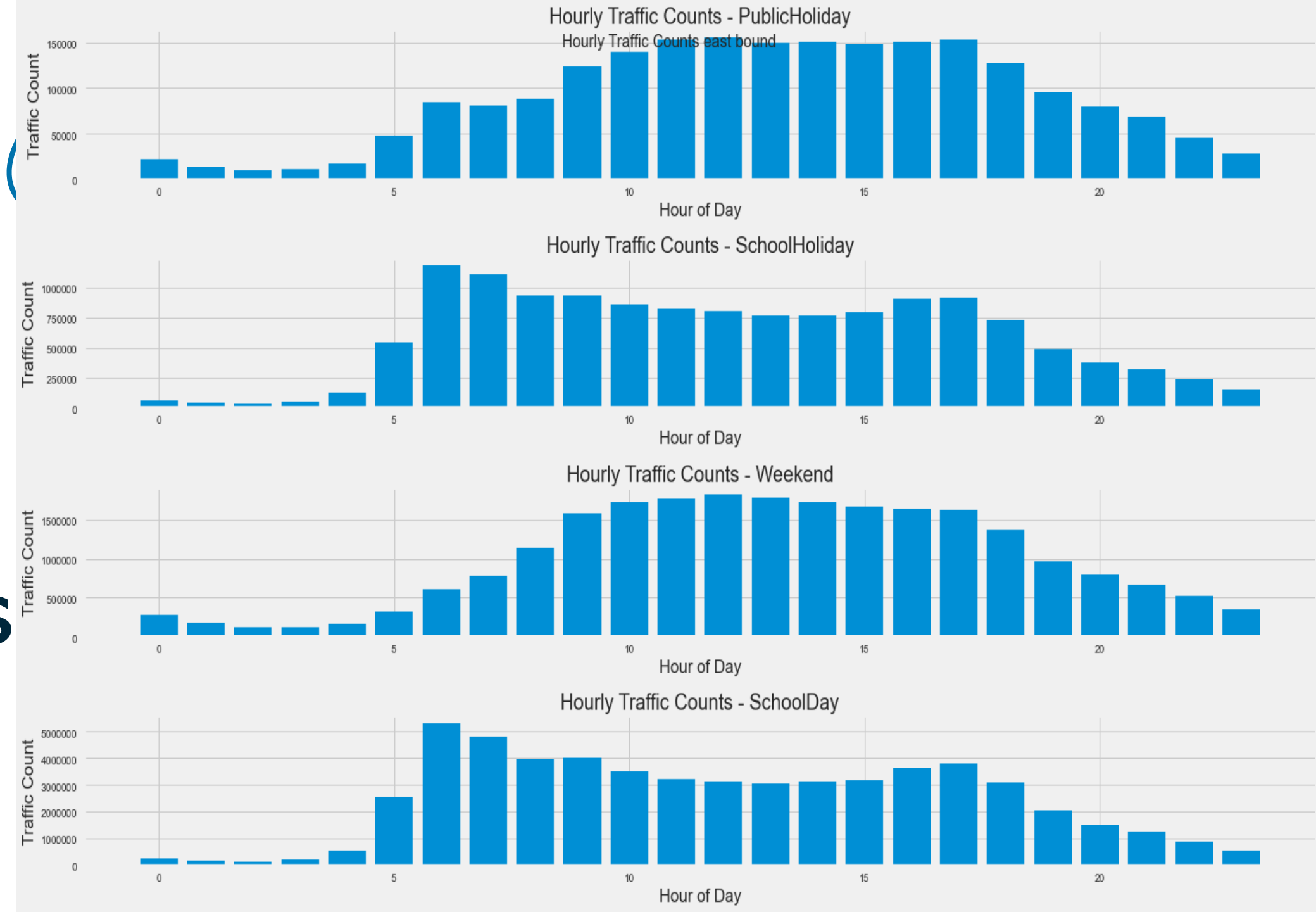
Traffic counts by weather conditions



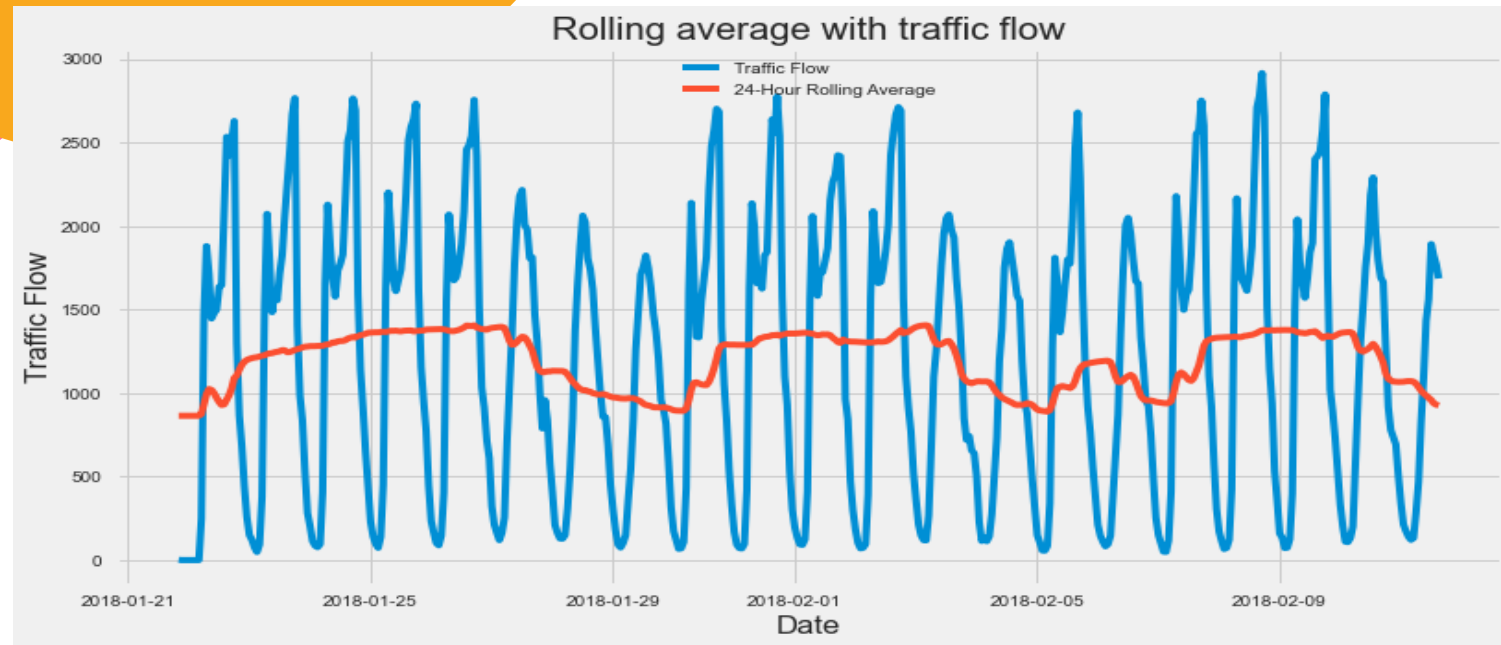
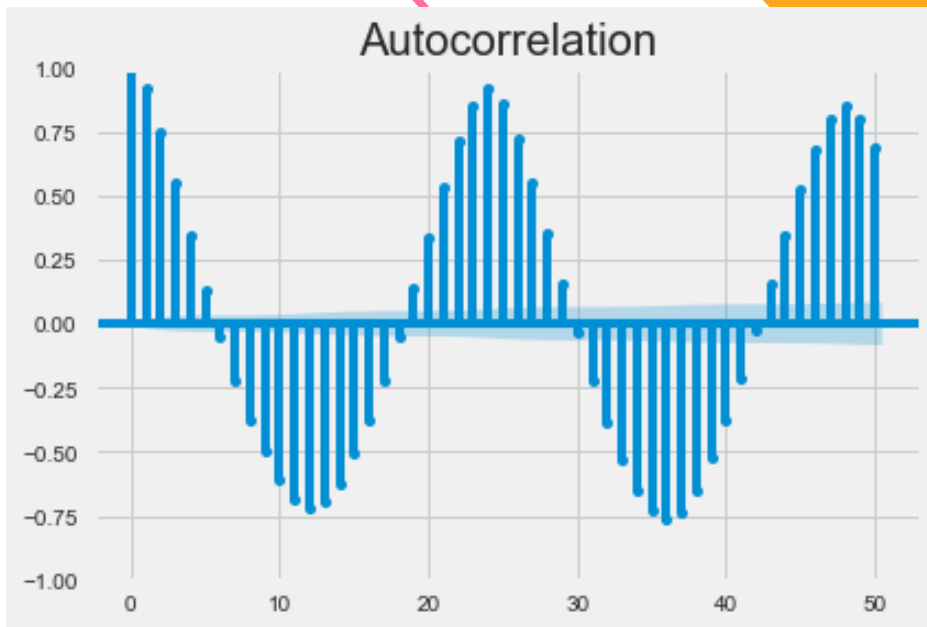
Traffic Counts by Type of Day



Traffic counts by the Hour



Time Series Analysis



ADF Statistic: -14.508013913379237

p-value: 5.786104273355456e-27

Critical Values:

1%: -3.431

5%: -2.862

10%: -2.567

Models Used

ARIMAX

Best model: `ARIMA(2,0,4)(0,0,0)[0] intercept`
Total fit time: 172.813 seconds
Model AIC: 195336.82040931657

SARIMAX

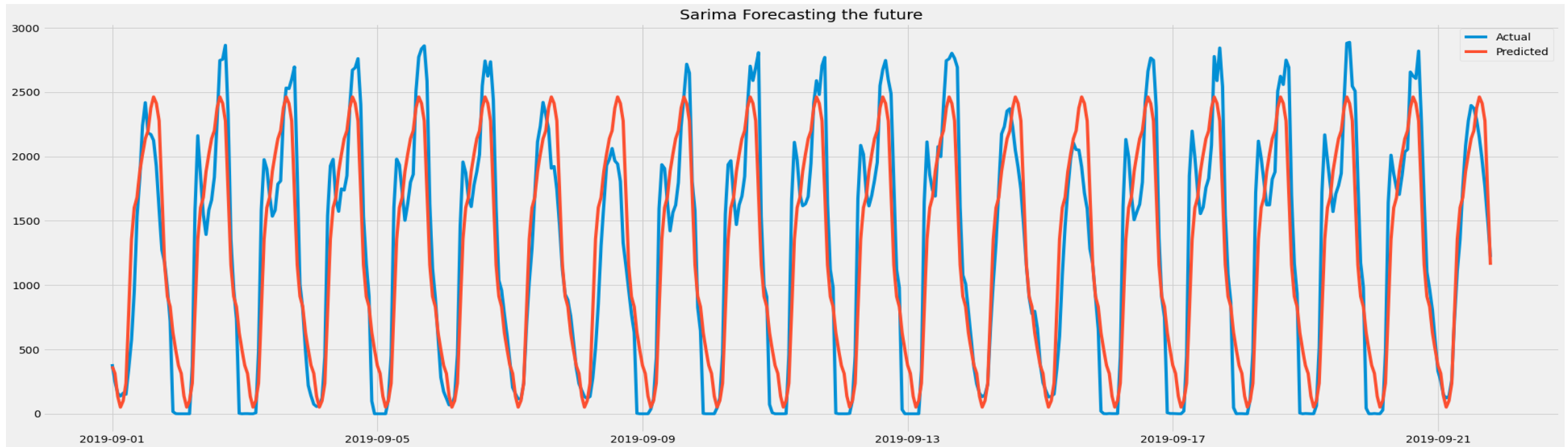
Mean squared error: $8.588884718677862e-09$
Standard deviation of squared error: $1.2146517254804674e-08$

LSTM

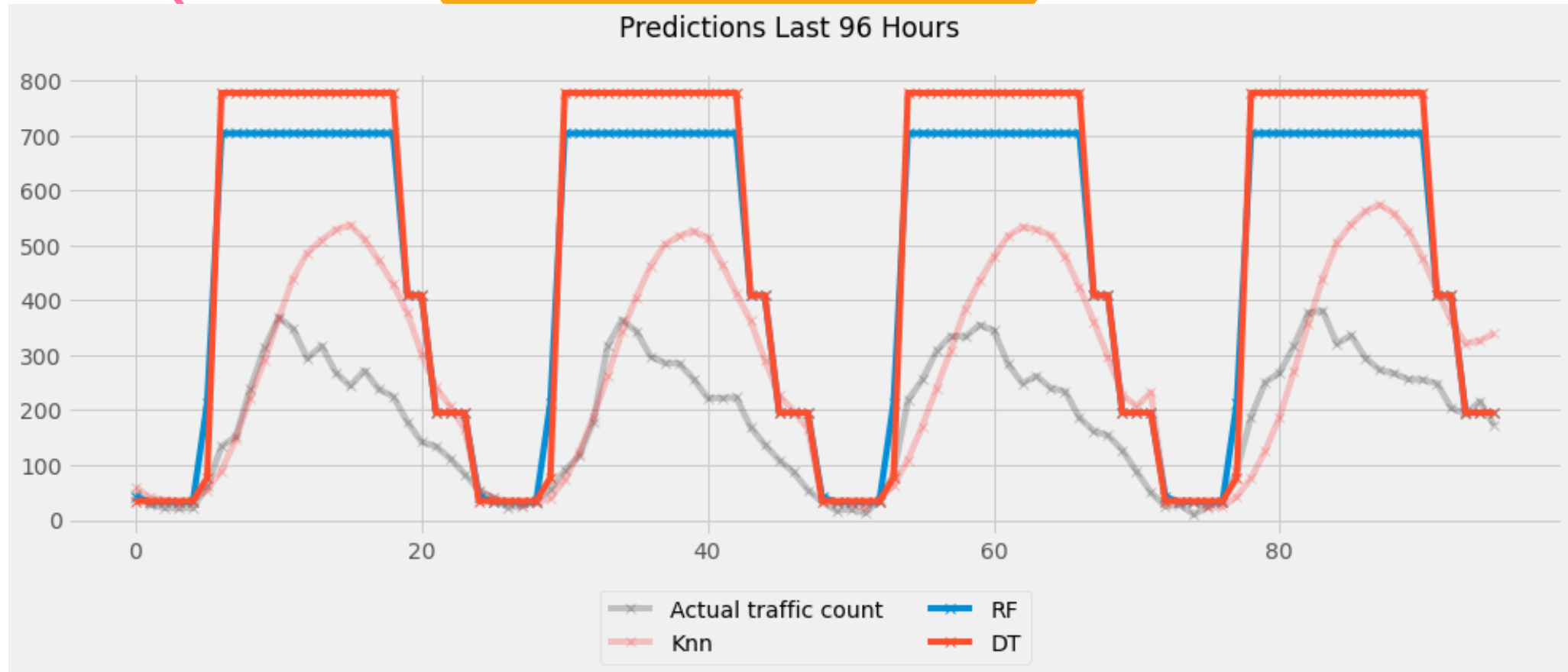
Root Mean Squared Error: 197.8687

GRU

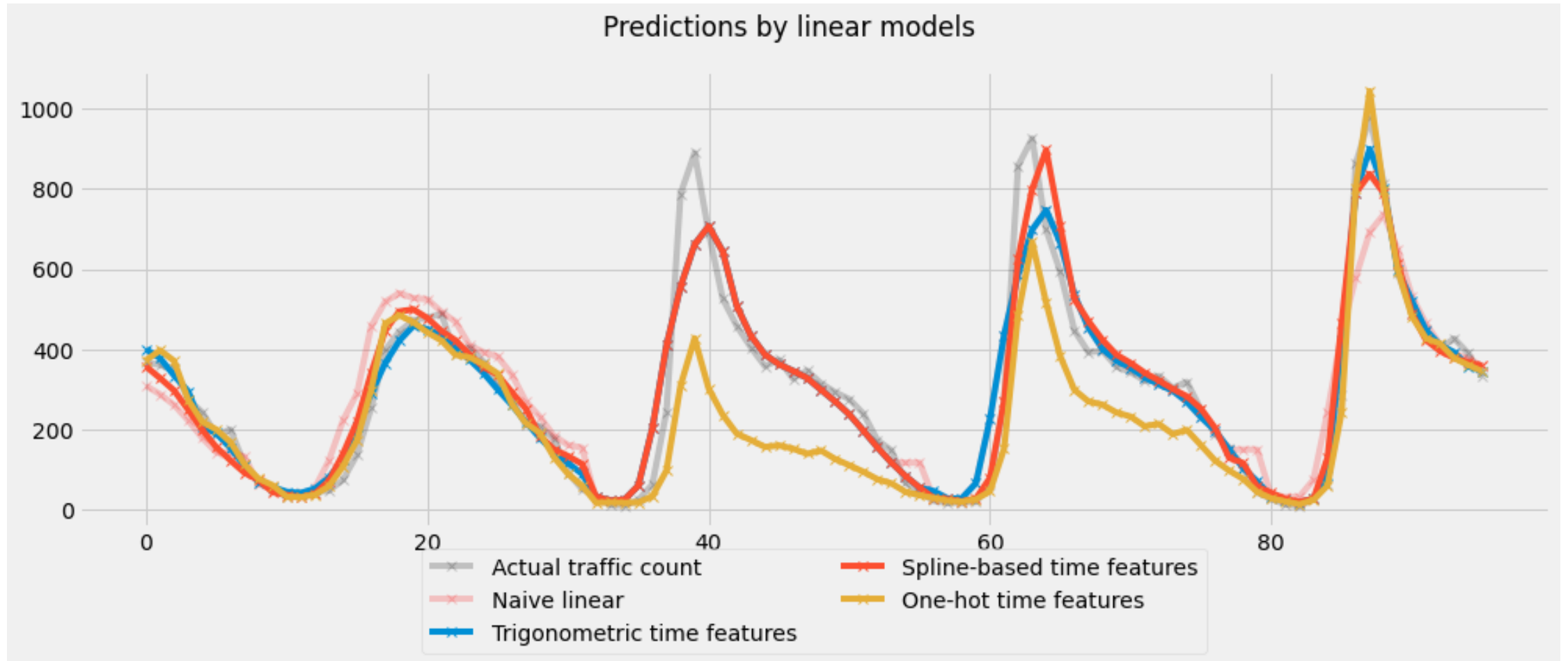
Root Mean Squared Error: 211.1265



Regression Models



Model performance



Performance metrics

Generalised Additive Model:1

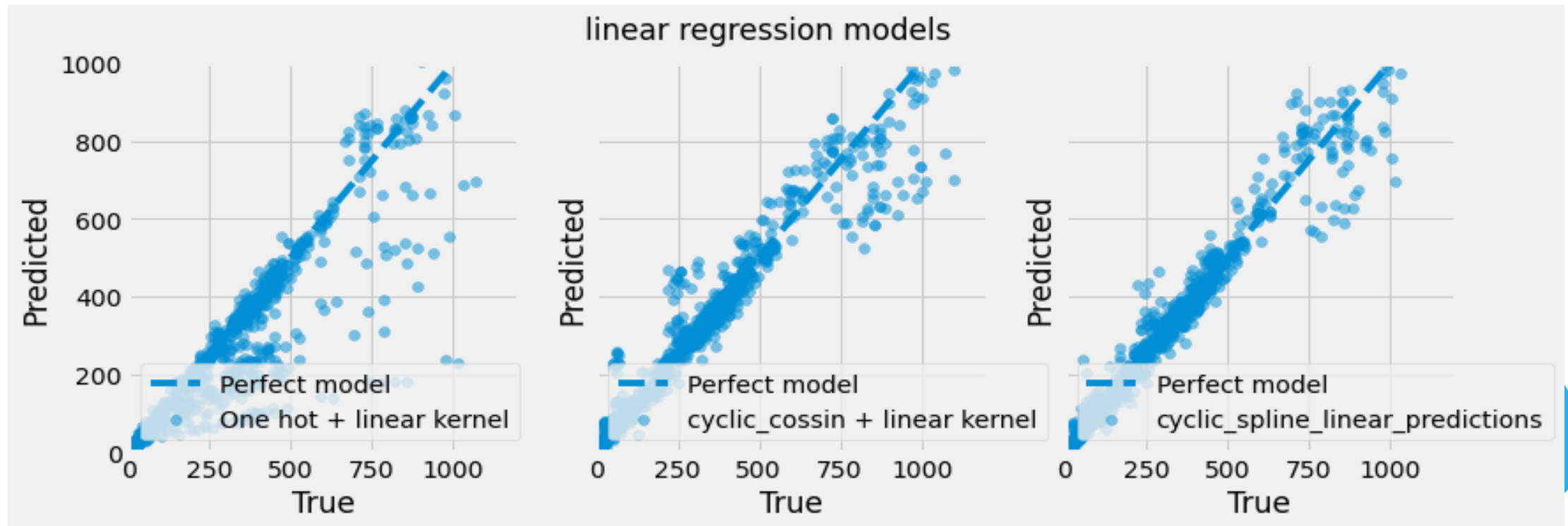
Mean Absolute Error: 57.774 +/- 8.435
Root Mean Squared Error: 111.198 +/- 9.399
One-Hot Linear MAPE: 0.163
One-Hot Linear NRMSE: 0.346
One-Hot Linear R2: 0.807

Generalised Additive Model:2

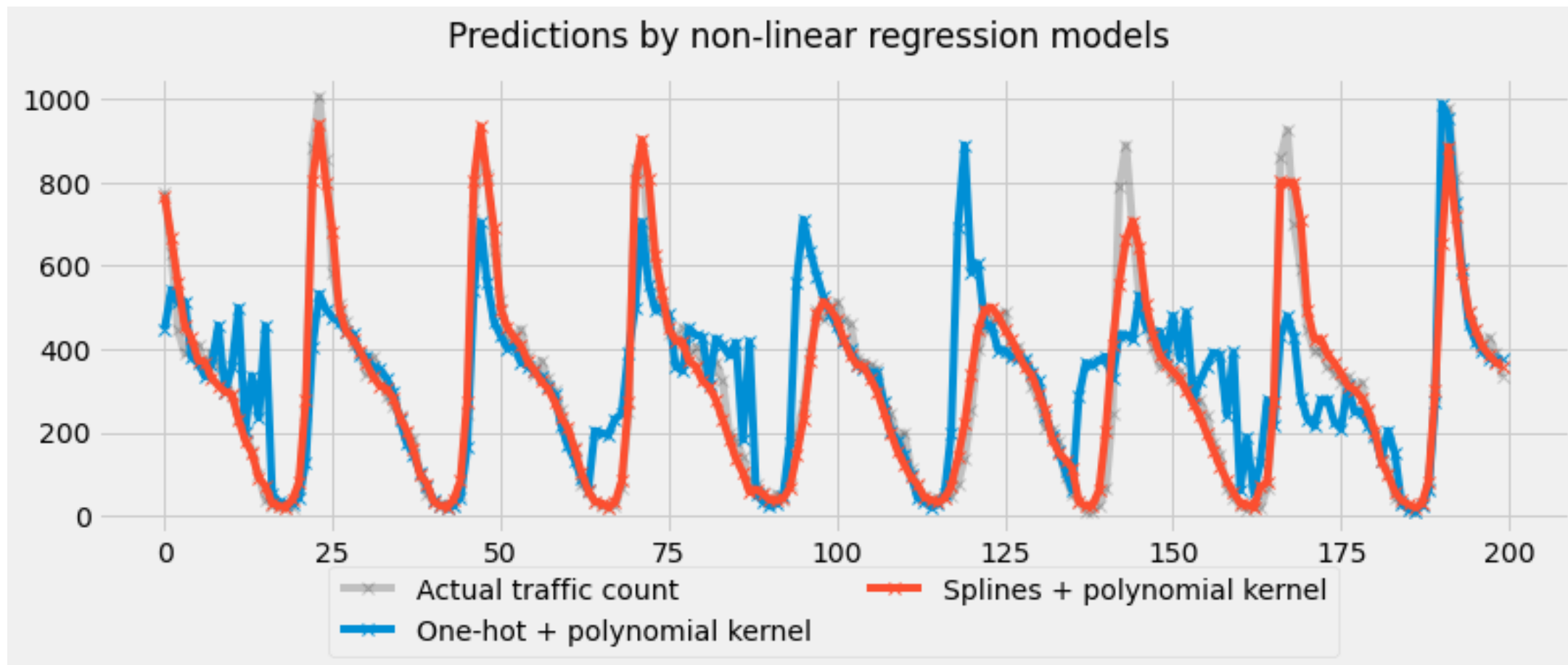
Mean Absolute Error: 41.892 +/- 5.708
Root Mean Squared Error: 77.313 +/- 8.812
Cyclic Cosine Linear MAPE: 0.205
Cyclic Cosine Linear NRMSE: 0.204
Cyclic Cosine Linear R2: 0.933

Generalised Additive Model:3

Mean Absolute Error: 32.437 +/- 5.643
Root Mean Squared Error: 55.618 +/- 10.615
Cyclic Spline Linear MAPE: 0.136
Cyclic Spline Linear NRMSE: 0.157
Cyclic Spline Linear R2: 0.961



Model performance



Model Metrics

Generalised Additive Model:4

Results for One Hot Poly Pipeline:

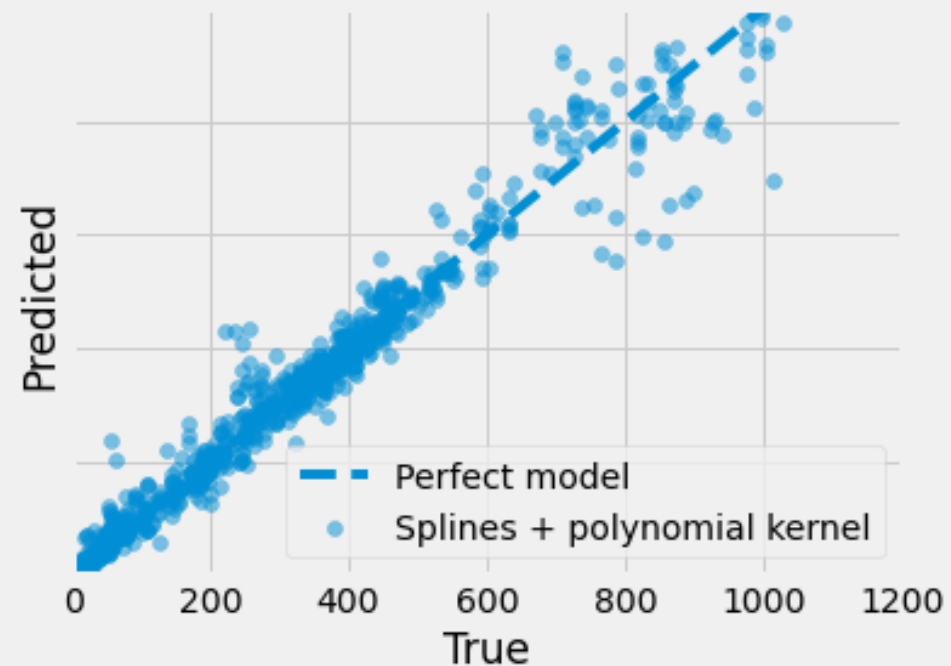
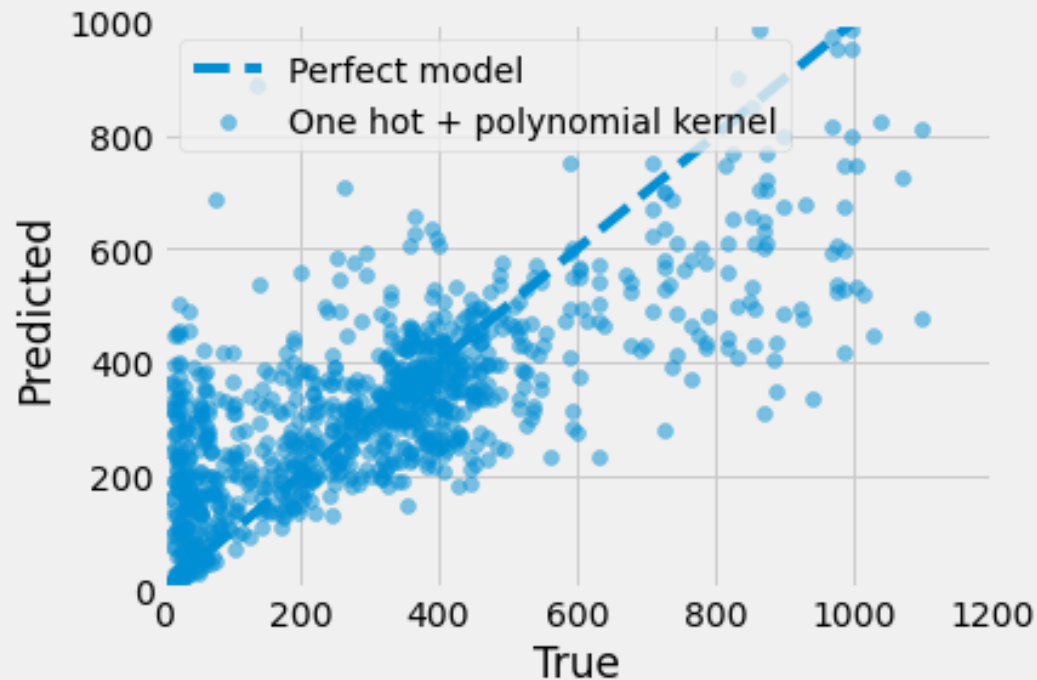
MAE: 113.538
RMSE: 155.718
MAPE: 146.229
NRMSE: 0.517
R2: 0.623

Generalised Additive Model:5

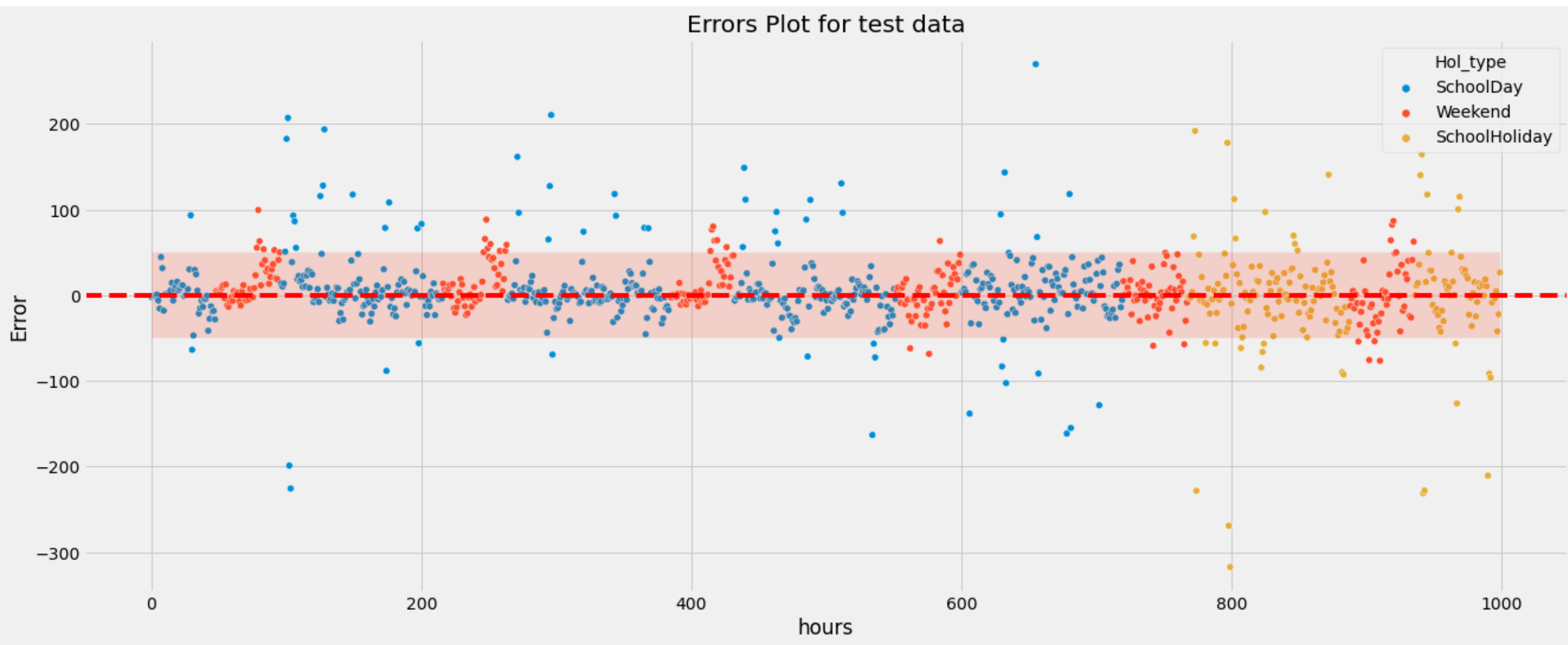
Results for Cyclic Spline Poly Pipeline:

MAE: 31.690
RMSE: 49.760
MAPE: 17.286
NRMSE: 0.165
R2: 0.961

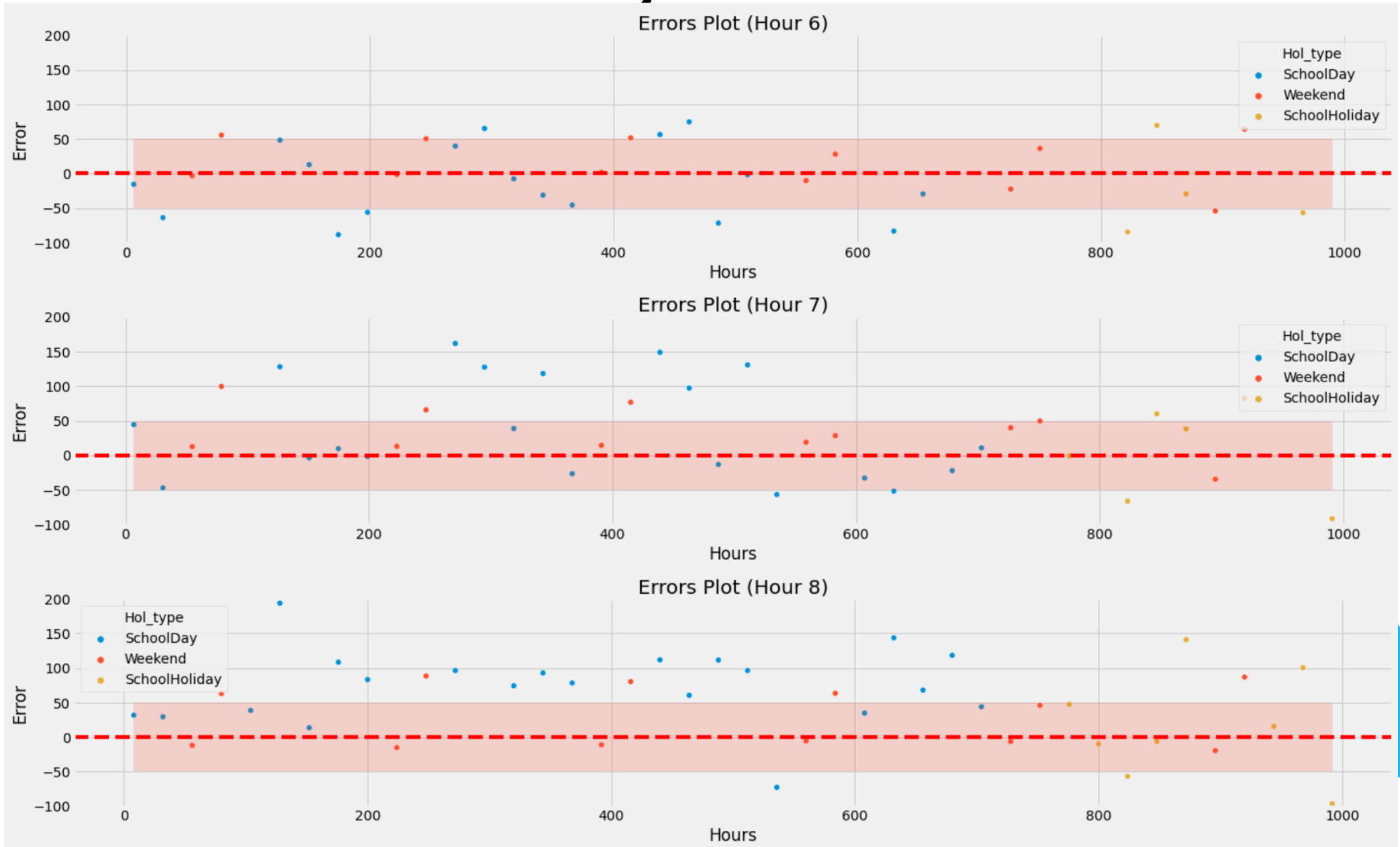
Non-linear regression models



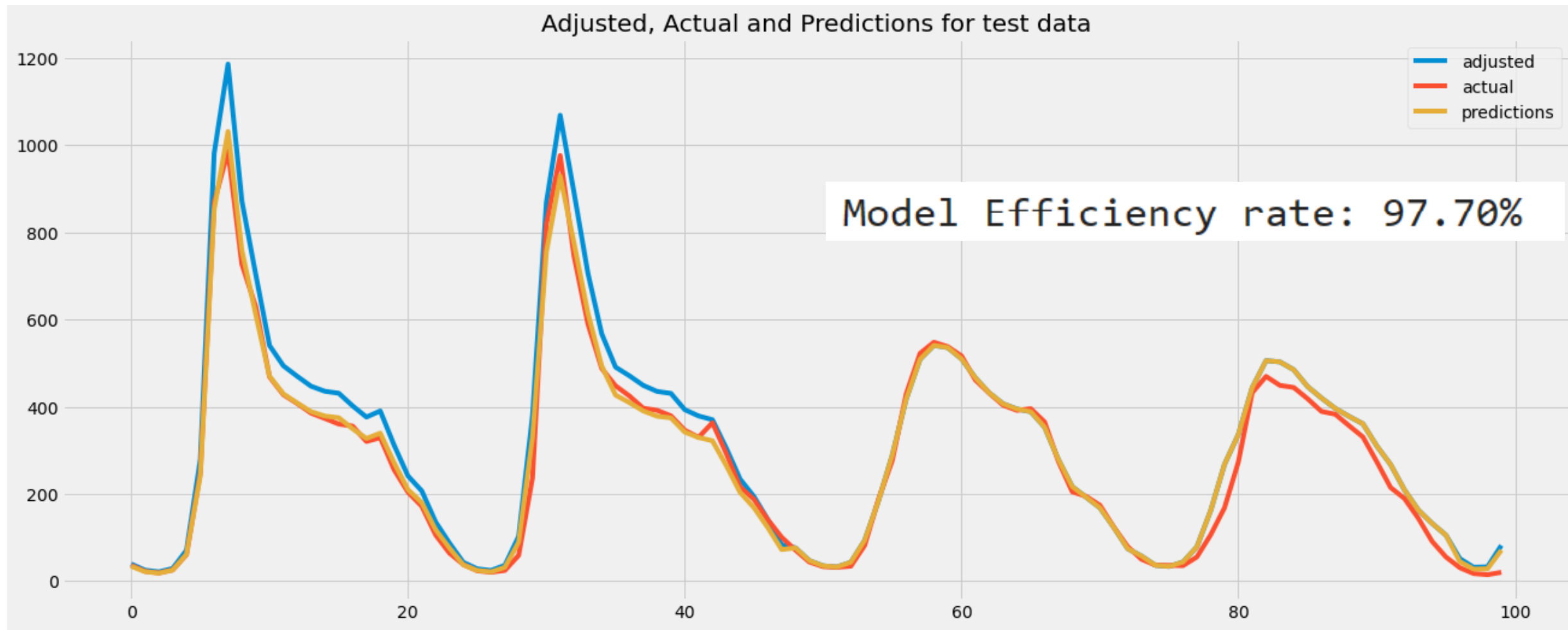
Analysing the errors



Errors by the hour



Adjusted prediction



SUMMARY

Model	Number of Features	R2 Score
Baseline Model - SARIMAX	No Features – Time Series only	0.816
Random Forest Regressor	14 features	0.537
Decision Tree Regressor	14 features	0.475
KNN Regressor	14 features	0.618
Generalised Additive Model - 1	14 features	0.807
Generalised Additive Model - 2	14 features	0.933
Generalised Additive Model - 3	14 features	0.961
Generalised Additive Model - 4	14 features	0.623
Generalised Additive Model - 5	14 features	0.961

Limitations

- Does not take into account incidents
- Working on a local machine
- Weather information was daily not hourly

Future Considerations

- More processing power
- Add more predictor variables



THANK YOU
Questions??