CAN COE PREMIUMS BE PREDICTED?

BACKGROUND

- **Certificate of Entitlement (COE)**
  - Right to own a vehicle for 10 years.
  - Separated into 5 Categories (A - E)
  - Quotas allocated by government.

- **Bidding Exercises**
  - Semi-Monthly (1st and 3rd week of each month)
  - Quota per category to be announced before each bidding exercise.

- **Purchasing Cars**
  - Before buying a car, there are various COEs packages to be bought.

PROBLEM STATEMENT

- As a car dealer in Singapore, I want to accurately predict COE premiums. This will allow the company to maximise profits and set prices that are competitive in the market.

- Successfully bid for COE as soon as possible.

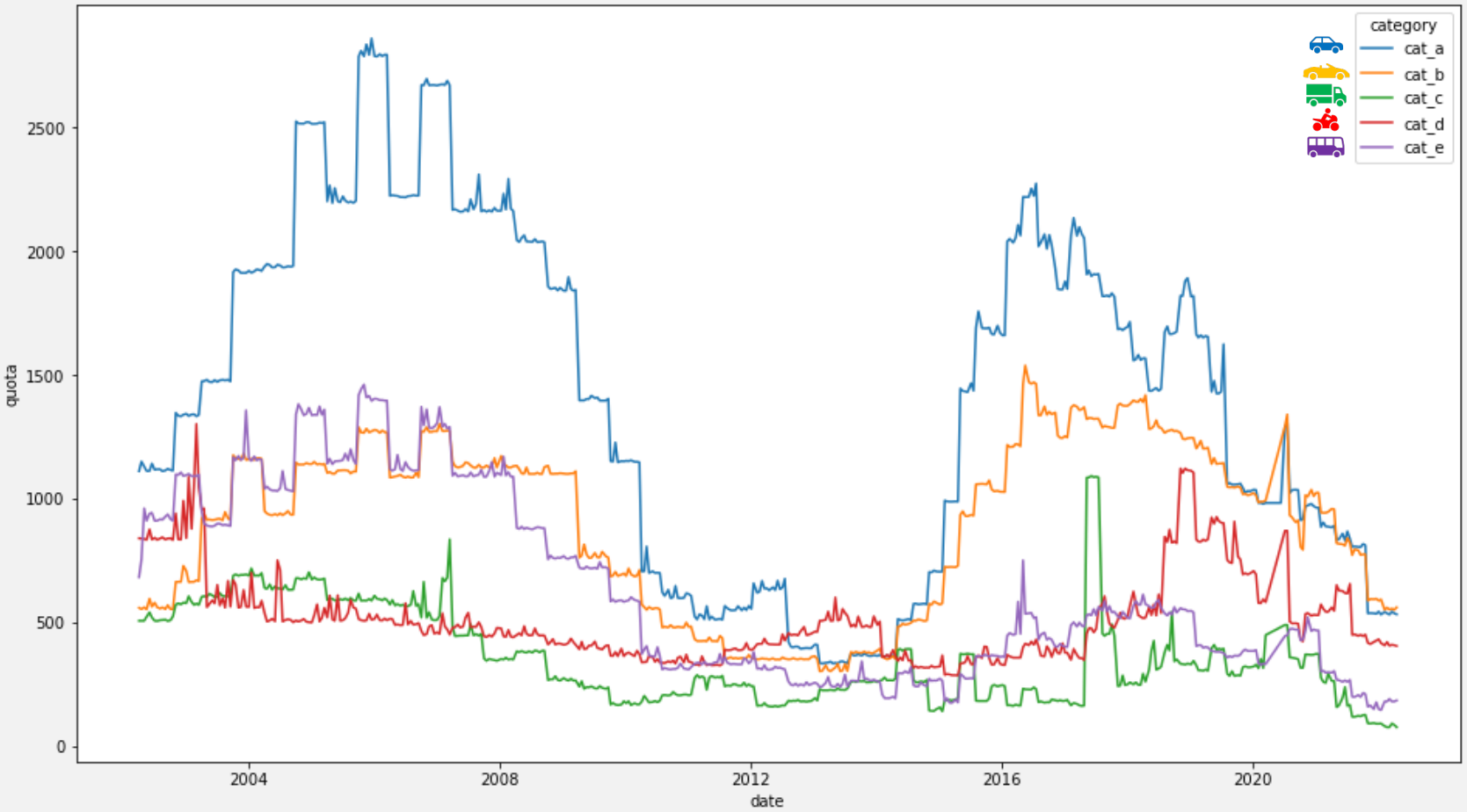
GOAL

- In order to be able to accurately bid for a COE, I will be focusing on reducing MSE of the predictions.

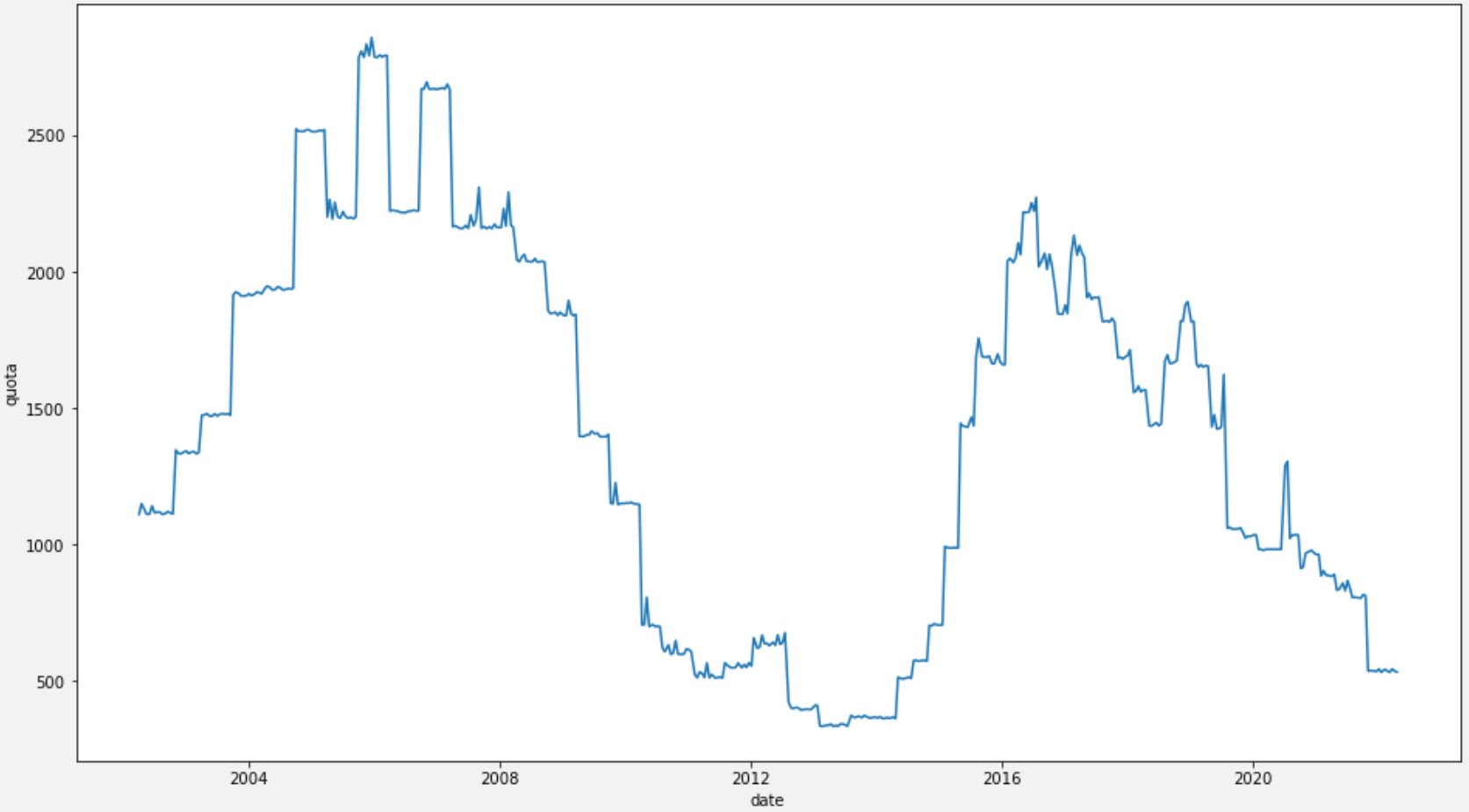
- High success rate in bidding for COE is also important.
  - Success: Predicted price >= premium price

EXPLORATORY DATA ANALYSIS

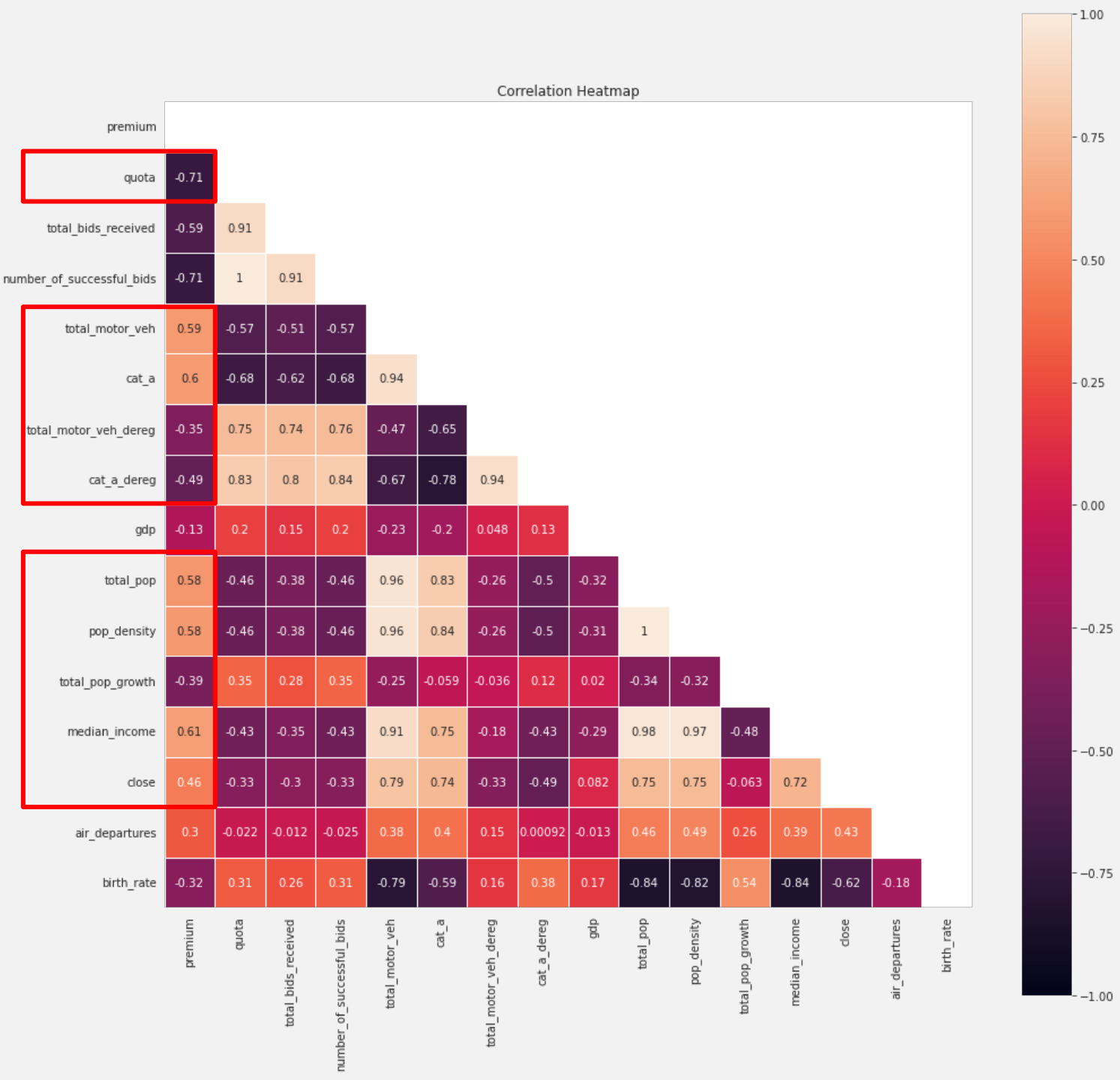
COE – QUOTAS BY CATEGORY



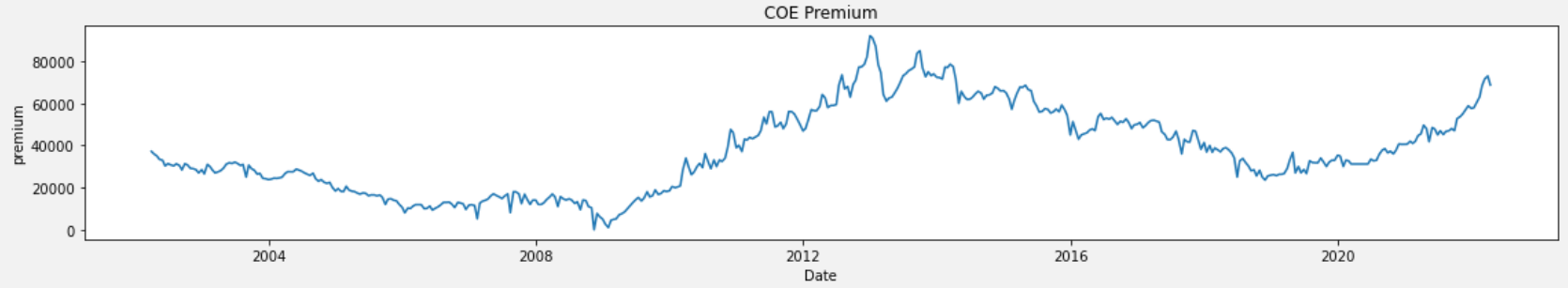
FOCUS – CATEGORY A PREMIUMS

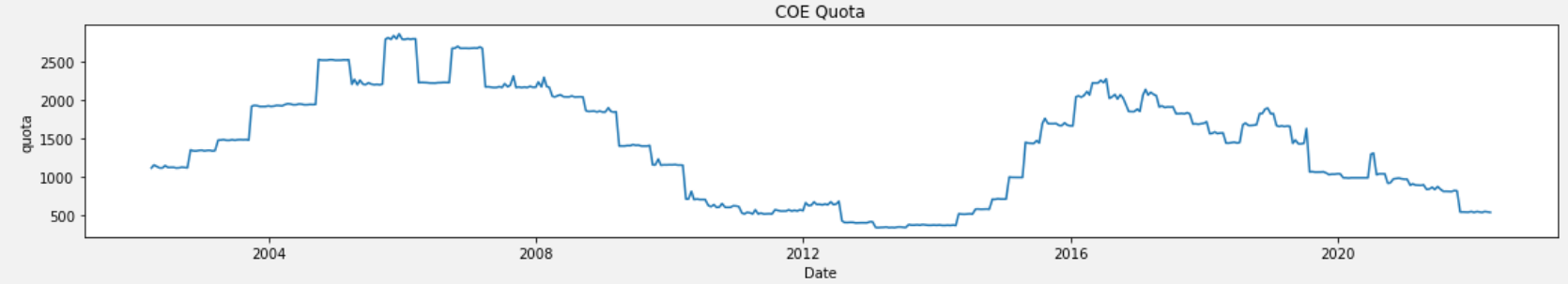


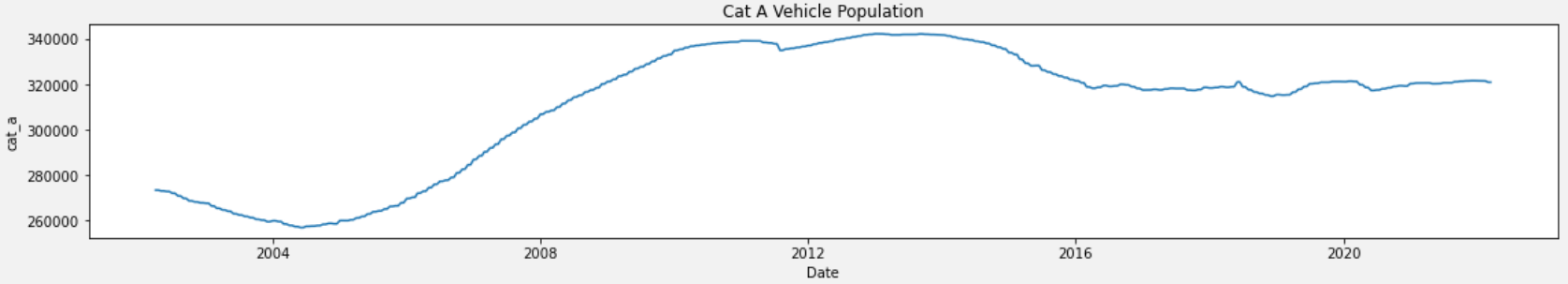
EDA - HEATMAP

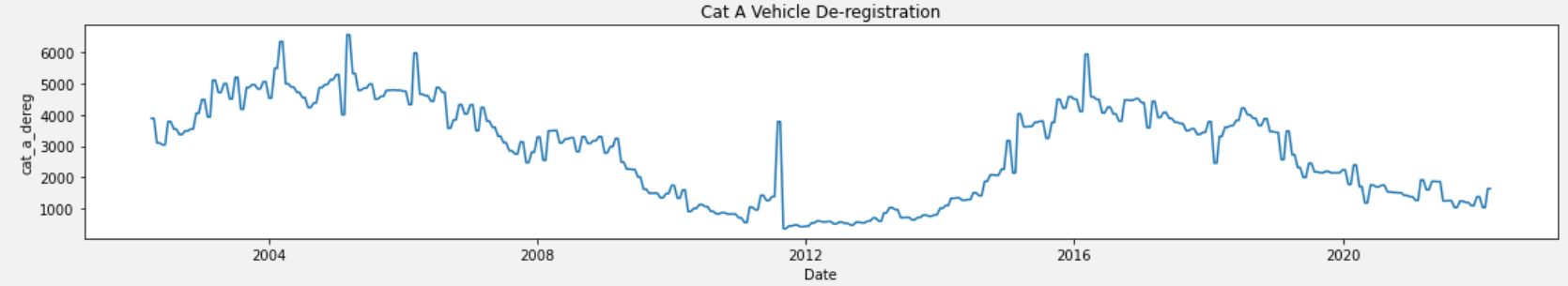

Correlation Heatmap
premium
quota -0.71
total_bids_received -0.59 0.91
number_of_successful_bids -0.71 1 0.91
total_motor_veh 0.59 -0.57 -0.51 -0.57
cat_a 0.6 -0.68 -0.62 -0.68 0.94
total_motor_veh_dereg -0.35 0.75 0.74 0.76 -0.47 -0.65
cat_a_dereg -0.49 0.83 0.8 0.84 -0.67 -0.78 0.94
gdp -0.13 0.2 0.15 0.2 -0.23 -0.2 0.048 0.13
total_pop 0.58 -0.46 -0.38 -0.46 0.96 0.83 -0.26 -0.5 -0.32
pop_density 0.58 -0.46 -0.38 -0.46 0.96 0.84 -0.26 -0.5 -0.31 1
total_pop_growth -0.39 0.35 0.28 0.35 -0.25 -0.059 -0.036 0.12 0.02 -0.34 -0.32
median_income 0.61 -0.43 -0.35 -0.43 0.91 0.75 -0.18 -0.43 -0.29 0.98 0.97 -0.48
close 0.46 -0.33 -0.3 -0.33 0.79 0.74 -0.33 -0.49 0.082 0.75 0.75 -0.063 0.72
air_departures 0.3 -0.022 -0.012 -0.025 0.38 0.4 0.15 0.00092 -0.013 0.46 0.49 0.26 0.39 0.43
birth_rate -0.32 0.31 0.26 0.31 -0.79 -0.59 0.16 0.38 0.17 -0.84 -0.82 0.54 -0.84 -0.62 -0.18
premium quota total_bids_received number_of_successful_bids total_motor_veh cat_a total_motor_veh_dereg cat_a_dereg gdp total_pop pop_density total_pop_growth median_income close air_departures birth_rate
1.00 0.75 0.50 0.25 0.00 −0.25 −0.50 −0.75 −1.00

EDA – CAT A VEHICLE ATTRIBUTES

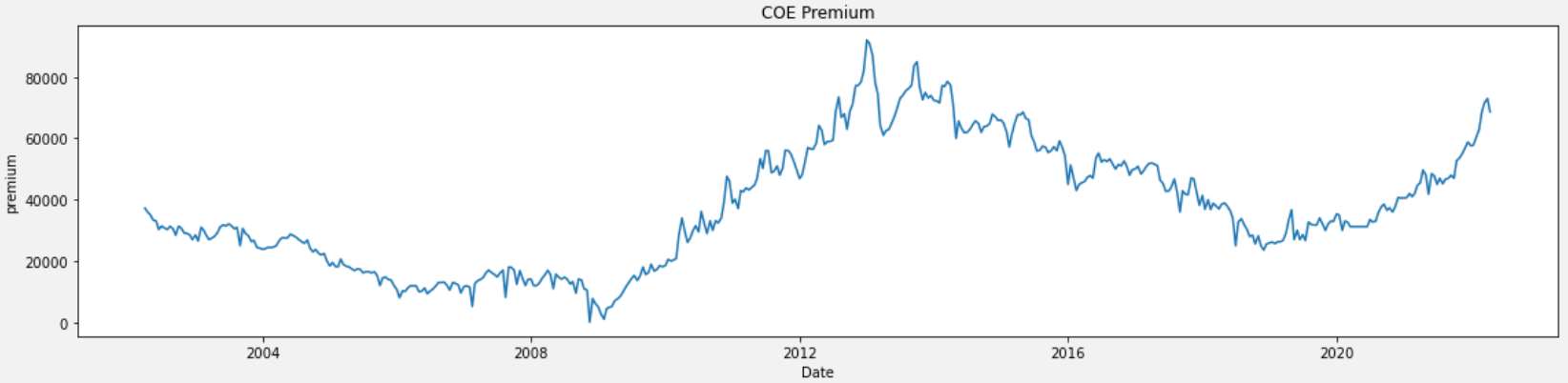


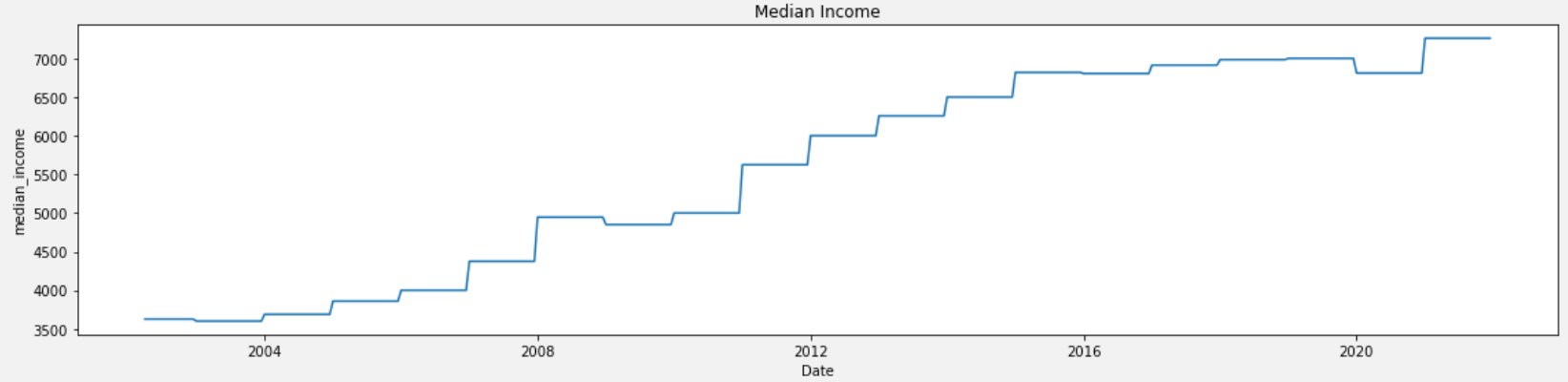


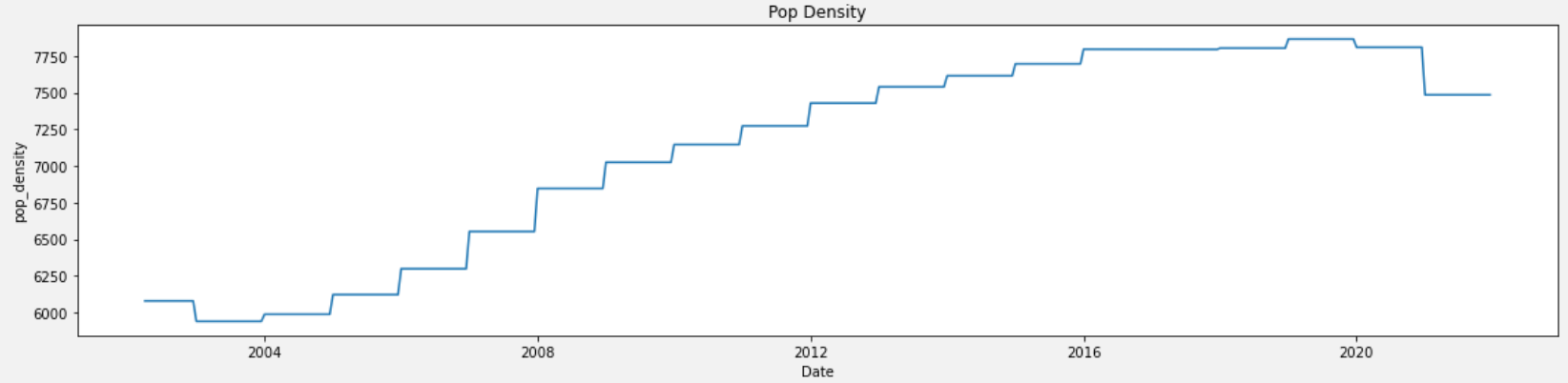




EDA – OTHER ECONOMIC VARIABLES







KEY FINDINGS FROM EDA

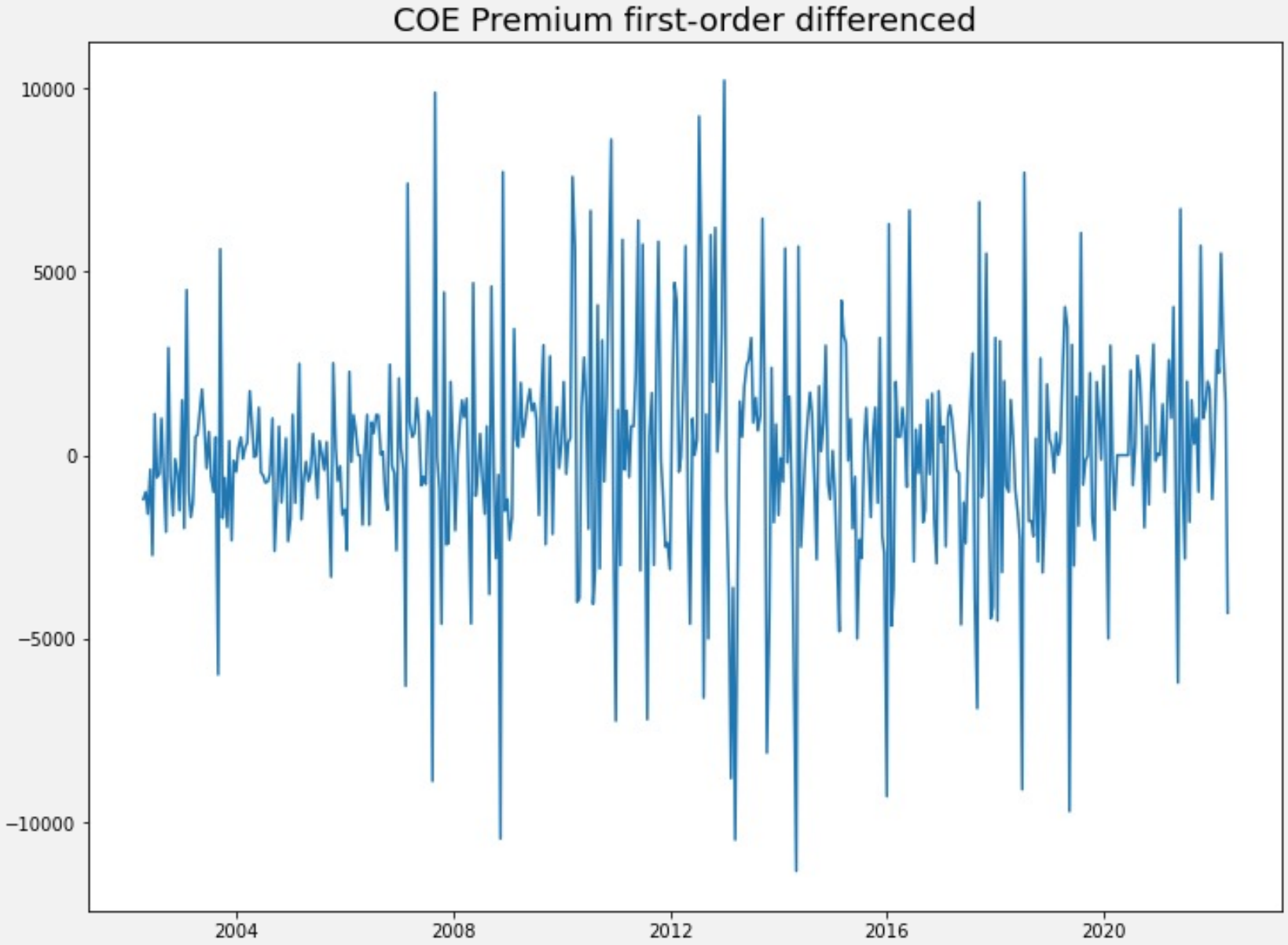
- COE premium has a general upward trend.
- COE's 10 years seasonality is not apparent.
- Overall population of vehicle remains the same but median income is increasing.

- Variables have a strong correlation to premium prices but poor estimator.
  - Resolution of variable information is too low vs premium price.

FEATURE ENGINEERING

- Scarcity – simulate behavioral demand
  - Quota / total bids received
- Lags for other variables
  - Yearly, Quarterly, Monthly data used to predict for the next cycle

CHECK FOR STATIONARITY



MODELING

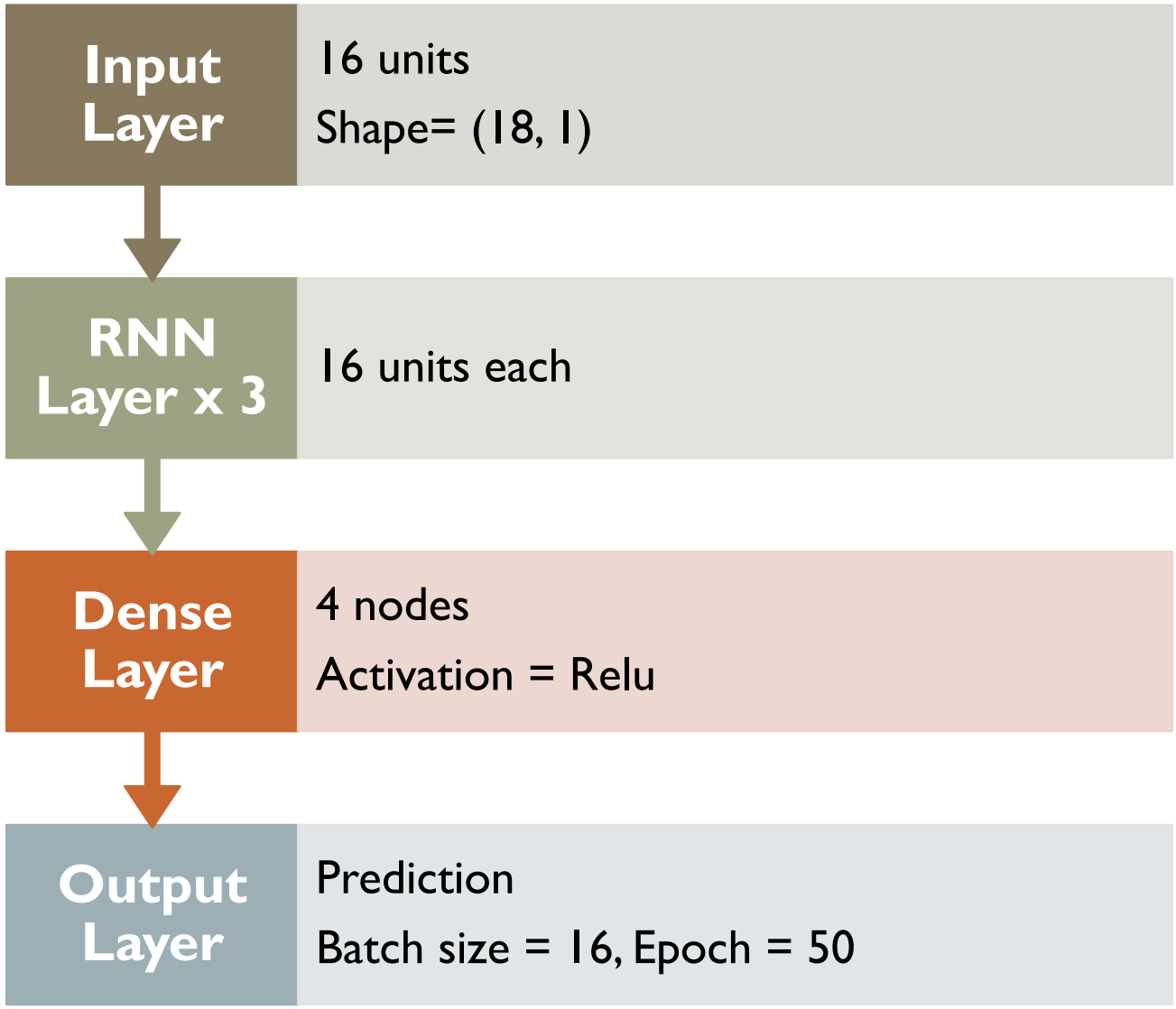
Model	Univariate / Multivariate	MSE	Success* %
ARIMA	Univariate	170,766,416.86	43.3
Linear Regression	Univariate	86,105,841.98	83.5
Linear Regression	Multivariate	627,776,111.23	100.0
Linear Regression w Lasso	Multivariate	629,132,479.33	100.0
Linear Regression w Ridge	Multivariate	629,122,991.38	100.0
SARIMA	Univariate	151,149,668.92	50.5

Variables used:
Linear Regression Model: Quota

Linear Regression Model 2: Quota, Deregistration of Cat A vehicle, Population Growth, STI close price

* Success: Predicted price >= premium price

STRUCTURE OF RECURRENT NEURAL NETWORK MODEL



MODELING – RECURRENT NEURAL NETWORK

Model	Univariate / Multivariate	MSE	Success %
RNN - LSTM	Univariate	15,133,745.46	40.5
RNN - GRU	Univariate	7,453,959.03	75.9
RNN - GRU 1	Multivariate	8,823,055.69	20.3
RNN - GRU 2	Multivariate	19,145,021.12	5.1
RNN - GRU 3	Multivariate	10,139,610.50	19.0

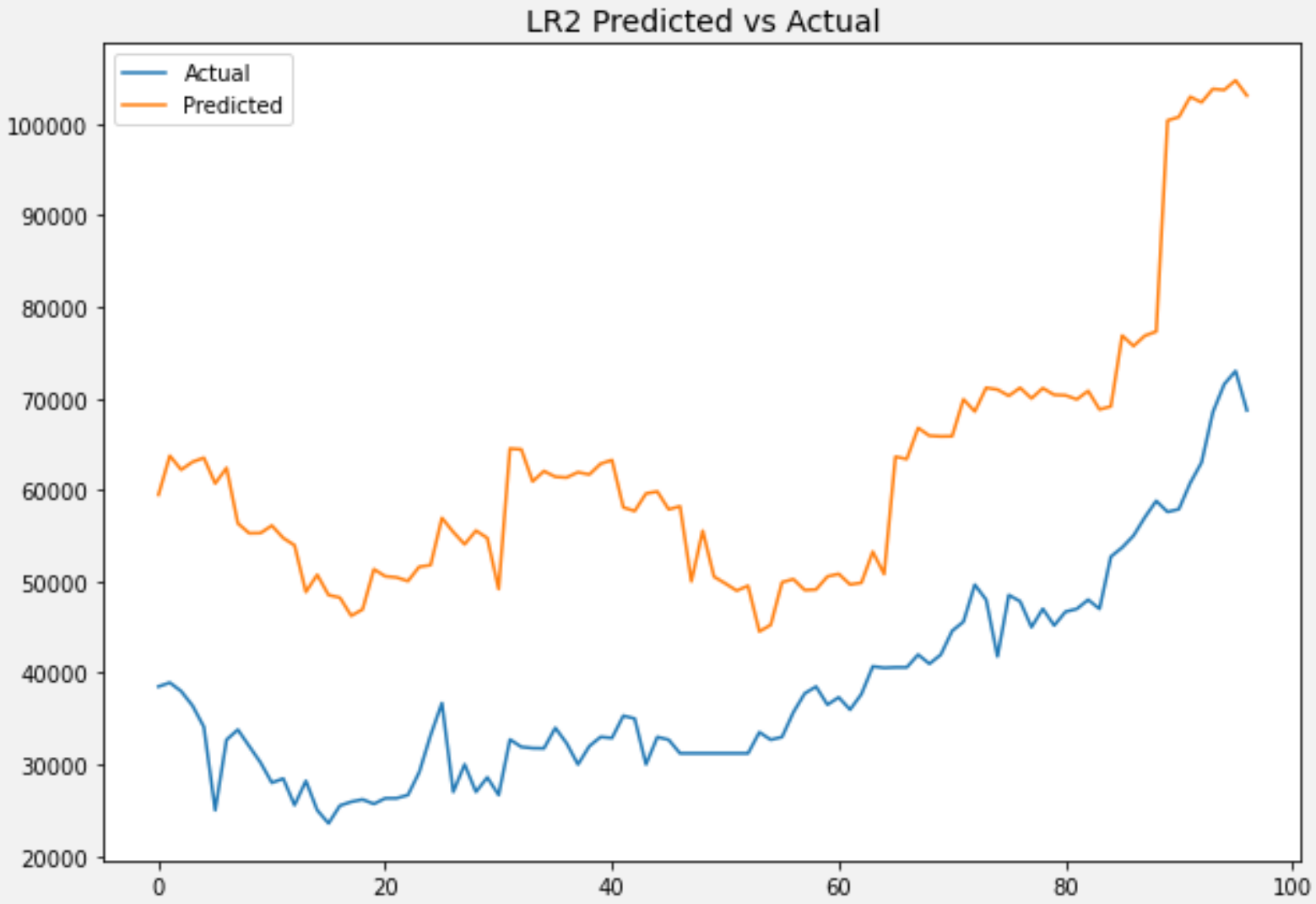
Variables for Multivariate models:

Model 1: Premium, Scarcity

Model 2: Premium, Quota, Deregistration of Cat A vehicle, Population Growth, STI close price

Model 3: Premium, Scarcity, Deregistration of Cat A vehicle, STI close price

ERROR ANALYSIS – SUCCESS RATE

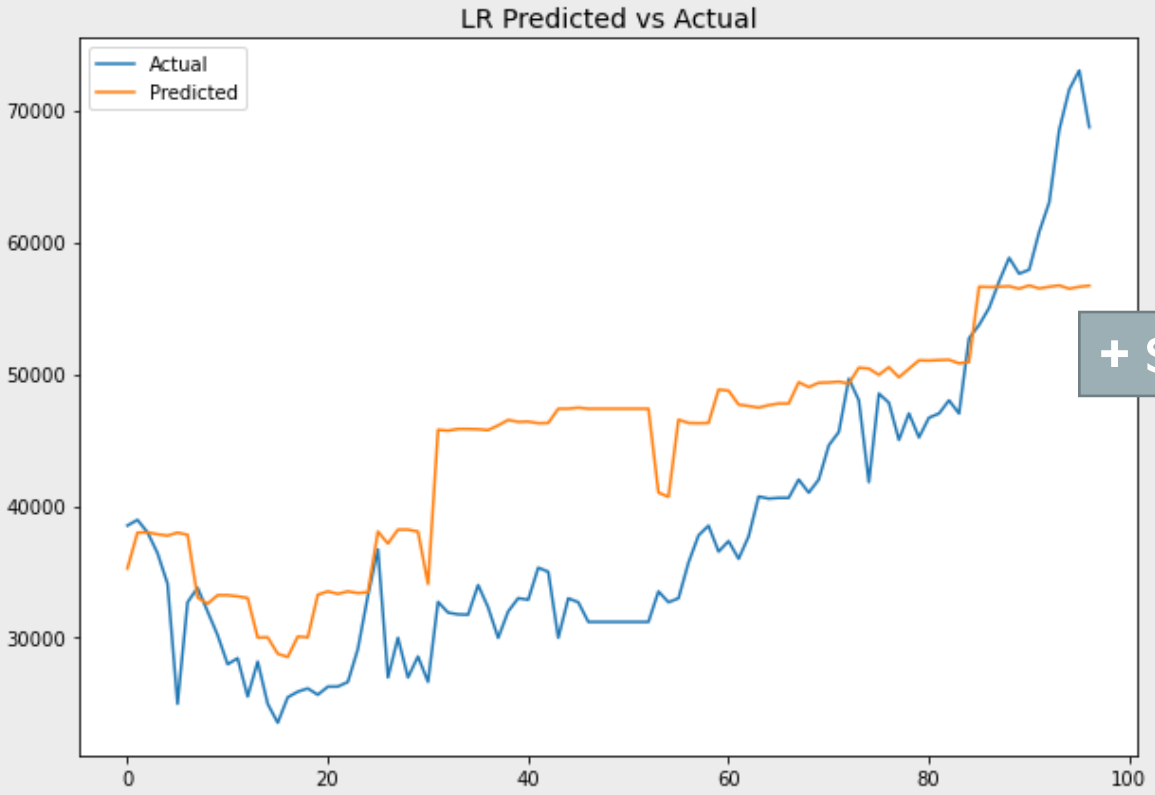


MSE	Success %
627,776,111.23	100

ERROR ANALYSIS – SUCCESS RATE

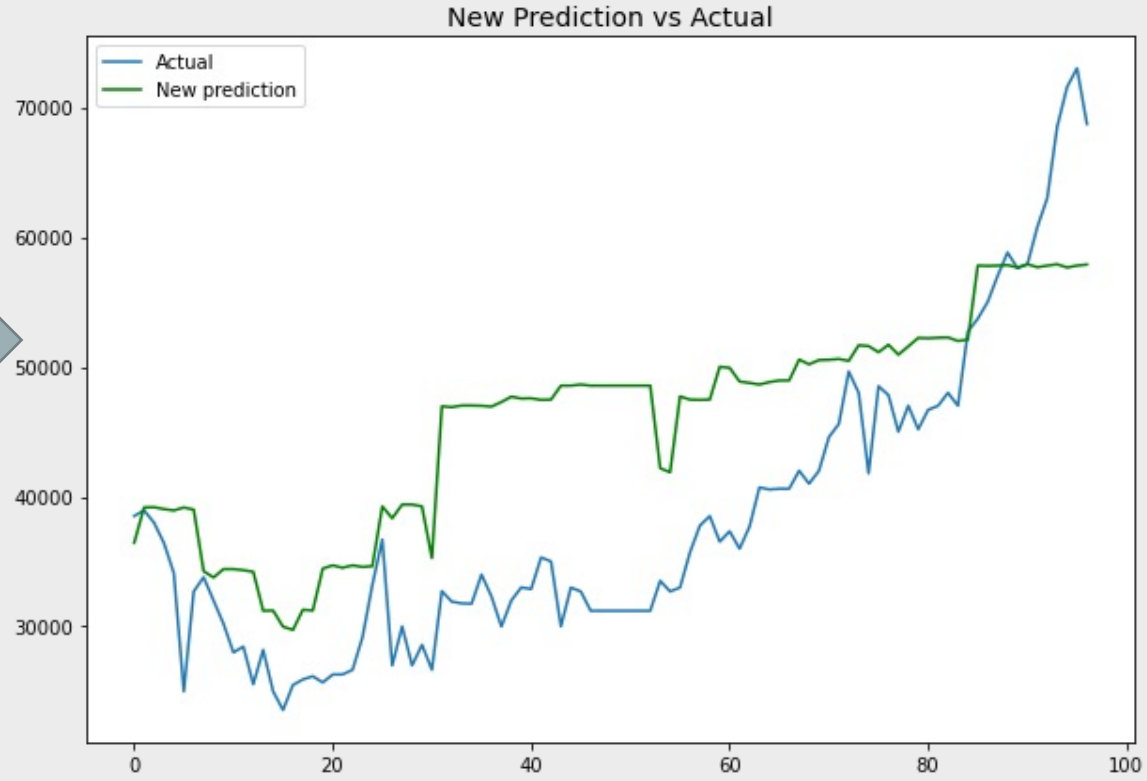
- Not accurate to compare across different success rates.
- Using a function, I added a different increment each model so that all models have an 90% success rate.

ERROR ANALYSIS – SUCCESS RATE



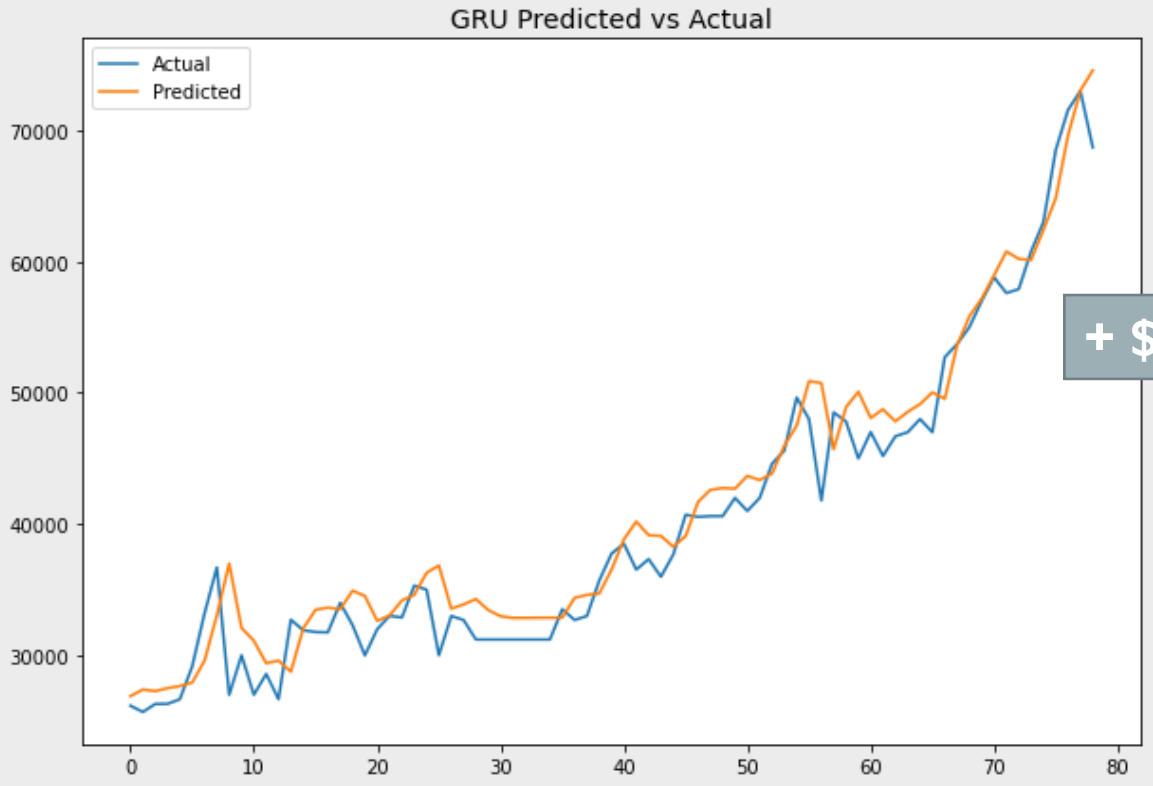
83.5%

+ $1190



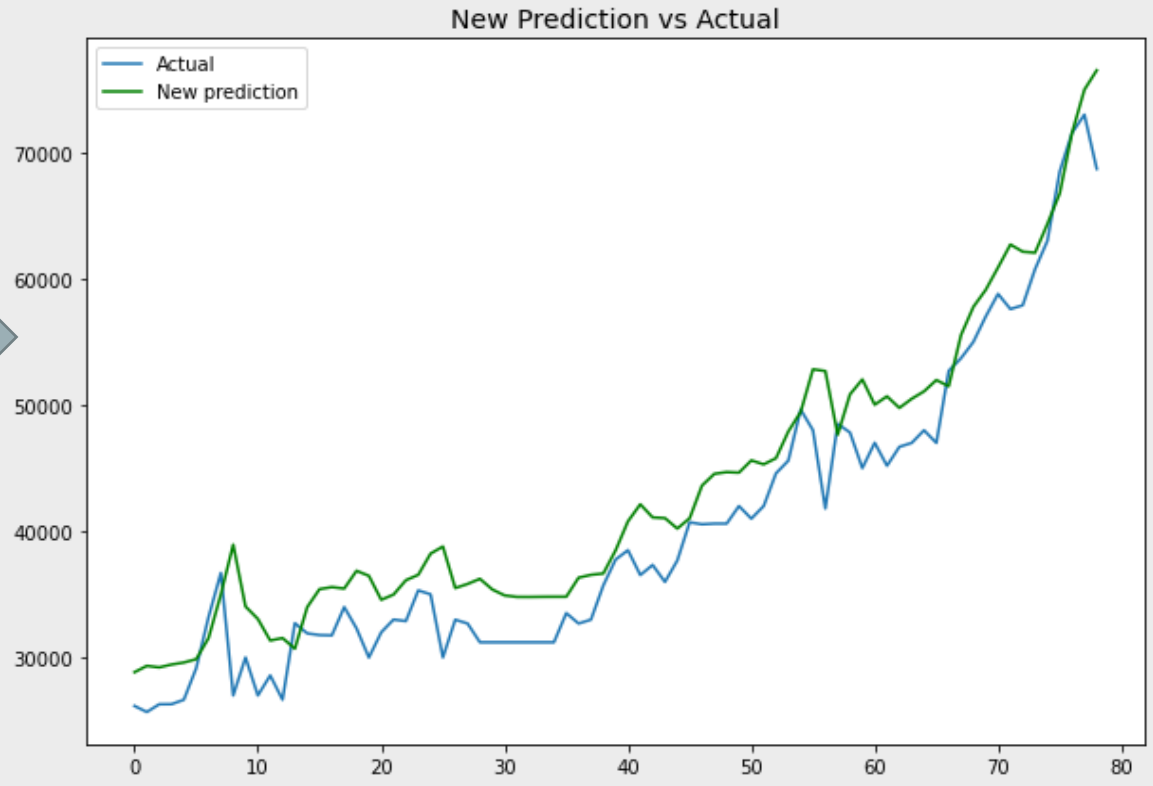
90.7%

ERROR ANALYSIS – SUCCESS RATE



75.9%

+ $1943



91.1%

ERROR ANALYSIS – SUCCESS RATE

Model	Average difference (Predicted vs Actual)	Standard Deviation
Linear Regression	8830.4	5205.9
Linear Regression 2	11499.5	5687.4
GRU - Univariate	3558.0	2133.1
GRU - Multivariate 1	3849.1	2134.4
GRU - Multivariate 2	4738.3	2216.6
GRU - Multivariate 3	3892.5	2142.2

MODEL SELECTION

- Lowest MSE
- High Success Rate with:
  - Low average difference to actual
  - Low standard deviation
- Eventual model selected:
  - Univariate GRU model

PREDICTION FOR MAY'S BIDDING

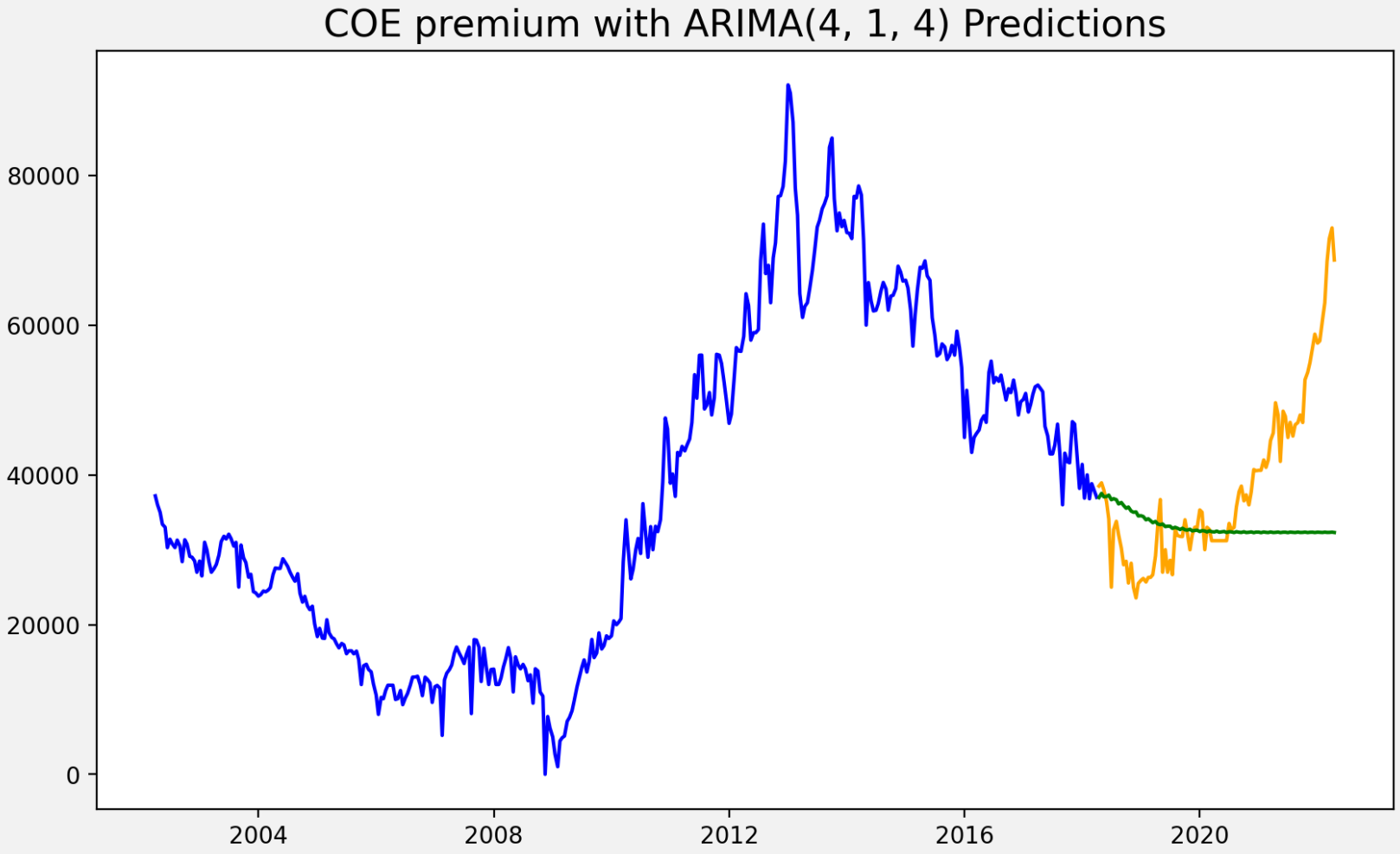
- Using the model, the predicted premium: $69,825.63
- To achieve 90% success rate: $1943 is added to the prediction
- Final prediction = $71,768.63
- Actual Premium = $70,901

FUTURE AREAS OF WORK

- Due to time constraints, other categories naively uses the same model for CAT A.
- COE's 10 years seasonality might have an effect but there are too little data points to show.
- Prediction for more than just the next cycle
  - Able to price COE packages according to predictions.

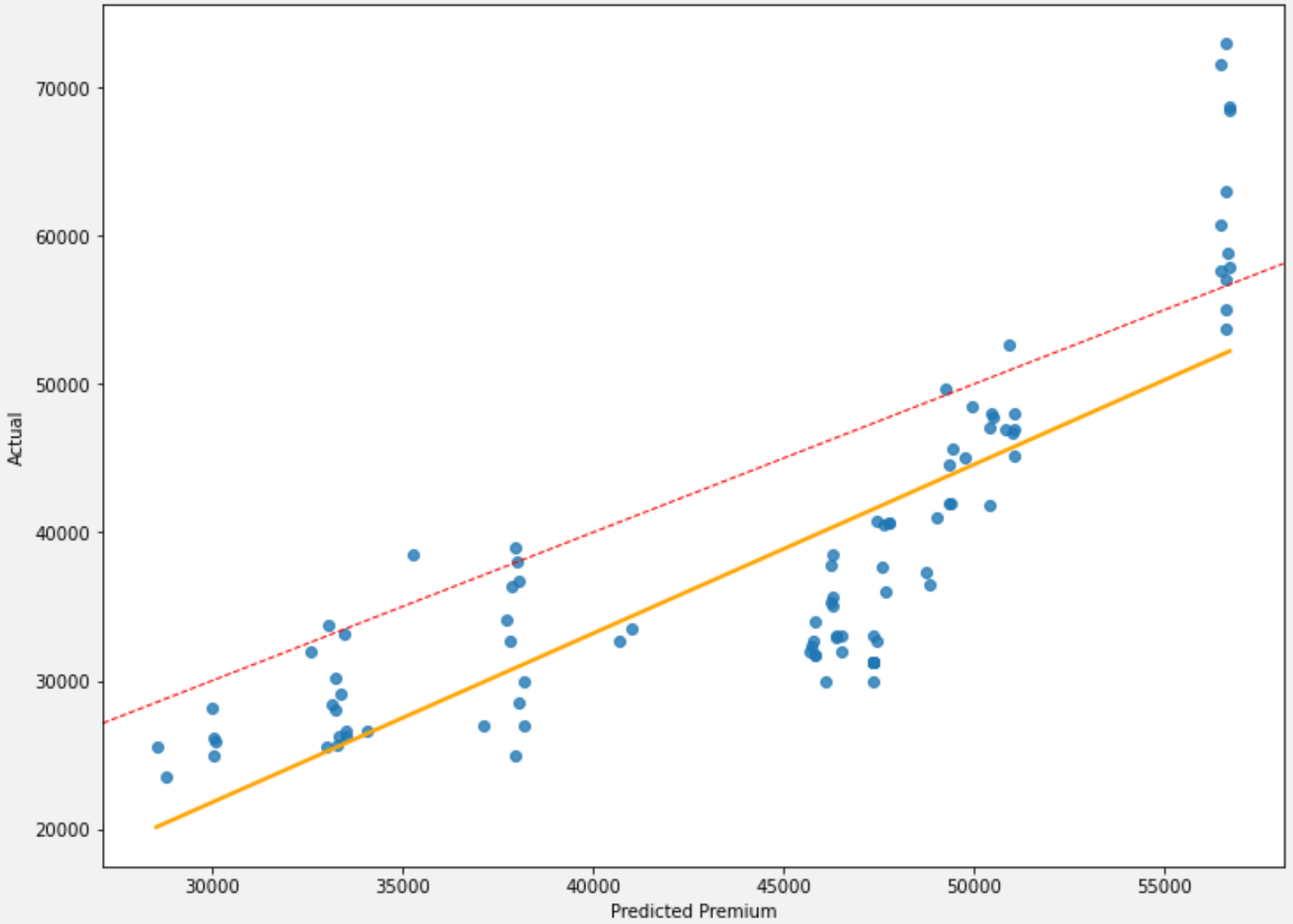
THANK YOU

ARIMA



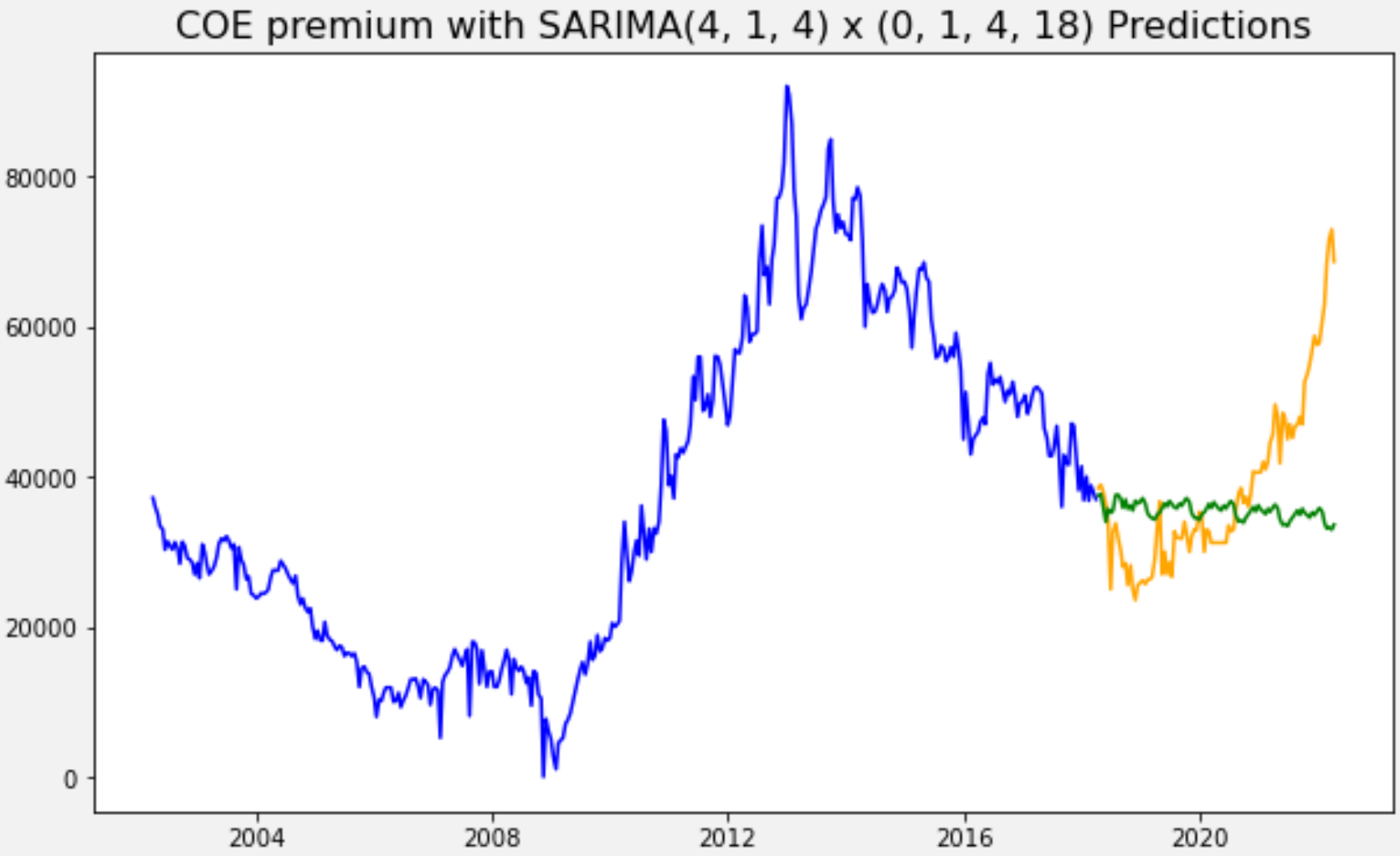
MSE	Success %
170,766,416.86	43.3

LINEAR REGRESSION



MSE	Success %
86,105,841.98	83.5

SARIMA



MSE	Success %
151,149,668.92	50.5

UNIVARIATE GRU

MSE	Success %
7,453,959.03	75.9

MULTIVARIATE GRU

MSE	Success %
8,823,055.69	20.3