CAR PRICE PREDICTION

MACHINE LEARNING

LINEAR REGRESSION

STEPS

01 PROBLEM ASSESMENT

It involves transforming business problem into suitable solvable form through ML techniques.

02 DATA PRE-PROCESSING

It involves collecting, curating preprocessing cleaning, removing missing data, outliers.

03 FEATURE SELECTION

It involves applying statistical principals to preprocessed data and selecting most influencing attributes to predict without degrading accuracy

04 MODEL TRAINING

We train and make model with selected features.

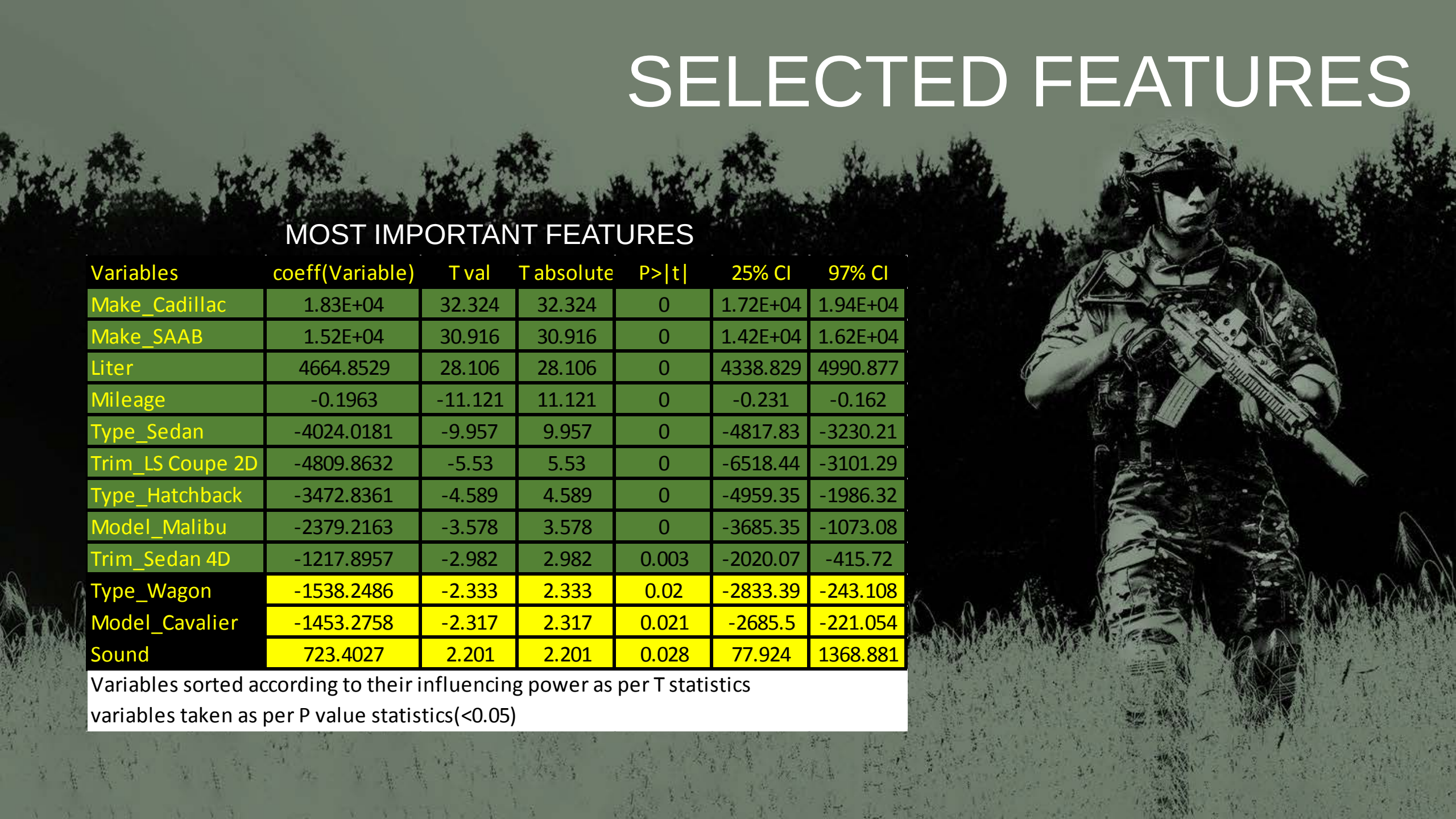
05 MODEL TESTING

It involves Validating trained model with test data, checking model accuracy .

06 VISUALISATION

It involves visualizing training error testing error trends.

SELECTED FEATURES



MOST IMPORTANT FEATURES

Variables	coeff(Variable)	T val	T absolute	P>|t|	25% CI	97% CI
Make_Cadillac	1.83E+04	32.324	32.324	0	1.72E+04	1.94E+04
Make_SAAB	1.52E+04	30.916	30.916	0	1.42E+04	1.62E+04
Liter	4664.8529	28.106	28.106	0	4338.829	4990.877
Mileage	-0.1963	-11.121	11.121	0	-0.231	-0.162
Type_Sedan	-4024.0181	-9.957	9.957	0	-4817.83	-3230.21
Trim_LS Coupe 2D	-4809.8632	-5.53	5.53	0	-6518.44	-3101.29
Type_Hatchback	-3472.8361	-4.589	4.589	0	-4959.35	-1986.32
Model_Malibu	-2379.2163	-3.578	3.578	0	-3685.35	-1073.08
Trim_Sedan 4D	-1217.8957	-2.982	2.982	0.003	-2020.07	-415.72
Type_Wagon	-1538.2486	-2.333	2.333	0.02	-2833.39	-243.108
Model_Cavalier	-1453.2758	-2.317	2.317	0.021	-2685.5	-221.054
Sound	723.4027	2.201	2.201	0.028	77.924	1368.881

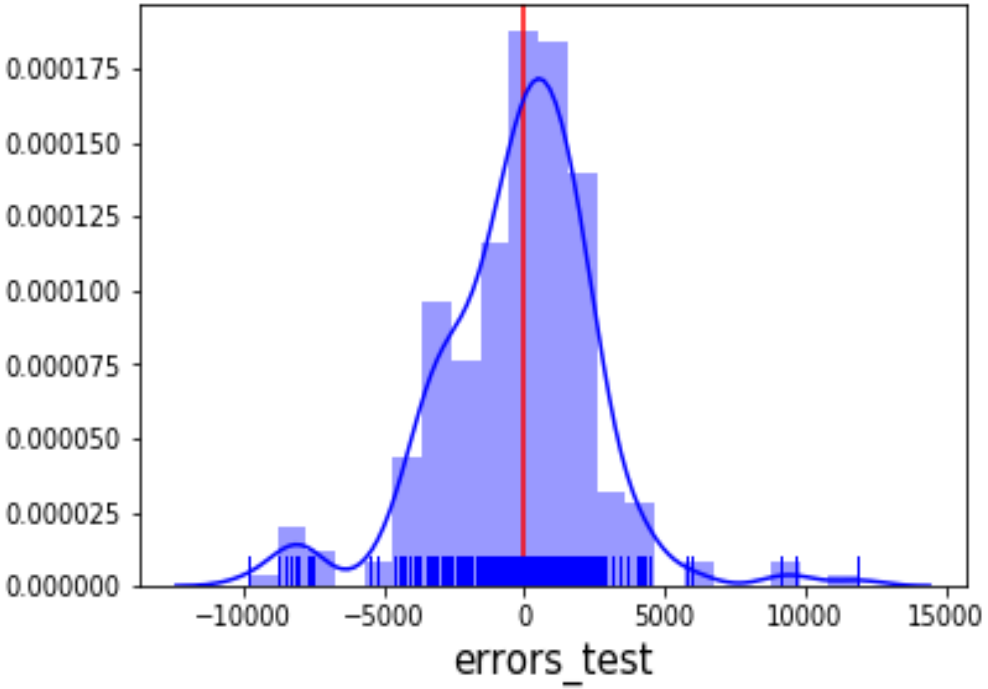
Variables sorted according to their influencing power as per T statistics

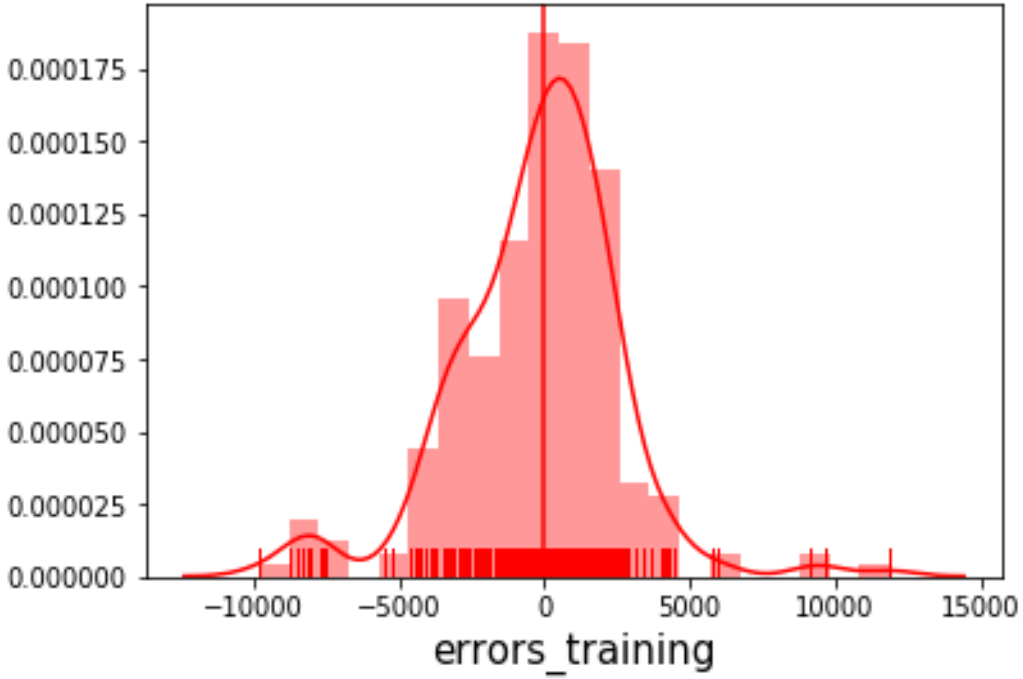
variables taken as per P value statistics(<0.05)

STATISTICAL COMPARISON

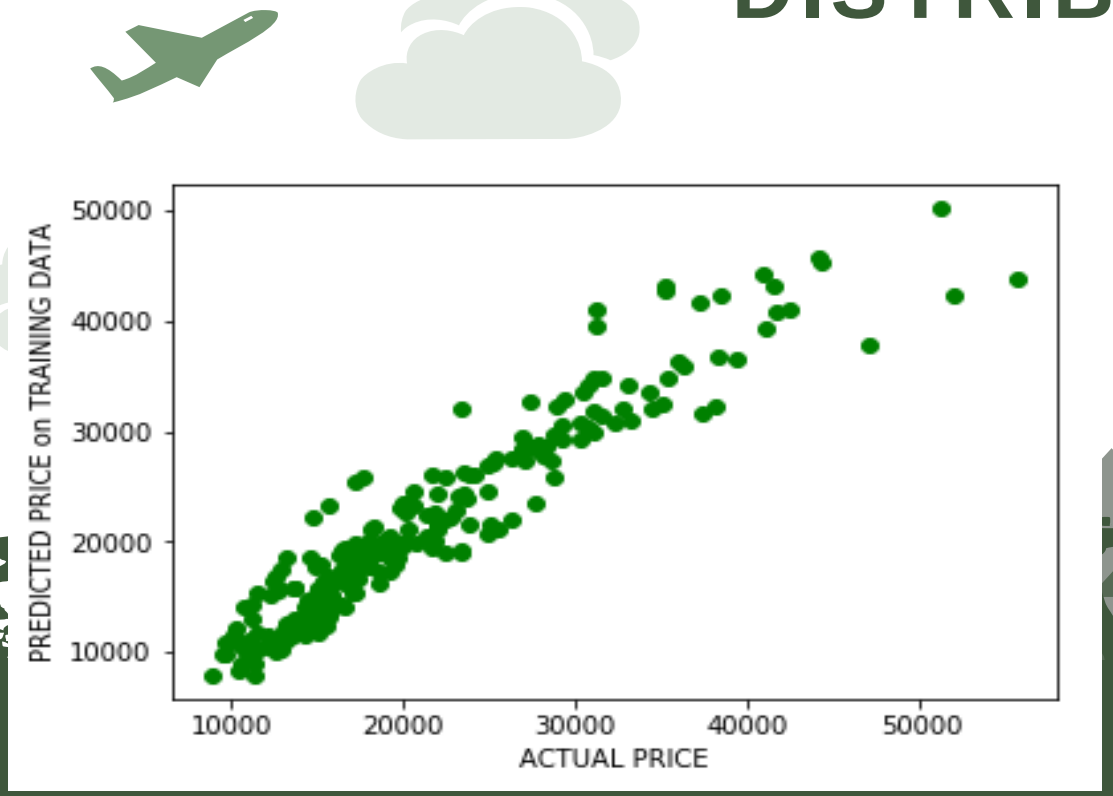
TRAINING	STATS	TESTING
20998.86051	ActualAverage Price	20998.861
21241.59553	Average Predicted Price	21241.596
8222330.321	Mean Squarred Error	8222330.3
2867.460605	Root Mean squarred error	2867.4606
13.50%	Error Percentage	13.50%
0.889	R-squared	0.8978699
0.886	Adj. R-squared	0.886

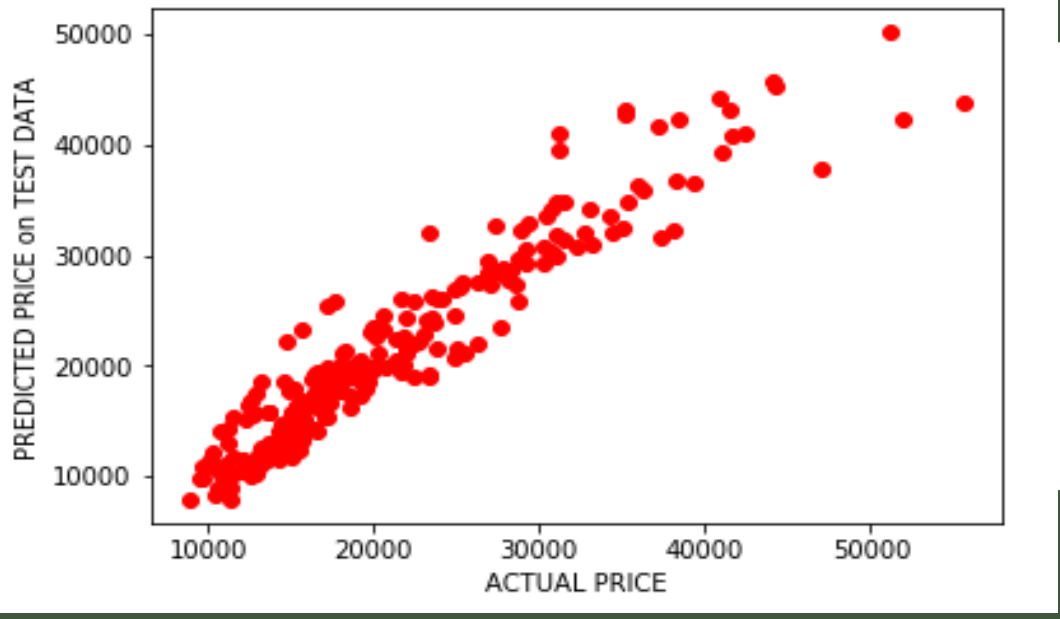
ERROR DISTRIBUTION



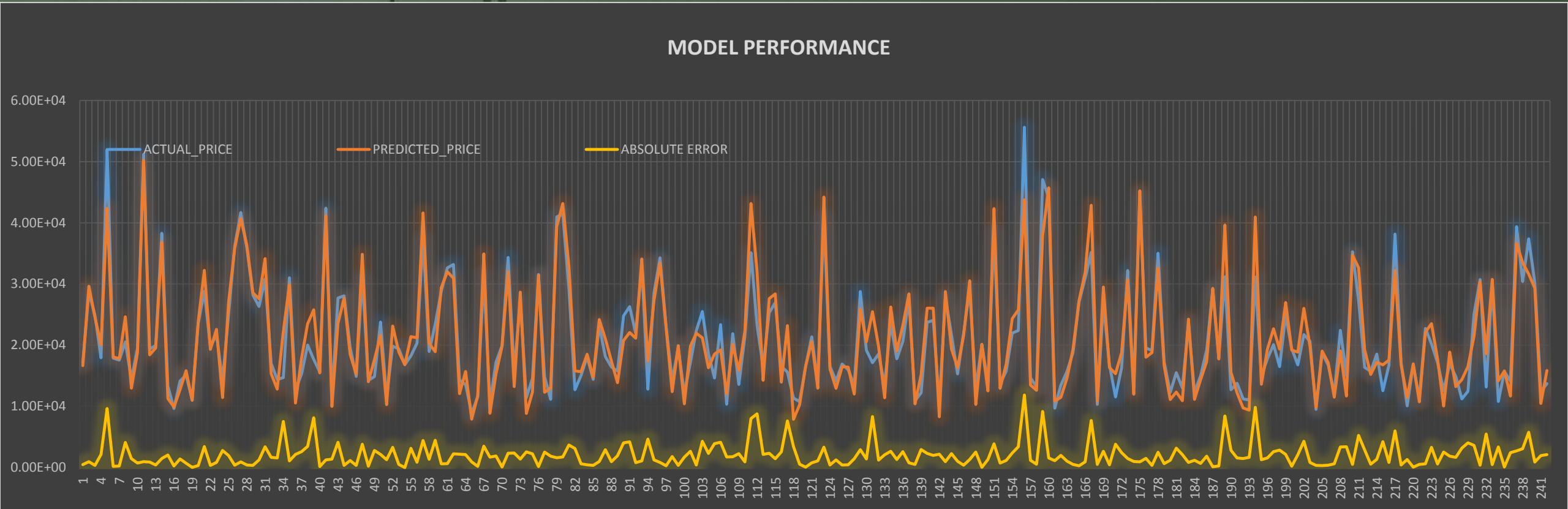


SCATTER PLOT DISTRIBUTION





MODEL PREDICTION PERFORMANCE TREND



FINAL VERDICT

1. Based upon analyzing model performance it can be assumed that model predicts correctly with 13.5% error accuracy.

2. Model so formed has low variance and low biasing