Capstone Project
Yes Bank Stock Closing Price Prediction

SHASHANK BHATIA

CONTENTS

AI

Problem Statement

Perform regression analysis using multiple models to predict the closing price of the stock and compare the evaluation metrics for all of them to find the best model.

DATA PIPELINE

- **Data overview**
- **EDA**
- **Data bifurcation**
- **Test-train data split**
- **Regression analysis**
- **Evaluation metrics comparison**

Data Summary

- **Data set - data_YesBank_StockPrices - contains observations regarding open, close, high and low prices of the yes bank stock from July 2005 - November 2020.**
- **rows: 185**
- **column: 5**
- **Date: It denotes the month and year for a particular price.**
- **Open: Open means the price at which a stock started trading that month.**
- **High: refers to the maximum price that month.**
- **Low: refers to the minimum price that month.**
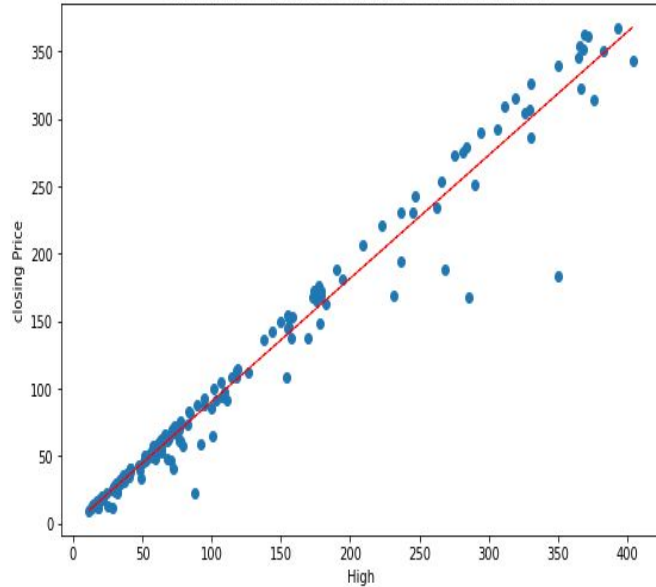- **Close: refers to the final trading price for that month, which we have to predict using regression techniques.**
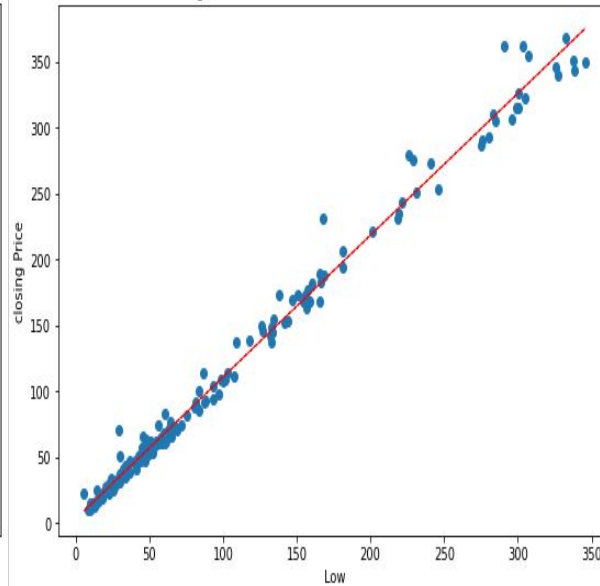
EDA
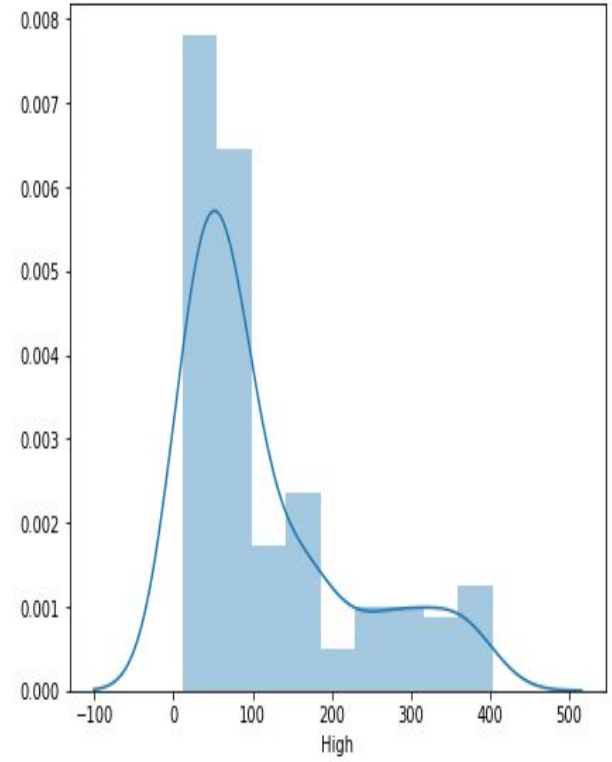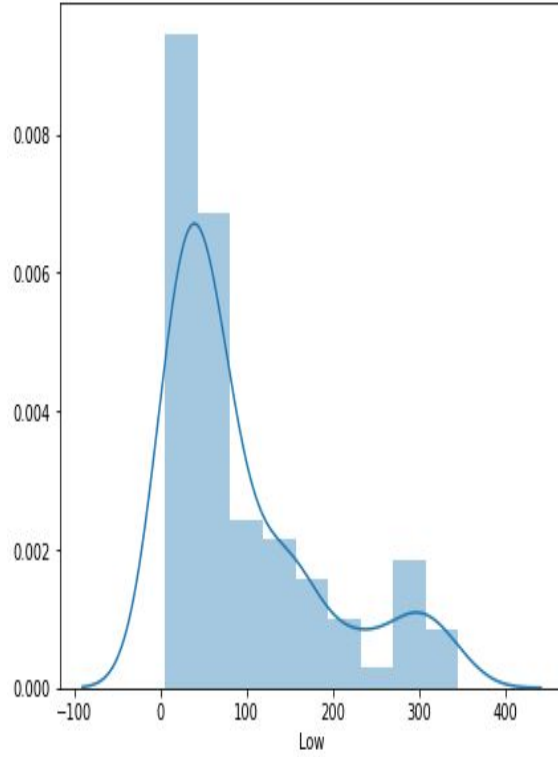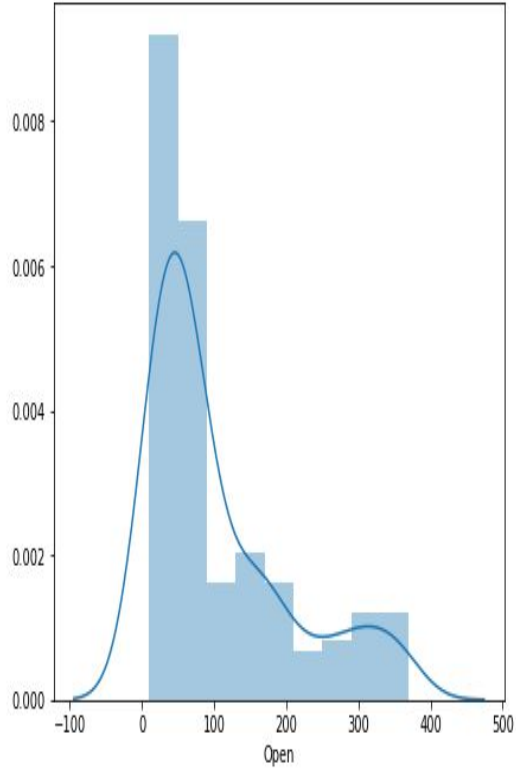
Data distribution visualization
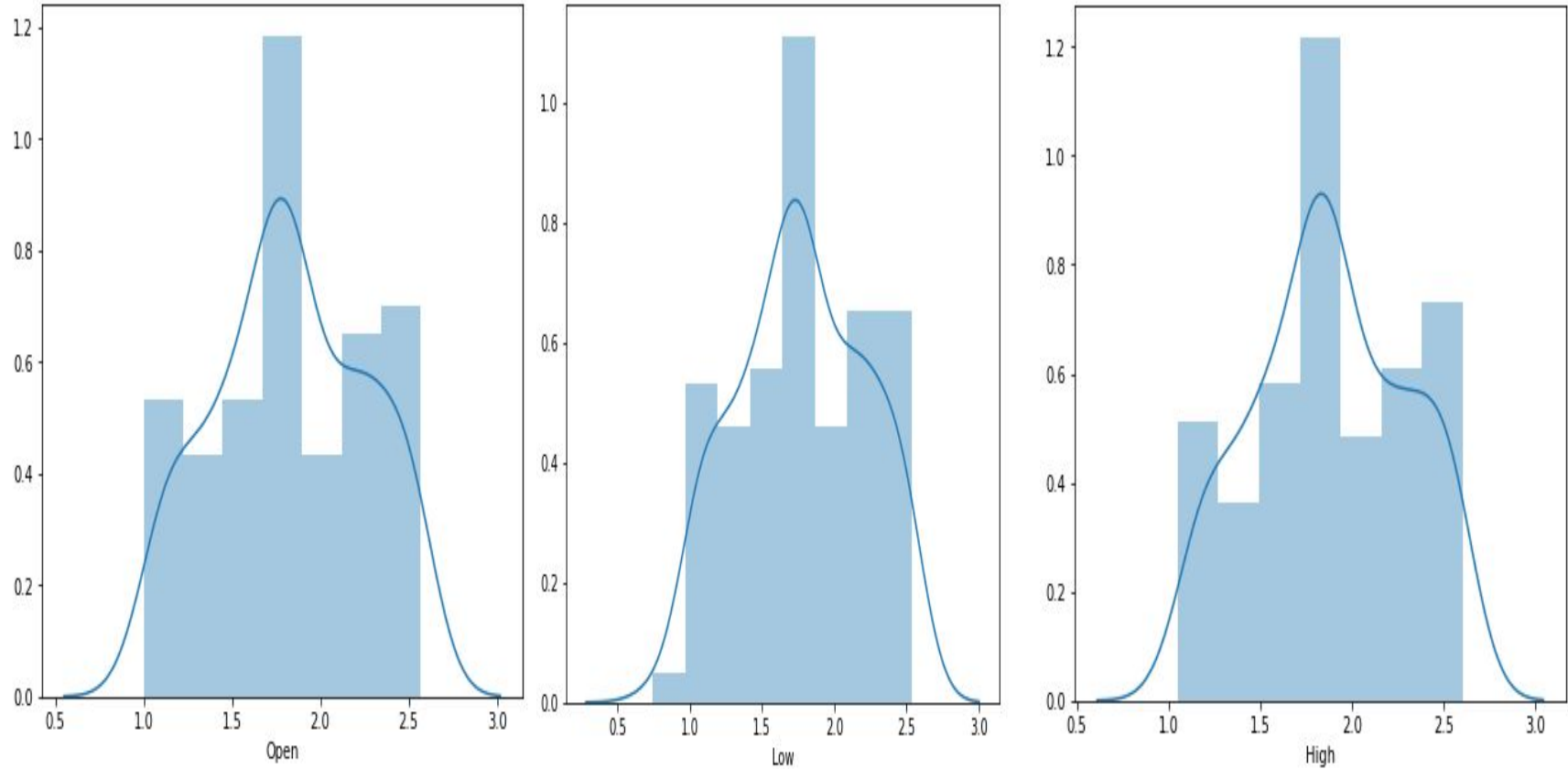
Data transformation visualization
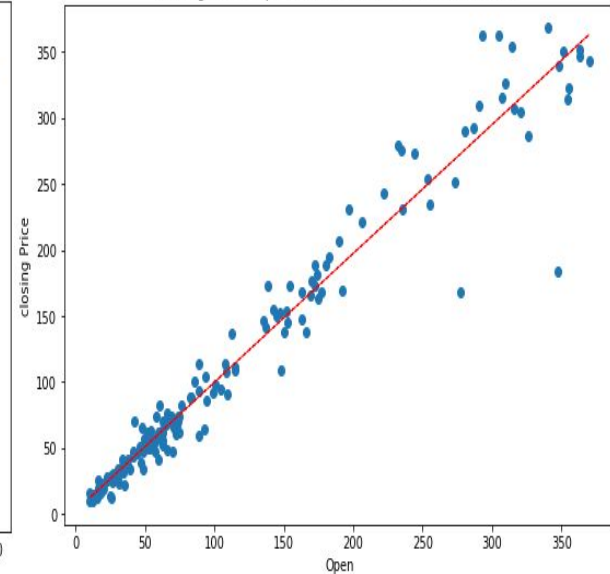
Bivariate Analysis plots

Correlation
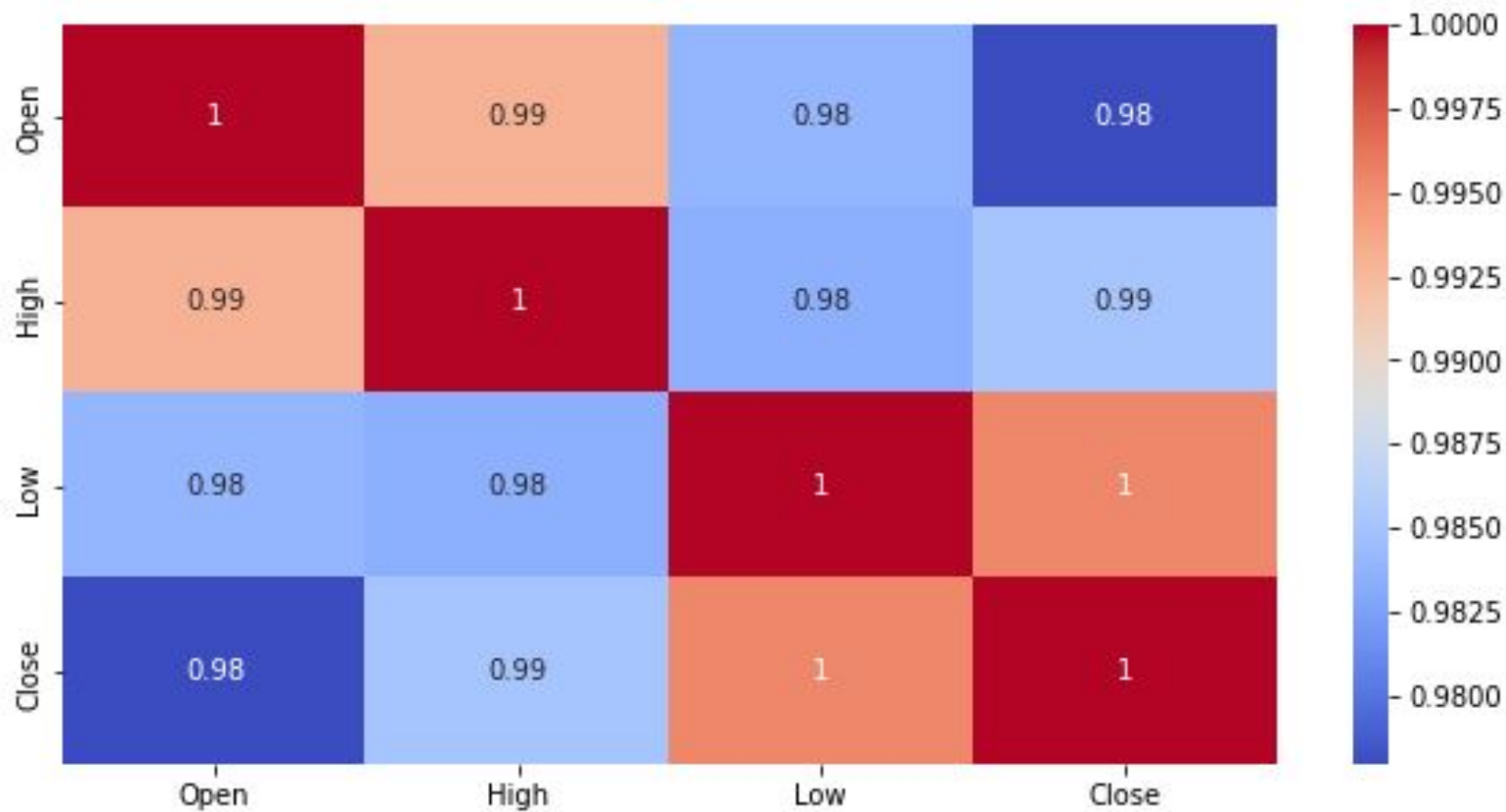
Linear Regression
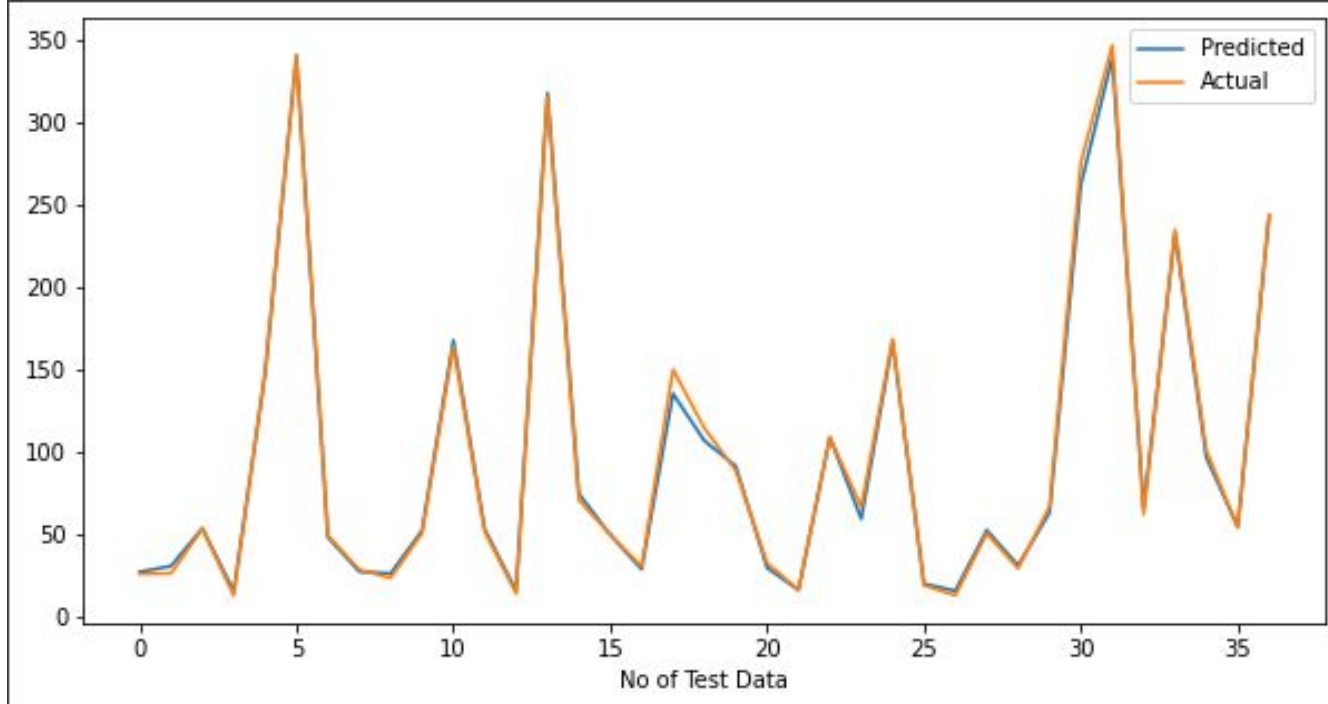
Lasso Regression
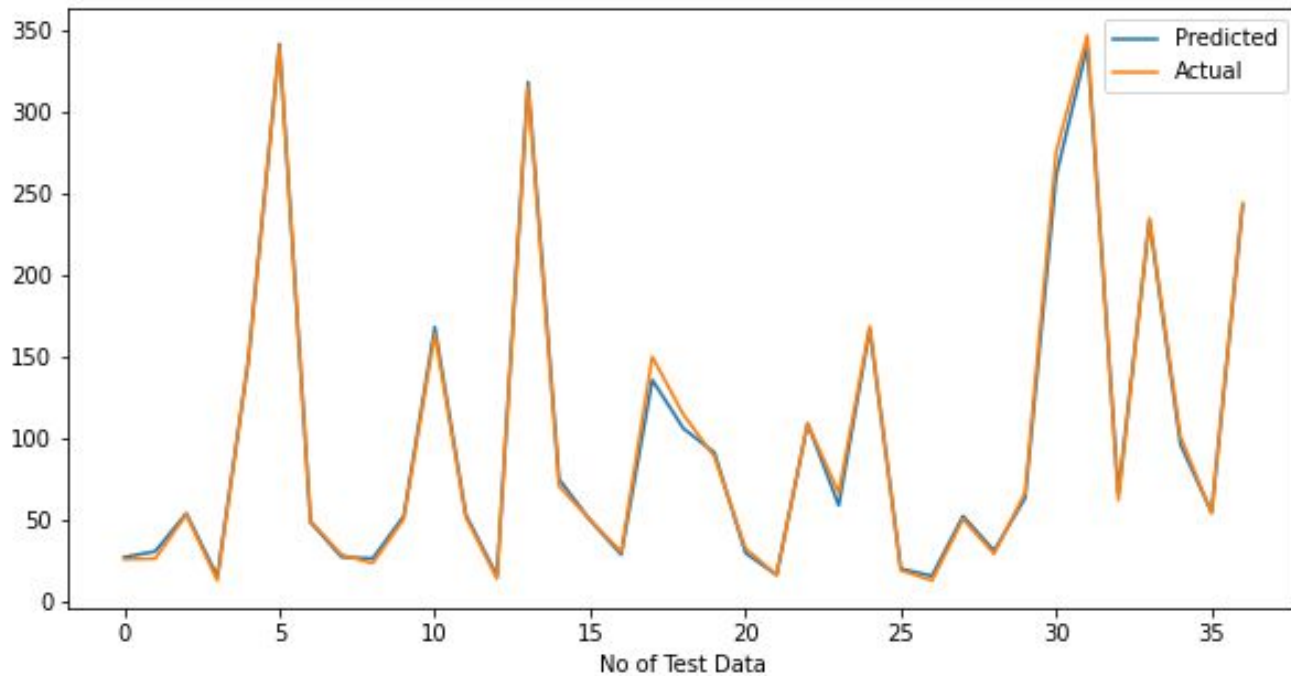

```
MAE  :  3.130856893860506
MSE  :  20.878651216190214
RMSE :  4.569316274475889
MAPE :  5.53404458516646
```

Ridge Regression
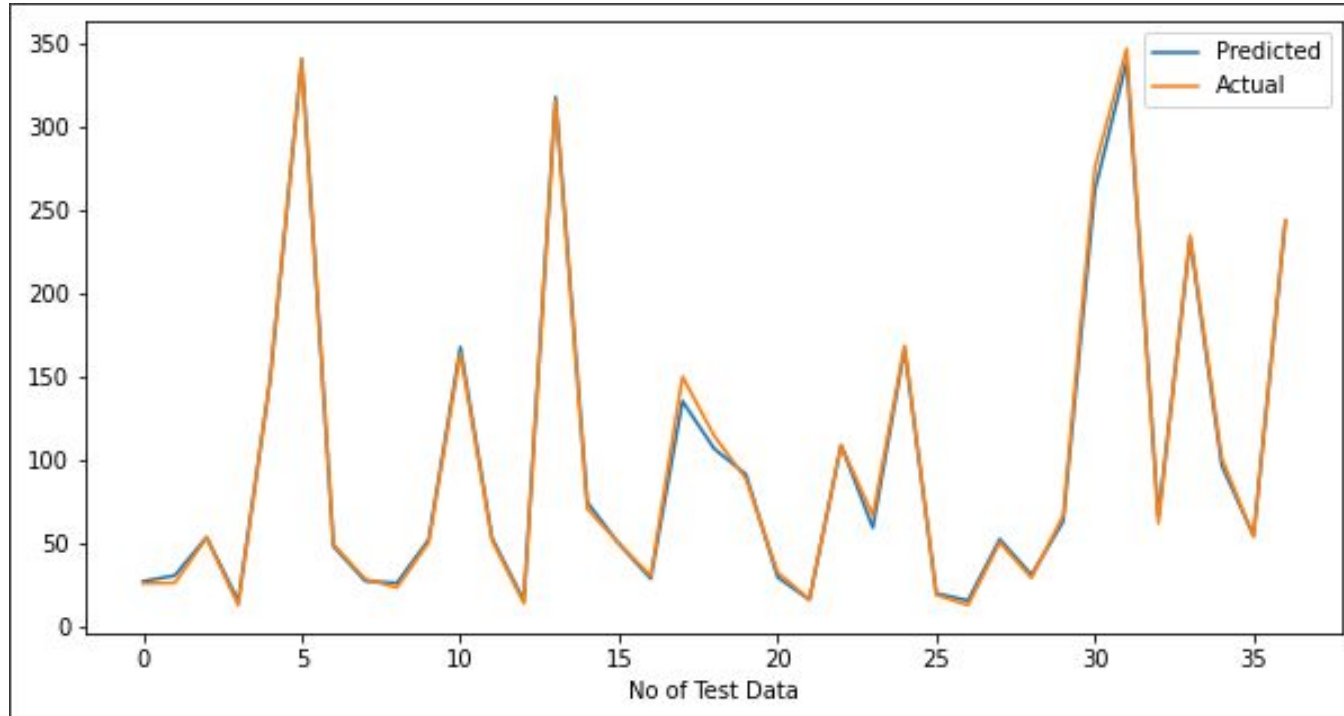
MAE : 3.061022460686153
MSE : 20.095425485603688
RMSE : 4.48279215284444
MAPE : 5.416737109823077

KNN
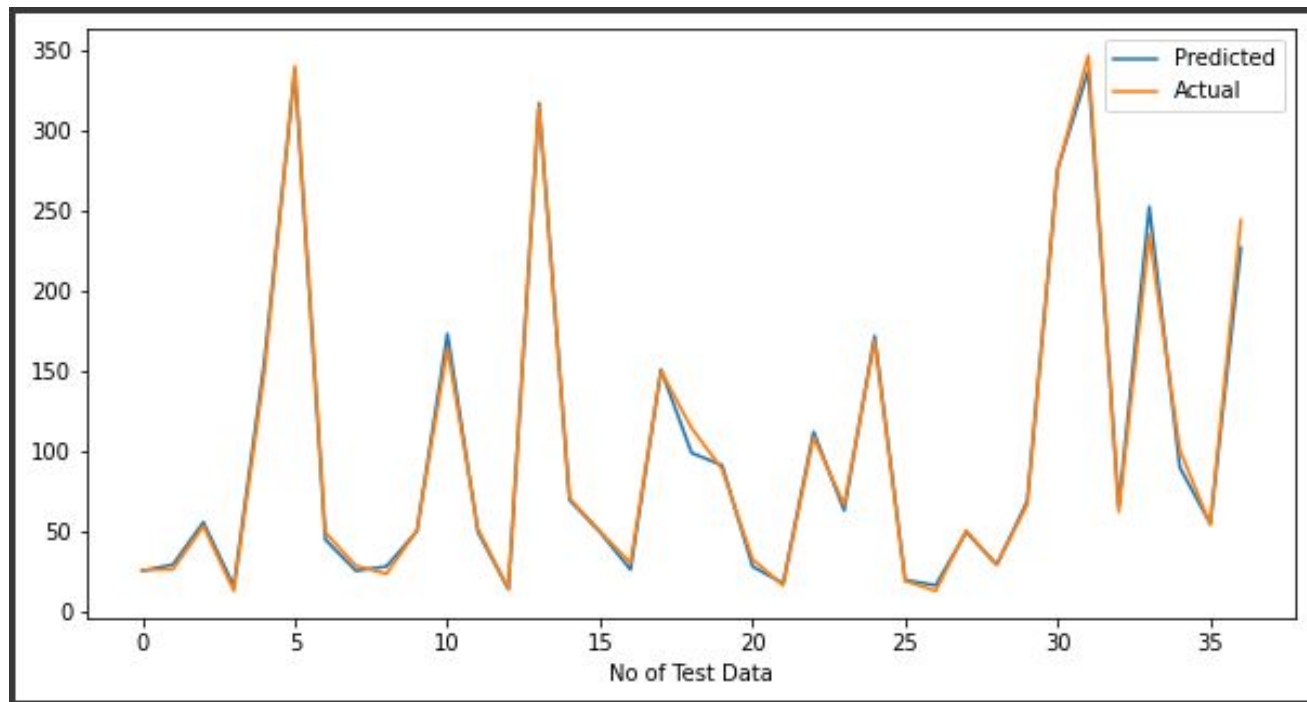
MAE : 4.104324324324326
MSE : 37.72864864864864
RMSE : 6.142365069633084
MAPE : 6.4409728288406365

XGBOOST

MAE : 4.25945175686398
MSE : 32.91328663051823
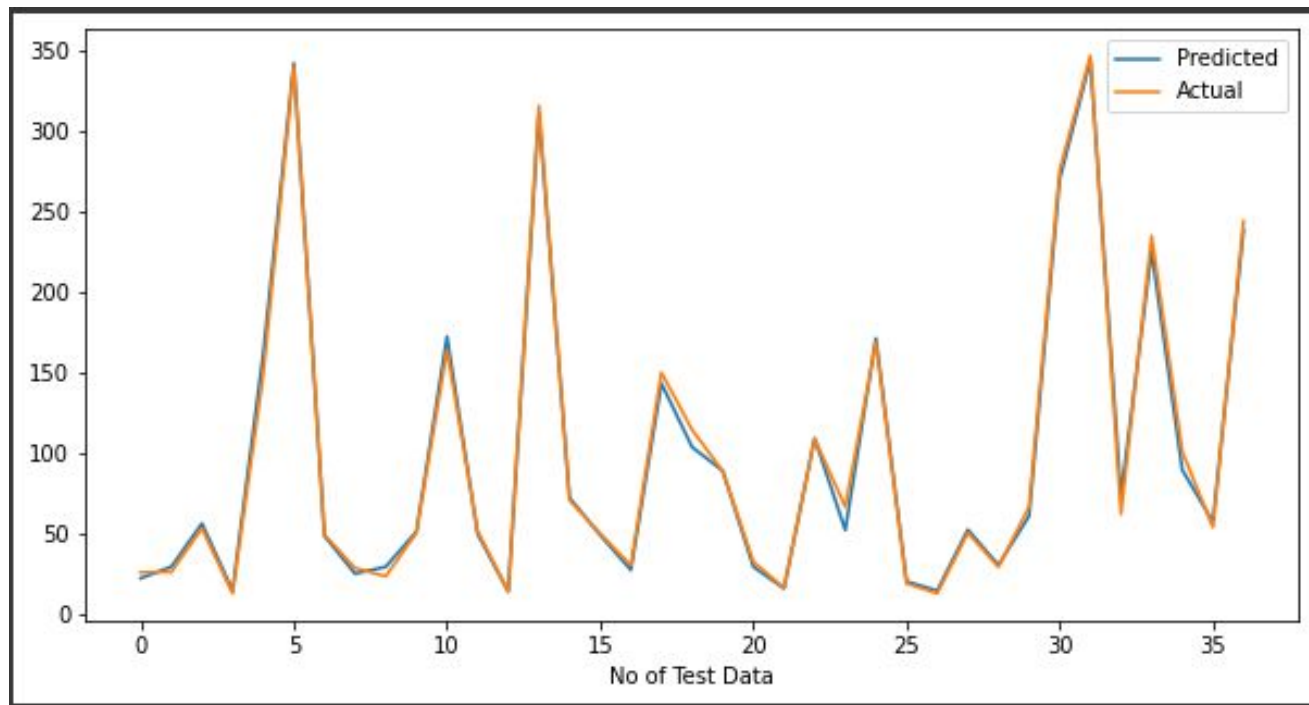RMSE : 5.737010251909807
MAPE : 6.8395711142342331

METRIC EVALUATION

	Model_Name	MAE	MSE	RMSE	MAPE
0	Linear regression	3.05	19.99	4.47	5.40
1	Ridge regression	3.06	20.10	4.48	5.42
2	Lasso regression	3.13	20.88	4.57	5.53
3	KNN regressor	4.10	37.73	6.14	6.44
4	XGBoost regressor	4.26	32.91	5.74	6.84

Conclusion

1. Target variable is highly dependent on input variables.
2. Linear Regression has given the best results with lowest MAE, MSE, RMSE and MAPE score.
3. Ridge regression shrunk the parameters to reduce complexity and multicollinearity, but ended up affecting the evaluation metrics.
4. Lasso regression did feature selection and ended up giving up worse result than ridge which again reflects to the fact that each feature is important(as previously discussed).
5. KNN AND XGBoost have given similar results.
6. The accuracy for each model is more than 90%.