Code Plateau Fellowship; Data Science Capstone Project 3

BY Franklyn Dimka

OBJECTIVE

▶ this project is to model the price of cars with the available independent variables. And this is expected to help the management to understand how exactly the prices vary with the independent variables. They can accordingly manipulate the design of the cars, the business strategy etc. to meet certain price levels. Further, the model will be a good way for management to understand the pricing dynamics of a new market
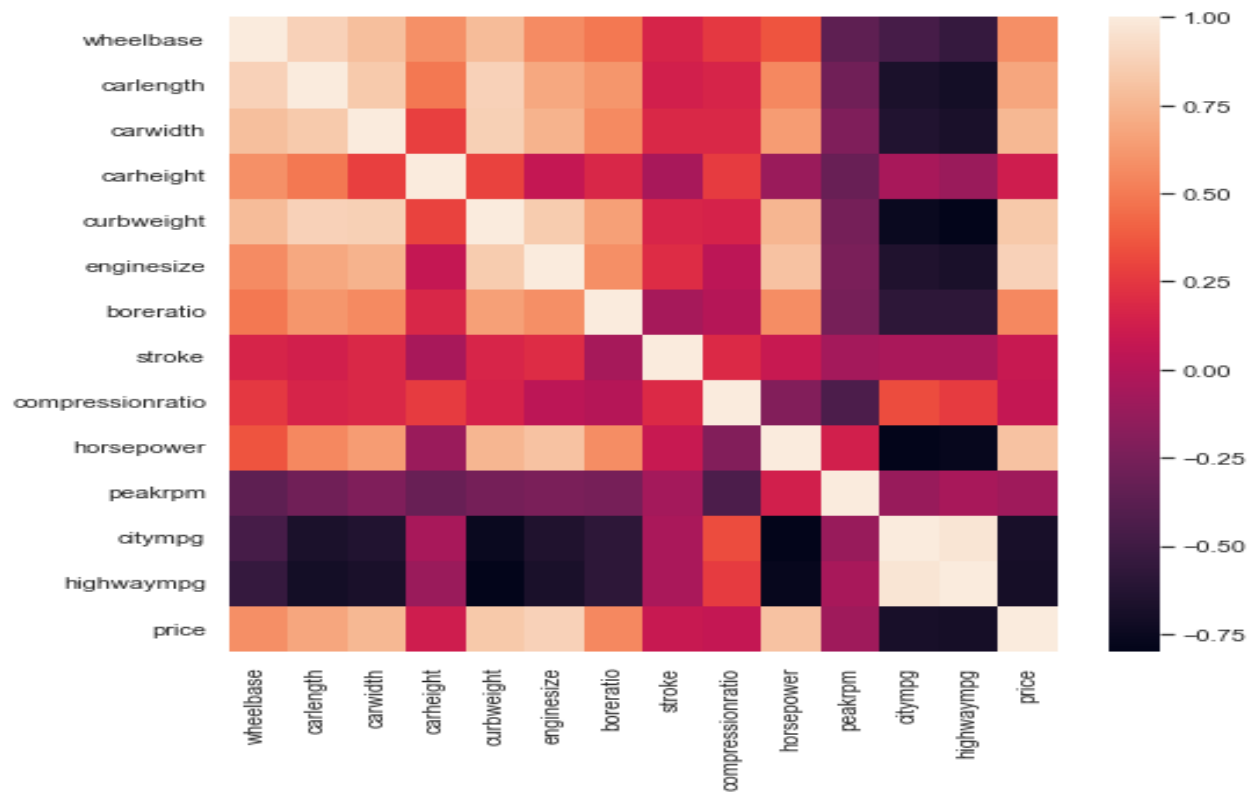
UNIVARIATE ANALYSIS

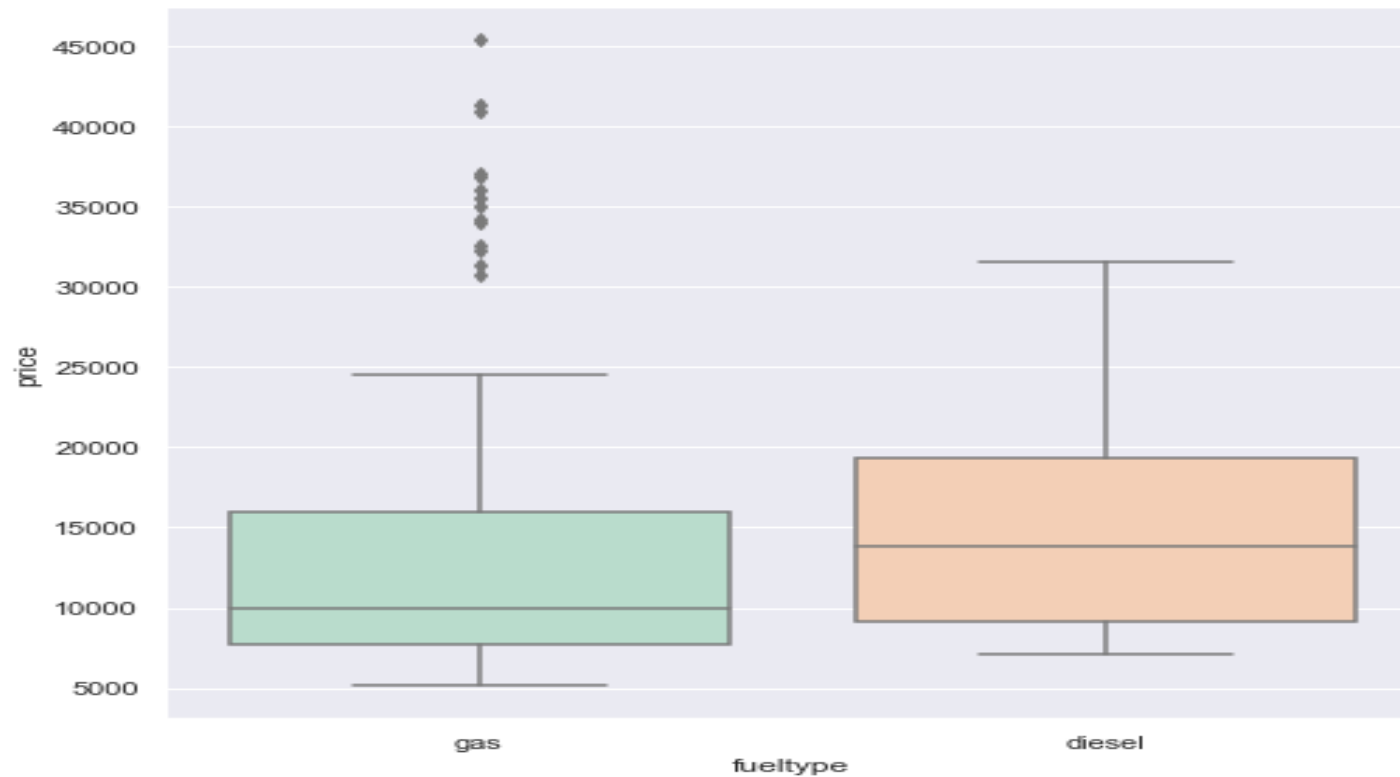▶ Toyota is the most suitable car, majprity of cars uses four cylinders
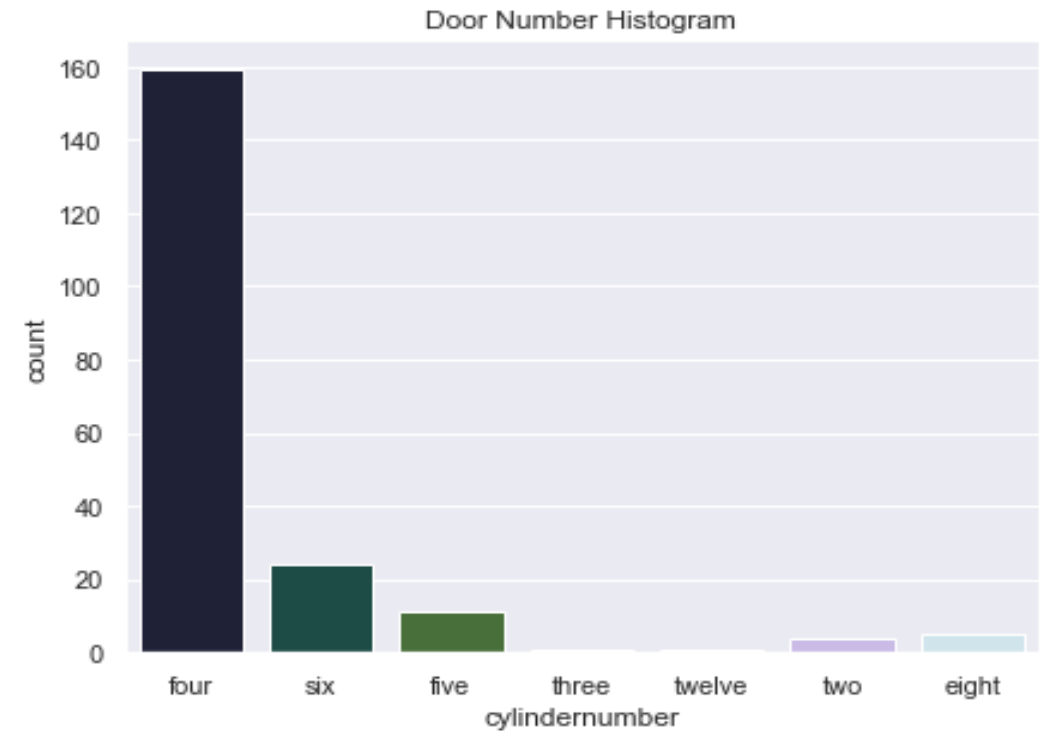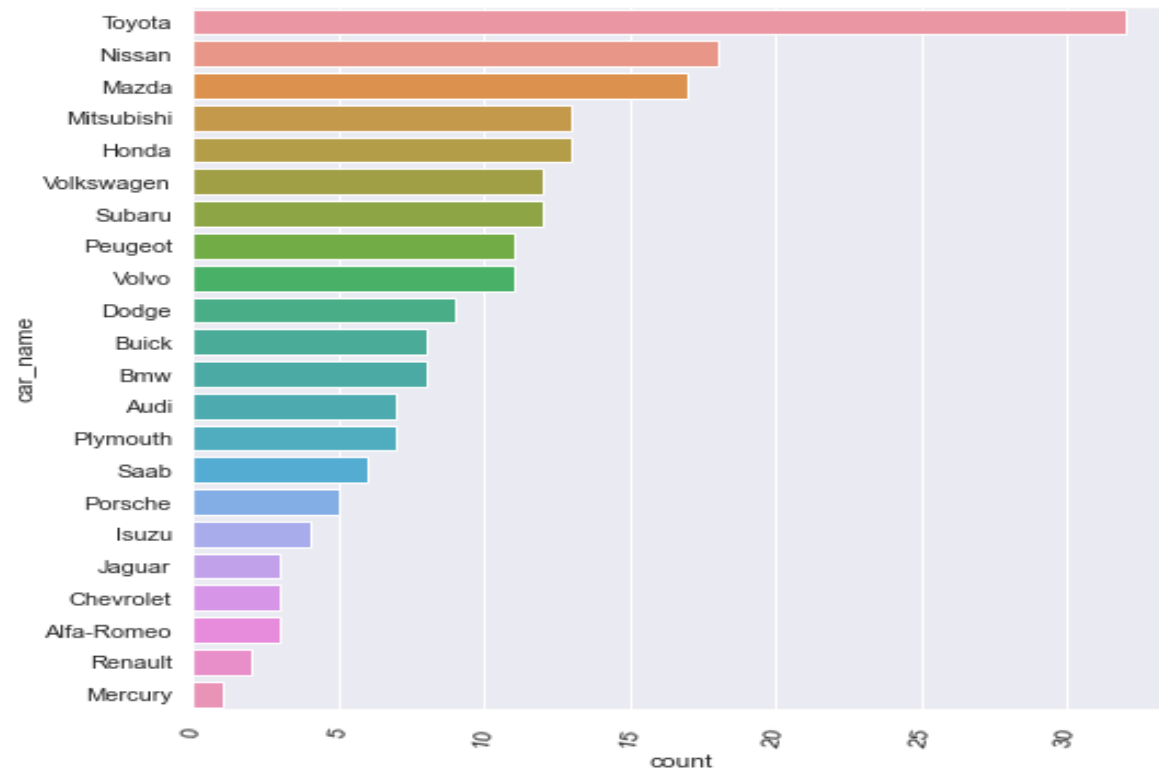
CORRELATION

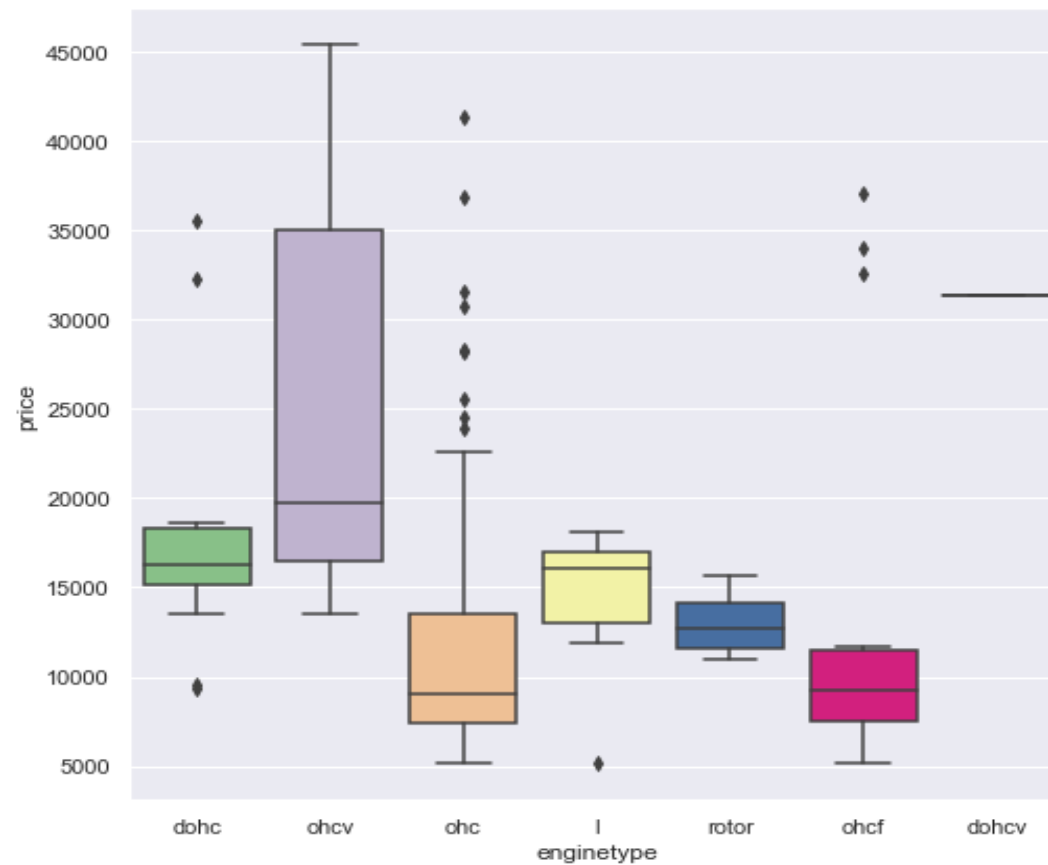▶ **This correlation is majorly between numeric features.**

BIVARIATE ANALYSIS

▶ Fuel vs price

ENGINE TYPE VS PRICE

conclusion

▶ Having applied three models that Linear Regression, Decision Tree Regressor and Random Forest Regressor and subjecting each to accuracy evaluation, we can now have a satisfied model from the random forest Regressor model based on its higher accuarcy performance to predict the prices of cars by Geely Auto in Nigerian market.

▶ Variables that are useful to describe the variances in car prices are present prices, Year, kms_driven, fuel type, seller type, and transmission. Our final model has satisfied the classical assumptions. The R-squared of the model is high, with 97,86% of the variables can explain the variances in the car price. The accuracy of the model in predicting the car price is measured with RMSE, with model has RMSE of 0.798 and prediction data has RMSE of 0.818, suggesting that our model may fit the prediction dataset.

▶ Most of the cars uses gas.

▶ Most of the cars have standard aspiration.

▶ Maximum number of engine type are of 'ohc'.

▶ Most of the cars have fwd drive wheel.