Machine Learning

Problem

A Chinese automobile company Geely Auto aspires to enter the US market by setting up their manufacturing unit there and producing cars locally to give competition to their US and European counterparts.

They have contracted an automobile consulting company to understand the factor on which the price of cars depends. Specifically, they want to understand the factors affecting the pricing of cars in the American market, since those may be very different from the Chinese market. The company wants to know:

- Which variables are significant in predicting the price of a car.
- How well those variables describe the price of a car.

Based on various market surveys, the consulting firm has gathered a large dataset of different types of cars across the American market.

Attributes

1. Car_ID - Unique ID for each observation.
2. Symboling - Its assigned insurance risk rating, value +3 indicates that the auto is risky, -3 that it is pretty safe.
3. carCompany - Name of company
4. fueltype - Car fuel type.
5. aspiration - Aspiration used in car
6. doornumber - Number of doors in a car
7. carbody - body of car
8. drivewheel - type of drive wheel
9. enginelocation - location of car engine
10. wheelbase - Wheelbase of car
11. carlength - length of car
12. carwidth - width of car
13. carheight - height of car
14. curbweight - The weight of a car without occupants or luggage.
15. enginetype - type of engine

16. cylindernumber - cylinder placed in the car.
17. enginesize - size of car.
18. fuelsystem - Fuel system of car.
19. boreratio - Boreratio of car
20. stroke - Stroke or volume inside the engine.
21. compressionratio - compression ratio of car.
22. horsepower - Horsepower
23. peakrpm - car peak rpm
24. citympg - Mileage in city
25. highwaympg - Mileage on highway
26. Price(Dependent Variable) - Price of car

Objective

You are required to model the prices of cars with the available independent variables. It will be used by 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 good for management to understand the pricing dynamics of the new market.

Questions -

Step 1 : Understand the Business Problem.

Step 2 : Import all the libraries and set up all the requirements that you will need.

Step 3: Import the data set and check the following
- dimension of the dataset.
- data types.
- Missing value available in the dataset.
- Descriptive statistics of data and write the observation.

Step 4: Data Cleaning
- Create the column as 'CompanyName' using 'CarName' Column. List down the unique 'CompanyName'.
- Check the correctness of data in the 'CompanyName' column.
- Check the duplicate data in the dataset.

Step 5: Exploratory Data Analysis

- Visualize the 'price' column using displot and boxplot. Write down the observations.
- Perform the appropriate transformation to make the target as a gaussian distribution.
- Check the linear relationship between the dependent variable "Price" and the numerical independent variables
- Checking the multicollinearity between the correlated independent variables above and Price
- Perform Univariate, Bivariate, and Multivariate analyses to find the factors that affect the Target variables.
- Perform feature engineering based on sound knowledge of the business problem and available dataset.

Step 6: Perform the preprocessing that is required for the model.

Step 7: Split the dataset into train and test data sets and perform the scaling on both sets if necessary.

Step 8: Build the base model.

Step 9: Understand how the model is performing, Perform feature engineering again if needed. Do feature selection. Try with various models like parametric and nonparametric models. Once you choose the final model, rebuild the model with the best parameters.
Note: If you are performing with Linear models, check the model is fulfilling the assumptions.

Step 10: Based on your understanding of the model and EDA analysis, Explain the business understanding.