Multiple Linear Regression

**Assignment Task:**

Your task is to perform a multiple linear regression analysis to predict the price of Toyota corolla based on the given attributes.

**Dataset Description:**

The dataset consists of the following variables:

Age: Age in years

KM: Accumulated Kilometers on odometer

FuelType: Fuel Type (Petrol, Diesel, CNG)

HP: Horse Power

Automatic: Automatic ( (Yes=1, No=0)

CC: Cylinder Volume in cubic centimeters

Doors: Number of doors

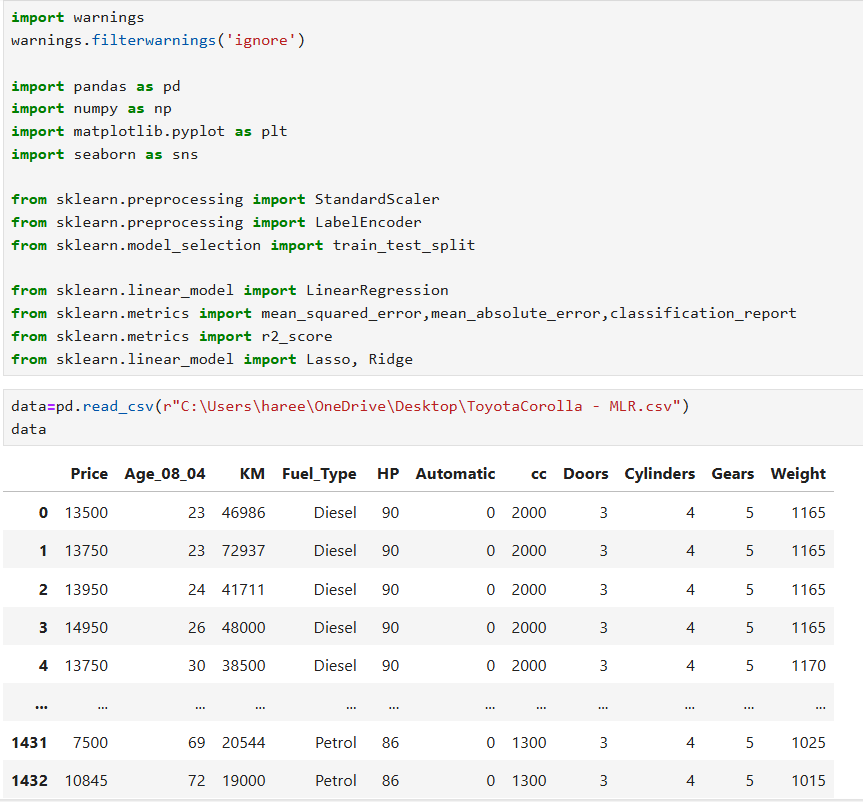
Weight: Weight in Kilograms

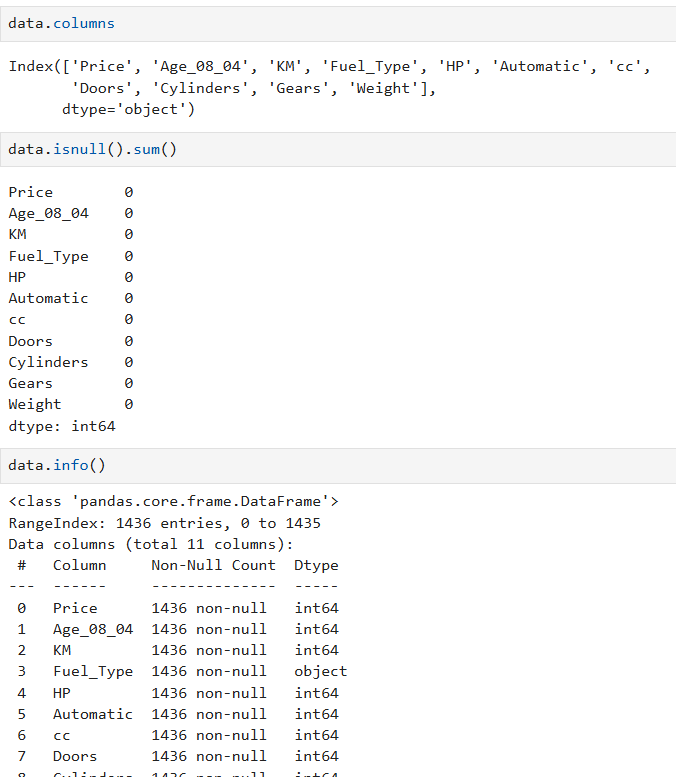
Quarterly\_Tax:

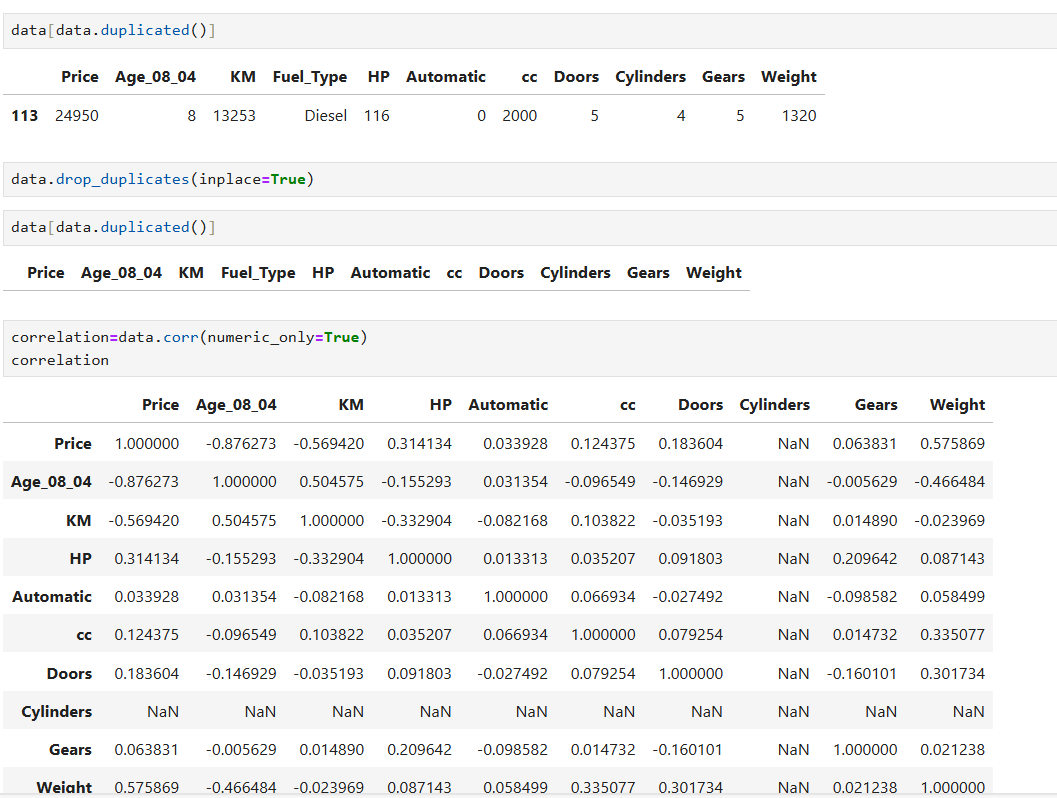
Price: Offer Price in EUROs

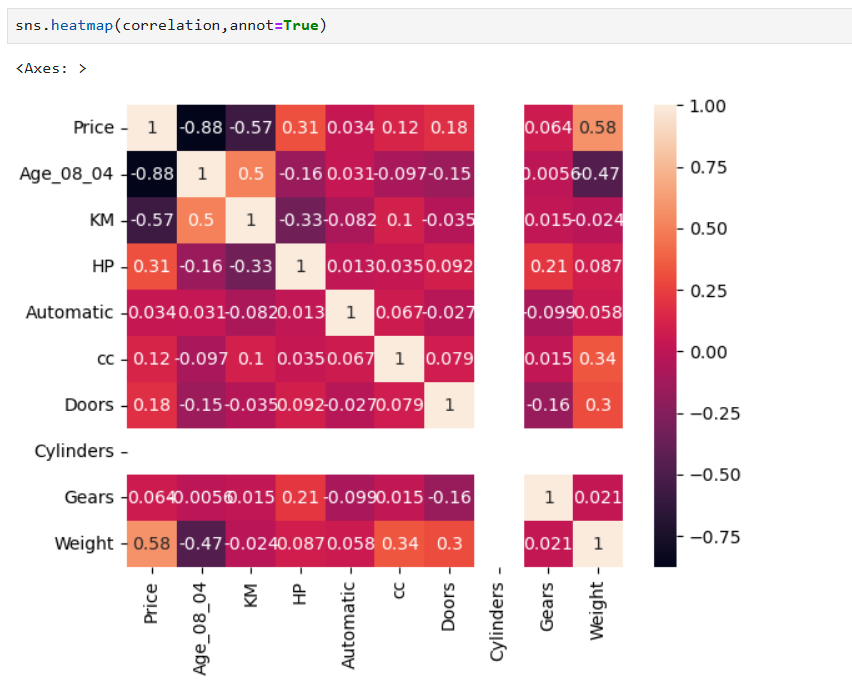
**Taskes:**

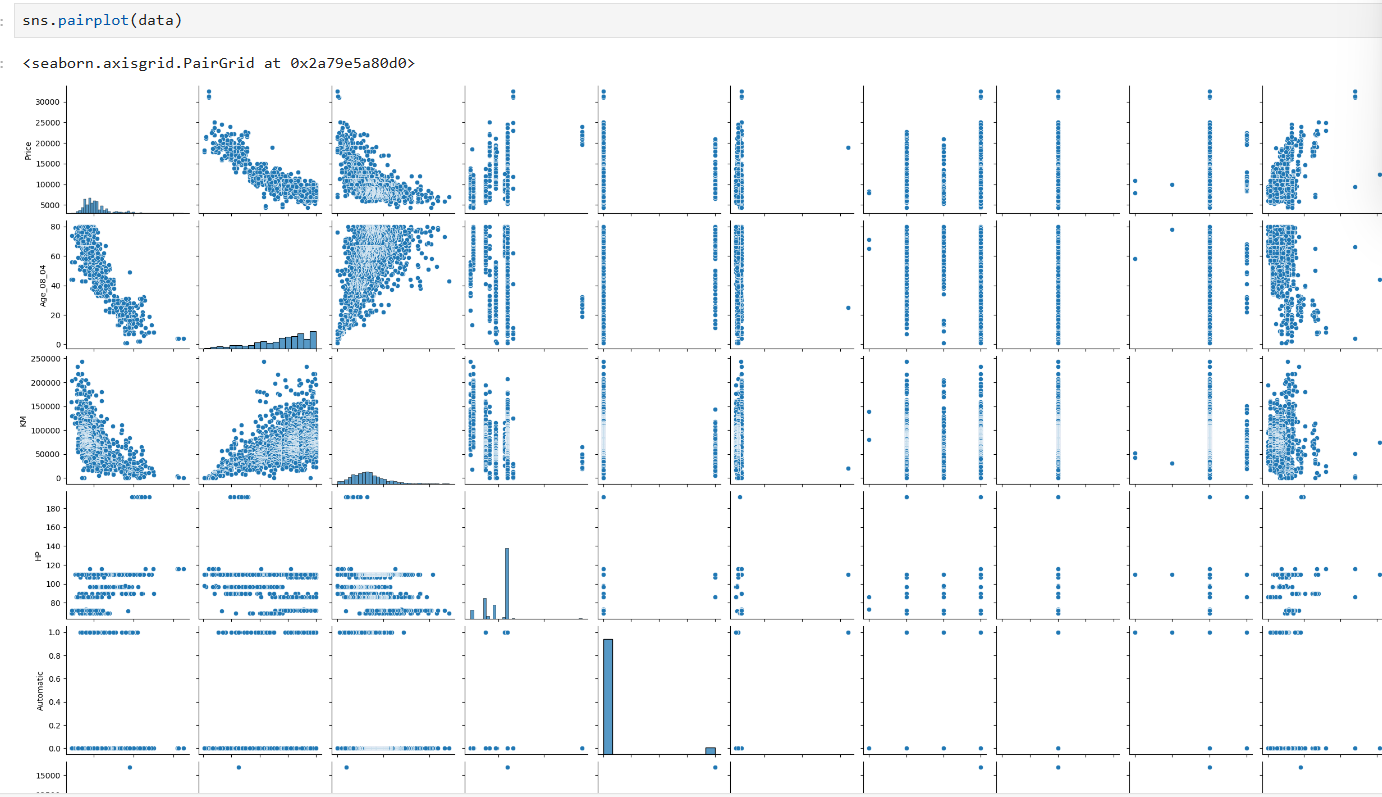
1.Perform exploratory data analysis (EDA) to gain insights into the dataset. Provide visualizations and summary statistics of the variables. Pre process the data to apply the MLR.

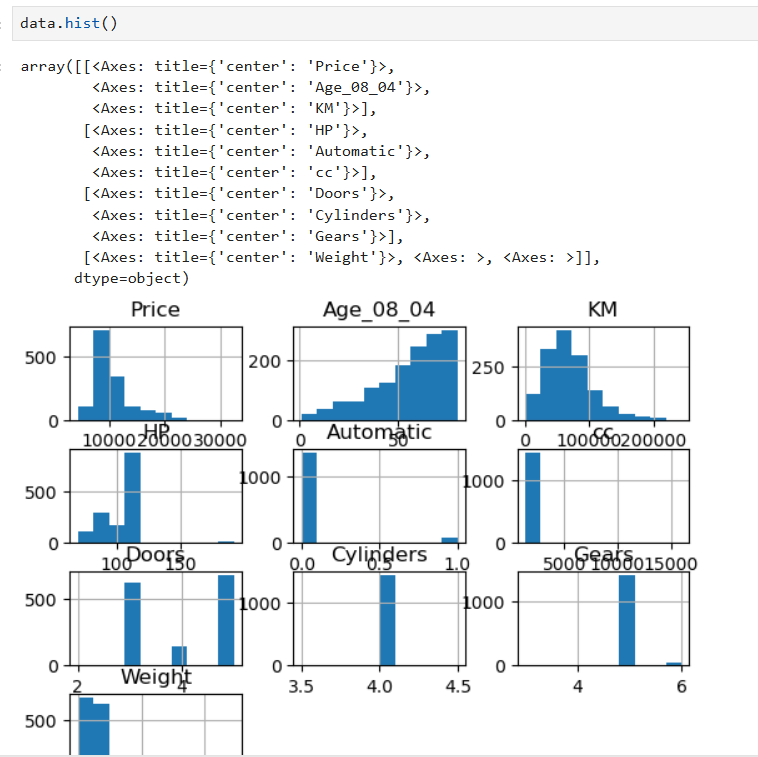


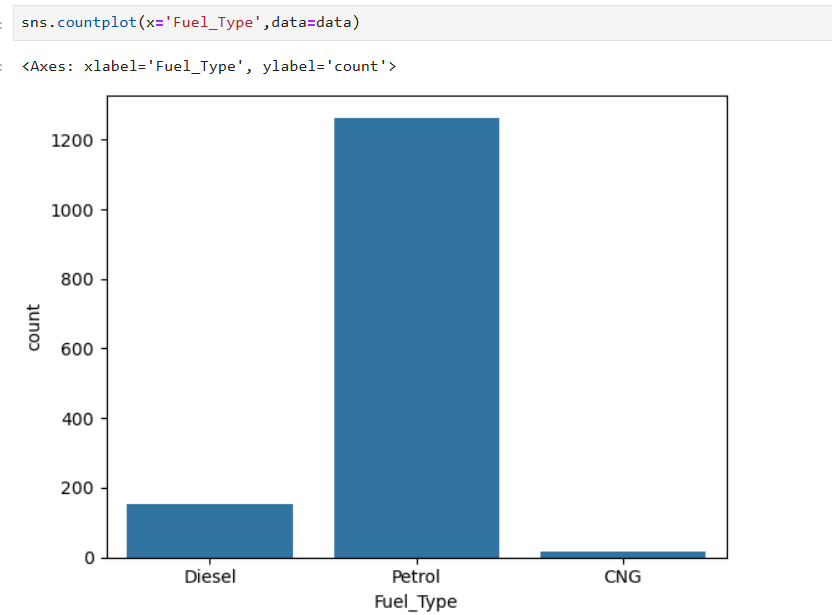


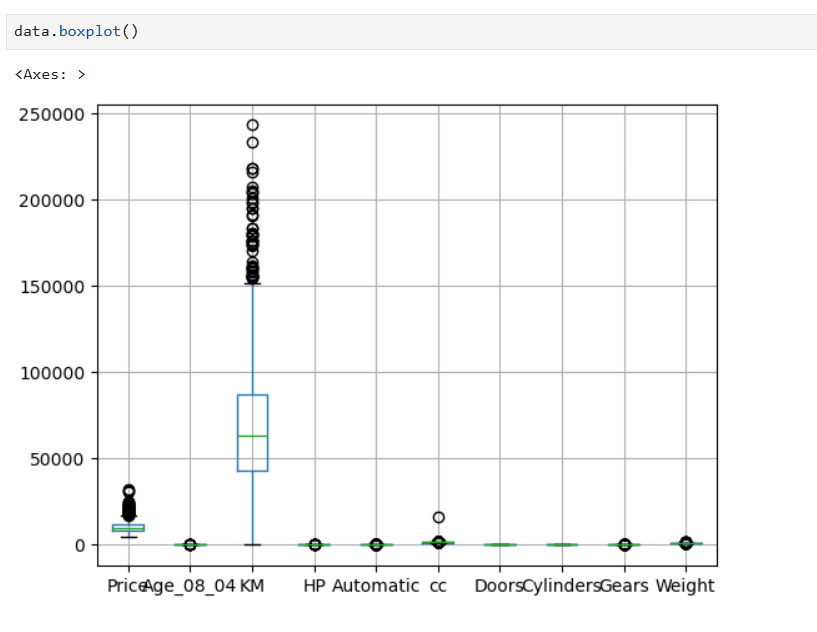


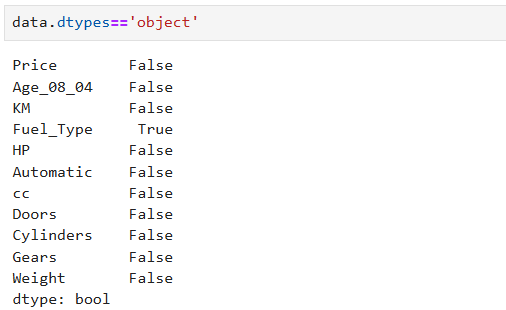




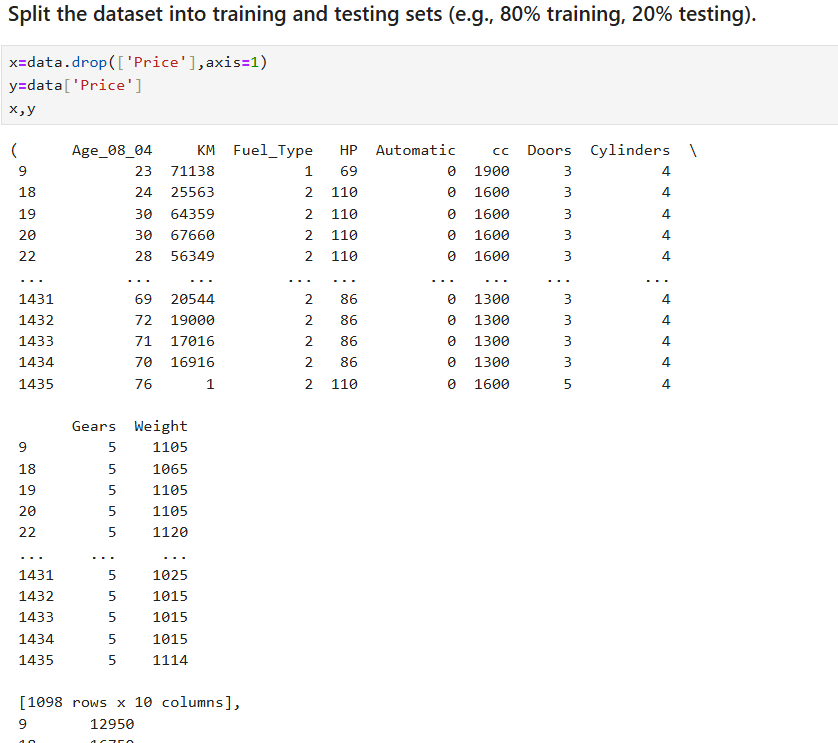


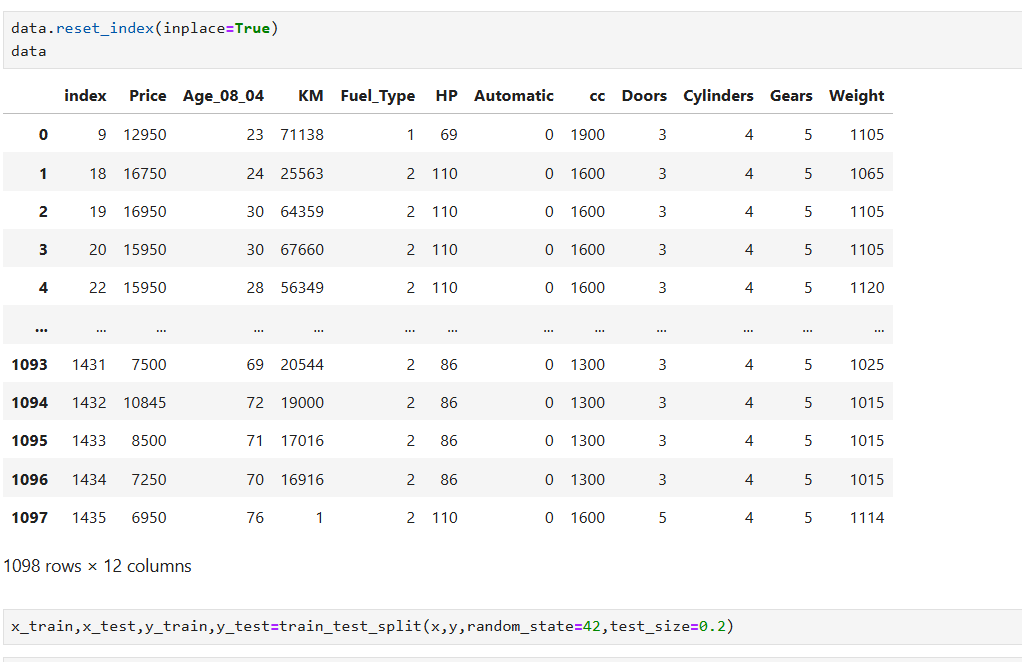


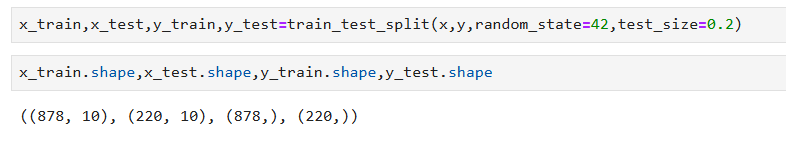




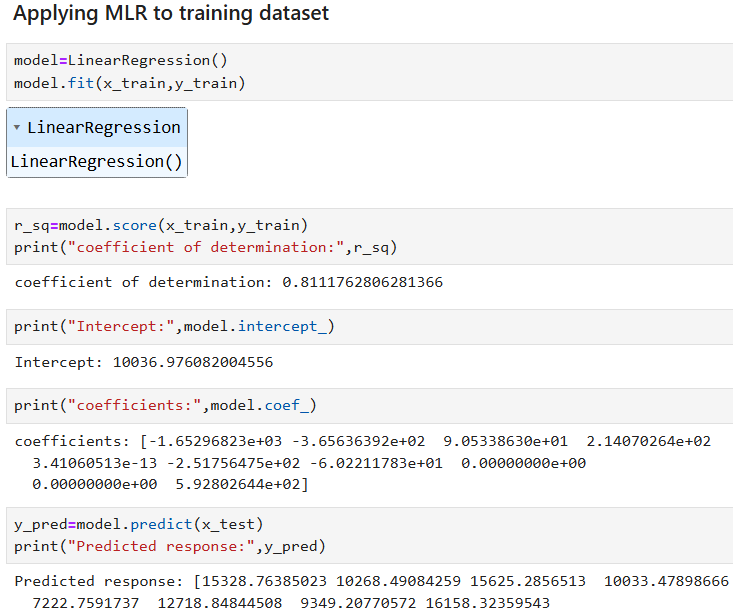
2.Split the dataset into training and testing sets (e.g., 80% training, 20% testing).

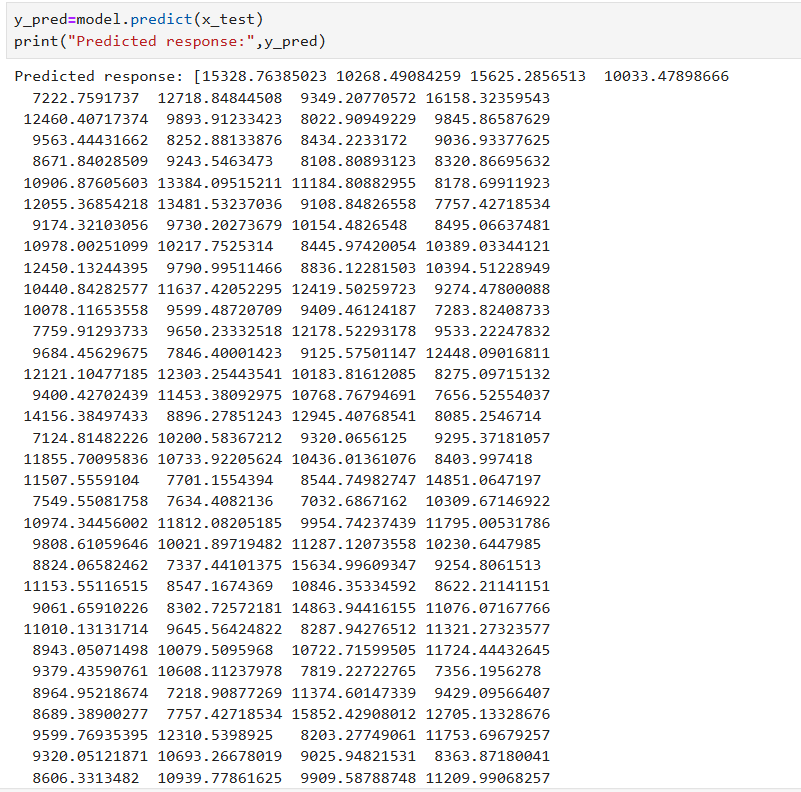


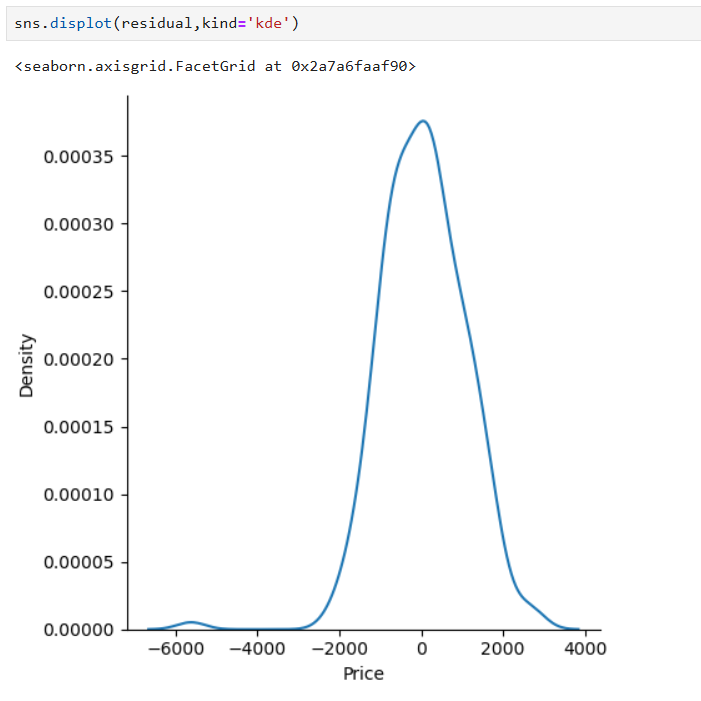


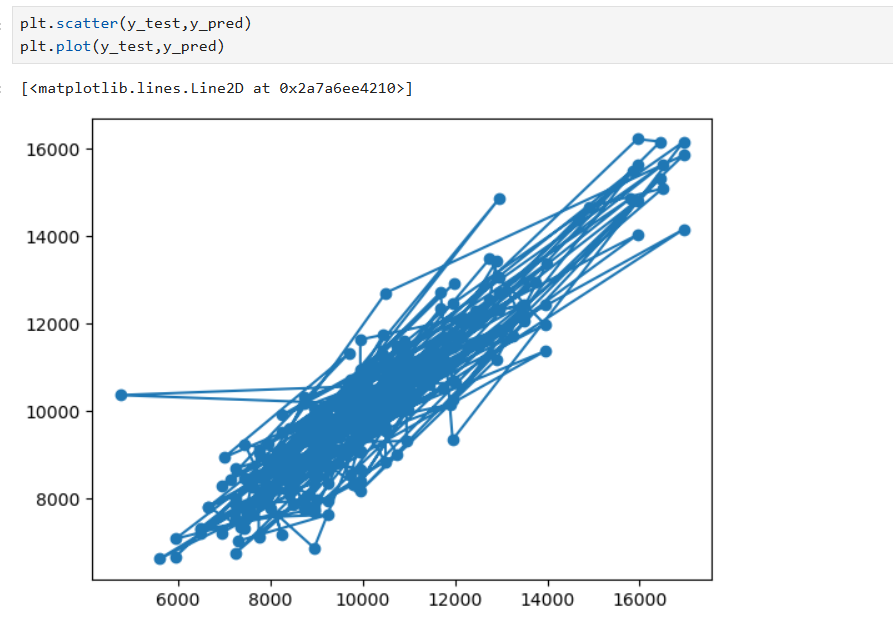


3.Build a multiple linear regression model using the training dataset. Interpret the coefficients of the model. Build minimum of 3 different models.



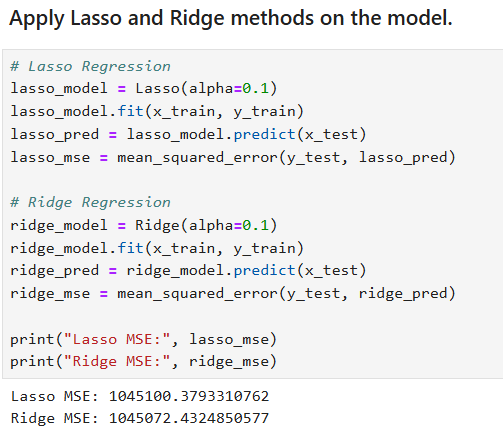






4.Evaluate the performance of the model using appropriate evaluation metrics on the testing dataset.

5.Apply Lasso and Ridge methods on the model.



**Interview Questions:**

1.What is Normalization & Standardization and how is it helpful?

Answer: Normalization in data-science refers to the process of scaling numeric features to a common range ,typically between 0 and 1. The goal of normalization is to bring all the features into a similar scale, which is particularly important for algorithms that are sensitive to the scale of the features

2.What techniques can be used to address multicollinearity in multiple linear regression?

Answer: Multicollinearity occurs when independent variables in a regression model are highly correlated with each other. This can cause issues in multiple linear regression, such as unstable coefficients estimates and difficulties in interpreting the importance of individual predictors.

Ensure to properly comment your code and provide explanations for your analysis.

Include any assumptions made during the analysis and discuss their implications.