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?

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

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

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

Answer: <https://colab.research.google.com/drive/1e6Jx7UjxdArcmttIKpQpYwdbDIiiFg_C?usp=sharing>

**Interview Questions-Answers**

**Ans :1)** Normalization and standardization are crucial data pre-processing techniques that scale and transform features in a dataset.

Normalization involves scaling individual samples to have unit norm. Standardization transforms data to have zero mean and unit variance, useful for features with different units or scales.

Common techniques include Z-score Standardization and Scaling to a Fixed Mean and Standard Deviation.

**Ans:2)** Multi-collinearity in regression models can lead to unstable coefficient estimates, inflated standard errors, and difficulty in interpreting the model.

Techniques to address this include feature selection, domain knowledge, Ridge Regression, Lasso Regression, Variance Inflation Factor (VIF) etc.