REPORT ON LINEAR REGRESSION

# Libraries Used

The notebook utilizes the following Python libraries:

* pandas: For data manipulation.
* matplotlib.pyplot: For data visualization.
* seaborn: For enhanced visualizations.
* scikit-learn modules:
  + train\_test\_split: To split the dataset into training and testing sets.
  + LinearRegression: To build and train the regression model.
  + metrics: For evaluating the model's performance.

## Dataset

The dataset is loaded from a file named cardata.csv. It contains the following columns:

* Car\_Name, Year, Selling\_Price, Present\_Price, Kms\_Driven, Fuel\_Type, Seller\_Type, Transmission, Owner.

A brief overview of the dataset shows:

* **Entries:** 301 rows.
* **Columns:** 9, with mixed data types (numerical and categorical).

## Linear Regression Model

The notebook includes the following steps:

1. **Data Splitting**:
   * Training set: 80% of the data.
   * Testing set: 20% of the data.
   * Shapes:
     + X\_train: (240, 8)
     + X\_test: (61, 8)
     + y\_train: (240,)
     + y\_test: (61,)
2. **Model Building**:
   * A LinearRegression model is trained using X\_train and y\_train.
   * Predictions are made on X\_test.
3. **Model Evaluation**:
   * **Coefficients**: [0.437, -5.3e-6, 0.345, -0.413, 2.23, 0.459, -1.209, -1.870]
   * **Intercept**: 7.074
   * **Mean Squared Error (MSE)**: 2.982

**Outputs**

The following results and visualizations are included:

* Dataset summary using info() and descriptive statistics.
* Model performance metrics:
  + MSE of predictions.
* Regression coefficients and intercept.