**Assignment 2**

**Saurabh Vijay Isane  
122B1F035**

**1. Introduction**

Linear Regression is a fundamental statistical technique used for predictive modeling. It establishes a relationship between an independent variable (features) and a dependent variable (target) using a linear equation.

**2. Concept of Linear Regression**

Linear Regression assumes a linear relationship between the dependent variable (Y) and independent variables (X). The equation is given by:

Y = b0 + b1X1 + b2X2 + ... + bnXn + e

* Y = Dependent variable (Target)
* X1, X2, ..., Xn = Independent variables (Features)
* b0 = Intercept
* b1, b2, ..., bn = Coefficients (weights)
* e = Error term

**3. Assumptions of Linear Regression**

1. **Linearity**: The relationship between independent and dependent variables is linear.
2. **Independence**: Observations are independent of each other.
3. **Homoscedasticity**: The variance of residuals is constant.
4. **No multicollinearity**: Independent variables should not be highly correlated.
5. **Normality of residuals**: Residuals should be normally distributed.

**4. Exploratory Data Analysis (EDA) Steps**

**4.1 Loading the Dataset**

* The dataset is imported using Pandas.
* Initial examination using .info(), .describe(), and .head() functions.

**4.2 Handling Missing Data**

* Missing values are checked using isnull().sum().
* Imputation techniques such as mean, median, or mode are used to fill missing values.

**4.3 Checking for Outliers**

* Box plots and scatter plots help identify outliers.
* Outliers can be handled using transformations or removal methods.

**4.4 Correlation Analysis**

* A heatmap is used to visualize the correlation between features and the target variable.

**5. Data Preprocessing**

**5.1 Encoding Categorical Variables**

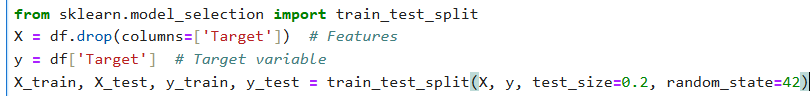
* Convert categorical variables using One-Hot Encoding or Label Encoding.

**5.2 Feature Scaling**

* Standardization (Z-score normalization) or Min-Max Scaling can be applied to numerical features to improve model performance.

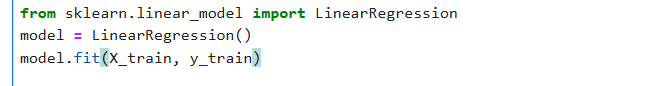
**5.3 Splitting the Data**

* The dataset is divided into training and testing sets using train\_test\_split().



**6. Applying Linear Regression**

**6.1 Model Training**



**6.2 Model Evaluation**

