Assignment 1

Create a dataset (in csv file) of at least 20 points for any two related variables (e.g., hours studied vs. marks scored).

Plot the data points on a graph and describe the trend you observe.

Using your dataset, calculate the slope mmm and intercept ccc of the best-fit line using the formulas.

Draw the regression line on a graph and label it.

Write a Python program to fit a linear regression model on your dataset using sklearn.

Print the slope and intercept obtained from the model.

Plot the regression line along with the data points.

Use your fitted model to predict the dependent variable y for at least two new values of x.

Compare your predictions with your expectations based on the trend.

Calculate the Mean Squared Error (MSE) of your model. Do not use built-in functions for MSE calculation.

Comment on how well your line fits the data.

Add one or two new data points to your dataset and observe how the regression line changes.

Assignment 2

Use a dataset containing information about houses. Link: https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/data

Plot scatter plots between Price and each independent feature. Comment on any observable trends or correlations.

Fit a linear regression model using only Area (considers all possible areas) as the independent variable.

Plot the regression line on a scatter plot of Area vs. Price.

Print the slope and intercept and interpret them.

Calculate the Mean Squared Error (MSE) and R^2 score of your model. You can use built-in functions.

Comment on how well your model fits the data.

Fit a multiple linear regression model considering other different features.

Print the coefficients and intercept, and interpret their meaning.