LAB ASSIGNMENT (Regression Analysis)

Perform the following analysis in python on the dataset provided.

- Create three well labelled scatterplots of this data with TV, Radio and News paper on the x-axis and Sales on the y-axis, and describe the relationship you see. The scatterplot colour should be red, blue and green respectively. Add suitable labels and title to the plot.
- In the scatterplot you made, what is the explanatory variable? What is the response variable? Why might you want to construct the problem in this way?
- Compute Pearson's correlation coefficient between sales and each of the independent variables. What is your observation?
- Split the data into train (80%) and test (20%) (without shuffling). Fit a simple linear regression model on the train data for the three independent variables separately and assess the accuracy of the model in terms of MSE (train and test). Which independent variable contributes to accurate prediction of Sales?
- Fit multiple linear regression model on the train data for the different possible combinations of the three independent variables and assess the accuracy of the model in terms of MSE (train and test). Which combination contributes to accurate prediction of Sales?
- What is the difference between R^2 and Adjusted R^2? Comment!
- Give your final comments on which model linear or multiple linear is apt for accurate prediction of sales based on MSE values for train and test and R^2, Adjusted-R^2 values.