SMDE FIRST ASSIGNMENT (40% OF THE FINAL MARK)

FIRST QUESTION: REGRESSION ANALYSIS (25% OF THE FIRST ASSIGNMENT).

Use the full laptop prices data set from the web site of kaggle (Laptop Price Prediction Cleaned Dataset). The data set consists of following variables: Company, Type Name, Ram, Weight, Price, Touch Screen indicator, IPS indicator, Ppi, Cpu_brand, HDD, SSD, Gpu_brand, Os (Operation System).

You had already assigned the type of the variables. Now fit a linear regression model to predict price of laptops.

a) Consider the numerical variables in the data set and find the best simple linear regression model to predict the prices (Test the assumptions and use transformations if it is required.) Explain why the model you find is the best simple linear regression model and interpret the coefficients of the model (25p)

b) Fit a multivariate linear regression model with two (numerical) independent variables. Choose the most significant regression model with two predictors. (Transform the variables if it is needed and test all the assumptions.) Then compare this model to the simple linear regression model that you fit in (a). Which one is a better model? Why? (25p)

c) Now add a factor to the regression model you have chosen in section (b). (You can write a loop to add factors one by one to the previous model and decide based on the results.) Interpret the coefficients and overall summary of the model. Test the model in section (b) with the model that has an additional factor. Which one would you choose? Why? (35p)

d) Test the validity of the final model. (15p)