

SMDE FIRST ASSIGNMENT (40% OF THE FINAL MARK)

FIRST QUESTION: VISUALISATION, CHI SQUARE AND T-TEST (25% OF THE FIRST ASSIGNMENT).

Download the laptop prices data set from the web site of kaggle ([Laptop Price Prediction Cleaned Dataset](#)) and read the data set in R.

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).

- a) Import data set to R assigning the type of each variable correctly. (5p)
- b) Create a dataset including only types of laptops: Ultrabook, Notebook and 2 in 1 Convertible. (5p)
- c) Summarize the variables Weight, Price and the categorical variables in the new created data set. (10p)
- d) Cross classify the variables type of computer and the touch screen indicator in a table. Compute and interpret the conditional probability tables. (15p)
- e) Is there an association between the type of computer and the touch screen characteristic of a computer. Analyze it by using proper statistical method. (10 p)
- f) Check the distribution of Price first for all observations then for subgroups of type of laptop in the data set created in section (b). Does it follow a normal distribution? (15p)
- g) Create a data frame just by including Ultrabooks and Notebooks. (5p)
- h) Make a boxplot to show the distribution of Price across two categories of Type: Ultrabook vs. Notebook. Interpret it. (10p)
- i) Compare the average price of Ultrabooks and Notebooks by using the appropriate method. Do not forget to test the assumptions. (25p)