DATA SCIENCE-(BSCS-2) ASSIGNMENT ONE

INSTRUCTIONS [TOTAL: 20 MARKS]

DATA DESCRIPTION

Wholesale customers data.csv (The description of the variables are shown in the sheet "Descriptors" within the excel workbook)

Questions

1. Check to see if all the continuous variables are normally distributed. (3 MARKS)
2. Transform the dataset to exclude any missing information (1 MARK)
3. Show the outliers in only the continuous variable in the dataset (3 MARKS)
4. Transform the dataset and handle the outliers in the continuous variables if any. (3 MARKS)
5. Save the newly transformed dataset as a CSV file called "Group Name.CSV (e.g. Group1.CSV) (1 MARK)
6. Import the above saved CSV file as a Python data frame called "Assignment1_ Group Name (e.g. Assignment1_ Group Name) (1 MARK)
7. Which of the continuous variables generates the highest annual spending? (2 MARKS)
8. Which retail channel do customers use the most? (2 MARKS)
9. Which Region has the lowest purchasing power? (2 MARKS)
10. Which Region spends most on Milk? (2 MARKS)

Submission Instructions

Please submit a Python notebook with the file extension '.ipynb'. Please use the naming convention for your work in this format: Group Name & YEAR_COURSE UNIT NAME (e.g., Group Name_Data Science).