

DAV 5400 Module 4 Assignment (30 Points)

This assignment provides us with our first opportunity to work with the Pandas library. The data set we'll be using is comprised of more than 10,000 observations, including 1 variable which indicates whether or not a customer of a company purchased a new product and 9 additional variables that provide us with some information about customers to whom the company offered that new product. You are given information about whether they did or did not sign up for the new product, together with some customer information and information about their buying behavior of two other products (Product A and Product B). A data dictionary for the dataset is provided below.

Attribute	Description
ID	Unique customer identifier
TARGET	Indicator of customer buying the new product (N = no, Y = yes)
Age	Customer age in years
LOR_m	Length of customer's relationship with company (in months)
Prod_A	Customer previously bought Product A (0=no, 1=yes)
Type_A	Type of product A
Turnover_A	Amount of money customer spent on Product A
Prod_B	Customer previously bought Product B (0=no, 1=yes)
Type_B	Type of product B
Turnover_B	Amount of money customer spent on Product B

For this assignment, you will need to load the data file into your online DAV 5400 GitHub repository and then read the data from your GitHub repository into a Pandas dataframe. You will then use your Python and Pandas skills to answer and complete the following questions and tasks:

1. **(3 Points)** What proportion of customers have an age of more than the average customer age but less than 56?
2. **(3 Points)** How many customers that have purchased Product B have spent more on Product B than 1.5x the median amount that customers typically spend on Product B?
3. **(3 Points)** How many customers have spent more on Product B than they have on Product A?
4. **(3 Points)** How many customers have purchased either type 6 of Product A or type 9 of Product B?
5. **(3 Points)** We are provided with the number of months for which each customer has been a client of the company (**lor_M**). Which customer has attained the highest average amount of monthly spending (i.e., across all products they have purchased from the company) during the time in which they have been a customer of the company? Answer by providing the dataframe row index, customer ID, and the total number of months the customer has been with the company, as well as that customer's total spending amount. Finally, display the complete content of the dataframe row corresponding to the customer you have identified.
6. **(3 Points)** Make bar plots of the '**type_A**' and '**type_B**' data. What do they tell us about the frequencies of the individual type_A and type_B values within the data set?
7. **(4 Points)** Make boxplots of the '**age**' attribute for **each type of Product B** and discuss any conclusions you can draw from the appearance of the boxplots regarding the ages of customers who purchase different types of Product B.
8. **(4 Points)** Make histograms of the **age** and **lor_M** attributes, and boxplots of the **turnover_A**, and **turnover_B** attributes and discuss any conclusions you can draw from the shape of the resulting distributions.

9. **(4 Points)** Make a scatter plot of **turnover_A** vs. **lor_M**. What can we say about the relationship between those two attributes?

Provide a short written narrative that explains your approach for each of the questions and tasks **using formatted Markdown cells in your Jupyter notebook**.

Your Jupyter Notebook deliverable should be similar to that of a publication-quality / professional caliber document and should include clearly labeled graphics, high-quality formatting, clearly defined section and sub-section headers, and be free of spelling and grammar errors. Furthermore, your Python code should include succinct explanatory comments.

Save all of your work for this assignment within **a single Jupyter Notebook** and submit it via the Module 4 Assignment page within Canvas. Be sure to save your Notebook using the following nomenclature : **first initial_last name_M4_assn**" (e.g., J_Smith_M4_assn).