R for Data Science - Assignment

**Problem Statement:**

Samaritan is a retail organization dealing with FMCG products. They have business in 3 Indian cities. They are planning to launch a marketing campaign to increase their customer base.

They have collected various data points of their customers. You work as a Data Analyst in Samaritan and you have been assigned with the task of analyzing the data.

**Data Description:**

|  |  |  |
| --- | --- | --- |
| **S.NO** | **Column Name** | **Description/ Comments** |
| 1 | ID | Customer's unique identifier |
| 2 | Year\_Birth | Customer's birth year |
| 3 | Education | Customer's education level |
| 4 | Marital\_Status | Customer's marital status |
| 5 | Income | Customer's yearly household income |
| 6 | Kidhome | Number of children in customer's household |
| 7 | Teenhome | Number of teenagers in customer's household |
| 8 | Dt\_Customer | Date of customer's enrollment with the company |
| 9 | Recency | Number of days since customer's last purchase |
| 10 | MntWines | Amount spent on wine in last 2 years |
| 11 | MntFruits | Amount spent on fruits in last 2 years |
| 12 | MntMeatProducts | Amount spent on meat in last 2 years |
| 13 | MntFishProducts | Amount spent on fish in last 2 years |
| 14 | MntSweetProducts | Amount spent on sweets in last 2 years |
| 15 | MntGoldProds | Amount spent on gold in last 2 years |
| 16 | Complain | 1 if customer complained in the last 2 years, 0 otherwise |
| 17 | Response | 1 if customer accepted the offer in the last campaign, 0 otherwise |

**Task:**

1. Find the top 3 Married customers based on their gold spending.
2. Find the number of customers who complained in the last 2 year and spent more than 300 on wine.
3. Get the month name from the date of customer joining. Find the total amount spend on meat based on the month of joining.
4. Based on the recency score, amount spent on sweet, and amount spent on fish, predict the income of the customers. What will be the income for a customer whose recency score is 60, amount spent on sweet is 200 and amount spent on meat is 400.
5. Plot a bar chart between the number of teenagers in home and the average amount spent on fish.
6. For the complete data take 25 random samples and find if there is a significant difference between the amount spent on wine and amount spent on meat, for the 25 data points. Frame the null and alternate hypothesis and attain the conclusion.

**Instructions:**

1. Once completed, save the file as R file and upload it in the given location.
2. Perform all the questions in a single R file.
3. Writing the inference is not mandatory, however doing it would fetch the participants extra marks.
4. More than the answers, it is the logical approach that you took to solve the questions matters.