Introduce about the data

We have a data namely **customers.csv** with the data on customers of a store with the following information:

Customer Information

ID: Customer's unique identifier
Year_Birth: Customer's birth year
Education: Customer's education level
Marital_Status: Customer's marital status
Income: Customer's yearly household income
Kidhome: Number of children in customer's household
Dt_Customer: Date of customer's first shopping at the store

Purchase History

MntWines: Amount spent on wine in last 2 years
MntFruits: Amount spent on fruits in last 2 years
MntMeatProducts: Amount spent on meat in last 2 years
MntFishProducts: Amount spent on fish in last 2 years
MntSweetProducts: Amount spent on sweets in last 2 years
MntGoldProds: Amount spent on gold in last 2 years

NumWebPurchases: Number of purchases made through the company's website
NumCatalogPurchases: Number of purchases made using a catalog
NumStorePurchases: Number of purchases made directly in stores
NumWebVisitsMonth: Number of visits to the company's website in the last month
Complain: 1 if the customer complained in the last 2 years, 0 otherwise

Questions

PART A: Some manipulation

First, import the data to R and name the dataframe as **Customer**.

1. From **Year_Birth**, make a new column (with the name **age**) to calculate the age of each customer in 2024.
2. Delete any missing rows on **Income**.
3. If the Marital_Status is "Alone", please change it to "Single". Keep only two statuses "Single" and "Married".
4. From column **Dt_Customer**, make a new column (with the name **relationship**) to calculate how many years he/she has been a customer until now (2024).

Hint: We want to get the year component from Dt_Customer (which is a character variable now in the data). You can use function **dmy()** to make the column into date format then use **year()** to get the year component of a date. Another way is to use **substr()** function to get a part of the string that you want.

5. Make a new column (with the name **NumPurchases**) which is the sum of three columns: NumWebPurchases, NumCatalogPurchases, and NumStorePurchases.

PART B: Test on one variable

6. Some people think that