# Retail Sales Case Study

You have been given a dataset from a leading retail giant. The dataset contains information on:

1. Demographics of the customers.
2. Purchase Behaviour.
3. Response to various marketing campaigns run.

The data dictionary for the variables is given below:

**demographics.txt**

- 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
- Teenhome->Number of teenagers in customer's household
- Dt_Customer->Date of customer's enrollment with the company
- Country->Customer's location

**behaviour.json**

- Recency->Number of days since customer's last purchase
- MntWines->Amount spent on wine in the last 2 years
- MntFruits->Amount spent on fruits in the last 2 years
- MntMeatProducts->Amount spent on meat in the last 2 years
- MntFishProducts->Amount spent on fish in the last 2 years
- MntSweetProducts->Amount spent on sweets in the last 2 years
- MntGoldProducst->Amount spent on gold in the last 2 years
- NumDealsPurchases->Number of purchases made with a discount
- NumWebPurchases->Number of purchases made through the company's web site
- NumCataloguePurchases->Number of purchases made using a catalogue
- NumStorePurchases->Number of purchases made directly in stores
- NumWebVisitsMonth->Number of visits to company's web site in the last month

**campaign.json**

- AcceptedCmp3->1 if customer accepted the offer in the 3rd campaign, 0 otherwise
- AcceptedCmp4->1 if customer accepted the offer in the 4th campaign, 0 otherwise
- AcceptedCmp4->1 if customer accepted the offer in the 5th campaign, 0 otherwise
- AcceptedCmp1->1 if customer accepted the offer in the 1st campaign, 0 otherwise
- AcceptedCmp2->1 if customer accepted the offer in the 2nd campaign, 0 otherwise
- Response->1 if customer accepted the offer in the last campaign, 0 otherwise

- Complain->1 if customer complained in the last 2 years, 0 otherwise

**Tools:**

- Python

**Business Problem**

The retailer wants to understand what kind of customers respond to different campaigns. To arrive at a reasonable answer to the above question, you've been tasked to analyze this dataset. Below are some pointed business questions you are required to answer.

**Data Quality and Check (Task 1)**

1. Create a consolidated view of data by joining the data present in three files.
2. Are there any variables where you will need to clean the raw data, what kind of cleaning will be needed?
3. Create a data quality report after doing the necessary cleaning and joining of the files by:
    - Doing univariates for continuous variables (compute: percentage of missing values, percentage of terms which are zero, mean, 25th, 50th, 75th, 90th and 95th percentile, min and max)
    - Doing univariates for categorical variables (compute:percentage of missing values, number of unique values)
4. Are there any extreme values of variables representing income, amount of money spent on various categories, recency of purchase?

**Business Analysis and Hypothesis (Task 2)**

1. Generate and check hypothesis around Amount Spent on different categories and response rate in different marketing campaigns.
2. Create a funnel analysis showing what percentage of unique customers accept campaign 1,2, 3,..etc
3. Find out how income impacts the amount spent on - Wine - Meat Products - Gold Products - Fish Products
4. Can you test the hypothesis that recent customers complain less in general compared to older customers?
5. Do people who accept the offer in the first campaign also accept in any other campaign?
6. Profile of people who respond vs. who don't.

**Deliverables**

- A well-designed deck outlining the conclusions and the analysis (.ppt)
- A well-structured code pushed on GitHub (Write an informative README, well-structured code/notebooks)
- *Optional*: A blog post on medium/personal blog/blogger/LinkedIn