CUSTOMER SEGMENTATION
DATA ANALYSIS REPORT

I. Introduction

1. Project discription

Each individual is different and so are his preferences. With such a diversity of needs and preferences, how do you serve all of them? Most importantly, how as a business do you know which customers to target and form the right marketing strategies for each of them? In this customer-centric world, Segmentation is your answer. Customer segmentation is the act of dividing a particular company's consumers into groups of consumers who have related needs and characteristics about them. It assists businesses to aim and adapt marketing communication, products and services to fit each group's requirements. Knowledge of customers' specific preferences implies improved satisfaction, higher levels of customers' loyalty and higher sales rates. Thus, a perfect segmentation should be able to support more effective resource management and more engaged consumers, putting the business relatively ahead of its competitors.

In this project, I will be performing an unsupervised machine learning - Clustering on a dataset contains the customer's records from a groceries firm's database. I will divide customers into segments to optimize the significance of each customer to the business. To modify products according to distinct needs and behaviours of the customers. It also helps the business to cater to the concerns of different types of customers.

2. Business task

The main objective of this project is using clustering algorithms to identify distinct groups of customers based on their purchasing behavior, demographics, or other relevant factors to uncover patterns and group customers with similar traits. This enables the business to create targeted marketing strategies, personalize offerings, and optimize customer engagement. By segmenting customers effectively, the company can enhance customer retention, improve product recommendations, and ultimately increase revenue through more efficient resource allocation.

II. Analysis

1. Data overview

This dataset for analysis was downloaded from Kaggle, contains 29 fields in total and 2240 rows. The dataset can be categorized into the following subsets:

Customer's 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
- Teenhome: Number of teenagers in customer's household
- Dt_Customer: Date of customer's enrollment with the company
- Recency: Number of days since customer's last purchase
- Complain: 1 if the customer complained in the last 2 years, 0 otherwise

Products

- 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

Promotion

- NumDealsPurchases: Number of purchases made with a discount
- 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
- 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
- AcceptedCmp5: 1 if customer accepted the offer in the 5th campaign, 0 otherwise
- Response: 1 if customer accepted the offer in the last campaign, 0 otherwise

Place

- NumWebPurchases: Number of purchases made through the company's website
- NumCatalogPurchases: Number of purchases made using a catalogue
- NumStorePurchases: Number of purchases made directly in stores
- NumWebVisitsMonth: Number of visits to company's website in the last month

Link dataset:
https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

2. Tools

This project mainly uses Python programming language with some libraries like Pandas, Numpy, Matplotlib, Seaborn for data import, data cleaning, data analysis, clustering and visualization.

3. Analysis

See full analysis project here: https://github.com/hoan110102/Customer-Segmentation/blob/main/customer_segmentation.ipynb

4. Result

We found 3 different groups of customer:

Group 1 - Average Income & High Spent: They are usually middle-aged people (over 30), mostly couples but most of them do not have children (maybe because they are not married), families mainly have only 1-2 people (maybe the family has children but the parents are divorced)

Group 2 - Low Income & Spent: They are usually younger than the above segment (under 50 years old), almost all have at least 1 child, and most of them are married, so the number of family members is mostly at least 2.

Group 3 - Average Income & Spent: They are usually older than the previous 2 groups (over 40 years old), almost all have only 1 child and most of them are couples (maybe married), so the number of family members is mainly 3 people.