Capstone Project
Customer Segmentation

Context

Customer segmentation is the process of dividing a dataset of customers into groups of similar customers based on certain common characteristics, usually for the purpose of understanding the population dataset in a better fashion. Understanding customer behavior and characteristics is usually a critical part of the marketing operations of any business or organization, with direct consequences on sales & marketing strategy. Customer segmentation is often viewed as a means to achieve a better return on investment from marketing efforts, and make organizations more efficient in terms of utilizing their money, time, and other critical resources in custom marketing strategies for different groups of customers based on their unique needs and motivations.

For example, it has been understood from various research that customer segmentation often has a huge impact on people's email engagement. Segmented campaigns often see over 100% more clicks than non-segmented campaigns, and email marketers who have segmented their audience before campaigning have reported a 6-7 times growth in their overall revenue. It has also been observed in various contexts that in today's world, individual customers prefer personalized communications and offerings that cater to their particular interests.

In the context of marketing analytics then, customer segmentation has a vital role to play in optimizing ROI. It typically involves analyzing metrics around customer engagement with various marketing activities including but not limited to, ATL (above the line) marketing activities, BTL (below the line) campaigns, and targeting personalized offers. Typically, the variables of interest are customer profiles, campaign conversion rates, and information associated with various marketing channels. Based on these feature categories, the target is to create the best possible customer segments from the given data.

Objective

Using Unsupervised Learning ideas such as Dimensionality Reduction and Clustering, the objective is to come up with the best possible customer segments using the given customer dataset.

Data Dictionary

The dataset contains the following features:

- ID: Unique ID of each customer
- Year_Birth: Customer's year of birth
- Education: Customer's level of education
- Marital_Status: Customer's marital status
- Kidhome: Number of small children in customer's household
- Teenhome: Number of teenagers in customer's household
- Income: Customer's yearly household income
- Recency: Number of days since the last purchase
- Dt_Customer: Date of customer's enrollment with the company
- MntFishProducts: The amount spent on fish products in the last 2 years
- MntMeatProducts: The amount spent on meat products in the last 2 years
- MntFruits: The amount spent on fruits products in the last 2 years
- MntSweetProducts: Amount spent on sweet products in the last 2 years
- MntWines: The amount spent on wine products in the last 2 years
- MntGoldProds: The amount spent on gold products in the last 2 years
- NumDealsPurchases: Number of purchases made with discount
- NumCatalogPurchases: Number of purchases made using a catalog (buying goods to be shipped through the mail)
- NumStorePurchases: Number of purchases made directly in stores
- NumWebPurchases: Number of purchases made through the company's website
- NumWebVisitsMonth: Number of visits to the company's website in the last month
- AcceptedCmp1: 1 if customer accepted the offer in the first campaign, 0 otherwise
- AcceptedCmp2: 1 if customer accepted the offer in the second campaign, 0 otherwise
- AcceptedCmp3: 1 if customer accepted the offer in the third campaign, 0 otherwise
- AcceptedCmp4: 1 if customer accepted the offer in the fourth campaign, 0 otherwise
- AcceptedCmp5: 1 if customer accepted the offer in the fifth campaign, 0 otherwise
- Response: 1 if customer accepted the offer in the last campaign, 0 otherwise
- Complain: 1 If the customer complained in the last 2 years, 0 otherwise