Description

Problem Statement: Customer segmentation is an important technique for developing marketing strategies. It helps in better targeting and positioning customers with the company's offerings. We can perform customer personality analysis by observing their demographics and preferences.

A detailed analysis of a company's preferred customers is known as a customer personality analysis. It enables a company to better understand its customers and makes it easier to modify products or services to meet the specific needs, behaviors, and demands of several types of customers.

Instead of spending money marketing a new product to every customer in a company's database, a company can analyze which customer segment is most likely to buy the product and then market the product only to that specific segment.

Objective:

Perform customer segmentation to better understand customer personality types

Dataset Description: Marketing Campaign

Variables Description

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 enrolment 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

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

NumDealsPurchases Number of purchases made with a discount

AcceptedCmp1 1, if the customer accepted the offer in the 1st campaign, 0 otherwise

AcceptedCmp2 1, if the customer accepted the offer in the 2nd campaign, 0 otherwise

AcceptedCmp3 1, if the customer accepted the offer in the 3rd campaign, 0 otherwise

AcceptedCmp4 1, if the customer accepted the offer in the 4th campaign, 0 otherwise

AcceptedCmp5 1, if the customer accepted the offer in the 5th campaign, 0 otherwise

Response 1, if the customer accepted the offer in the last campaign, 0 otherwise

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

Z-CostContact Category value for cost to contact

Z-Revenue Category value for revenue

Note:

Download the “Marketing\_campaign.csv” using the link given in the business strategy - customer segmentation project statement.

# Steps to perform:

1. Import the input file. Check if the variable with date values has been imported appropriately
2. Find the variables with missing values. If the proportion of missing values is less than 5%, then delete the rows
3. Calculate the newest and oldest customer's enrolment date in the records
4. Create a feature "Customer\_For" of the number of days the customers started to shop in the store relative to the last recorded date
5. Find the "Age" of customers by the "Year\_Birth" indicating the birth year
6. Create a feature "Spent" indicating the total amount spent by the customer in various categories over two years
7. Create another feature "Living\_With" out of "Marital\_Status" to extract the living situation of couples. Consider a value ‘Partner’ for the variable "Living\_With" for the instances where "Marital\_Status" is either “Married” or “Together”. Rest all can be taken as ‘Alone’
8. Create a feature "Children" to indicate the total number of children in a household that is, kids and teenagers
9. To get further clarity on a household, create a feature indicating "Family\_Size"
10. Create a feature "Is\_Parent" to indicate the parenthood status
11. Keep only two categories in the field – ‘Education’ – Undergraduate, Graduate
12. For clarity, change the name of the variables as:

MntWines Wines

MntFruits Fruits

MntMeatProducts MeatProducts

MntFishProducts FishProducts

MntSweetProducts SweetsProducts

MntGoldProds GoldProds

1. Drop the redundant columns: "Marital\_Status", "Dt\_Customer", "Z\_CostContact", "Z\_Revenue", "Year\_Birth", and "ID"
2. Create box plots and histograms for age and income. Identify the outliers and delete rows with outliers
3. Find out the correlation between variables. Create a heatmap to visualize the correlation plot
4. Now you should prepare the data for cluster analysis and Categorical variables must be incorporated in clustering. Perform necessary encoding techniques to transform the categorical variables
5. Find the appropriate number of clusters using the elbow method
6. Perform cluster analysis using k-means
7. After a cluster analysis is performed, it is important for the business to define the clusters. Perform cluster profiling, describe each cluster in detail, and use appropriate visualizations to support your views