

Project 1: Customer Segmentation via Cluster Analysis

Data Science and Machine Learning for Engineering Applications

Politecnico di Torino

Introduction

Customer behavior analysis is a comprehensive examination of an organization's ideal customers. It enables a company to gain a deeper understanding of its customer base. It facilitates the customization of products or services to meet the specific needs, behaviors, and concerns of different customer segments. This analysis plays a crucial role in helping companies refine their strategies and enhance their overall customer experience.

Objective

Perform clustering to identify and summarize various customer segments. Then, characterize the clusters of customers found by analyzing the characteristics of the various clusters through distributions and statistics of features or association rules on the customers of each group.

Dataset and Task

Task

The dataset consists of a single CSV file where no ground-truth labels are provided since it is an unsupervised learning task.

Dataset Attributes

The following attributes are present in the dataset:

People

- 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

Reference

The original dataset can be found in the following URL: <https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis>

Note

The dataset used in the project is a re-sampling of the original dataset (i.e., it is a modified version of the original dataset). The data collection is the result of manual collection, so it is necessary to pay attention to possible errors, missing values, etc.