

Artificial Intelligence for Communication
and Marketing

CLUSTER ANALYSIS MODEL.

for multichannel retail company

Francesca Sotgia - 513067

Business Context and Objectives

Business Context and Objectives

Business Context

A multichannel retail company regularly runs **targeted promotional campaigns**.

Due to a lack of data-driven foundation, the response rate of the marketing campaigns remained low, meaning that most customers didn't accept the offers presented in the campaigns.

To deal with this problem, the CMO of the company wants **segment the customers** with the objective of identify groups of customers with similar behaviors to provide better **personalized communications** and **maximize the ROI** of the marketing campaigns.

Project Objectives

The aim of the project is to segment the customers in different clusters based on:

- Demographic attributes
- Behavioral patterns
- Historical campaign response

Then, analyze the segments to propose marketing strategies that will result in personalized communications and in a maximization of the Return of Investment (ROI).

Dataset description

Dataset description & Variables

The original dataset contains 2240 entries and 29 features. Each entry represents a customer of the company, who is described in terms of:

- **Demographic:** Description of the customer and his household
 - *Year_Birth, Education, Marital_Status, Income, Kidhome, Teenhome*
- **Purchase behaviors:** Amount spent in different categories of products in the last 2 years
 - *MntWines, MntFruit, MntMeatProducts, MntFishProducts, MntSweetProducts, MntGoldProducts*
- **Purchase channels:** Number of purchases made using different channels
 - *NumWebPurchases, NumCatalogPurchases, NumStorePurchases, NumWebVisitsMonth*
- **Campaign response:** Whether the customer accepted or note the offers in the last campaigns
 - *NumDealsPurchase, AcceptedCmp1-5, Response*

Dataset description

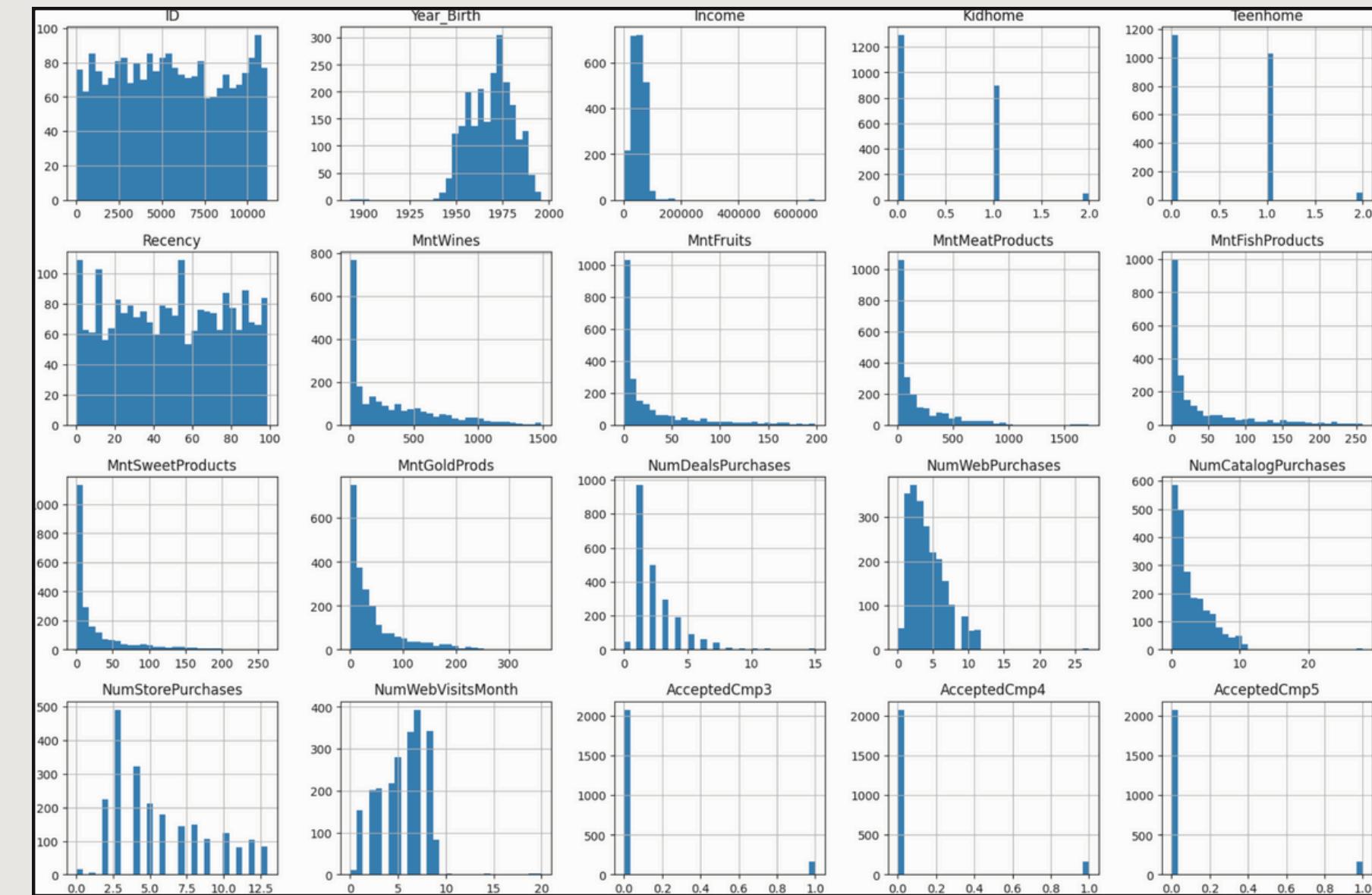
Dataset description & Variables

Once the dataset is imported, data exploration is performed.

It consists of:

- Inspection of **missing values**
- Extraction of **unique values** and their count for each feature
- Analysis of the **distribution of the features**

This is a crucial step that enables a **deeper understanding of the data**, needed in later steps.



Analytical methods for model development

Processing

Once the dataset is imported and explored, preprocessing is implemented to have consistent and more meaningful data.

- **Extraction of year of enrollment** from '*Dt_Customer*'
- Adjustment of '*Marital_Status*' labels to have more **meaningful labels**
- **Drop unnecessary columns**: '*Z_CostContact*' and '*Z_Revenue*' because they have constant value and are not informative
- Handling of **missing values**: fill the missing values in '*Income*' with mean imputation
- Conversion of **categorical features** into numerical features: use *LabelEncoder* to encode categorical labels with value between 0 and number_classes - 1
- **Scaling** of the data: normalize the features to be in range [0, 1] for consistency purposes with *MinMaxScaler*

Analytical methods for model development

Cluster analysis with K-means

Perform cluster analysis to **segment the customers** based on their demographic attributes, behavioral patterns and historical response data.

Cluster analysis is implemented by a **K-means algorithm**: partitioning methods that iteratively reallocates the customers in the clusters in order to improve their representativeness.

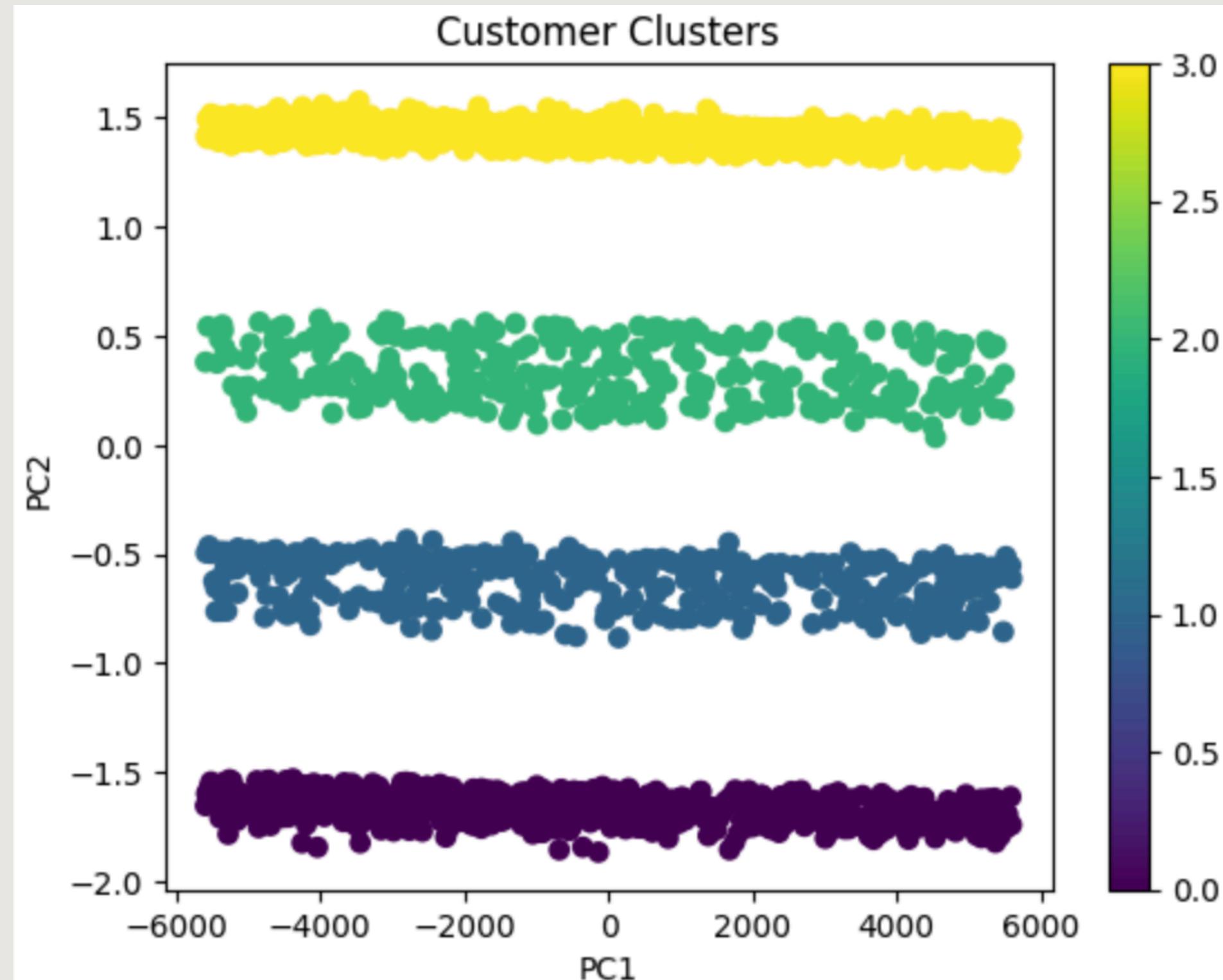
Start by selecting the optimal number of clusters through:

- **Elbow Method**: how the SSE varies across the number of clusters.
- **Silhouette Score**: how much a customer is similar to the others present in its cluster.

Once the best number of clusters is identified (in this case K=4), tune the K-means algorithm with that value. Through this algorithm each customer is assigned to a cluster based on its characteristics.

Analytical methods for model development

Clusters visualized through PCA



To visualize the clusters, **Principal Components Analysis** (PCA) is applied. PCA is a Dimensionality Reduction technique that enables the visualization of the clusters in a 2-dimensional plot.

The four clusters are **distinguishable** and **compact**, suggesting a good performance of the algorithm.

Business Insights

Business Insights

Analyze the characteristics of each cluster to implement data-driven marketing strategies.

Cluster 0 - Mature Wealthy Customers

Older, married customers with high income (~68K). They have high spending with store, web and catalog purchases. Not responsive to last campaigns.

Goal: Increase loyalty and purchase of luxury products.

Strategy: premium loyalty programs, personalized catalogs with high-end products, promote bundled packages with exclusive products.

Cluster 1 - Young mid-income single customers

Single customers with lower income (~49K) and product spending. They prefer to shop in store and they have moderate number of website visits. Not responsive to last campaigns and lower number of purchases.

Goal: Increase frequency of purchase and up-selling.

Strategy: Digital discounts and website promotions. offer monthly promotions on essential products bundled with more expensive items.

Business Insights

Cluster 2 - Mature stable customers

Single customers with moderate income (~59K) and spending. Purchases balanced across channels. Low number of children at home. They accepted offers in last marketing campaign.

Goal: Increase up-selling and cross-selling.

Strategy: cross-channels discounts, personalized product recommendations and middle-range bundle offers with more expensive products.

Cluster 3 - Young limited-budget customers

Younger married customers with higher number of children at home. They have lowest income (~38K) and product spending. They prefers in store and online purchases and have highest number of website visits. Not responsive to last campaigns.

Goal: Increase frequency of purchase.

Strategy: Highly discounted promotions for essential products and offer bundle packages for young families. Focus on online promotions, family loyalty programs and seasonal promotions

Practical Implications

Increased personalization

Customers in the clusters receives promotions and communications tailored to their needs and behaviors.

Increase engagement

Offering loyalty programs and personalized offers helps achieve higher engagement and purchases.

Maximize ROI

Understanding how the customers in the clusters respond to marketing campaigns helps the company to reallocate resources to achieve a higher ROI.