

# Performing Psychographic Segmentation Based on Customer Behaviour

Priyatham Sai Chand Bazaru

Computer Science Department

California State Polytechnic University

Pomona

pbazaru@cpp.edu

Ritika

Computer Science Department

California State Polytechnic University

Pomona

nritika@cpp.edu

Siva Charan Mallena

Computer Science Department

California State Polytechnic University

Pomona

smallena@cpp.edu

**Abstract**—Customer personality is assessed to find the ideal customer for a business. This is analyzed by using the dataset that contains various attributes pertaining to an individual and what makes him an ideal customer. This paper offers a solution how to find such customers by analyzing various aspects with respect to an individual. Customers are clustered into various groups based on attributes and recognized how they can be targeted. Various metrics and types of clustering were used in order to identify such clusters and how the customers in the data can be fit accordingly. It helps a business to modify its product based on its target customers from different types of customer segments.

**Index Terms**—attributes, ideal, clustered, metrics, segments

## I. INTRODUCTION

In order to stay competitive in the open market, businesses today require them to understand customers and their behaviors thoroughly so they can produce even better products and tailor experiences. This requires analysis of customer's personality and concerns of different types of customers. Using Data Science, Artificial Intelligence and Machine Learning, Individuals can be classified based on their personality traits. Every human being is unique in their own right. Due to multi-modal data capture of users' online activities, a huge corpus of data enabled increased effectiveness of marketing campaigns on a per individual basis like click through rate, resulting in more revenue. The personalized marketing messages and communications are shown to be highly effective and increase not only the demand of products, but also their usage and customer satisfaction. Since recommender systems also depend heavily on knowing each customer's personality, this analysis can be of use in such systems like product recommendations or music recommendations. The predictive relationship of a large and comprehensive set of personal descriptors to aspects of product and brand use is examined. The descriptors comprise demographic and general psychographic variables frequently used in segmentation studies and studies of consumer purchase behavior. The evidence is overwhelming that the covariates are related to brand use in an identical way for all brands, indicating that they are not useful for predicting relative brand preference. The covariates are shown to be predictive of product use. Discussion of the explanatory content of the variables is offered. [1]

## II. DATASET TO BE USED

The dataset consists of 2240 instances upon which clustering can be performed. Based on the given data, Clustering of clients in dataset, we will define the segments of the clients by using 4 equally weighted customer segments such as:

- **Stars:** Old customers with high income and high spending nature
- **Need Attention:** New customers with below-average income and low spending nature.
- **High potential:** new customers with high income and high spending nature.
- **Leaky bucket:** old customers with below-average income and a low spending nature.

The dataset defined the following attributes to identify and segment the customers:

- **ID:** Customer's unique identifier
- **YearBirth:** Customer's birth year
- **Education:** Customer's education level
- **MaritalStatus:** 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
- **DtCustomer:** 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

#### Target

Need to perform clustering to summarize customer segments.

### III. METHODOLOGY

In this process we will separate customers based on common characteristics such as demographics, behaviors or purchasing habits to get a better understanding of the market more effectively. We will go over the data in detail like which attributes are more important or discard any attributes not necessary for the project. We will also check for null/not null values instances. In clustering models these unsupervised learning methods we will draw references from datasets consisting of input data without labeled responses. To look at the clustering of clients in the dataset, we will define the segments of the clients. Based on spending of customers, segments are created like non-buyer, low-buyer, frequent-buyer, biggest-buyer.

Various Clustering algorithms could be used in order to create multiple clusters based on the different attributes that are present to better segment the customers. Algorithms such as the K-means as it is simple and very fast, so in many practical applications, the method is proved to be a very effective way that can produce good clustering results. But it is very suitable for producing globular clusters [2], Agglomerative clustering since this is a pair-group method: at each iteration exactly two clusters are agglomerated into a single cluster [3]. This can be helpful in our case with multiple attributes where many clusters can be formed. To get a better understanding of the data and produce a more generalized view with fewer clusters is possible with this method, spectral clustering because Spectral Clustering solely related to the number of data points, but has nothing to do with the dimension [4] could be potentially used after pre-processing has been performed on the given data. It is north-worthy that all these algorithms requires number of clusters be defined beforehand. It is safe to assume that the number of clusters is above 2 but not more than 8 to be computationally capable to train the data.



Fig. 1. Methodology flow chart diagram

### REFERENCES

- [1] Fennell, G., Allenby, G.M., Yang, S. et al. The Effectiveness of Demographic and Psychographic Variables for Explaining Brand and Product Category Use. *Quantitative Marketing and Economics* 1, 223–244 (2003). <https://doi.org/10.1023/A:1024686630821>
- [2] S. Na, L. Xumin and G. Yong, "Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm," 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, 2010, pp. 63-67, doi: 10.1109/IITSI.2010.74.
- [3] Day, W.H.E., Edelsbrunner, H. Efficient algorithms for agglomerative hierarchical clustering methods. *Journal of Classification* 1, 7–24 (1984). <https://doi.org/10.1007/BF01890115>
- [4] Shifei Ding, L. Zhang and Y. Zhang, "Research on Spectral Clustering algorithms and prospects," 2010 2nd International Conference on Computer Engineering and Technology, 2010, pp. V6-149-V6-153, doi: 10.1109/ICET.2010.5486345.