

“Data Analysis of Customer Segmentation”

¹Sakshi Balpande¹, Janvi Kolhe², Aparna Neware³, Pramila Gharjale⁴

^{1,2,3}Student, ⁴Teacher
Computer Science & Engineering

Abstract:

In today's highly competitive business landscape, understanding customers' needs, preferences, and behaviours has become paramount for businesses aiming to stay ahead. Traditional, one-size-fits-all marketing strategies are no longer sufficient to meet the diverse and evolving demands of modern consumers. This realization has led to the widespread adoption of customer segmentation techniques, allowing businesses to divide their customer base into distinct groups with similar characteristics and needs. Amidst this paradigm shift, the concept of customer segmentation has emerged as a powerful tool, enabling businesses to delve deeper into the diverse mosaic of their customer base and unearth invaluable insights.

Customer segmentation, a vital component of marketing analytics, enables businesses to tailor their products, services, and marketing efforts to specific customer segments, thereby enhancing customer satisfaction, loyalty, and ultimately, profitability. By identifying and targeting the right customer segments, businesses can optimize resource allocation, improve customer engagement, and drive revenue growth.

Customer segmentation represents a departure from the traditional, monolithic approach to marketing, where broad strokes were used to target entire markets. Instead, it advocates for the disaggregation of customer populations into distinct groups, or segments, based on shared characteristics, behaviors, and needs. By segmenting customers, businesses can unlock a treasure trove of actionable insights, allowing for more targeted and personalized marketing efforts, product offerings, and customer experiences.

Keywords: Customer Segmentation, Machine Learning, Python, K-Means Clustering, Data Analytics, Data Visualization.

1. INTRODUCTION

Customer segmentation is the process of dividing a business's customer base into distinct groups based on specific characteristics. These characteristics can include demographics, geographic location, psychographic and Behavior: -

- Demographics: Age, gender, income, location, education, etc.
- Geographics: Location, region, climate, etc.
- Psychographics: Lifestyle, values, interests, attitudes, etc.
- Behavioral: Purchase history, website activity, engagement, etc.

Customer segmentation is a vital technique in modern marketing that involves dividing a customer base into distinct groups based on shared characteristics and behaviors. This project aims to implement customer segmentation using machine learning algorithms, primarily K-Means Clustering, to identify meaningful customer segments. By analyzing variables such as annual income, spending score, age, and purchasing patterns, businesses can better understand their customers and tailor marketing strategies accordingly. The project leverages data visualization techniques to interpret the clusters and provides actionable insights to enhance customer engagement, optimize product offerings, and improve overall business performance. The project customer segmentation focuses to support data-driven marketing decisions. By leveraging machine learning techniques on customer demographic and behavioral data, the study identifies distinct segments that share similar traits. The K-Means clustering algorithm is used to group customers based on features like age, income, and spending score. The insights gained allow businesses to personalize marketing campaigns,

improve customer satisfaction. Nowadays Customer segmentation became very popular method for dividing company's customers for retaining customers and making profit out of them.

Customer segmentation involves grouping customers based on shared characteristics to tailor marketing efforts and improve customer experiences, ultimately leading to more effective and personalized interactions.

OBJECTIVE OF RESEARCH:-

1. Personalized Marketing :- To tailor marketing messages and campaigns to specific customer groups, increasing relevance and engagement.
2. Improved Customer Retention :- To identify high-risk customers and create strategies to retain them by understanding their behaviors and preferences.
3. Enhanced Customer Experience :- To provide more relevant products, services, and support to each segment, leading to better customer satisfaction.
4. Product Development :- To understand customer needs and preferences, guiding the design or improvement of products and services.
5. Efficient Resource Allocation :- To focus marketing and sales efforts on high-value or high-potential segments, optimizing time and budget.

2. LITERATURE REVIEW

1. Gong and Ardesiri 2020 [1] describes, with lower prices and more accessible options, the EV manufacturer can tap into a larger segment of the market. This includes customers who are concerned about the environment but prioritize the price more highly. The goal is to boost the volume of sales and to increase market share by making EVs more affordable and appealing to a wider audience.

2. Carvache – Franco 2020 [2] explains markets can be segmented in various ways, including geographically, demographically (e.g., age, gender, income), behaviorally (e.g., prior product experience, frequency of purchase), and psychographically (e.g., lifestyle, beliefs, interests). Different types of market segmentations are Socio-demographic characteristics such as age, gender, income, and education, are additional criteria for segmentation. As with geographic segmentation, socio-demographic categories can easily be determined for every consumer. In some instances, the socio-demographic criterion may also explain specific product preferences. However, in many instances, the socio-demographic criterion is not the cause for product preferences and does not provide sufficient market insight for optimal segmentation decisions.

3. Bloem 2020 [3] states the third group of criteria are psychographic characteristics, which include beliefs, interests, preferences, aspirations, or benefits sought when purchasing a product.

4. Asadi 2021 [4] describes the land transportation industry has undergone significant developments during the past several decades. Electric vehicles (EV) are an innovative, automotive technology that are fully electricity-powered via an onboard battery pack.

5. Huang 2021 [5] explains due to their potential to improve the effectiveness of energy utilization, energy security, and reduced greenhouse gas emission, EVs are becoming popular among customers and policymakers.

6. Policarpo and Aguiar 2020 [6] who prefer environmentally friendly products are considered to be of a higher social status and, in our particular analysis, share the characteristics of altruists who are willing to pay a higher price for a product that sometimes underperforms compared to its conventional counterparts but is better for society as a whole.

7. Özkan and Evrim 2020 [7] loyal customers with low price sensitivity, who demand a high-quality, longer-lasting, sophisticated product, and (2) deal-prone consumers, who are sensitive to price changes and prefer a basic, shorter-lasting, lower quality product.

3. METHODOLOGY

Segmentation Algorithm:- In our project we implemented segmentation algorithm k-means clustering. K-Means Clustering is one of the most popular unsupervised machine learning algorithms used to identify and group similar data points into clusters. The goal is to divide a dataset into K distinct, non-overlapping clusters, where each data point belongs to the cluster with the nearest mean (also called the centroid). K-Means is widely used due to its simplicity and efficiency, especially for large datasets where labeling is not available. However, choosing the right value of K and handling outliers are crucial for its success.

Problem Definition:-

The Main Objective of customer segmentation is to better understand and serve customers by tailoring marketing strategies, products, and services to meet the specific needs of each group. By identifying these segments, companies can create targeted campaigns, improve customer satisfaction, and ultimately drive growth and profitability.

Data Collection:-

- **Demographic Segmentation:-** Grouping customers based on demographic variables such as age, gender, income, etc.
- **Behavioral Segmentation:-** Segmenting based on purchasing behavior, usage patterns, brand interactions, etc.
- **Psychographic Segmentation:-** Segmenting based on lifestyle, values, attitudes, interests, etc.
- **RFM Analysis:-** Analyzing customers based on Recency, Frequency, and Monetary value of their transactions.
- **Cluster Analysis:-** Using statistical techniques to identify naturally occurring clusters within the data.
- **Machine Learning Algorithms:-** Utilizing algorithms like k-means clustering, decision trees, or neural networks to automatically identify segments.

Data Preprocessing:- In data preprocessing, two steps are done initially Data Cleaning and Preparation is done. Cleansing the data to remove duplicates, errors and inconsistencies, and then prepare the data for analysis in a structured format.

Secondly, Data Preprocessing- it preprocess the data by cleaning, transforming, and normalizing for analysis.

Database:- In our project, Data Analysis of Customer Segmentation, an SQLite database is used to store and managed structured customer data like: Customer details (name, age, gender, income), Purchase history, Demographics.

Programming Language:- We have used Python programming language. Python is very useful in data analysis and machine learning. Also, K-Means Clustering Algorithm is used.

Libraries/Framework:- pandas, numPy, scikit learn, matplotlib or seaborn.

4. RESULT

The work described in this notebook is based on a database providing details on purchases made on an E-commerce platform over a period of one year. Each entry in the dataset describes the purchase of a product, by a particular customer and at a given date. In total, approximately ~4000 clients appear in the database. Given the available information,

I decided to develop a classifier that allows to anticipate the type of purchase that a customer will make, as well as the number of visits that he will make during a year, and this from its first visit to the E-commerce site.

The first stage of this work consisted in describing the different products sold by the site, which was the subject of a first classification. There, I grouped the different products into 5 main categories of goods. In a second step, I performed a classification of the customers by analyzing their consumption habits over a period of 10 months.

I have classified clients into 11 major categories based on the type of products they usually buy, the number of visits they make and the amount they spent during the 10 months. Once these categories established, I finally trained several classifiers whose objective is to be able to classify consumers in one of these 11 categories and this from their first purchase. For this, the classifier is based on 5 variables which are:

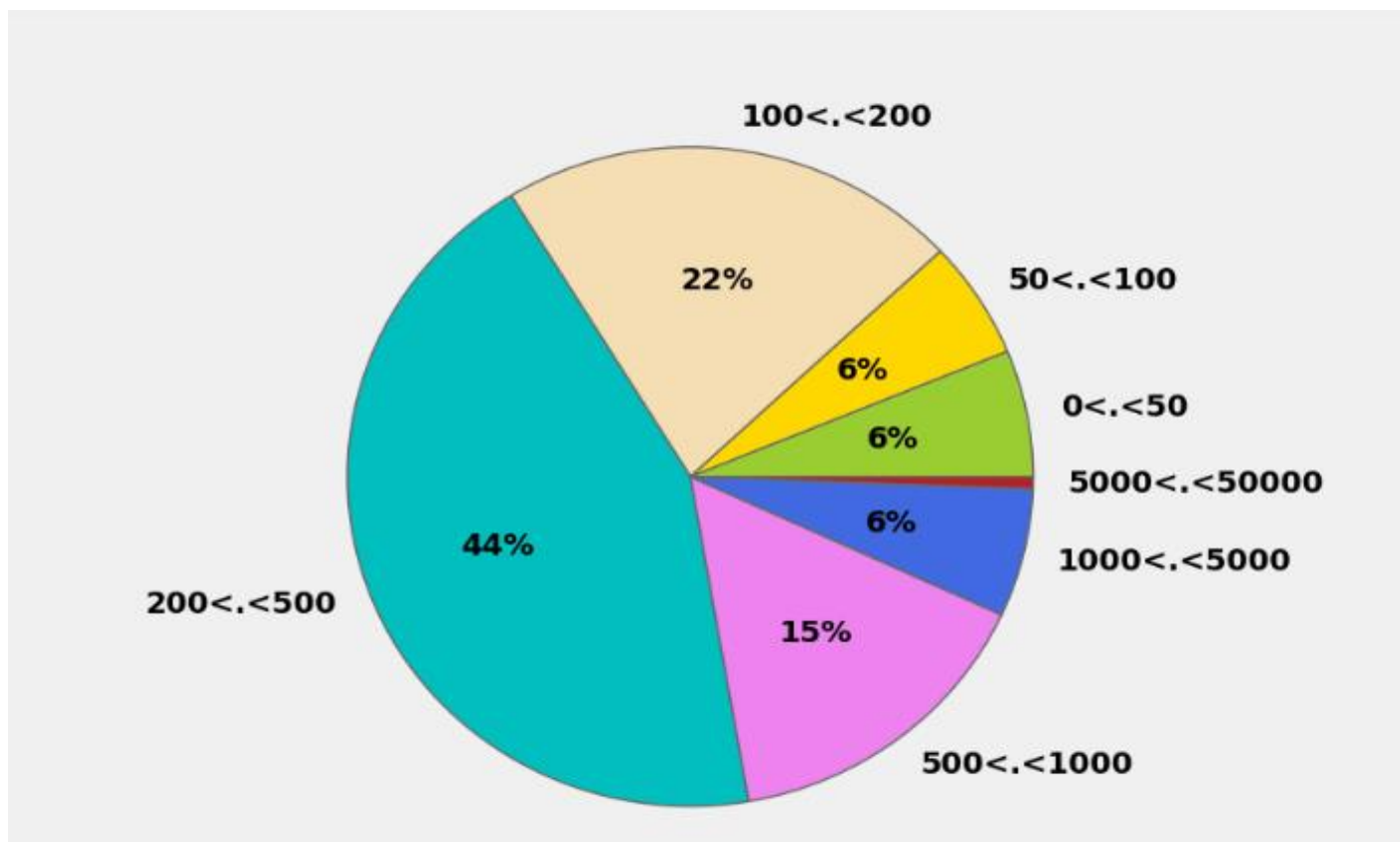
- **Mean** : amount of the basket of the current purchase
- **Categ_N** with $N \in [0:4]$ $N \in [0:4]$: percentage spent in product category with index N

Finally, the quality of the predictions of the different classifiers was tested over the last two months of the dataset.

The data were then processed in two steps: first, all the data was considered (over the 2 months) to define the category to which each client belongs, and then, the classifier predictions were compared with this category assignment. I then found that 75% of clients are awarded the right classes.

The performance of the classifier therefore seems correct given the potential shortcomings of the current model. In particular, a bias that has not been dealt with concerns the seasonality of purchases and the fact that purchasing habits will potentially depend on the time of year (for example, Christmas).

In practice, this seasonal effect may cause the categories defined over a 10-month period to be quite different from those extrapolated from the last two months. In order to correct such bias, it would be beneficial to have data that would cover a longer period of time.



It can be seen that the vast majority of orders concern relatively large purchases given that $\sim 65\%$ of purchases give prizes in excess of £ 200.

5. CONCLUSION

The customer segmentation project successfully identified distinct groups within our customer base using unsupervised machine learning techniques, primarily K-Means clustering. Through analysis of key features such as annual income, spending score, and other behavioral and demographic factors, we categorized customers into clearly defined segments.

These insights enable more targeted marketing strategies, personalized customer experiences, and optimized resource allocation. For example, high-value customers with high spending scores can be prioritized for loyalty programs, while low-engagement groups may benefit from re-engagement campaigns. Overall, this segmentation provides a data-driven foundation for improving customer retention, enhancing satisfaction, and ultimately driving business growth. Future improvements can include dynamic segmentation based on real-time data and incorporating additional variables like online behavior or purchase history. Customer segmentation helps businesses gain a deeper understanding of their customer base by identifying distinct groups with shared characteristics, behaviors, and needs. Customer segmentation is a vital strategy for businesses to understand their customers better, personalize marketing efforts, and ultimately drive growth and profitability by tailoring products, services, and communications to specific groups.

REFERENCES:

1. Gong, S.; Ardeshiri, A.; Rashidi, T.H. Impact of government incentives on the market penetration of electric vehicles in Australia. *Transp. Res. Part D Transp. Environ.* 2020.
2. Carvache-Franco, M.; Carvache-Franco, W.; Carvache-Franco, O.; Hernández-Lara, A.B.; Buele, C.V. Segmentation, motivation, and sociodemographic aspects of tourist demand in a coastal marine destination: A case study in Manta (Ecuador). *Curr. Issues Tour.* 2020.
3. Bloem, S.; Stalpers, J.; Groenland, E.A.G.; van Montfort, K.; van Raaij, W.F.; de Rooij, K. Segmentation of health-care consumers: Psychological determinants of subjective health and other person-related variables. *BMC Health Serv. Res.* 2020.
4. Asadi, S.; Nilashi, M.; Samad, S.; Abdullah, R.; Mahmoud, M.; Alkinani, M.H.; Yadegaridehkordi, E. Factors impacting consumers' intention toward adoption of electric vehicles in Malaysia. *J. Clean. Prod.* 2021.
5. Huang, Y.; Qian, L.; Soopramanien, D.; Tyfield, D. Buy, lease, or share? Consumer preferences for innovative business models in the market for electric vehicles. *Technol. Forecast. Soc. Change* 2021.
6. Policarpo, M.C.; Aguiar, E.C. How self-expressive benefits relate to buying a hybrid car as a green product. *J. Clean. Prod.* 2020.
7. Özkan, P.E.; Evrim, D. The mediating role of price sensitivity in the effect of trust and loyalty to luxury brands on the brand preference 2020.