



Market Segmentation: Understanding The Behavior Of Customers And Launch A Targeted Market Campaign

¹ Dr.V.Sreenivas, ² Manoj Tilakchand, ³ Adi Vikram Dileep, ⁴ Appikatla Bhargav Charan,

⁵ Chippala Devi Chand

¹Associate Professor(Guide) , ² Student, ³ Student,⁴ Student, ⁵ Student

Department of CSE-Data Science,

SRK Institute of Technology, Vijayawada, India

Abstract: The project aims to enhance sales by pinpointing customers who successfully make purchases within their budget, whether through installments or upfront payments, and then launching targeted marketing campaigns toward this specific segment. A dataset has been gathered containing various customer attributes such as ID, balance, purchase behavior, installment plans, cash advances, credit limits, payments, and tenure. By analyzing these features among existing customers, the marketing strategy can be tailored effectively to increase company sales by focusing on specific clusters of customers who demonstrate responsible spending behavior within their financial means. Finally, an analysis of each cluster's characteristics is conducted to gain insights into the distinct segments of customers. This method allows companies to customize their marketing plans and offerings to specific customer groups, thereby enhancing customer satisfaction and maximizing profitability.

Index Terms – Market Segmentation, Customer Segmentation, K-Means Clustering,PCA.

I. INTRODUCTION

In today's tough business world, it's really important to have good marketing plans to stay ahead and do well. A big part of this is understanding who your customers are and how to reach them effectively. In this project, we'll start by looking at a bunch of information about customers. Then, we'll organize this data and use a special method called K-Means to group customers into different categories based on things like how much they spend and how often they buy. Once we've sorted customers into groups, companies can use this info to make targeted ads and promotions. They can also focus on groups with fewer customers to try to get more sales from them. We'll also use a method called Principal Component Analysis to make sure our grouping is accurate. This project aims to help businesses understand their customers better and use strategies to increase sales.

K-means Clustering:

K-Means clustering is a technique used in data analysis to group similar data points together. Imagine you have a bunch of points on a graph, and you want to organize these points into different 'clusters' based on how close they are to each other. K-Means does this by first guessing where the center of these clusters could be. It then assigns each point to the nearest cluster center. After that, it recalculates the centers based on the points assigned to them. This process repeats until the centers don't move much anymore. The 'K' in K-Means stands for the number of clusters you want to create. This method is great because it's simple and fast, but it works best when the clusters are distinct and well separated. If the data is too mixed up or there's a lot of noise, K-Means might not be the best choice

PCA:

PCA is a statistical technique used for dimensionality reduction and Data Visualization. It works by transforming high-dimensional data into a lower-dimensional space, while still preserving the most important information in the dataset. PCA achieves this by identifying the directions, or principal components, along which the data varies the most. These principal components are orthogonal to each other, meaning they are uncorrelated, and they capture the maximum

variance in the data. By retaining only the most significant principal components, PCA reduces the complexity of the dataset while minimizing information loss. This process helps in simplifying data analysis and interpretation, as well as in identifying patterns and trends within the data. PCA is widely applied in various fields, including finance, image processing, and genetics, to uncover underlying structures and relationships within large datasets.

Elbow Method:

Elbow method is a strategy used in data analysis to find the best number of clusters for K-Means clustering. It's like trying to find the right balance between not having too many or too few groups. When you plot the sum of squared distances of points from their cluster centers against the number of clusters, it forms a curve. The point where the rate of decrease sharply changes, like an elbow on an arm, is considered the optimal number of clusters.

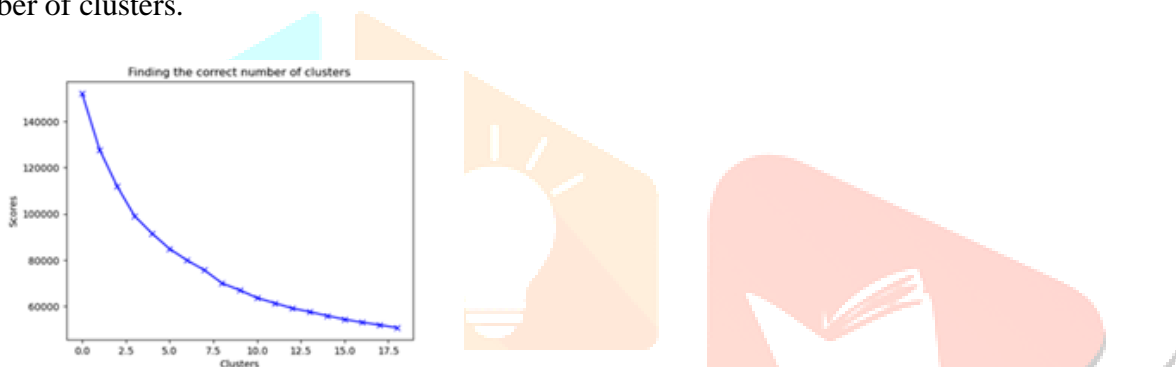


Fig. 1: Graph for Values of K VS WCSS

From the above graph it is clearly be seen that from number of cluster = 7.5 to number of cluster = 10 there has been substantial decrease hence, we choose the K value for our dataset as 8.

II. RESEARCH METHODOLOGY

Customer Segmentation:

Dhiraj et al. [1] states that Customer segmentation is simply means grouping your customers according to various characteristics (for example grouping customers by age).

Rudrendu Paul et al. [2] Segment customers based on buying behavior by applying k-means clustering algorithm to calculate the optimal number of customer segments with similar buying habits (features).

He says there are different methods for customer segmentation. They are:

1)Geographic Segmentation: customer segmentation is very simple, it is all about the user's location. This can be implemented in various ways. You can group by country, state, city, or zip code.

2)Demographic Segmentation: Demographic segmentation is related to the structure, size, and movements of customers over space and time. Many companies use gender differences to create and market products. Parental status is another important feature. You can obtain data like this from customer surveys.

3)Behavioral customer segmentation: Behavioral customer segmentation is based on past observed behaviors of customers that can be used to predict future actions. For example, brands that customers purchase, or moments when they buy the most. The behavioral aspect of customer segmentation not only tries to understand reasons for purchase but also how those reasons change throughout the year.

4)Psychological segmentation: Psychological segmentation of the customers generally deals with things like personality traits, attitudes, or beliefs. This data is obtained using customer surveys, and it can be used to gauge customer sentiment

Clustering:

Patankar et al. [3] states that K-means Clustering is a clustering Algorithm in which we are given with data points with its data set and features and the mechanism is to categories those data points into clusters as per their similarities.

Aman Banduni et al [4] states that K-means that an algorithm is one of the most popular classification algorithms. This clustering algorithm relies on centroid, where each data point is placed in one of the overlapping ones, which is pre-sorted in the K-algorithm.

The algorithm forms K clusters based on its similarity. To calculate the similarity K-means uses Euclidean distance measurement method.

Principal Component Analysis:

Principal Component Analysis (PCA) is a statistical procedure that uses an orthogonal transformation that converts a set of correlated variables to a set of uncorrelated variables. PCA is the most widely used tool in exploratory data analysis and in machine learning for predictive models. Moreover,

Principal Component Analysis (PCA) is an unsupervised learning algorithm technique used to examine the interrelations among a set of variables. It is also known as a general factor analysis where regression determines a line of best fit.

The main goal of Principal Component Analysis (PCA) is to reduce the dimensionality of a dataset while preserving the most important patterns or relationships between the variables without any prior knowledge of the target variables.

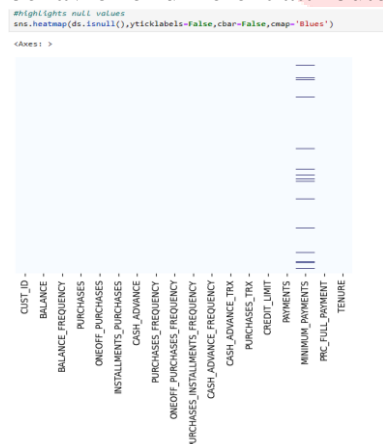
Our project mainly focuses on customer behavior of customer based on data of each customer. We analyzed about the balance of customer, purchases he makes, one off purchases he makes, whether he purchases instantly i.e. in one go or purchases in installments. We also saw that how much cash advance the customer has paid, frequency of the purchases made by installments, cash advance, limit of the credit card, tenure of the credit card.

By analysis this data we are trying to predict that if the customer purchases any product with credit card what are the chances that he will repay the debt so that they can purchase other product in future.

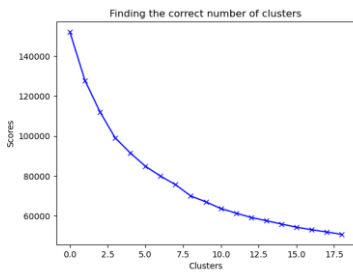
So that the company mainly focuses on those customers to launch a targeted campaign which increases the revenue of the company and sales of the product which the company is trying to sell.

We start by preprocessing the data trying to remove the unwanted values and filling missing values which are required for further analysis. We have visualized the

behavior of different attributes using distribution plot.



Then we have used K-Means algorithm to create clusters containing the customers of different groups which are targeted for different market campaigns. We have used elbow method to correctly guess the optimal number of clusters required for our project.

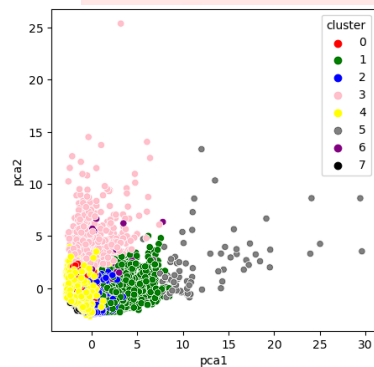


We cannot visualize all the 17 columns at a time so we use principal component analysis so that the number of dimensions will reduce to 2.

pca_df			
	pca1	pca2	cluster
0	-1.682220	-1.076451	0
1	-1.138295	2.506447	3
2	0.969684	-0.383455	1
3	-0.873627	0.043166	0
4	-1.599434	-0.688582	0
...
8945	-0.359629	-2.016149	4
8946	-0.564368	-1.639099	4
8947	-0.926203	-1.810786	4
8948	-2.336551	-0.657978	4
8949	-0.556422	-0.400455	4

8950 rows x 3 columns

Then we visualize the clusters of various customers using scatter plot.



We are not only trying to increase sales by targeting customers belonging to large clusters but also try to focus on people belonging to small clusters so that they can make strategies to increase purchases made by customers of small clusters so that the revenue of the company increases.

Principal Component Analysis

PCA is used for dimensionality reduction.

Previously we had 17 columns so we reduced to 2 dimensions which makes it easy to visualize without data loss.

Step-1: Standardize the data: This ensures all features are on the same scale.

Step-2: Calculate the covariance matrix: This matrix captures how each pair of features varies together.

Step-3: Eigenvalue decomposition: Decompose the covariance matrix to find eigenvectors (directions of greatest variance) and eigenvalues (amount of variance explained by each eigenvector).

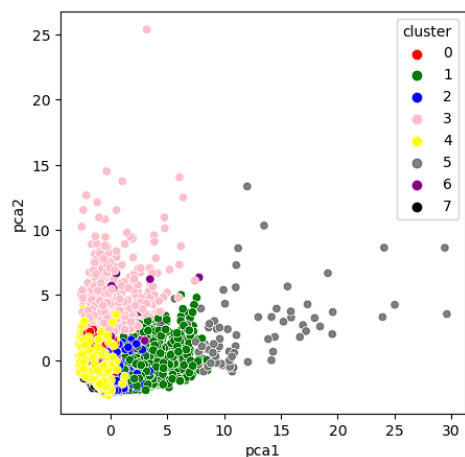
Step-4: Choose principal components: Select the eigenvectors with the highest eigenvalues. These correspond to the directions of greatest variance in the data.

Step-5: Project the data: Transform the original data points onto the new axes defined by the chosen eigenvectors. This creates the principal components, which are linear combinations of the original features.

	pca1	pca2	cluster
0	-1.682220	-1.076451	0
1	-1.138295	2.506447	3
2	0.969684	-0.383455	1
3	-0.873627	0.043166	0
4	-1.599434	-0.688582	0
...
8945	-0.359629	-2.016149	4
8946	-0.564368	-1.639099	4
8947	-0.926203	-1.810786	4
8948	-2.336551	-0.657978	4
8949	-0.556422	-0.400455	4

III. RESULTS

We have got clusters containing different customers in each cluster which are differentiated by each color. By seeing the clusters the company can launch targeted campaign so that the sales of the company will increase and revenue of the company will increase.



IV. CONCLUSION

Improving the sales of the products can be done by applying different strategies to identify the required customers. Required customers can be formed by dividing them into different clusters based on their behavior.

V. REFERENCES

- [1]Dhiraj Kumar “Implementing Customer Segmentation Using Machine Learning”,Year: 2023.
- [2]Rudrendu Paul, MBA+MS (Analytics), Boston University (May, 2019), <https://github.com/RudrenduPaul/Python-Customer-segmentation-and-consumer-behavior-analysis-using-machine-learning/blob/master/Consumer%20buying%20behavior%20-%20-%20Paul.ipynb>
- [3]Prof. Nikhil Patankar,Soham Dixit,Akshay Bhamare,Ashutosh Darpel,Ritik Raina “Customer Segmentation Using MachineLearning”, Sanjivani College of Engineering,Year: 2021.
- [4]Aman Banduni, Prof Ilavendhan A.“Customer Segmentation Using Machine Learning”, IJIRT,Year: 2020..

