

# Customer Segmentation Based on Online Shopping Using K-Means Algorithm

<sup>1</sup>Km Vandna, <sup>2</sup>Mr. Pawan Yadav

<sup>1</sup>M. Tech Scholar, <sup>2</sup>Asst. Professor  
Department of Computer Science Engineering,  
Eshan College of Engineering,  
Mathura India

## 1. Abstract

Customer segmentation is a crucial strategy for businesses seeking to optimize marketing efforts and enhance customer experiences. This study explores the use of the K-means clustering algorithm to segment customers based on their online shopping behavior. By analyzing data from a large dataset of online transactions, we identify key features such as purchase frequency, total spending, recency of purchases, and product category preferences for clustering. Using the K-means algorithm, we group customers into distinct segments, each exhibiting unique shopping behaviors and preferences. The analysis reveals valuable insights into customer profiles, enabling targeted marketing strategies and personalized recommendations. We evaluate the clustering results using metrics such as silhouette score and visualize the segments for a comprehensive understanding of customer groups. The findings provide actionable strategies for businesses to engage with customers more effectively, improve customer satisfaction, and drive sales growth. Overall, this study demonstrates the potential of leveraging the K-means algorithm for customer segmentation in the context of online shopping behavior, offering businesses a pathway to achieve better customer engagement and optimize marketing campaigns.

**Keywords:** introduction Customer Segmentation; overview K-Means Algorithm, implementation

## 2. Introduction

The primary goal of customer segmentation is to tailor marketing strategies and offerings to the specific needs and preferences of each segment, leading to more personalized and effective customer interactions. This approach not only enhances customer satisfaction but also drives higher sales and loyalty.

The K-means algorithm is one of the most widely used clustering methods for customer segmentation due to its simplicity and efficiency. It clusters customers based on their shopping behavior, grouping them into distinct segments with similar characteristics. By applying the K-means algorithm, retailers can gain actionable insights into different customer profiles, such as high-value shoppers, frequent buyers, or those with specific category preferences.

Through effective customer segmentation using the K-means algorithm, businesses can:

- **Personalize Marketing:** Craft targeted campaigns that resonate with the specific needs and interests of each segment.
- **Optimize Inventory:** Stock products that cater to the preferences of different customer groups.
- **Improve Customer Retention:** Offer personalized promotions and loyalty programs that keep customers engaged.

## 3. Literature Review

Segmentation is the division of a market into clusters of consumers with similar needs, such that the more closely the needs match up, the smaller the segment tends to be. Market segmentation is the process of dividing the

- **Increase Revenue:** Drive higher sales through more relevant product recommendations and promotion

So that objects within a cluster are similar to one another and dissimilar to objects in other clusters.

Customer Segmentation is the process of division of customer base into several groups called as customer segments such that each customer segment consists of customers who have similar characteristics. This segmentation is based on the similarity in different ways that are relevant to marketing such as gender, age, interests, and miscellaneous spending habits.

The customer segmentation has the importance as it includes the ability to modify the programs of market so that it is suitable to each of the customer segment, support in business decision; identification of products associated with each customer segment and to manage the demand and supply of that product; identifying and targeting the potential customer base, and predicting customer defection, providing directions in finding the solutions.

The thrust of this paper is to identify customer segments using the data mining approach, using the partitioning algorithm called as K-means clustering algorithm. The elbow method determines the optimal clusters.

market into subsets of customers who share common characteristics. The four pillars of segmentation marketers use to define their ideal customer profile (ICP) are demographic, psychographic, geographic and behavioral budgets, gain a competitive edge over their rival companies, demonstrating the better knowledge of the needs of the custo

mer. It also helps an organization in increasing their market efficiency, determining new market opportunities, making better brand strategy, identifying customer retention.

**Overview**

The K-means algorithm is a popular clustering method used in customer segmentation due to its efficiency and simplicity. It is widely applied to divide a customer base into distinct groups based on similar shopping behaviors and preferences. Here's an overview of the K-means algorithm and how it is used for customer segmentation

**K-means Algorithm?**

**K-means Clustering:** K-means is an unsupervised machine learning algorithm used to partition a dataset into K clusters, where each cluster consists of data points that are similar to each other. The goal is to minimize the within-cluster variance and maximize the between-cluster variance.

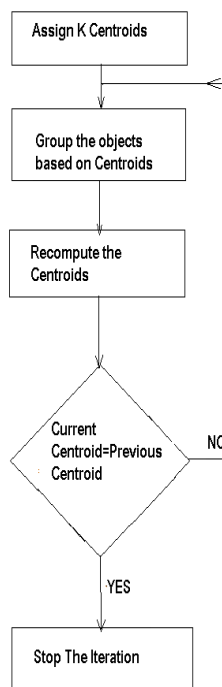
**Algorithm Process**

**Initialization:** Choose K initial centroids (cluster centers) either randomly or using a specific initialization method.

**Assignment:** Assign each data point to the cluster whose centroid is closest to it based on a distance metric (commonly, Euclidean distance).

**Update:** Recalculate the centroids as the mean of all data points assigned to each cluster.

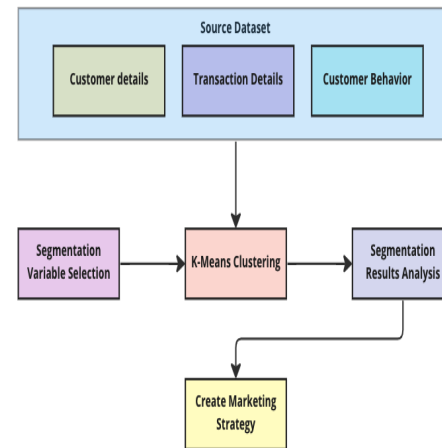
**Iteration:** Repeat the assignment and update steps until convergence (i.e., centroids no longer change significantly or a maximum number of iterations is reached).



**4. Methodology**

The dataset used to implement clustering and K-means algorithm was collected from a store of shopping mall. The data set contains 5 attributes

and has 200 tuples, representing the data of 200 customers. The attributes in the data set have Customer, gender, age, annual income (k\$), spending score on the scale of



**Steps for Implementing K-means for Customer Segmentation**

**1. Data Collection and Preparation:**

- **Collect Data:** Gather data on online shopping behavior such as purchase history, product categories, total spending, recency of purchases, and any other relevant customer attributes.

```

Before Label Encoder
0      Male
1      Male
2      Female
3      Female
4      Female
...
195    Female
196    Female
197    Male
198    Male
199    Male
Name: Gender, Length: 200, dtype: object
    
```

```

After Label Encoder
0      1
1      1
2      0
3      0
4      0
...
195    0
196    0
197    1
198    1
199    1
Name: Gender, Length: 200, dtype: int32
    
```

- **Data Cleaning:** Handle missing values, outliers, and duplicate data entries.
- **Data Transformation:** Depending on the data, transformations such as encoding categorical variables or normalizing numerical features may be required.

- 

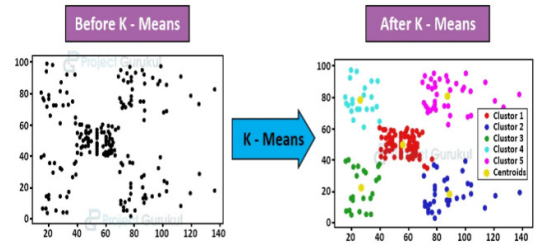
X data before PCA:

```
[[ 1 19 15 39]
 [ 1 21 15 81]
 [ 0 20 16 6]
 [ 0 23 16 77]
 [ 0 31 17 40]]
```

X data after PCA:

```
[[-0.40638272 -0.52071363]
 [-1.42767287 -0.3673102 ]
 [ 0.05076057 -1.89406774]
 [-1.6945131 -1.63190805]
 [-0.31310838 -1.81048272]]
```

customers are grouped and the characteristics of each group. By interpreting the results of clustering, businesses can identify actionable insights for targeted marketing, customer engagement, and strategic decision-making.



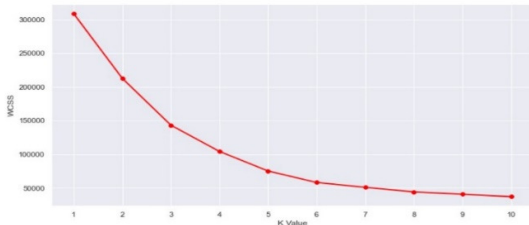
**.Data Analysis**

Data analysis in the K-means algorithm for customer segmentation provides an in-depth understanding of how

**4.4 Visualize the clusters**

**Visualize Clusters:**

- **Scatter Plots:** Use scatter plots to visualize how the clusters are distributed across different feature spaces



- **Cluster Centroid Visualization:** Show the positions of the centroids within the data

**Conclusion**

The process of customer segmentation ensures that your brand is customer-centric and helps you serve them better. It boosts conversions, brings your marketing efforts to fruition, and also helps build everlasting customer relationships. K-means clustering is a popular machine learning algorithm that can be used to segment customers based on their similarities. The algorithm aims to find k clusters in the data, where each cluster represents a group of customers that are similar to each other.

**References:**

[1] [https://www.researchgate.net/publication/355587534\\_Customer\\_Segmentation\\_Using\\_K-Means\\_Clustering\\_Algorithm](https://www.researchgate.net/publication/355587534_Customer_Segmentation_Using_K-Means_Clustering_Algorithm)

[2] <https://www.ijert.org/research/customer-segmentation-using-k-means-clustering-IJERTV11IS030152.pdf>

[3] <https://medium.com/@ugursavci/step-by-step-customer-segmentation-using-k-means-and-pca-in-python-5733822295b6>

[4] [https://www.researchgate.net/publication/355587534\\_Customer\\_Segmentation\\_Using\\_K-Means\\_Clustering\\_Algorithm](https://www.researchgate.net/publication/355587534_Customer_Segmentation_Using_K-Means_Clustering_Algorithm)

[5] <https://www.ijert.org/research/customer-segmentation-using-k-means-clustering-IJERTV11IS030152.pdf>

[6] <https://ieeexplore.ieee.org/document/8769171>

[7] [https://ijcrt.org/papers/IJCRT\\_196650.pdf](https://ijcrt.org/papers/IJCRT_196650.pdf)

[8] [https://www.google.com/search?q=+customer+segmentation+using+k+means+algorithm+example&sca\\_esv=125395e02bc1fefe&sca\\_upv=1&udm=2&biw=1366&bih=633&ei=0QwiZuzqHZGXseMP3ZiX0A8&ved](https://www.google.com/search?q=+customer+segmentation+using+k+means+algorithm+example&sca_esv=125395e02bc1fefe&sca_upv=1&udm=2&biw=1366&bih=633&ei=0QwiZuzqHZGXseMP3ZiX0A8&ved)

