**CUSTOMER SEGMENTATION**



**REPORT BY,**

|  |  |
| --- | --- |
| **Anushka R Chavan** | **Adinath Mahangare** |
| DATA SCIENCE INTERN AT | DATA SCIENCE INTERN AT |
| EXPOSYS DATA LABS | EXPOSYS DATA LABS |

**Abstract**

Customer segmentation is one of the key aspects business support system, In this Project we are using python language, clustering Techniques, by which the several segments of customers are allowing them to target the possible users. In this Data science project, it will make use of K-means Clustering which is the essential algorithm for clustering the dataset. we are using one way to segment the customer as demographics like age, gender, customer id, annual income and spending score.

Customer Segmentation is the subdivision of a market into discrete customer groups that share similar characteristics. Customer Segmentation can be a powerful means to identify unsatisfied customer needs. Using the above data companies can then outperform the competition by developing uniquely appealing products and services.

You are owing a supermarket mall and through membership cards, you have some basic data about your customers like Customer ID, age, gender, annual income and spending score. You want to understand the customers like who are the target customers so that the sense can be given to marketing team and plan the strategy accordingly.

**Table of Contents**

1. Introduction
2. Existing Method
3. Architecture
4. Methodology
5. Conclusion

**Introduction**

Clustering helps marketers improve their customer base, work on target areas, and segment customers based on  purchase history, interests, or activity monitoring.

The scope of relationship marketing is to retain customers and win their loyalty. This can be achieved if the companies’ products and services are developed and sold considering customers ’demands. Fulfilling customers’ demands, taken as the starting point of relationship marketing, can be obtained by acknowledging that the customers’ needs and wishes are heterogeneous.

The segmentation of the customers’ base allows operators to overcome this because it illustrates the whole heterogeneous market as the sum of smaller homogeneous markets. The concept of segmentation relies on the high probability of persons grouped into segments based on common demands and behaviors to have a similar response to marketing strategies.

**Existing Method**

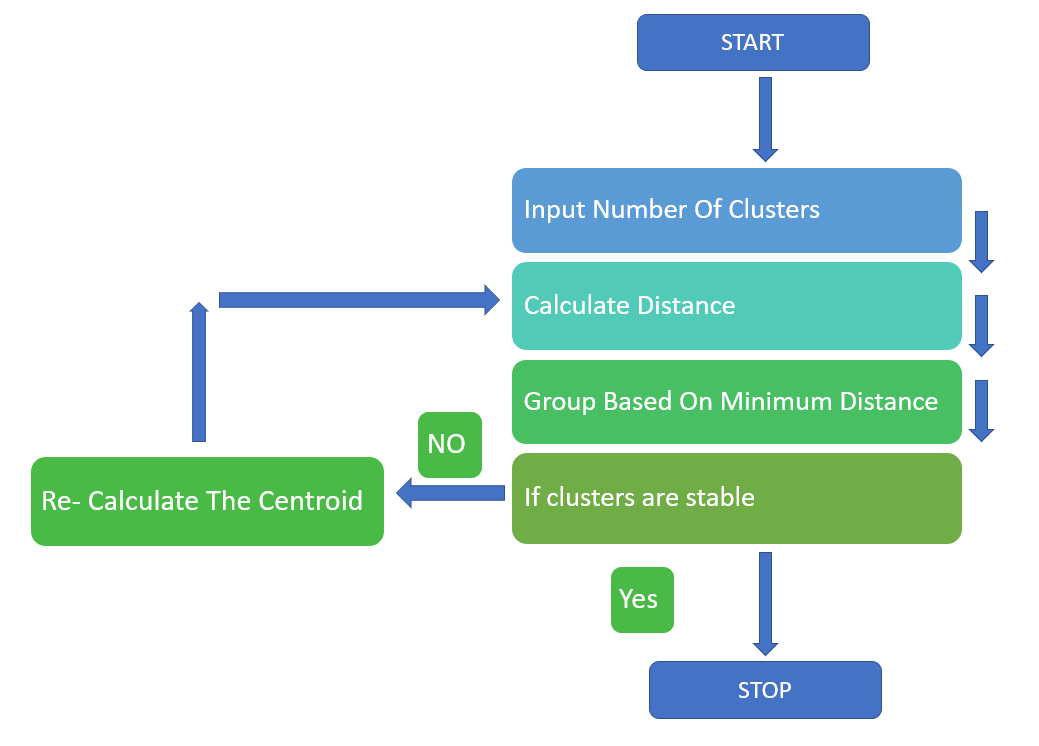
Hierarchical clustering, as the name suggests is an algorithm that builds hierarchy of clusters. This algorithm starts with all the data points assigned to a cluster of their own. Then two nearest clusters are merged into the same cluster. In the end, this algorithm terminates when there is only a single cluster left.

This algorithm has been implemented using bottom up approach. It is also possible to follow top-down approach starting with all data points assigned in the same cluster and recursively performing splits till each data point is assigned a separate cluster.

The decision of merging two clusters is taken on the basis of closeness of these clusters. There are multiple metrics for deciding the closeness of two clusters :

* + Euclidean distance: ||a-b||2 = √(Σ(ai-bi))
  + Squared Euclidean distance: ||a-b||22 = Σ((ai-bi)2)
  + Manhattan distance: ||a-b||1 = Σ |ai-bi|
  + Maximum distance:||a-b||INFINITY = maxi |ai-bi|
  + Mahalanobis distance: √((a-b)T S-1 (-b))   {where, s : covariance matrix}

**Proposed Method with Architecture**



1. Input the Number of Clusters
2. Calculate the Distance of the Data Point from the Cluster Centroid
3. Assign the Data Point to the two nearest Cluster
4. Re- Calculate the Centroid
5. Repeat the process for all the Data Points

**Methodology**

Data set contains 200 entries representing data of 200 peoples and a total of 5 attributes. After proper visualization of the data, clustering needs to be done. But, to perform clustering knowing the optimal number of clusters is necessary. Elbow method is used to calculate the optimal number of clusters. Once the number of cluster are calculated, all that is required is to implement the K – Means clustering algorithm and providing it with the cluster number and visualizing it.

**Conclusion**

Thus, the given datasets of customer segmentations is successfully processed by using K-means clustering algorithm. And it helps to improve customer service.

K means algorithm is one of the most popular clustering algorithms and normally the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K-means is to group data points into distinct non-overlapping subgroups. One of the major application of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.

K means clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset. The goal of K means is to group data points into distinct non-overlapping subgroups. One of the major application of K means clustering is segmentation of customers to get a better understanding of them which in turn could be used to increase the revenue of the company.