SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

Course Level Project Report

MACHINE LEARNING

(213CSE2304)

Team Number	14
Project Title	Mall Customer Segmentation
Team Members	K. Nikhil (9921004374)
	K. V. S. Sai Ram Santosh Babu (9921004355)
	K. Soma Sekhar (9921004385)
	K. Hari krishna (99210041214)
Faculty Guide	Dr. R. Sumathi

MALL CUSTOMER SEGMENTATION

Abstract: -

Customer segmentation process is a separation of the types of consumers/customers are visited to the mall/market/shopping complex. i, e., segregating multiple distinct groups of customers who shares their similar characteristics. The segmentation of mall is the potent way of representing and defining the customer needs.

Managing the customer relationship will always play the important/crucial role to supply business intelligence to build, manage and develop valuable interminable consumer/buyer relationship/connections. This will help in the business for ideas to develop the relationship with customer with smooth manner.

Introduction: -

The business intelligence has the important role to take part in the permitting companies/organizations to use technical expertise to earn good consumer/customer knowledge and programs for overstep with the use of clustering methods, customers with similar criteria are clustered together.

The widespread use of data mining techniques in extracting meaningful and strategic information from an organization's database has resulted from the increased competition among businesses over the years, and the large historical data that is available has resulted from the widespread use of data mining techniques in extracting meaningful and strategic information from an organization's database. Data mining is a process in which methods are used to extract data patterns and present them in a human-readable format that can be used for decision-making.

The customer segmentation which helps in the marketing team to recognize and exploring the different customer segments i.e., each customer has different behavior. That develops different purchasing strategies done by the customer to the marketing team.

In the customer segmentation, clustering is an iterative procedure of knowledge/intelligence from huge amount of vast huge amounts of raw data, previous data and unorganized data.

Clustering is an iterative approach for extracting knowledge from large amounts of unstructured data. Clustering is a sort of exploratory data mining that is utilized in a variety of applications, including machine learning and predictive analytics.

Data: -

The dataset consists of customer data with following variables

Variable	Description
Customer ID	Unique id
Gender	Male / Female
Age	Age
Annual Income (k$)	Annual Income
Spending Score (1-100)	Customers spend rate in percentage

Table: - Dataset

Data Source: - Mall Customer Dataset

Sample data: -

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40

Clustering: -

Clustering is a technique used in data mining and machine learning to group together similar data points or objects based on their characteristics or attributes. The goal of clustering is to partition a dataset into subsets or clusters such that the objects within a cluster are more similar to each other than they are to objects in other clusters.

There are several different algorithms used for clustering, each with its strengths and weaknesses. Some of the most commonly used algorithms include k-means clustering, hierarchical clustering, and density-based clustering

Clustering has many applications, including customer segmentation, image segmentation, anomaly detection, and bioinformatics. It can also be used as a pre-processing step for other machine learning algorithms, such as classification or regression

Clustering Algorithms: -

- K- Means Clustering
- Hierarchical Clustering
- Density- Based Clustering

K- Means Clustering: -

K-means clustering is an algorithm which is used to perform the mall basket analysis which comes under the category Unsupervised learning. This will help in the mall basket analysis to be carried out to predict the final/Target customer that who can easily converged, among all the customers visited. The main agenda is to companies need the customer data to know the better feature of the customer. Also, companies need to know the customers area of interests in their needs and shops for their buying aspects. Using K-means clustering segregating the customers with the similarities and differences of predicting the behaviour, introducing better options and things to customer

By the K-means clustering customer segmentations will helps in figuring out the consumers who differ in the terms of expectations, desires, attributes and preferences. The important use of customer segmentation is to be grouping the customers with similar interests and needs which helps the marketing team to implement effective marketing strategy plan.

Elbow Method:

The elbow approach is a tool for analysing the clusters created by our dataset and assisting in interpreting the appropriate number of ideal clusters in the dataset. The best number of clusters for our dataset is determined by this method to be five.

The elbow technique is based on the finding that increasing the number of clusters can assist lower each cluster's total within-cluster variation. This is due to the fact that larger clusters allow for the capture of finer groups of data objects that are more comparable to one another.
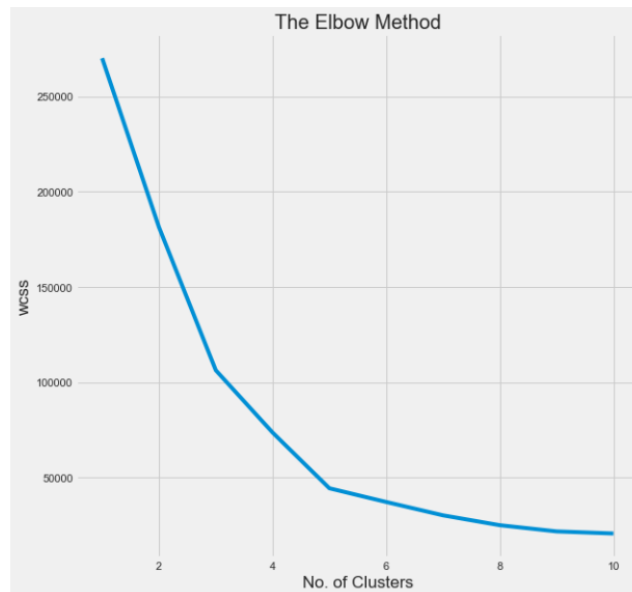
Elbow Method

Elbow method which shows the optimal number of clusters
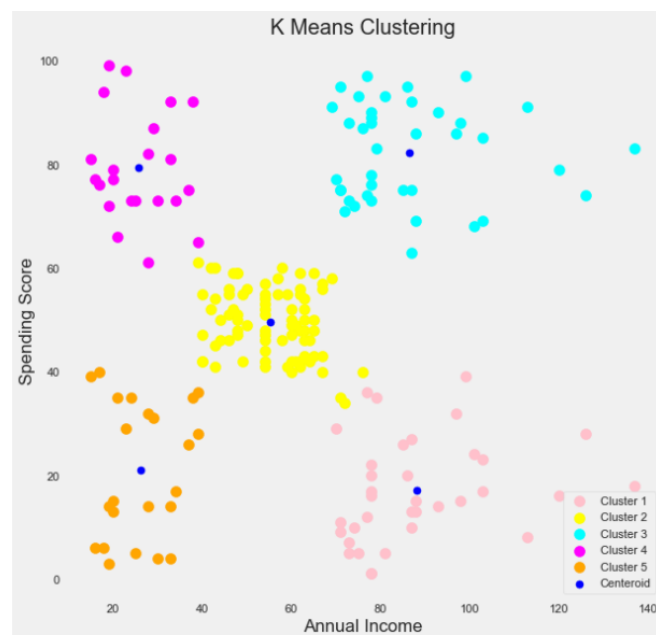
Clustering based on Annual income

Cluster 1 (Red) implies earning a lot while spending less.

Cluster 2 (blue) reflects the mean in terms of earnings and spending.

Cluster 3 (Green) shows both high earnings and significant spending. [Prospective customers]

Cluster 4 (blue) denotes earning less but spending more.

Cluster 5Earning less and spending less is represented by (magenta colourA).

Clustering based on Age

Hierarchical Clustering: -

Hierarchical clustering is a method of clustering analysis that groups similar objects into clusters based on their similarities and differences. In hierarchical clustering, the data objects are organized into a hierarchy of nested clusters, where each object belongs to a cluster at the bottom level, and clusters at higher levels group clusters from lower levels.

The two main types of hierarchical clustering are agglomerative clustering and divisive clustering.

In agglomerative clustering, each data point is initially considered as a separate cluster, and then, at each iteration, the two closest clusters are merged together, until all the data points are grouped into a single cluster.

In divisive clustering, all the data points are initially grouped into a single cluster, and then, at each iteration, the cluster is split into two smaller clusters, until each data point is in a separate cluster.

The result of hierarchical clustering is typically represented as a dendrogram, which is a tree-like diagram that shows the hierarchy of clusters. The dendrogram illustrates the order in which the clusters were formed, and the

distance between them, which can be used to determine the optimal number of clusters to use in subsequent analysis.

Hierarchical clustering is commonly used in fields such as biology, sociology, and marketing, to group similar data points together and identify meaningful patterns in the data.
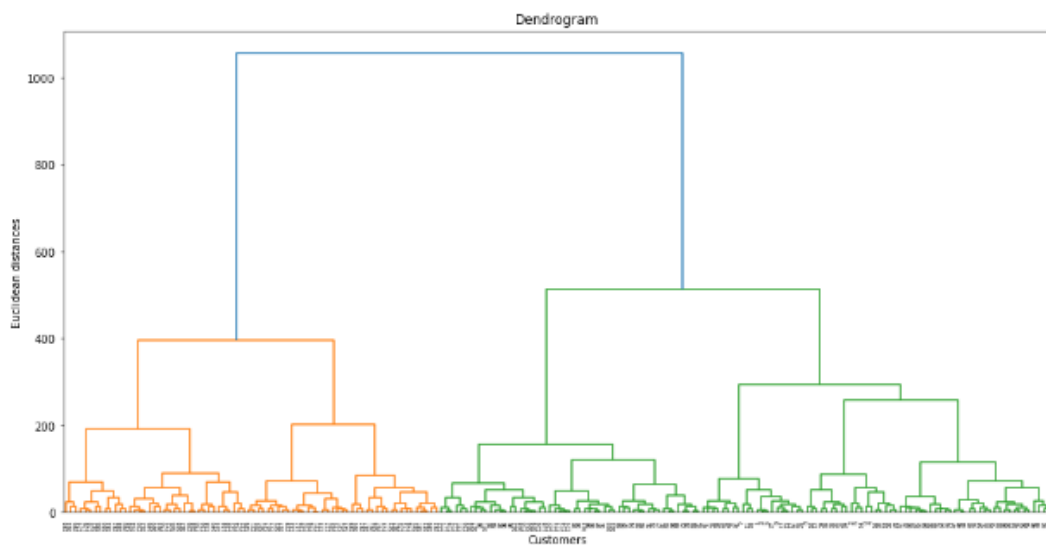

Dendrogram
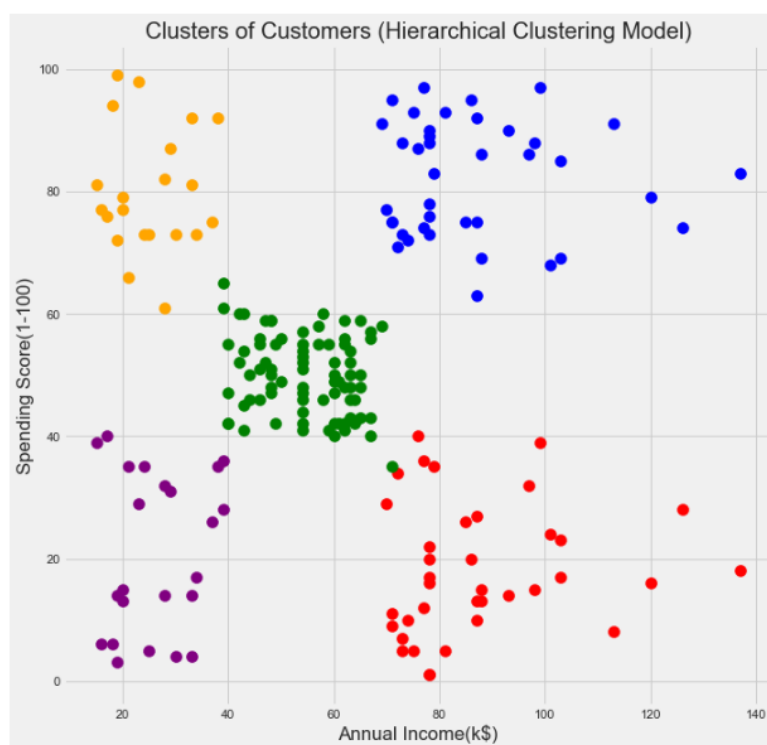
Clustering based on Annual income

Density- Based Clustering: -

Density-based clustering is a clustering technique that groups together data points based on their density in a high-dimensional space. It is commonly used to identify clusters of arbitrary shape in a dataset and is particularly useful when the clusters have varying densities or are separated by noisy or sparse regions.

One of the most popular density-based clustering algorithms is DBSCAN (Density-Based Spatial Clustering of Applications with Noise). DBSCAN works by defining two parameters: epsilon (ε) and minimum number of points (minPts). The algorithm first identifies "core" points that have at least minPts points within a distance of ε. These core points form the center of a cluster, and any points within ε distance of them are added to the same cluster. Points that are not core points and are not within ε distance of any core points are considered noise points.

Another density-based clustering algorithm is OPTICS (Ordering Points To Identify the Clustering Structure). OPTICS is similar to DBSCAN but uses a hierarchical ordering of points based on their density to identify clusters of varying density and shape.

One advantage of density-based clustering is that it can handle noise and outliers well. However, it can be sensitive to the choice of parameters such as ε and minPts, which can greatly affect the resulting clusters
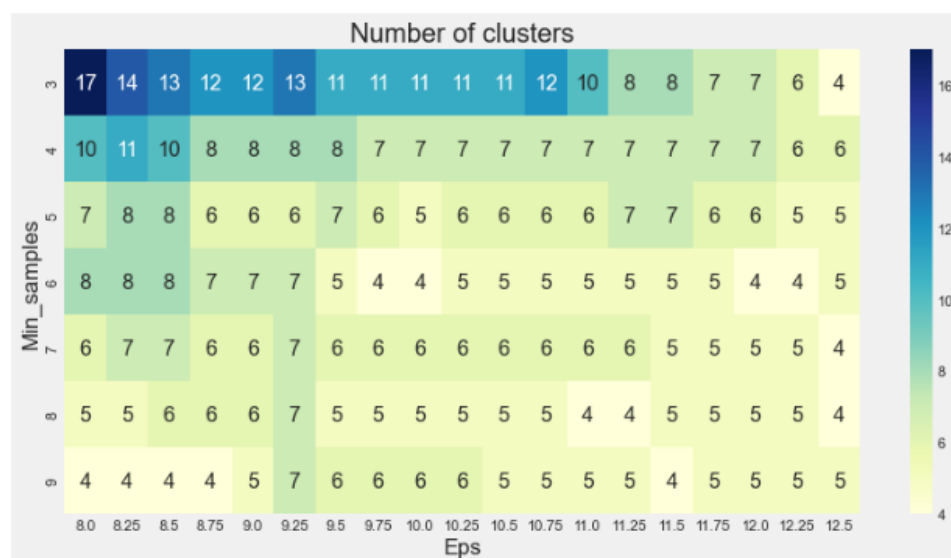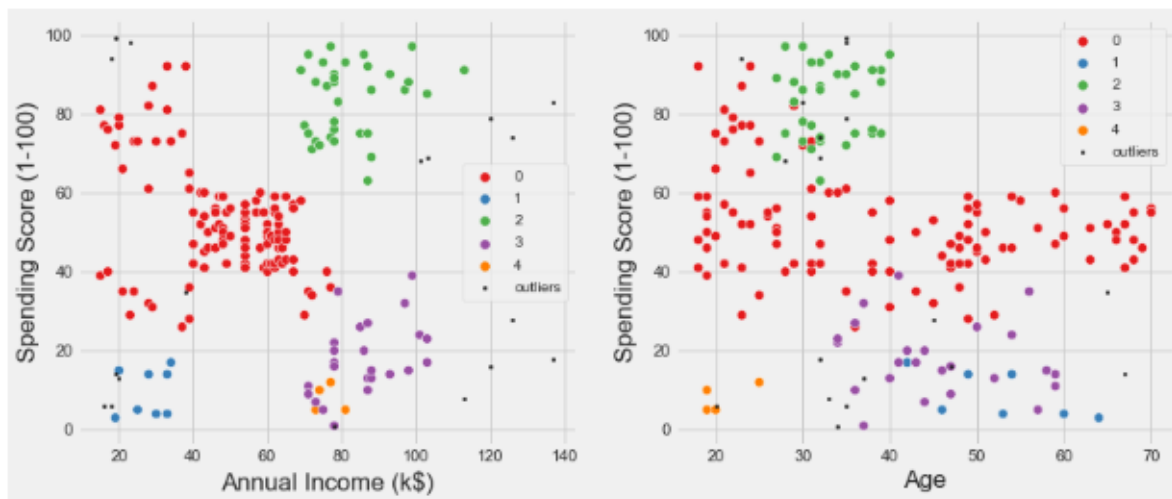

Heat map based on number of clusters
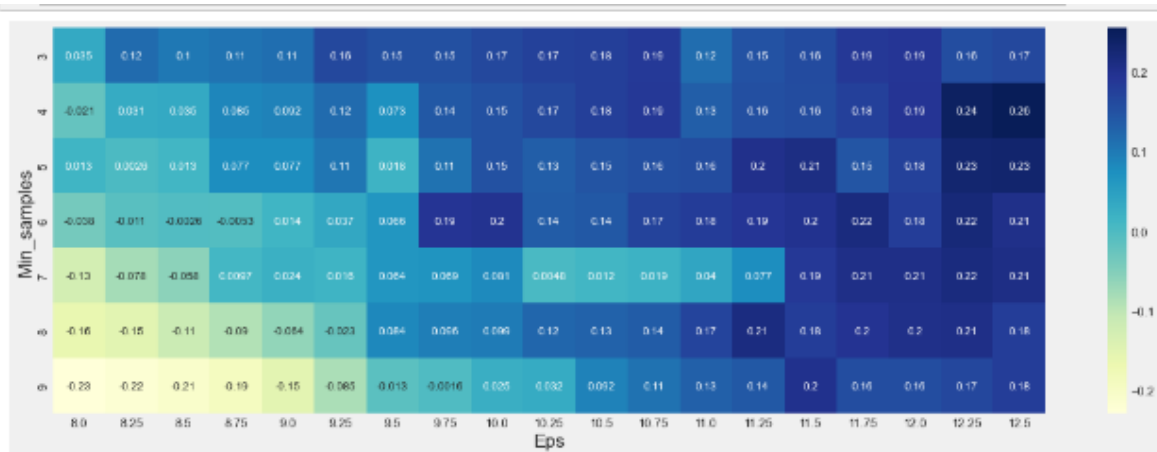
Heat map based on Score

Clusters based on annual income   Clusters based on age

Conclusion: -

As a result of this massive data volume, consumer data is growing tremendously. These clustering models must be able to process this massive amount of data properly.

Cluster 1 denotes the customer with a high annual income as well as a high annual spend, as seen in the above visualization. Cluster 2 denotes a group with a high annual income but a low annual expenditure. Cluster 3 represents customers who have a low annual income and spend a small amount each year. Cluster 5 indicates a modest annual income but a high annual expenditure. Customers with a medium income and a medium expenditure score fall into clusters 4 and 6.

References: -

- C. Calvo-Porral and J. P. Lévy-Mangin, "Profiling shopping mall customers during hard times," J. Retail. Consum. Serv., vol. 48, no. November 2018, pp. 238–246, 2019, doi: 10.1016/j.jretconser.2019.02.023.
- G. Parsons, "Assessing the effectiveness of shopping mall promotions: Customer analysis," Int. J. Retail Distrib. Manag., vol. 31, no. 2, pp. 74–79, 2003, doi: 10.1108/09590550310461976.
- L. Lucia-Palacios, R. Pérez-López, and Y. Polo-Redondo, "Does stress matter in mall experience and customer satisfaction?," J. Serv. Mark., vol. 34, no. 2, pp. 177–191, 2020, doi: 10.1108/JSM-03-2019-0134.
- M. F. Diallo, F. Diop-Sall, S. Djelassi, and D. Godefroit-Winkel, "How Shopping Mall Service Quality Affects Customer Loyalty Across Developing Countries: The Moderation of the Cultural Context," J. Int. Mark., vol. 26, no. 4, pp. 69–84, 2018, doi: 10.1177/1069031X18807473.
- L. Kouhalvandi, O. Ceylan, and S. Ozoguz, "Automated Deep Neural Learning-Based Optimization for High Performance High Power Amplifier Designs," IEEE Trans. Circuits Syst. I Regul. Pap., pp. 1–14, 2020, doi: 10.1109/tcsi.2020.3008947.
- S. Hidayat, M. Matsuoka, S. Baja, and D. A. Rampisela, "Object-based image analysis for sago palm classification: The most important features from high-resolution satellite imagery," Remote Sens., vol. 10, no. 8, 2018, doi: 10.3390/RS10081319.
- L. L. Rego, N. A. Morgan, and C. Fornell, "Reexamining the market share-customer satisfaction relationship," J. Mark., vol. 77, no. 5, pp. 1–20, 2013, doi: 10.1509/jm.09.0363.
- X. Luo and C. B. Bhattacharya, "Social Responsibility , Corporate Customer and Market Satisfaction , Value," Am. Mark. Assoc., vol. 70, no. 4, pp. 1–18, 2006.
- Doulamis and N. Doulamis, "Customer Experience Survey," pp. 3–6, 2016, doi: 10.3390/technologies8040076.