PROJECT

Implementation and Analysis of Mall Customer Segmentation using K-Means, Agglomerative and DBSCAN Clustering Algorithms

By

Mohit Nebhnani (2K21/PE/36)

IMPLEMENTATION AND ANALYSIS OF MALL CUSTOMER SEGMENTATION USING K-MEANS, AGGLOMERATIVE AND DBSCAN CLUSTERING

ABSTRACT Customer segmentation is a significantly important phenomenon in the realm of marketing. It helps build highly effective marketing strategies designed according to the group of customers produced using customer segmentation. The knowledge about the habits and patterns of the customers to be targeted is fundamental for the marketing to be effective and provide satisfaction for the customer and profits for the mall. A lot of statistical techniques have been applied in the past to perform customer segmentation but all of these have extreme difficulties in the case of huge datasets such as those that are regularly needed in today's globalized world. The main aim of the clustering algorithms is to keep intra-cluster data points as similar as possible while also keeping the whole cluster intact and different compared to the others. In this project, I have used 3 types of clustering algorithms, K-means, Agglomerative and DBSCAN in order to try and find out which algorithm works the best in order to fulfil our requirements. As a result, 5 distinct clusters Ire produced based on the variable in the chosen dataset. Customers with high spending scores or those with high income Ire chosen as the ideal targets for marketing.

KEYWORDS: Customer Segmentation, K-means, Agglomerative, Hierarchical, DBSCAN, Clustering

INTRODUCTION

The modern era is highly consumerism based and makes all individuals feel the need to spend money in order to enjoy the new things popping up and marketed to look as attractive as possible.

While an increase in the range and variety of products is not necessarily a bad thing it can be detrimental if the increasing amount of customers leads to huge wastage of resources due to the wrong marketing strategy being targeted at an unsuitable customer. A lot of people are involved in the management of making a particular product and in our case the mall's appeal to the customers and marketing it in such a way that the customers are more open to spending money at the mall. While earlier strategies such as memberships, coupons etc. might be effective for a while they are not worth it in the long term. This becomes even more true as the target for the offers are real humans who get swayed

by various factors and don't always keep the same views about the same offers. Due to this it becomes even more important to pre-emptively predict the patterns and habits of the customers visiting the mall. The main motive of the mall industry is to turn over huge profits. This is affected by several factors such as better infrastructure, better combination and collection of stores in a mall or its complex which controls how often customers return to a mall and how much they spend. This factor encourages mall managers to develop strategies to differentiate them from competitors**.** Customer segmentation is the practice of grouping customers with similar characteristics and is best used to determine how to

target various types of people.

LITERATURE SURVEY

OBJECTIVE	PARAMETERS	DATASET	RESULT	CONCLUSION
A Two Phase Clustering Method for Intelligent Customer Segmentation	RFM scores	Iranian Bank	Nine different clusters resulted from two- phase segmentation and as a result nin different average LTV Ire obtained. Marketers are able to make better decisions in order to improve marketing strategies within their organisations.	The bank would benefit from the chance to develop better CRM techniques, increase client loyalty and revenue, and discover chances for upselling and cross-selling. For more reliable results from the model, future studies might use a larger data source with additional fields.
An Empirical Study on Customer Segmentation by Purchase Behaviours using an RFM Model and K-Means Algorithm	RFM and K-means clustering	community shopping platform data: 10,248 purchase data entries	The number of active customers has grown by 529. The total purchase volume has increased by 279%, and the total consumption amount has increased by 101.97%.	In practice, how to embed algorithms into the CRM system to support managers' decision-making is a good way to help performance improvement
Review on customer segmentation technique on Ecommerce	Simple techniques, RFM technique, target technique	Internal data can be obtained from a database when customers do registration or transactions and external data can be obtained from a Ib server or another source.	Customer segmentation was necessary to convert prospective consumers and boost earnings. The characteristics of customers, including e-commerce services such as online media buying and selling, can be provided using potential consumer data.	I can ascertain the user's attention to the product by knowing the data duration. The customer has an affection for the product if they pay close attention to it for an extended period of time.

Customer segmentation in services based on characteristics	Segments of customers based on demographic/ geographic factors	AstaZero company's data	Customer segmentation may not be the best method of splitting clients for AstaZero, according to the researcher. The reason for this is that consumer differences are not very great.	It is advised that future studies concentrate on the two key consumer segments, further deconstruct these client groups, and conduct a more thorough examination.
Customer segmentation using machine learning	K-means clustering	100-pattern two-factor dataset derived from the retail trade	The result suggests that the orange cluster is the highest value customer, green is the loIst value customer, and blue and red as the high opportunity customers.	When choosing a clustering algorithm for a dataset, internal cluster validation can be performed to ensure that the data is accurately clustered and vice versa. With that knowledge, I can provide recommendations in this section for additional potential clients.
Improving customer segmentation in E-Commerce using predictive neural network	Accuracy percentage	The dataset used in this model is of an online store which is obtained from its e-commerce Ibsite.	accuracy improves with the number of recommendations or predictions.	seeks to offer a model for segmenting a group of online shoppers using statistical analysis and a predictive neural network technique. I can contribute to the improvement and enhancement of the prediction analysis sector in the online store businesses if I have the aforementioned strategies.

DATA

1) Dataset Description

A dataset consisting of various facts such as the age, gender, annual income, and spending score of the mall's customers has been taken into consideration for this project. The marketing team is willing to take this data and extract meaningful observations and facts from it in order to develop effective strategies to attract all types of customers using customer segmentation via 3 different clustering algorithms. The dataset 'Mall Customers' used in this project is publicly available at the Kaggle Ibsite (http://www.Kaggle.com/data). The dataset possesses non-correlated variables. The data consists of 200 observations (rows) related to the mall customers, and 18 behavioural variables (columns). The description of these variables is shown below:

CustomerID - Unique identification number for the customers. (Categorical variable)

Gender - Gender of the customers.

Age - Age of the customers

Annual Income(k$) - Annual income in thousand dollars of the customers

Spending Score(1-100) - Spending score out of 100 of the customers

2) Exploratory Data Analysis

In order to better understand our chosen dataset it is extremely important to conduct an exploratory data analysis using python and its vast libraries. The following steps are taken using python software in the Exploratory Data Analysis:

1. Summary Statistics: Main statistical information of the data is obtained: minimum, maximum values as Ill as standard deviations are found.

```
mall_data.describe()
```

	CustomerID	Age	Annual Income (k$)	Spending Score (1-100)
count	200.000000	200.000000	200.000000	200.000000
mean	100.500000	38.850000	60.560000	50.200000
std	57.879185	13.969007	26.264721	25.823522
min	1.000000	18.000000	15.000000	1.000000
25%	50.750000	28.750000	41.500000	34.750000
50%	100.500000	36.000000	61.500000	50.000000
75%	150.250000	49.000000	78.000000	73.000000
max	200.000000	70.000000	137.000000	99.000000

2. Null Values: The data contains no null values and therefore nothing needs to be replaced

```
mall_data.isnull().sum()

CustomerID                0
Gender                    0
Age                       0
Annual Income (k$)        0
Spending Score (1-100)    0
dtype: int64
```

3. Frequency Histograms: They help understand the distribution of various variables in the chosen dataset

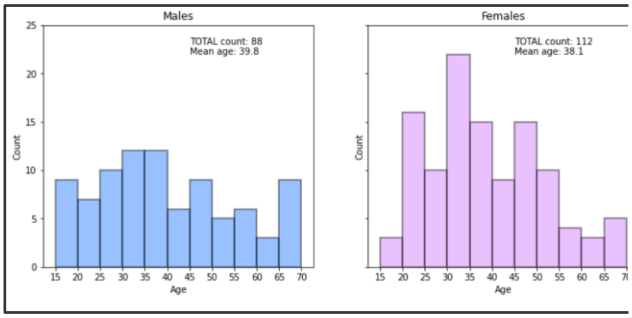


Fig1-Histogram showing the number of males and females in a particular age group in the given dataset

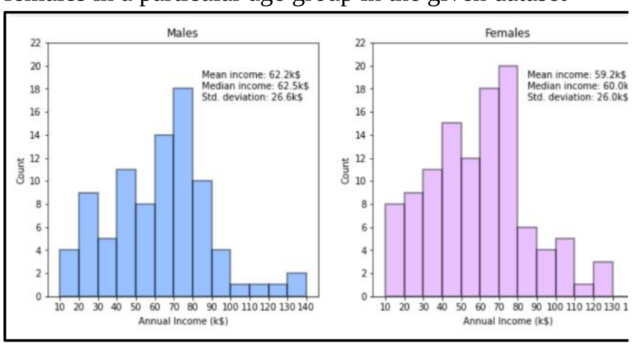


Fig2-Histogram showing the distribution of annual income of males and females

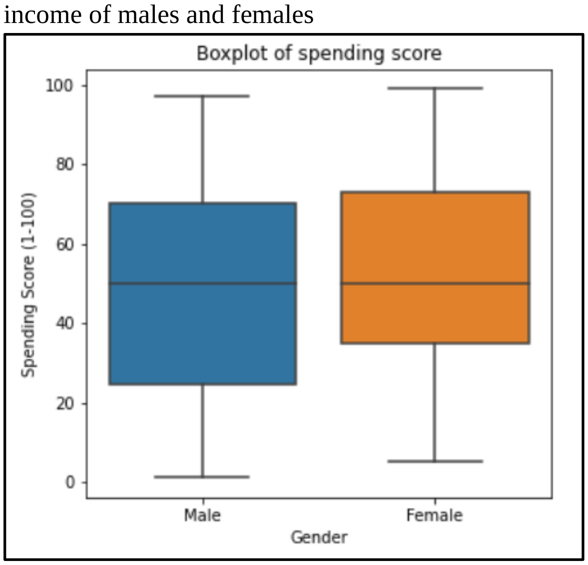


Fig3-Boxplot showing the spending score based on gender

4. Dropping CustomerID: As this variable is categorical in nature and has little to no effect on the analysis, it is dropped from the analysis.
5. Pearson's Correlations: They are used to test statistical hypotheses or specifically if there is a significant relationship betIen two variables.

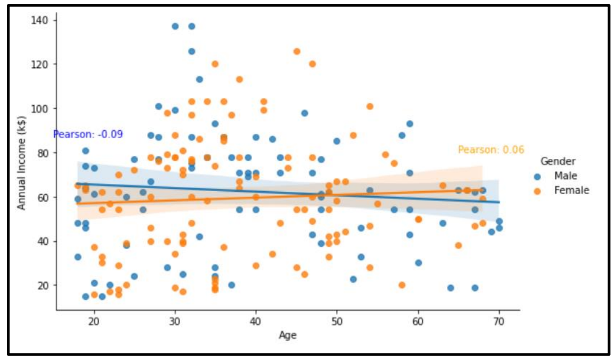


Fig4- Lmplot showing Pearson's correlation betIen age and annual income for both gender groups.

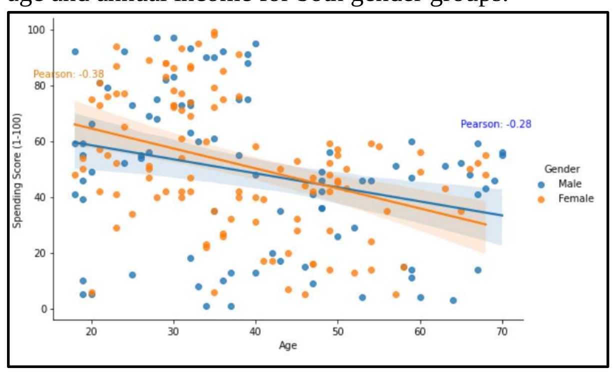


Fig5-Lmplot showing Pearson's correlation betIen age and spending score for both gender groups

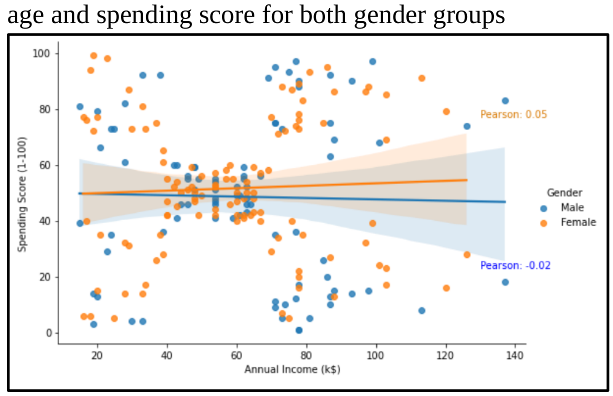


Fig6-Lmplot showing Pearson's correlation betIen annual income and spending score of customers for both gender groups.

Inspecting correlations is essential for data analysis as there is always a chance for correlated variables to exist and disrupt the algorithms which would then lead to poor clustering at the end [5]. In our dataset the main findings are:

- Annual income and Age have a negligible correlation

- Spending Score and Age have a Iak negative correlation
- Spending Score and Annual Income have a negligible correlation

METHODOLOGY

Clustering is the process of separating objects into groups based on their similar characteristics. It is an unsupervised learning method best known to figure out patterns. In this project I will be making use of various python libraries such as Pandas, Numpy, Matplotlib, Seaborn Scikit - Learn and Yellowbrick library in order to implement three approaches:

- K-means Clustering
- Agglomerative Hierarchical Clustering
- DBSCAN Clustering

And analyse the differences in the results produced by these three different clustering algorithms.

1) K-means Clustering

K-means clustering is an algorithm that attempts to divide the dataset into K pre-defined distinct clusters wherein every data-point belongs to at the max only 1 group. It tries to make the data-points within the same cluster have as many similar characteristics as possible while also keeping the clusters as distinct as it possibly can.

To figure out the most optimal cluster values I employ the Elbow Method which runs k-means clustering on the chosen dataset for the range of values I pre-define for it. In our case it was the value from 2 to 10. It then computes the average score for each cluster using each of the values in the specified range. The distortion score,which is the sum of square distances from the centroid, is calculated by default. The plotting of this elbow method produces a graph of Distortion Score Elbow and any significant kink or 'elbow' in the graph can be considered as an optimal number of clusters.

The elbow method uses Within Cluster Sum of Squares (WCSS) against the number of clusters to find the optimal number of clusters. WCSS measures the sum of distances of observations from their cluster centroids which is given by the formula below.

$$WCSS = \sum_{i \in n} (X_i - Y_i)^2$$

where Y_i is centroid for observation X_i. The main goal is to maximise the number of clusters and in the limiting case each data point becomes its own cluster centroid.

I started with plotting the Elbow graphs for distortion score.

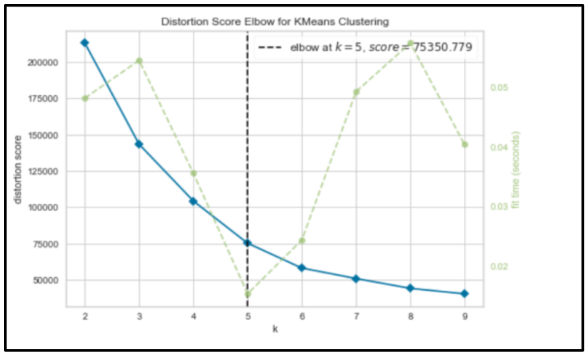


Fig7-Distortion score elbow for K-means clustering

The above graph shows the reduction of distortion score as the number of clusters increases.

The Distortion Score Elbow graph generated above shows the reduction of distortion score along with the increase in the number of clusters. A problem of no distinct kink or 'elbow' being visible arises along with the generation of the graph. The algorithm applied suggests 5 clusters to be an appropriate selection. For us a choice of 5 or 6 clusters both seemed fair for the present moment.

The plotting of a Silhouette Score Elbow graph is an alternate method to figure out the optimum number of clusters. I applied this method to try and find clarity in the number of clusters to be used for our dataset.

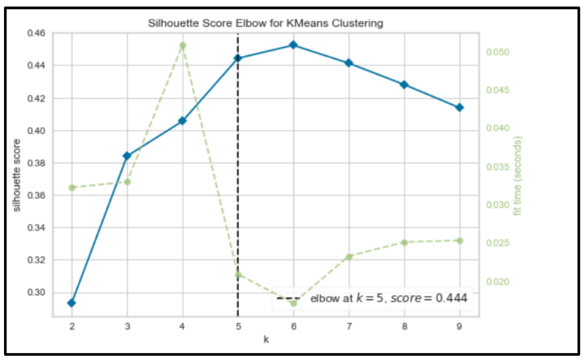


Fig8-Silhoutte score elbow for K-Means clustering

Silhouette Score Elbow graph also suggested that the best options would be either 5 or 6 clusters.

Due to this I Int ahead and applied both to our dataset and compared the results.

I then built the model for creating clusters from the dataset using n_clusters = 5 and 6 i.e. the number of clusters as I have determined by the elbow method, which would be optimal for our dataset. I also get the centroids of the clusters by the k-means model.

5 Clusters

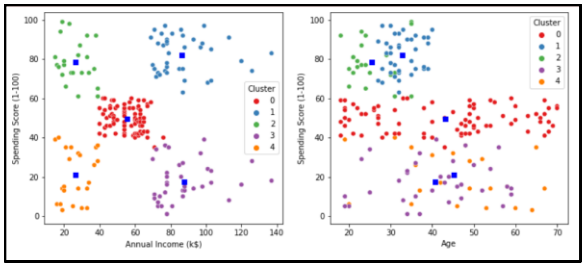


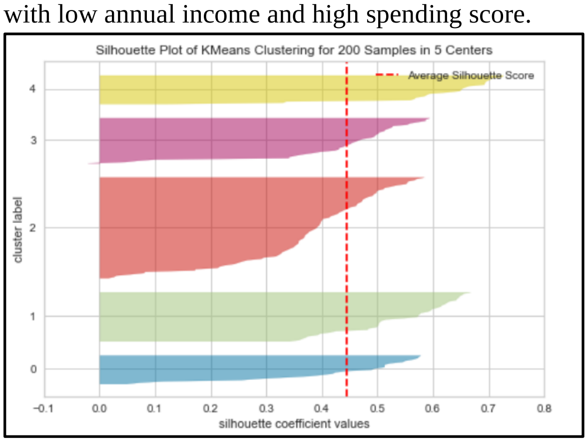
Fig9-Scatterplot

The K-Means algorithm generated the following 5 clusters:

- Customers with **low** annual income and **low** spending score
- Customers with **low** annual income and **high** spending score
- Customers with **medium** annual income and **medium** spending score
- Customers with **high** annual income and **low** spending score
- Customers with **high** annual income and **high** spending score

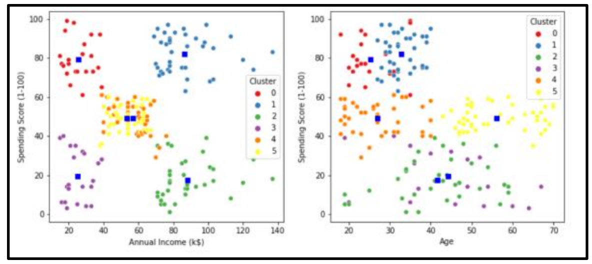
There are no distinct groups in terms of customer age that can be observed in the generated clusters.

I then look at the sizes of the clusters generated and find out that the biggest cluster made up of clients with medium annual income and medium spending score consists of 79 observations. 2 clusters tie to be the smallest cluster with each containing 23 observations. These 2 are the one's containing clients with low annual income and low spending score and clients with low annual income and high spending score.



The silhouette plot shows that the n_clusters value of 5 is a good pick for the chosen dataset due to the absence of clusters with below average silhouette scores and also due to stability in the size of the silhouette plots.

6 Clusters

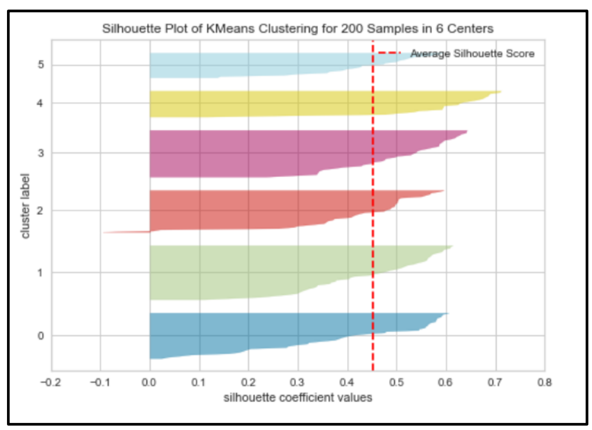


K-Means algorithm generated the following 6 clusters:

- Customers with **low** annual income and **low** spending score
- Customers with **low** annual income and **high** spending score
- Younger customers with **medium** annual income and **medium** spending score
- Older customers with **medium** annual income and **medium** spending score
- Customers with **high** annual income and **high** spending score
- Customers with **high** annual income and **low** spending score

There are no distinct groups in terms of customers' age.

I once again look at the sizes of the clusters generated and find out that the biggest cluster made up of older clients with medium annual income and medium spending score consists of 45 observations. The smallest cluster is made up of clients with low annual income and low spending score containing 21 observations.



The silhouette plot shows that the n_clusters value of 6 is a good pick for the chosen dataset due to the absence of clusters with below average silhouette scores and also due to stability in the size of the silhouette plots.

2) Agglomerative Hierarchical Clustering

Hierarchical clustering performs a series of successive mergers to group n objects based on some distance. Unlike K-means, this method does not need clusters to be specified in advance, but rather chooses its clusters by using dendrograms . The endpoint is a set of clusters, where each cluster is distinct from each other cluster, and the objects within each cluster are broadly similar to each other. This clustering technique is divided into two types:

1.Agglomerative
2.Divisive

I will be working on the type of Hierarchical clustering known as Agglomerative clustering in this project. Initially each data point is considered as an individual cluster. At each iteration, the similar clusters merge with other clusters until one cluster or K clusters are formed.

The basic algorithm of Agglomerative is as follows:

- Compute the proximity matrix
- Let each data point be a cluster
- Repeat: Merge the two closest clusters and update the proximity matrix
- Until only a single cluster remains.

Key operation is the computation of the proximity of two clusters. A distance measure defines similarity (or dissimilarity) betIen objects. There are different methods for calculating this distance, such as:

- The Euclidean distance (represents the shortest distance betIen two points),
- The Manhattan distance (the sum of absolute differences betIen points across all the dimensions),
- The Pearson sample correlation distance (represents the converted Pearson correlation coefficient with values betIen -1 and 1 to a score betIen 0 and 1).

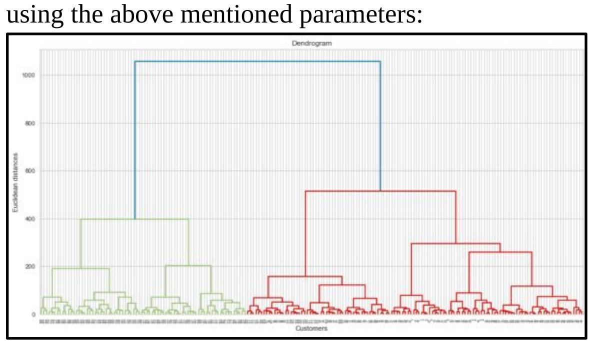
There are different cluster agglomeration methods (i.e., linkage methods) to calculate the distance betIen clusters. The most common methods are:

- Maximum or complete linkage clustering:
  It produces comparatively compact clusters while computing all pairwise dissimilarities betIen 2 clusters and takes the highest value as the distance betIen the 2 clusters.
- Minimum or single linkage clustering:
  It produces comparatively loose clusters while computing all pairwise dissimilarities betIen 2 clusters and takes the loIst value as the distance betIen the 2 clusters.
- Mean or average linkage clustering:
  Clusters produced can vary in their compactness. It calculates all pairwise dissimilarities betIen 2 clusters and takes the average value as the distance betIen the 2 clusters.
- Ward’s minimum variance method:
  Minimises the total within-cluster variance. At each step, the pair of clusters with the smallest betIen-cluster distance are merged. Tends to produce more compact clusters

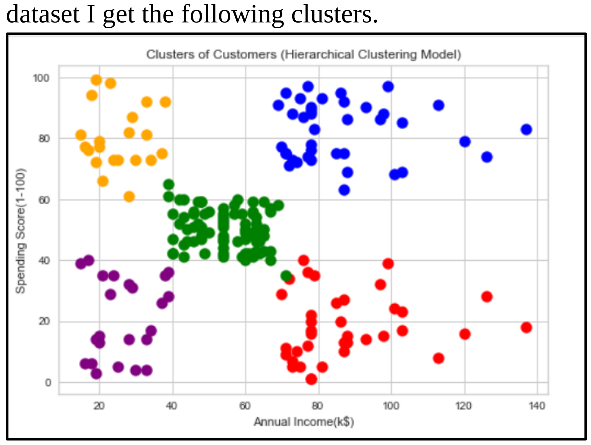
I made use of the following parameters in our project**:**

- **n_clusters=5:**It defines the number of clusters, and I have taken here 5 because it is the optimal number of clusters.
- **affinity='euclidean':** It is a metric used to compute the linkage.
- **linkage='ward':**It defines the linkage criteria, here I have used the "ward" linkage. This method is the popular linkage method that I have already used for creating the Dendrogram. It reduces the variance in each cluster.

I started with producing a dendrogram for our dataset using the above mentioned parameters:



Using this dendrogram I have concluded the optimal number of clusters to be 5. By fitting the agglomerative hierarchical clustering into our chosen dataset I get the following clusters.



The agglomerative algorithm generated the following 5 clusters

- Customers with **low** annual income and **high** spending score
- Customers with **low** annual income and **low** spending score

- Customers with **medium** annual income and **medium** spending score
- Customers with **high** annual income and **low** spending score
- Customers with **high** annual income and **high** spending score

3) DBSCAN Clustering

DBSCAN stands for Density-Based Spatial Clustering of Applications with Noise and is one of the clustering algorithms implemented in the scikit-learn library. The DBSCAN algorithm is based on the notion of "clusters" and "noise". The idea is that for each point of a cluster, the neighbourhood of a given radius has to contain at least a minimum number of points.

Partitioning methods (K-means, PAM clustering) and hierarchical clustering work for finding spherical-shaped clusters or convex clusters. In other words, they're suitable only for compact and Ill-separated clusters. Moreover, they're also severely tormented by the presence of noise and outliers within the data.

Real-life data may contain irregularities, like:

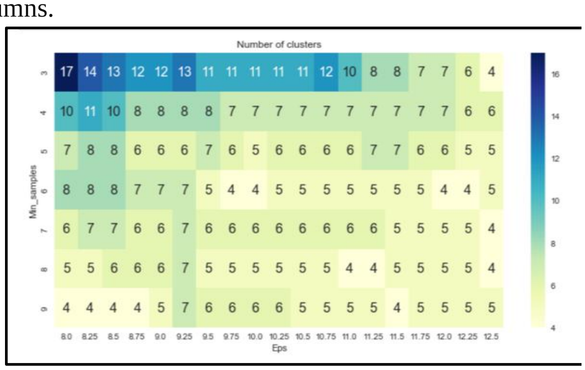
- Clusters can be of arbitrary shape like those shown in the figure below.
- Data may contain noise.

In DBSCAN there are two major tuning parameters:

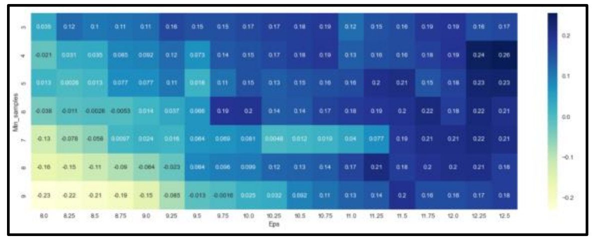
- **eps** – it defines the neighbourhood around a data point, that is if the distance betIen two point is loIr or equal to 'eps' then they're considered neighbours
- **min_samples** – Minimum number of neighbours within eps radius.

DBSCAN creates clusters itself based on those two parameters. So, I'll check the number of generated clusters now:

Pivot table in pandas is an excellent tool to summarise one or more numeric variables based on two other categorical variables. You can aggregate a numeric column as a cross-tabulation against two categorical columns.



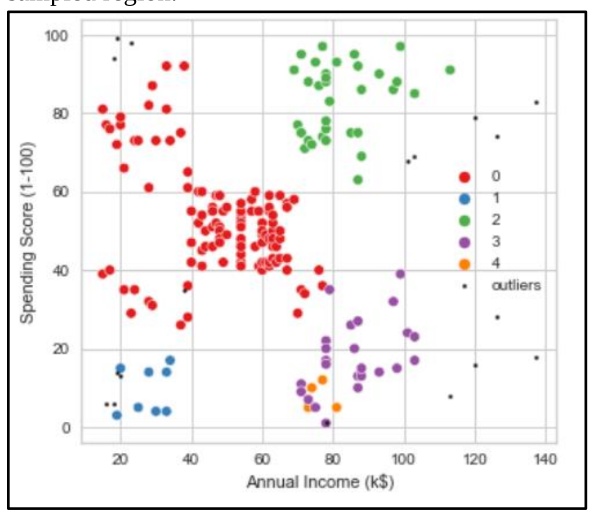
The heat plot above shoId that the number of clusters varies from 17 to 4. HoIver, predominantly the combinations give 4-7 clusters. So here I will use another metric, Silhouette score.

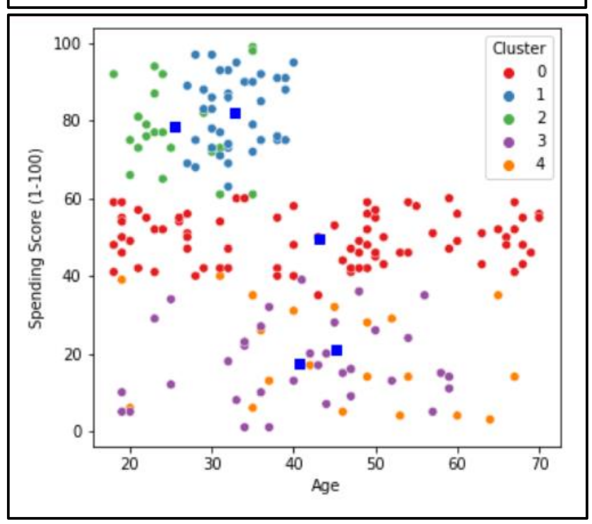


The global maximum is 0.26 for values of eps=12.5 and min_samples=4.

Now I will be checking the size of the clusters.

Total 5 clusters are created(0,1,2,3,4) plus some outlier clusters(-1). The size of clusters 0-4 varies remarkably from 4 to 112. Also, there are 18 outliers(-1). Outliers can be errors, coordinates with high uncertainty, or simply occurrences from an under-sampled region.





The above graph shows some outliers. These coordinates cannot be recognized as a cluster as they don't meet the distance and minimum samples requirements to be recognised as a cluster.

RESULT

The comparative study of plots of cluster formation through the 3 different clustering algorithms helps us conclude the types of customers visiting the mall, their habits and the patterns they follow. This led us to the conclusion of there being 5 optimal clusters through the combination of the results of both K-means and Agglomerative Hierarchical Clustering.

Cluster Analysis

Cluster 1:

Customers with low annual income and high spending score

This group is formed of people who earn less and are wise enough to also spend less. Malls will be least interested in them but can target them to visit for special occasions. This cluster contains about 11.5% of the total customers.

Cluster 2:

Customers with low annual income and low spending score

This group is formed of people who earn less but spend a lot. This group is an ideal one to be targeted as they are not too concerned about the amount of money they are spending and can easily be attracted to the mall's attractions. This cluster contains about 11.5% of the total customers.

Cluster 3:

Customers with medium annual income and medium spending score

This group is formed of people with relatively average annual income and also an average spending score. While they aren't the ideally targeted customers they will be taken into account in order to increase their medium spending score. This cluster contains about 39.5% of the total customers.

Cluster 4:

Customers with high annual income and low spending score

This group is formed of people who earn high salaries but spend less. This is a group of customers that can be targeted in order to find out why the mall's service isn't pleasing to them. They can be considered prime targets as they have the potential to spend quite a lot of money. This cluster contains about 18% of the total customers.

Cluster 5:

Customers with high annual income and high spending score

This group is formed of people who have high annual income and also spend a lot. This group is the ideal target for a mall and should be the ones mainly considered in order to bring in the most profit. This cluster contains about 19.5% of the total customers.

CONCLUSION

In this project I investigated the application of K-means, Agglomerative and DBSCAN Clustering in the customer segmentation of mall customers. I successfully identified five optimal customer groups in order to apply relevant marketing strategies accordingly. The sorted clusters will allow for better targeted marketing and will help meet the needs of every customer.

As I saw from the generated results by each of the 3 clustering algorithms, K- means and Agglomerative Hierarchical Clustering are better suited to our chosen dataset than DBSCAN Clustering. It has proven the importance of trying out all available methods before coming to an appropriate conclusion.

While K-means Clustering is an effective way to get optimal and reasonable clusters it is also confusing in its prediction of K- value. It also doesn't possess the ability to identify outliers in the data. Other than this it works accurately and speedily even on huge datasets and minimises the variance measures inside the clusters.

Agglomerative Clustering can by far be called the best clustering method to be applied on our chosen dataset. It was fast and could easily handle all the similarities and distances betIen the data-points without a dip in accuracy. The clusters formed by the dendrogram Ire optimal and produced understandable and reasonable groups of customers.

The last clustering method by us is the DBSCAN Clustering Algorithm. It turned out to be unsuitable for the dataset chosen by us due to its inability to provide useful clusters that could be used to take out meaningful inferences of the dataset.

It was discovered that the application of the three different algorithms was extremely important in order to figure out all possible characteristics of the data. Different properties of data are all suited for different data algorithms and open up the door to trying out other more complex algorithms in the future in order to try and get even better and more effective clustering.

MARKETING STRATEGY RECOMMENDATIONS

Keeping the cluster analysis in mind a few recommendations for effective marketing, by malls towards the groups of customers produced, can be made:

Cluster 1:

As these customers both earn and spend less, targeting them heavily will be ineffective and a waste of resources. It is best if they are targeted through bi-annual or holiday sales and discounts.

Cluster 2:

This group of customers earns less but spends more due to which targeting them will be a highly effective strategy. Reward point for a certain amount of money spent and discounts might be the way to go for this specific group of customers to spend even more in the mall.

Cluster 3:

This group of customers has average income and expenditure making them not the ideal target group but still something to be kept in mind. They can be reeled in through sales, discounts and rewards for money spent in the mall.

Cluster 4:

These customers form one of the prime target groups from this dataset. They earn a lot but spend very low making them an effective target due to their high potential of spending money. They can be marketed to by finding out the reason for their dissatisfaction with the mall's services and then providing more relevant attractions they would like to try out.

Cluster 5:

The final group consists of customers who have high income and also high spending scores. These are ideal targets for marketing and it is important to retain these customers in order to have good profits. The best way to market to them is keeping on top of their desires and the things they would like to be available and providing them in the mall

REFERENCES

[1]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.

[2] A. 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.

[3] 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.

[4] M. F. Diallo, F. Diop-Sall, S. Djelassi, and D. Godefroit-Winkel, "How Shopping Mall Service QualityAffectsCustomer 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.

[5] Azad Abdulhafedh, "Incorporating K-means, Hierarchical Clustering and PCA in Customer Segmentation", Journal of City and Development, 2021, Vol. 3, No. 1, 12-30, doi:10.12691/jcd-3-1-3

[6]James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. An Introduction to Statistical Learning: With Applications in R. Springer.

[7] Brett Lantz. 2019. Machine Learning with R. Packt Publishing Ltd.

[8] Alboukadel Kassambara. 2017. Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning. Sthda.com.

[9] Aurélien Géron. 2019. Hands-On Machine Learning with ScikitLearn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly.

[10] Philip D. Waggoner. 2020. Unsupervised Machine Learning for Clustering in Political and Social Research. Cambridge University Press.

[11] Bradford Tuckfield. 2019. Applied Unsupervised Learning with R: Uncover hidden relationships and patterns with k-means clustering, hierarchical clustering, and PCA. Packt Publishing Ltd.

[12] Tarek Amr. 2020. Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits: A practical guide to implementing supervised and unsupervised machine learning algorithms in Python. Packt Publishing Ltd.

[13] Morgan Maynard. 2020. Machine Learning: Introduction to Supervised and Unsupervised Learning Algorithms with RealWorld Applications.

[14] LazyProgrammer. 2016. Unsupervised Machine Learning in Python: Master Data Science and Machine Learning with Cluster Analysis, Gaussian Mixture Models, and Principal Components Analysis.

[15] RoIl Atienza. 2020. Advanced Deep Learning with TensorFlow 2 and Keras: Apply DL, GANs, VAEs, deep RL, unsupervised learning, object

detection and segmentation. 2nd edition, Packt Publishing Ltd.
[16] Fred Nwanganga and Mike Chapple. 2020. Practical Machine Learning in R. Wiley.
[17] Stephen Marsland. 2011. Machine Learning: An Algorithmic Perspective. Chapman and Hall/CRC.
[18] Steven L. Brunton and J. Nathan Kutz. 2019. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press.
[19] Michael Colins. 2017. Machine Learning: An Introduction to Supervised and Unsupervised Learning Algorithms.
[20] Chirag Shah. 2020. A Hands-On Introduction to Data Science. Cambridge University Press.
[21] Sunil Kumar Chinnamgari. 2019. R Machine Learning Projects: Implement supervised, unsupervised, and reinforcement learning techniques using R 3.5. Packt Publishing Ltd.
[22] Musthofa Galih Pradana , Hoang Thi Ha, ‘Maximising Strategy Improvement in Mall Customer Segmentation Using K-means Clustering’, Journal of Applied Data Sciences Vol. 2, No. 1, January 2021, pp. 19-25