Customer Segmentation Analysis Using Clustering Algorithms

Biyyapu Sri Vardhan Reddy, C. A. Rishikeshan, VishnuVardhan Dagumati, Ashwani Prasad, and Bhavya Singh

Abstract Customer segmentation has been deployed as a prudent marketing strategy by companies to ensure that their investments are less risky and more judicious. Segmenting customers helps the companies to divide the customers into groups that reflect similarity and maximize the value of each customer to the business. The main goal of this research is to use a machine learning clustering approach called K-means clustering to accomplish consumer segmentation. Besides, the research work is also focused on performing exploratory data analysis on the given dataset. To group the customers into clusters, a K-means clustering algorithm is performed on the customer dataset. To achieve optimization and validation of the clusters, popular heuristic, interpretation, and approximation methods have been included in this paper. Further, for analyzing and visualizing the important facets of the customer dataset and the operation of the K-means algorithm, the paper presents some colorful and informative representations. The implementation of this research work has been done in the R programming language. The outcome of this work includes visualizing the segments of the mall customers in the form of clusters based on their spending scores and annual incomes. Furthermore, a better customer segmentation could be achieved by taking product reviews and customer feedback into consideration. Nevertheless, customer segmentation remains a prospective topic for many researchers and companies due to dynamic customer behavior.

B. S. V. Reddy · C. A. Rishikeshan · A. Prasad · B. Singh
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, India
e-mail: biyyapu.srivardhan2019@vitstudent.ac.in

C. A. Rishikeshan
e-mail: rishikeshan.ca@vit.ac.in

A. Prasad
e-mail: ashwani.prasad2019@vitstudent.ac.in

B. Singh
e-mail: bhavya.singh2019@vitstudent.ac.in

V. Dagumati (✉)
Department of Computer Science and Engineering, School of Computing, Amrita Vishwa Vidyapeetham, Chennai, India
e-mail: ch.en.u4cse19004@ch.students.amrita.edu

1 Introduction

In the present day and age, the significance of dealing with the customers as the most prominent resource of a conglomerate is becoming greater in usefulness. Establishments are expeditiously providing capital for finer consumer accession, perpetuation, and evolution. The notion of corporate brainpower has a climacteric role to play for the probability of establishments using methodological prowess to take possession of preferable understanding for capacity-building programs. Hence through establishing it, in this framework the idea of customer segmentation analysis accumulates much attention due to it being an all-inclusive procedure of gaining and preserving consumers, making use of business intellect, with the purpose of augmenting the customer desirability for a corporate endeavor. One of the two most paramount intents of customer segmentation is consumer advancement with the aid of awareness. This intention of customer segmentation necessitates the employment of systematic outlook for the sake of accurately evaluating customers' particulars and exploration of the worth of consumers for greater customer insight. Keeping abreast of ever-changing times, establishments are altering their corporate drift prototypes through the deployment of systems engineering, change administration and planning out information technology responses that would be helping in gaining brand-new consumers, alongside keeping, and maintaining the existing consumer base and enhancing the customer's long-lasting merit. Because of the variety of products and facilities that are within an easy reach in the market and besides fierce competitiveness prevailing amongst diverse group of organizations, this form of customer relationship management has now come off as playing out quite an important role for the purpose of recognition and probing of a particular company's finest clients as well as the acquisition of superlative marketing methodologies in order to obtain and preserve competitive precedence [1]. Unsupervised learning is used for a variety of purposes, including customer segmentation. Customer segmentation would go on to enable organizations in finding copious categories of clients who think and operate distinctively and pay heed to diverse perspectives in their expenditure and procurement habits. Companies can discover numerous segments of clients with the utilization of K-means clustering technique, allowing them to target the possible user base by splitting it by gender, age, interests, and other buying habits. These techniques divulge inside correlative and outside miscellaneous categories. Consumers differ in their characteristics, needs and the primary objective of all types of clustering techniques is to find out consumer sets and classify the base into classes of like profiles for the purpose of carrying out target marketing in a much more well-organized manner. Data visualization would be done through differing visual elements in the form of plots and graphs for the provision of accessible ways to infer the trends, outliers, and patterns in the data [2]. K-means clustering, which is a vector

quantization method, would be partitioning the observations into clusters with each observation belonging to the cluster with the nearest mean. We would be finding out the optimal clusters using the following three methods- elbow, silhouette, and Gap Statistic. Finally, the visualization of the optimized clustering results would be performed with the utilization of the principal component analysis [3].

The paper is put in order as follows. Section 2 contains the related work i.e., the literature survey of the works corresponding to this paper's topic. Section 3 explains the data preparation and visualization part. Section 4 elaborates the proposed method. Section 5 summarizes the results and discussion. Section 6 has the conclusions in it and finally Sect. 7 enumerates all the references.

2 Related Work

Numerous works have been published in the field of customer segmentation. Different approaches yield different results with their unique advantages and disadvantages. For instance, Kansal et al. [3] have used three distinct approaches to compartmentalize the customers and those clustering techniques are- K-Means, bottom-up approach/ hierarchical agglomerative clustering (represented by dendrogram) and mean shift algorithm. In this work; random, unlabeled data has been used which would hence explain the employing of internal clustering validation that would pick out the most suitable technique which could accurately cluster the input data into its contrary cluster. Syakur et al. [5] have used the K-Means clustering technique in combination with the elbow method. The objective of picking out the Elbow method is to supplement the K-Means execution for the refining of the extensive quantity of data utilized. This would occur by deciding the best number of clusters from those rendered by the K-Means. Ezenkwu et al. [6] have introduced the application of MATLAB programming language for the purpose of carrying out the K-Means clustering method on the basis of data gathered from a MegaCorp clothes vocation. This research paper would conclude in the business orchestrating particular market schemes that would be appropriate for each one of its consumer sections. Kashwan et al. [7] have used the K-Means clustering algorithm and Statistical Package for the Social Sciences (SPSS) tool with the agenda of forecasting the sales of a specific market in varied yearly recurrent cycles. An analysis of variance was also carried out for the purpose of testing the stability of the clusters. The computing established set-up proved to be intuitive and provided outcomes to the managers for making swift and fast resolutions. Aryuni et al. [8] have fitted K-Means and KMedoids methods based on recency, frequency and monetary (RFM) analysis score in relation to an Internet banking system data. And the output showed that the former method has outperformed the latter one on the grounds of intra cluster i.e., AWC distance. Datta et al. [9] has used the K-Nearest Neighbor (KNN) algorithm. The K-nearest neighbors' technique is a supervised Machine Learning algorithm that may be used to predict both classification and regression issues. It's a straightforward notion that uses the concept of viewing surrounding data to classify any set of inputs. If K = 1, only

one neighbor is cross-checked, and the input is classified based on the class category of that neighbor. When K = 4, four of the neighbors are examined, and the appropriate class is assigned. If an appropriate value of K can be roughly calculated, this is a proper approach. Rizki et al. [10] has used two methods for Customer Loyalty Segmentation on a Point-of-Sale System; they are Recency-Frequency-Monetary (RFM) model and K-Means algorithm.

Besides K-Means Clustering and related approaches, other techniques for segmentation were also used in several research works. For instance, Song et al. [11] considered feature selection to be an important aspect of customer segmentation. They used one of the advanced techniques to perform customer segmentation, called hydrological cycling optimization (HCO). This method was based on a meta-heuristic approach. The proposed method was able to evolve a set of non-dominated solutions with a smaller number of features which yielded highly accurate results. Wu et al. [12] integrated churn prediction with customer segmentation. Their area of focus lies in a telco industry, where churn management becomes crucial. They implemented multiple ML classification algorithms to perform the churn prediction including K-means clustering. Further, to conduct the factor analysis and identify some key features for turnover customer segmentation, Bayesian Logistic Regression was used. Manjunath et al. [13] proposed a multi-layer hierarchical super peer P2P network architecture to perform the distributed clustering problem involving customer segmentation. Their methodology was in contrast to the centralized clustering approach. The method allowed the flexibility to use different datasets of varying sizes. José J. López et al. [14] aim at making provisions of electric efficacies with a bulk of data on the basis of customer classification to facilitate them to set up distinct sorts of excise tariffs. They have used the ensuing methods to classify the electricity clients- hierarchical clustering, modified to follow the leader and K-Means. Their proposition eradicates the volatility of the preliminary solution and advances towards the global optimum. Maree et al. [15] discovered that the clusters created aren't discriminating enough for micro-segmentation. As a result, they concentrated on extracting temporal features with continuous values from hidden states of neural networking in order to forecast client spending behavior in terms of transactions. They created micro-segments and course segments using Long Short-Term Memory (LSTM) and feed-forward neural networks, respectively.

The motivation for this research work includes the following:

1. Major emphasis on data visualization of the attributes required for customer segmentation.
2. Re-defining the use of K-Means Clustering algorithm with optimization so that the results are comparable with recent approaches for customer segmentation. If K-Means clustering algorithm is used all by itself, as seen in some of the related works, then there are highly probable chances of it producing high fallacies and mediocre cluster results, that is due to it being a localized maximization method.
3. We have some rudimentary details about the customers like age, Customer ID, gender, yearly earnings and spending score. We ought to understand the customers like who the target customers are, so that the perception can be made

Table 1 Mall Customer Dataset

Customer ID	Gender	Age	Annual Income (k$)	Spending Score (1–100)
1	Male	19	15	39
2	Male	21	15	81
3	Female	20	16	6
4	Female	23	16	77
5	Female	31	17	40
6	Female	22	17	76
…	…	…	…	…

as a provision to the marketing team who can eventually sketch their strategy correspondingly to maximize their profits.

3 Data Preparatıon and Vısualızatıon

In this section, the dataset collected for performing customer segmentation is explored. Further, the visualizations of some fundamental attributes related to customers at a typical shopping complex is carried out using insightful and varied forms of graphs.

3.1 Data Exploration

The dataset is collected from a typical mall. There were 200 customers in total, who were analyzed based on their gender, age, annual income (per 1000 $ or k $) and based on their spending behavior, each customer was given a spending score out of 100. The dataset does not contain any redundant values or null values. The completed dataset consists of 200 rows excluding the header row and 5 columns in total. The first 6 records of the dataset are shown in Table 1 for reference.

3.2 Visualization of Attributes

In this section we use each attribute of our dataset for visualization purposes like Age, Gender, and annual income in 1000$ or k$. In this section, there will be four divisions (subsection) each section consists of an attribute.
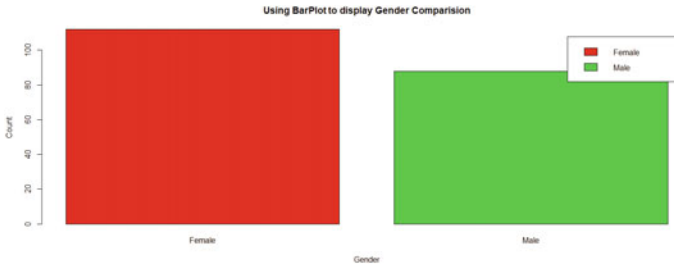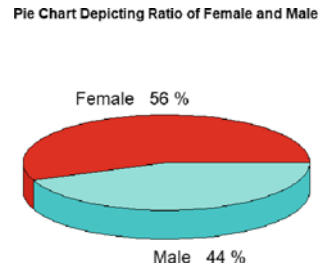
Fig. 1 Bar plot to display gender comparison

Fig. 2 Pie chart to display
ratio of male and female

3.2.1 Data Visualization Using Customer Gender

A box plot and pie chart is create to to illustrate using mall_customer datset. For the bar graph (as show in Fig. 1), we are using red color for female and green color for male it will be useful for gender comparison. Then, for the pie chart (as show in Fig. 2), we are using red to show female percentage and sky blue for male percentage.

Now, from bar plot (as show in Fig. 1), we come to know that the females are more in number than male persons. Then From the above pie-chart (as show in Fig. 2), we can say that the female's percentage is 56%, and the male percentage in the mall customer dataset is 44%.

3.2.2 Data Visualization Using Customer Age

A histogram is plotted using (as show in Fig. 3) using Age attribute and a box plot to know which age group members are usually coming to the mall and the age of customers coming minimum and maximum to the mall using histogram and a box plot. We are using pink color for histogram and deep pink for box plot. For knowing the hex value of deep pink, we use color picker tool from the internet, then we get "#ff00ee" we use this to get a deep pink color in the box plot.

From the above two plots (as show in Fig. 3 and Fig. 4), i.e., the box plot and the histogram, we can deduce that the maximum customer ages are between 30 and 35, while the minimum and maximum client ages are 18 and 70.

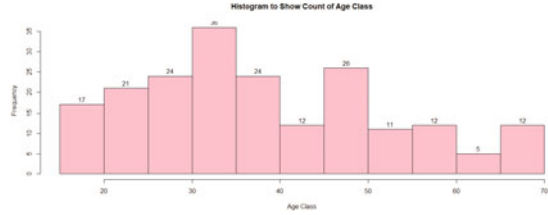**Fig. 3** Histogram plot to display count of age class

Fig. 4 Boxplot for displaying detail analysis of age a

3.2.3 Data Visualization Using Customer Annual Income (k$)

In addition, we will construct visualizations to analyze the annual income of our customers. So, first, we'll make a histogram plot, and then we'll move on to a density plot. The red color is used in the histogram (as show in Fig. 5). We are getting hex value in the same way above as we used in age histogram (as show in Fig. 3). Then we get "#ff002f". Then for density graph (as show in Fig. 6) we are using light yellow color for that we get "#f2ff00".

From the above analysis (as show in Fig. 5 and Fig. 6), we can say that the minimum and maximum customers annual income is 15 and 137. Customers (people) with an average annual income of $70 are the most frequent in the histogram. We can also state that the average salary of clients is $60.56. In the preceding Density Plot, we can see that the annual income of the consumers follows a normal distribution.
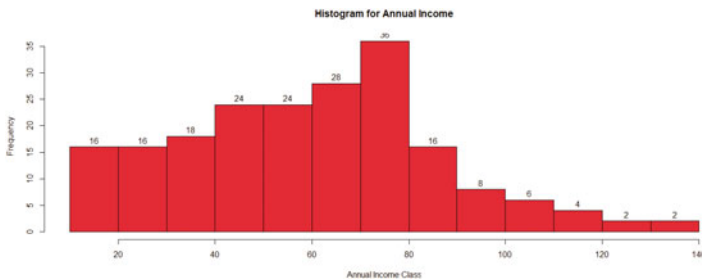

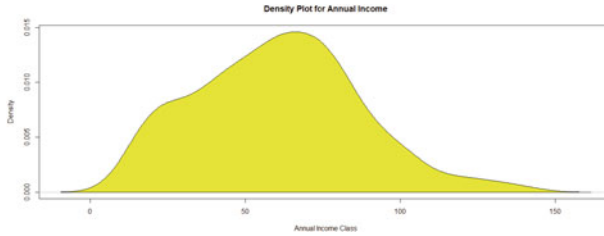**Fig. 5** Histogram plot to display count for annual income

Fig. 6 Density Plot display detail analysis annual income

3.2.4 Data Visualization Using Customer Spending Score (1–100)

A histogram is plotted using using the spending score attribute and a box plot to know the minimum and maximum spending score of customers. We are using orange color for boxplot (as show in Fig. 7) and blue for histogram plot. For knowing the hex value of orange, we use color picker from net then we get "#ffa200" for orange and "#0099ff" for blue which we are going to use in the histogram (as show in Fig. 8).

Now, from the above analysis (Fig. 7 and Fig. 8), we can say that the minimum, maximum and average spending score is 1, 99 and 50.20. And we learned from the histogram plot that clients in the 40–50 age range have the highest spending score of all the classes, which is 40.
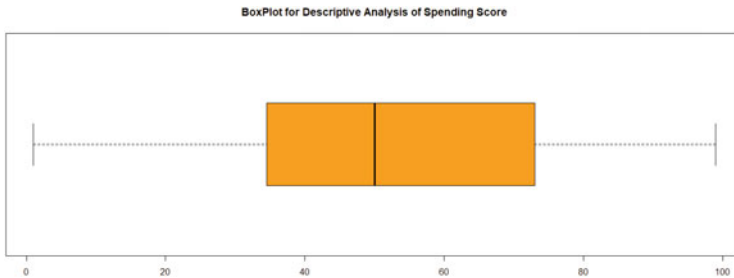

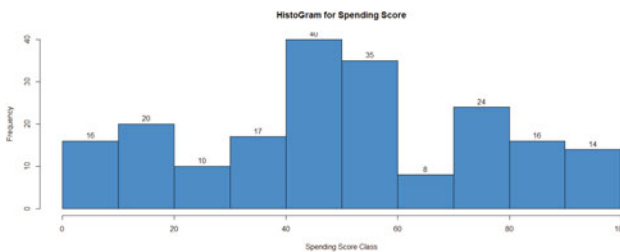**Fig. 7** Boxplot for detail analysis of spending score

Fig. 8 Histogram plot to show detail analysis of spending score

4 The Proposed Method

First, after collecting the dataset and data preparation, we would be performing data exploration. Then we would go on to visualize and analyze the data in R to get some necessary insights. Henceforth, we would be making use of the K-Means clustering algorithm in order to determine the optimal clusters representing different customer bases. Finally, the visualization of the clustering outputs would be done by making use of the Principal Components Analysis. The proposed methodology is presented in the form of a flow chart as show in Fig. 9.

4.1 K-means Clustering Algorithm

While we are in the phase of the utilization of the K-Means clustering algorithm, the pre-eminent step is to depute the quantity (number) of clusters that we are preparing to produce in the eventual output. The algorithm would begin by randomly selecting 'n' objects from the dataset to serve as early cluster centers. These picked objects are the cluster means, which are also known as centroids in mathematics. The selected objects would then be assigned to the next closest centroid in the next stage, cluster assignment. The Euclidean distance, or the distance of a line segment between two

Fig. 9 Flow chart representing the proposed method
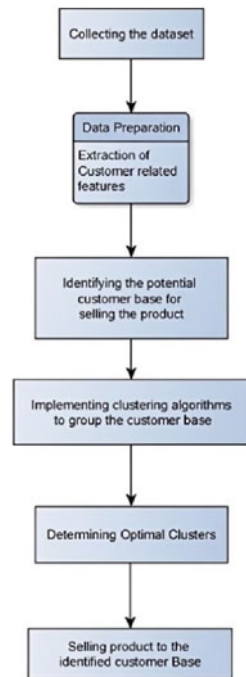
points in Euclidean space, reveals the closest centroid. When this step is completed, the algorithm will compute the new mean value for each cluster in the data. The observations are analyzed to see if they are closer to a distinct cluster after all of this re-computation of the centers' values. The items will be reassigned after the cluster mean has been renewed. This step would be repeated several times until the cluster assignments stop getting changed i.e., the clusters existing in the present iteration are equivalent to those which we received in the former iteration.

The mathematical expressions used, corresponding to the above elaborated K-Means clustering algorithm are as follows:

$$Euclidean\ Distance, d(p, q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

where p and q are the two points in the Euclidean space.

Also, we know the formula of mean is given by:

$$\bar{x} = \frac{(x_i + x_j)}{2}$$

4.2 Cluster Optimization

One needs to enumerate the number of clusters to be made use of beforehand itself. For this we would be using the optimal number of clusters. Clustering techniques' major goal is to define clusters in such a way that intracluster variation is kept to a bare minimum.

$$Minimum\ (sum\ C(W_n))\ having\ n = 1, 2, 3 \ldots n$$

where, W_n is the nth cluster and $C(W_n)$-intracluster variation.

For the stated-out purpose we would be making use of the following three prominent methods:

- Average Silhouette method
- Gap Static method
- Elbow method

4.2.1 Average Silhouette Method

Utilizing this method, we would be calculating the attribute of the clustering performance. Good clustering is indicated by a high average silhouette breadth. For different 'n' values, this approach computes the average of all the silhouette observations. We would maximize the average silhouette over noteworthy values for k clusters. We

utilize the silhouette function and k-mean function for the purpose of calculating the average silhouette width.

4.2.2 Gap Statistic Method

With the usage of this method, we can contrast the aggregate intracluster disparity for distinct values of 'n' besides their contemplated values coming under the null reference distribution of the whole of the data. By using the Monte Carlo simulations, we can generate the trial dataset. Then we can compute the range between minimum and maximum for every variable in the generated dataset. Through this process one can generate values evenly starting from the lower bound up till the upper bound. We are using the "clusgap" function for the purpose of producing Gap Statistics alongside the standard error for a given result.

4.2.3 Elbow Method

First and foremost, we'll compute the clustering technique for a variety of 'n' values. The former can be accomplished by varying the number of clusters inside n from one to 10. The aggregate intra-cluster sum of squares is then computed (s). We would be plotting it on the basis of the number of 'n' clusters which would be indicating the suitable number of clusters needed in our representation. The optimum number of clusters is designated by a bend's location.

5   Results and Discussion

As discussed in the above section there are 3 methods, we are using in this optimal clustering so, here in this section we are going to discuss the results of those methods in 3 sub divisions and another subdivision for visualizing the clusters.

5.1   Elbow Method

In this section we are going to show the result for the Elbow method, so that we are taking the number of clusters on the x-axis and total intra-clusters sum of squares on the y-axis and then start the plotting. The resultant graph is shown in Fig. 10.

Now, based on Fig. 10, we may deduce that k = 6 is the correct number of clusters because it appears at the elbow bend in the graph above.
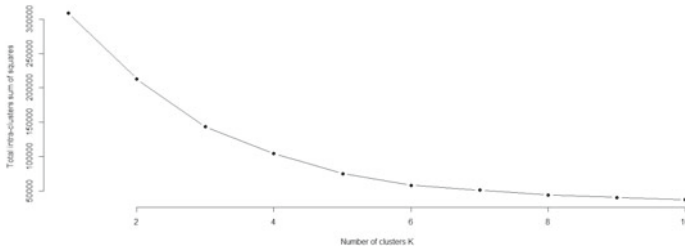
Fig. 10 Line graph using Elbow method

5.2 Average Silhouette Method

In this section we are going to show the result for the Average Silhouette method. So, first we need to draw silhouette plots (as show in Fig. 11 and Fig. 12) for different numbers of clusters to know the maximum average. So, we start with k = 2 to k = 10 because we used up to 10 clusters in the above elbow graph as well and record the average values.
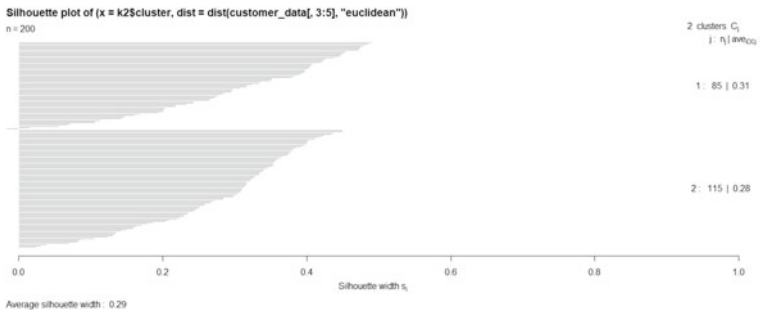

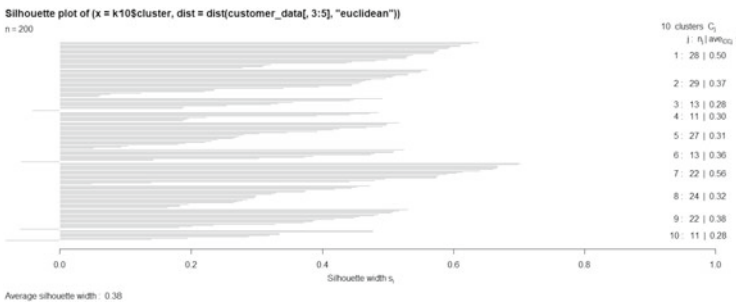**Fig. 11** For k = 2 clusters, the average silhouette plot is shown

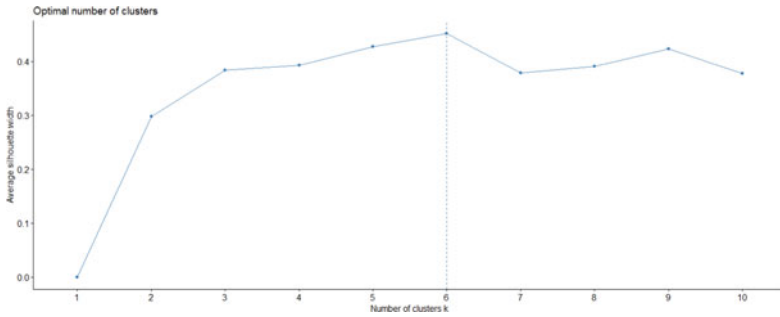**Fig. 12** For k = 10 clusters, the average silhouette plot is shown

Fig. 13. Optimal clusters graph using Average Silhouette method

Here, the averages we got are 0. 29, 0.38, 0.41, 0.44, 0.45, 0.44, 0.43, 0.42, 0.38 respectively from k = 2 to = 10.

Now, by using them we are going to visualize the optimal clusters, so, for that on the x-axis we are taking the number of clusters(k) and on the y-axis we are taking average silhouette width. The resultant graph is illustrated in Fig. 13.

From the above graph (as show in Fig. 13), we observe that the k = 6 is seeming to be the correct value for the number of clusters, because it is having the highest average silhouette width.

5.3 Gap Statistic Method

In this section we are going to display the results for the gap static method. For plotting, we take the number of clusters on the x-axis and the gap static(k) on the y −axis. Then we get the following graph as show in Fig. 14.

From Fig. 14, we can conclude that there are 6 optimal clusters. The trial datasets are genrated using by Monte Carlo simulations.
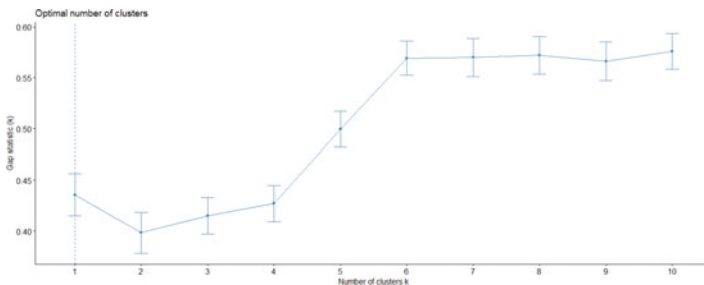

Fig. 14 Optimal clusters graph using Gap Static method

5.4   Visualization and Analysis of the Clustering Results

From the above results we know that the optimal clusters are 6. So, we are going to plot the 6 clusters in a scatter plot using the "gg-plot" package in RStudio. So, we are taking annual income on x-axis and spending score on y-axis (as show in Fig. 15). To draw segmentations of mall customers using K-Means clustering, we are taking annual income on x-axis and age on y-axis to draw segmentations of mall customers (as show in Fig. 16). Then finally we are taking classes on x-axis and k-means on y-axis to show final K-Means clusters, which is visualized in the form of a scatter plot as show in Fig. 17.

From Fig. 15, we observe that:

1) Cluster 1 - Customers with a high annual income and a high annual expenditure make up this cluster.
2) Cluster 2 - This cluster is characterized by a high annual income and a low annual outlay.
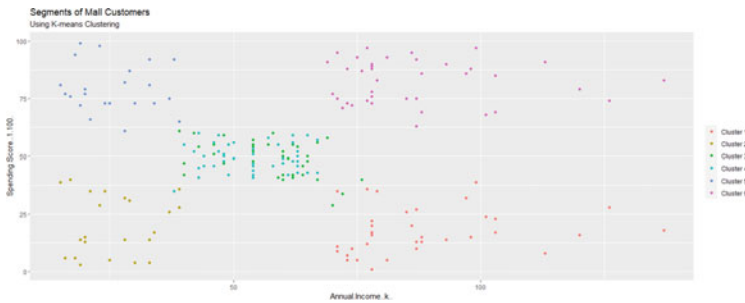3) Cluster 3 - Customers with a low yearly income and a low annual income spend are represented in this cluster.

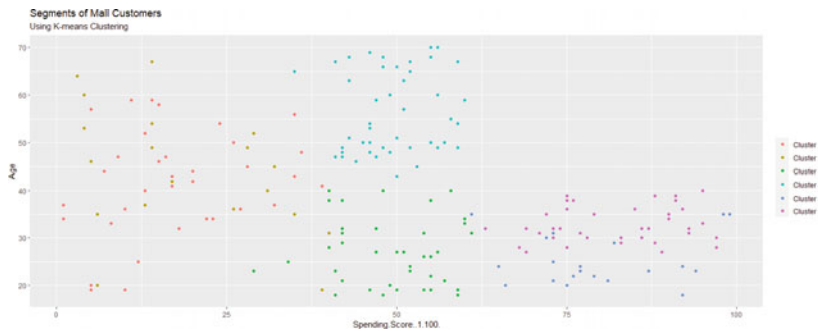**Fig. 15.** 6-optimal clusters using spending score and annual income

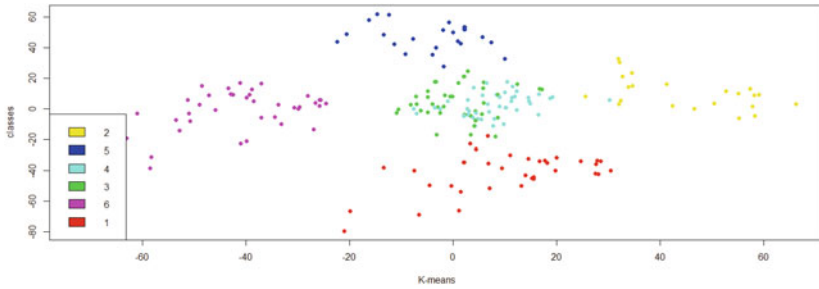**Fig. 16.** 6-optimal clusters using age and spending score

Fig. 17 Final 6-optimal clusters using k-means and classes

4) Cluster 5 - People in this cluster have a low annual income but a high annual spending.
5) Customers in Clusters 6 and 4 have a medium annual salary spend as well as a medium annual salary income.

Further, in Fig. 16, we see that the data is scattered randomly all over the scatter plot. Strong clustering is therefore difficult to achieve as the points are not close together. We can also conclude that there are several people coming to shop at malls having different ages (both young and old) with varying economic status.

From Fig. 17, we can analyze the clusters in terms of Principal Component Analysis (PCA) score. PCA score is of two types: PCA1 and PCA2. PCA1 is the linear combination with the largest possible explained variation, and PCA2 is the best of what's left.

1) Customers with a medium PCA1 and medium PCA2 score make up Clusters 4 and
2) Clients in Cluster 2 have a high PCA2 but a low yearly income spending.
3) Cluster 3 is made up of customers with a high PCA1 and PCA2 income.
4) Customers in Cluster 5 have a medium PCA1 score but a bad PCA2 score.
5) Cluster 6 - This cluster represents customers with a high PCA2 and a low PCA1.

6 Conclusion and Future Work

In this research work, we were successfully able to perform a customer segmentation on a group of 200 customers. We gave the visualization flavor to our research work by incorporating colorful and insightful data visualization graphs. They helped us gaining insights into the dataset at a glimpse. Further, the principal approach used for the segmentation of the mall customers was the K-Means clustering. Clusters were visualized in the form of scatter plots and thus, we were able to group the customers into six groups based on certain attributes used in the dataset. To improve the clustering process, we also ingeniously implemented a few optimization techniques for achieving better results.

For future work, we can work upon a more complex dataset consisting of thousands of customers. Further, the dataset to be worked upon can also include several other realistic attributes like customer survey data and customer staying time in malls.

References

1. Kashwan KR (2013) Customer segmentation using clustering and data mining techniques article in. Int J Comput Theory Eng. https://doi.org/10.7763/IJCTE.2013.V5.811
2. Li Y, Chu X, Tian D, Feng J, Mu W (2021) Customer segmentation using K-Means clustering and the adaptive particle swarm optimization algorithm. Appl Soft Comput 113:107924. https://doi.org/10.1016/J.ASOC.2021.107924
3. Kansal T, Bahuguna S, Singh V, Choudhury T (2018) Customer segmentation using K-means clustering. In: Proceedings of the international conference on computational techniques, electronics and mechanical systems, CTEMS 2018, pp 135–139. https://doi.org/10.1109/CTEMS.2018.8769171
4. https://www.kaggle.com/datasets/shwetabh123/mall-customers
5. Syakur MA, Khotimah BK, Rochman EMS, Satoto BD (2018) Integration K-means clustering method and elbow method for identification of the best customer profile cluster. In: IOP conference series: materials science and engineering, vol 336, no 1. https://doi.org/10.1088/1757-899X/336/1/012017
6. Ezenkwu C, Ozuomba S, Kalu C (2021) Application of K-Means algorithm for efficient customer segmentation: a strategy for targeted customer services. Accessed 21 Nov 2021
7. Kashwan KR, Velu CM (2013) Customer segmentation using clustering and data mining techniques. Int J Comput Theory Eng 856–861. https://doi.org/10.7763/IJCTE.2013.V5.811.
8. Aryuni M, Madyatmadja ED, Miranda E (2018) Customer segmentation in XYZ bank using K-means and K-medoids clustering. ieeexplore.ieee.org. https://doi.org/10.1109/ICIMTech.2018.8528086
9. Datta D, Agarwal R, David PE (2020) Performance enhancement of customer segmentation using a distributed python framework, ray. papers.ssrn.com. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3733832. Accessed 21 Nov 2021
10. Rizki B, Ginasta NG, Tamrin MA, Rahman A (2020) Customer loyality segmentation on point of sale system using recency-frequency-monetary (RFM) and K-Means 5(2):130–136. https://doi.org/10.15575/join.v5i2.511. join.if.uinsgd.ac.id
11. Song X, Liu MT, Liu Q, Niu B (2021) Hydrological cycling optimization-based multiobjective feature-selection method for customer segmentation. Int J Intell Syst 36(5):2347–2366. https://doi.org/10.1002/INT.22381
12. Wu S, Yau WC, Ong TS, Chong SC (2021) Integrated churn prediction and customer segmentation framework for telco business. IEEE Access 9:62118–62136. https://doi.org/10.1109/ACCESS.2021.3073776
13. Kuruba Manjunath YS, Kashef RF (2021) Distributed clustering using multi-tier hierarchical overlay super-peer peer-to-peer network architecture for efficient customer segmentation. Electron Commer Res Appl 47. https://doi.org/10.1016/J.ELERAP.2021.101040
14. López JJ, Aguado JA, Martín F, Muñoz F, Rodríguez A, Ruiz JE (2011) Hopfield–K-Means clustering algorithm: a proposal for the segmentation of electricity customers. Electric Power Syst Res 81(2):716–724. https://doi.org/10.1016/J.EPSR.2010.10.036
15. Maree C, Omlin CW (2021) Clustering in recurrent neural networks for micro-segmentation using spending personality. Accessed 21 Nov 2021