Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

MALL CUSTOMER SEGMENTATION USING K-MEANSCLUSTERING

Narayanam Prudhvish , Bharath kumar Uppala
, L.V.Tharun Kumar , Jayanthi J
Department of Computer Science andEngineering,
Kalasalingam Academy of Researchand Education,
Virudhnagar, Tamil nadu ,India

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 wayof representing and defining the customer needs. K-means clustering is an algorithm which is used to performthe 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 predictingthe behavior, introducing better options and things to customer

Key words: Target Customers, Clusters, Segmentation, Mall Basket AnalysisIntroduction.

I. Introduction:

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 smoothmanner. Managing the customer relationship develops the business intelligence. The significance of treating the consumers / Customers as an organization is important asset is growing in value in the present day and era. Now-a-days Organizations are layout money quickly in the development of the customer acquisition, Maintenance, Features and development master plan.

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 foroverstep.

With the use of clustering methods, customers withsimilar 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 hasresulted 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.

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 marketingteam to implement effective marketing strategy plan.

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.

II. Related Work

Sanjana et al [1] offered different clustering algorithms that took into account Big Data properties such size, noise, dimensionality, algorithm calculations,and cluster structure, as well as a brief introduction ofthe field.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

Partitioning, hierarchical, and other clustering techniques are listed here. Algorithms based on density, grids, and models Merged clustering of fuzzy c-means and genetic clustering, Azarnoush Ansari et al [2],consumers in the cluster steel sector with algorithms The LRFM variables are used. Customers were segmented using the (length, recency, frequency, monetary value) approach. Into two groups.

Pedro Quelhas Brito et al. [3] looked into several methods to data. Customer segmentation, clustering, and sub-group identification are all possible using data mining. The Six market groups and 49 rules were built using the models that were generated in a highly efficient manner-tailoring (customised fashion maker)enabled for a better fit. Consumer preferences are understood.

Shreya Tripathi et al [4] looked into the importance of client segmentation utilising clustering techniques as a major CRM feature. The pros and disadvantages of the most commonly used K-Means and Hierarchical Clustering algorithms were examined. Finally, the ideaof developing a hybrid strategy is addressed by combining the two strategies above, which has the potential to outperform the individual approaches.

Kishana R. Kashwan et al [5] presented a complete report using k means clustering approach with the SPSS Tool to construct a real-time and interactiveframework for forecasting sales in multiple annual seasonal cycles for a given supermarket. The prototype created was a smart gadget that took inputs from sales data records at the end of the day and automatically changed segmentation statistics. To testthe cluster's stability, an ANOVA study was also done. Actual sales figures Day-to-day figures are compared to the model's predicted statistics.

The results were good and showed a high degree of precision.

III. Methodology

a. Clustering:

Clustering is one of the most utilized approaches in data exploration to gain a clear grasp of the data structure. It can be defined as the task of locating subgroups within a large dataset. Similar data is grouped together in the same subgroup. A cluster is a collection of aggregated data elements that share characteristics. Clustering is a technique used in market basket research to categorize customers based on their behavior.

b. K-means Clustering Algorithm:

K Means Clustering is perhaps the most frequent andsimple Machine Learning algorithm, and it employs aniterative strategy to partitioning the dataset into various "k" number of specified and non-overlapping subgroups, with each data point belonging to just one of them.

IV. Proposed System:

It is a web tool for mall customer segmentation that allows retailers to promote their products based on apredetermined strategy. The image format is used tostore the cluster generated by the application.

a. Login Module

The marketing team logs into the programme with theusername and password, and the marketing team must register with the system each time they want tosee the details. Stimulus: A member of the marketing team inputs the username and password. Response: Opens the Mall Customer SegmentationRegistration page.

b. Register Mode:

The online application has a registration option for the marketing team. After successfully logging into the system, the marketing team must register with the system each time they want to see the details. Stimulus: The marketing team enters the registrationinformation. Response: Navigates to the Upload Dataset Module.

c. MCS-Data set upload Module

The Marketing team is given the MCS dataset feature, which allows the dataset to be imported and routed to the K-Means execution module. The MCS dataset willbe imported in this module, and after the display imported successfully message, it will be routed to theK-Means Algorithm execution module.
The marketing team imports the dataset as astimulant.
Response: The system successfully imported themessage dataset.

d. MCS-K-Means Algorithm Module

The K-Means method will be used, which produces fivedistinct groups that represent customers based on their spending score and annual income. Stimulus: The marketing team activates the K-Means algorithm. It navigates to the visualization page as a result of theresponse.

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

e. Customer segmentation

Because of the intense rivalry in the business sector, businesses have had to improve their profitability andbusiness throughout time by satisfying client requestsand attracting new customers based on their wants. Customer identification and meeting each customer's needs is a difficult and time-consuming task. This is because clients differ in terms of their wants, tastes, and preferences, among other things. Customer segmentation, as opposed to a "one-size-fits-all" strategy, separates customers into groups with similarfeatures or behavioral characteristics.

The information employed in the customer segmentation technique, which divides customers intogroups, is based on a variety of elements, including data geographical circumstances, economic conditions, demographical conditions, and behavioral tendencies. The customer segmentation technique enables a company to make better use of its marketing dollars, acquire a competitive advantage over competitors, and demonstrate a deeper understanding of the client's needs. It also aids a company in improving marketing efficiency, recognizing new market opportunities, developing a stronger brand strategy, and assessing client retention.

f. Elbow Method

The elbow approach is a tool for analyzing the clusterscreated by our dataset and assisting in interpreting theappropriate number of ideal clusters in the dataset. The best number of clusters for our dataset isdetermined 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 dueto the fact that larger clusters allow for the capture offiner groups of data objects that are more comparable to one another.

g. Visualization Module

The visualization module returns results based on theclusters listed below. The findings are generated as agraph and saved as an image, which is then retrievedby the marketing team.

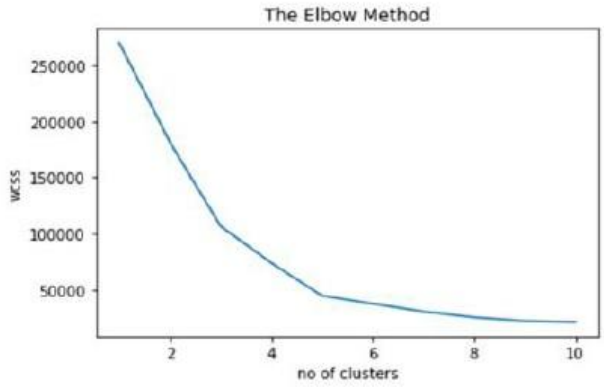


Figure 1: Elbow Method

Elbow method which shows the optimal number ofclusters.

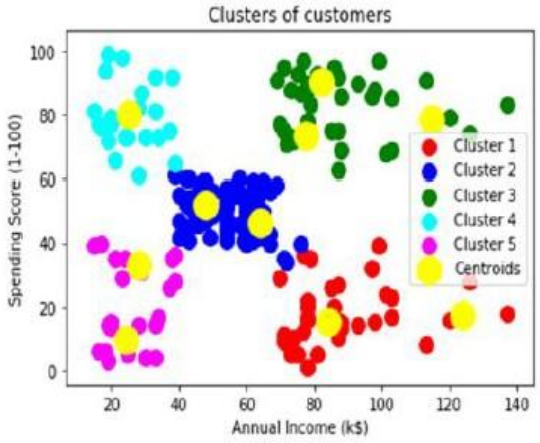


Figure 2: visualization customer

h. Algorithm

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

Cluster 1 (Red) implies earning a lot while spendingless.
Cluster 2 (blue) reflects the mean in terms of earningsand spending.
Cluster 3 (Green) shows both high earnings and significant spending. [Prospective customers] Cluster 4 (blue) denotes earning less but spendingmore.
Cluster 5Earning less and spending less is representedby (magenta color).
Stimulus: The marketing team activates the K-Meansalgorithm.
Response: The findings are generated as a graph andsaved as a picture. K=5 for clusters
Result

Visualize the gender of customers:

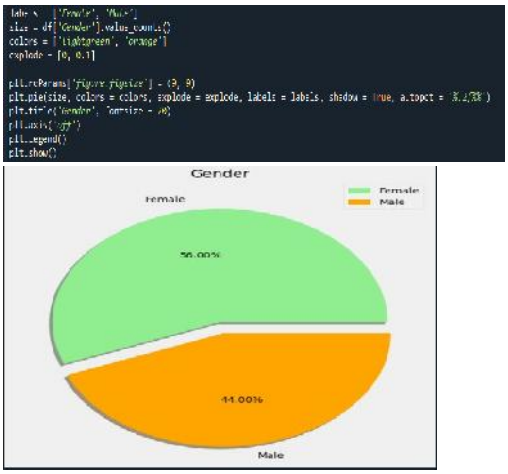


Figure 3: Gender visualization

By glancing at the above pie chart, you can see how gender is distributed in the mall.
Interestingly, Females have a 56 percent part of the vote, while Males have a 44 percent share. This is a significant disparity, especially given that Males have abigger population than Females**.**

Distribution of Age and Annual income:

The distribution pattern of annual income and age canbe seen in the plots above. We may deduce one thing from the plots: there are few persons who earn more than $100 US Dollars. Most folks earn between $50 and $75 per month. We can also state that the lowestincome is roughly $20 USD.

The graph implements the annual income of the customer with range and count of the customer. The graph implements the distribution of age with count and range of age. Each graph shows kinds of people are vising to the mall by their needs.so company sales execute implement the needs based on their count and annualincome

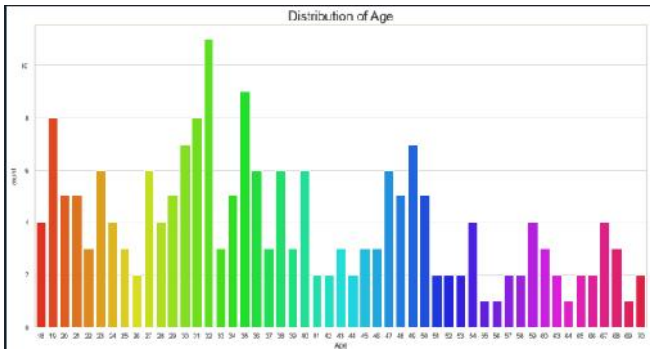
```
import warnings
warnings.filterwarnings('ignore')

plt.rcParams['figure.figsize'] = (18, 8)

plt.subplot(1, 2, 1)
sns.set(style = 'whitegrid')
sns.distplot(df['Annual Income (k$)'])
plt.title('Distribution of Annual Income', fontsize = 20)
plt.xlabel('Range of Annual Income')
plt.ylabel('Count')
plt.subplot(1, 2, 2)
sns.set(style = 'whitegrid')
sns.distplot(df['Age'], color = 'red')
plt.title('Distribution of Age', fontsize = 20)
plt.xlabel('Range of Age')
plt.ylabel('Count')
plt.show()
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

.



Representation of distribution of age and annualincome by bar graph.

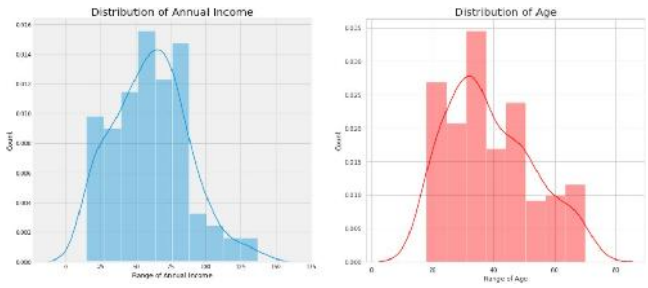


Figure-4: Distribution of annual income and Age

Distribution of Age:

```
plt.rcParams['figure.figsize'] = (15, 8)
sns.countplot(df['Age'], palette = 'hsv')
plt.title('Distribution of Age', fontsize = 20)
plt.show()
```

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

This Graph depicts an Interactive Chart depicting the distribution of each Age Group in the Mall for a betterunderstanding of the Mall's Visitor Age Group.
Looking at the graph above, it can be observed that theages of 27 to 39 are relatively common, but there is no apparent pattern; instead, we can only detect some group-wise patterns, such as the older age groups being less prevalent in contrast. The Mall has an equal number of visitors aged 18 and 67, which is an interesting fact. Malls are far less frequented by people aged 55, 56, 69, and 64. The Mall's most frequent visitors are people between the ages of 32 and 34.

Figure-5: Distribution of AgeDistribution of annual income:

```
plt.rcParams['figure.figsize'] = (20, 8)
sns.countplot(df['Annual Income (k$)'], palette = 'rainbow')
plt.title('Distribution of Annual Income', fontsize = 20)
plt.show()
#########
```

Again, here is a chart to better show the Distribution of Each Income Level. It's interesting to see that there areconsumers in the mall that have quite similar annual incomes ranging from 15 to 137 thousand dollars.
There are more customers in the mall with annualincomes of 54, 000 or 78, 000 dollars.

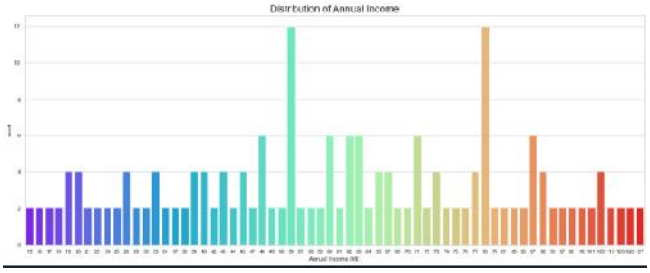


Figure-6: Distribution of IncomeDistribution of Spending Score:

This is the Most Important Chart from the Mall's Point of View, as it is critical to have some intuition and concept about the Mall's Customers' Spending Score. On a broad level, we can deduce that most customers have a Spending Score of 35-60. Clients with an I am spending score and a 99-spending score are also present, demonstrating that the mall caters to a diverse range of customers with varying demands andpreferences.

```
#########
plt.rcParams['figure.figsize'] = (20, 8)
sns.countplot(df['Spending Score (1-100)'], palette = 'copper')
plt.title('Distribution of Spending Score', fontsize = 20)
plt.show()
#####
```

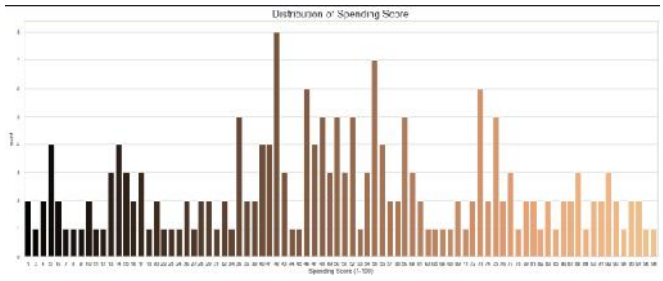


Figure-7: Distribution of spending score

Gender V/s Spending Score:

```
# Gender vs Spendscore

plt.rcParams['figure.figsize'] = (18, 7)
sns.boxenplot(df['Gender'], df['Spending Score (1-100)'], palette = 'Blues')
plt.title('Gender vs Spending Score', fontsize = 20)
plt.show()
```

Gender and Spending Score Bi-variate Analysis, It is apparent that the majority of males have a Spending Score of around 25k to 70k US Dollars,

Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN : 2581 - 611X)
Volume 5- Issue 2, Paper 44 August 2022

whereas the majority of females have a Spending Score of about 35k to 75k US Dollars. This emphasizesthe fact that women are shopping leaders once again.

Figure-8: Bi-variate graph of Gender v/s spendingscore

Cluster of Ages:

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.

Conclusion

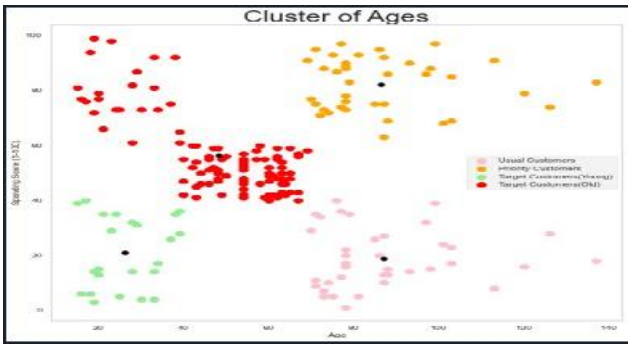
As a result of this massive data volume, consumer data is growing tremendously. These clustering models must be able to process this massive amount of data properly. Cluster 1 denotes the customer with a high annual income as well as a high annual spend, as seen in the above visualization. Cluster 2 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.

```
kmeans = KMeans(n_clusters = 5, init = 'k-means++', max_iter = 300, n_init = 10, random_state = 0)
ymeans = kmeans.fit_predict(x)

plt.rcParams['figure.figsize'] = (10, 10)
plt.title('Cluster of Age', fontsize = 30)

plt.scatter(x[ymeans == 0, 0], x[ymeans == 0, 1], s = 100, c = 'pink', label = 'Usual Customers' )
plt.scatter(x[ymeans == 1, 0], x[ymeans == 1, 1], s = 100, c = 'orange', label = 'Priority Customers')
plt.scatter(x[ymeans == 2, 0], x[ymeans == 2, 1], s = 100, c = 'lightgreen', label = 'Target Customers(Young)')
plt.scatter(x[ymeans == 4, 0], x[ymeans == 4, 1], s = 100, c = 'red', label = 'Target Customers(Old)')
plt.scatter(kmeans.cluster_centers_[:, 0], kmeans.cluster_centers_[:, 1], s = 50, c = 'black')
plt.style.use('fivethirtyeight')
plt.xlabel('Age')
plt.ylabel('Spending Score (1-100)')
plt.legend()
plt.grid()
plt.show()
```

I divided the clients into four categories based on the clustering plot between their age and their respective spending scores: Usual Customers, Priority Customers, Senior Citizen Target Customers, and Young Target Customers. Then, after we get the results, we may devise various marketing tactics and policies to maximize the customer's spending scores in the mall. The cluster of ages represents the group of age peoplespent the money in the mall by the center denoted by blue color.



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. 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.
3. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. 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.
9. Doulamis and N. Doulamis, “Customer Experience Survey,” pp. 3–6, 2016, doi: 10.3390/technologies8040076.