CUSTOMER SEGMENTATION

EXPOSYS PROJECT

TEAM

1. Mukul Rayana - aimukulrayana@gmail.com
   3rd Semester in B.TECH Artificial Intelligence & Data Science of Easwari Engineering College, Chennai

2. Anil Kumar D - dhana5169@gmail.com
   3rd Semester in B.TECH Artificial Intelligence & Data Science of Easwari Engineering College, Chennai

3. Kanishkaa R- aikanishkaa01@gmail.com
   3rd Semester in B.TECH Artificial Intelligence & Data Science of Easwari Engineering College, Chennai

4. Harish Rangasamy - aiharishr@gmail.com
   3rd Semester in B.TECH Artificial Intelligence & Data Science of Easwari Engineering College, Chennai

ABSTRACT:

We face a daily reality such that an enormous and immense measure of information is gathered every day. Investigating such information is a significant need. In the cutting edge period of development, where there is a huge rivalry to be superior to everybody, the business system should be as per the advanced conditions. The business done today runs based on imaginative thoughts as there are an enormous number of potential clients who are confounded with regards to what to purchase and what not to purchase. The organizations doing the business

are likewise not ready to analyze the objective expected clients. This is the place where AI comes into picture, the different calculations are applied to distinguish the secret examples in the information for better direction. The idea of which client fragment to target is finished utilizing the client division process utilizing the bunching method. In this paper, the bunching calculation utilized is K-implies calculation which is the dividing calculation, to fragment the clients as indicated by the comparative qualities. To decide the ideal groups, the elbow technique is utilized.

Introduction:

Throughout the long term, the opposition among organizations is expanded and the huge recorded information that is accessible has brought about the far and wide utilization of information mining methods in separating the significant and vital data from the data set of the association. Information mining is the cycle where strategies are applied to extricate information designs to introduce it in the intelligible arrangement which can be utilized for the motivation behind choice help. Concurring to, Bunching procedures consider information tuples as articles.

They segment the information objects into gatherings or clusters,so that articles inside a bunch are like one another and unlike articles in different groups. Client Division is the course of division of the client base into a few gatherings called client portions to such an extent that every client fragment comprises clients who have comparative attributes. The division depends on the likeness in various ways that are pertinent to promoting like sexual orientation, age, interests, and different ways of managing money.

The client division has the significance as it incorporates, the capacity to adjust the projects of market so it is reasonable to every one of the client portion, support in business choice; recognizable proof of items related with every client fragment and to manage the interest and supply of that item; recognizing and focusing on the potential client base, and foreseeing client absconding, giving headings in viewing as the arrangements.

The push of this paper is to recognize client sections utilizing the information mining approach, utilizing the dividing calculation called as K-implies bunching calculation. The elbow strategy decides the ideal groups.

Clustering and K-Means Algorithm:

Grouping algorithms creates bunches to such an extent that inside the bunches are comparative dependent on certain qualities. Likeness is characterized as far as how close the items are in space. K-implies calculation in one of the most well known centroid based algorithm. Assume informational index, D, contains n objects in space. Partitioning techniques disseminate the items in D into k groups, $C1,\ldots,Ck$, that is, $Ci \subset D$ and $Ci \cap Cj = \varnothing$ for $(1 \leq I, j \leq k)$. A centroid-based parceling procedure utilizes the centroid of a bunch, Ci , to address that group. Adroitly, the centroid of a bunch is its middle point. The distinction between an article $p \in Ci$ and ci , the delegate of the group, is estimated by dist(p,ci), where dist(x,y) is the Euclidean distance between two focuses x and y.

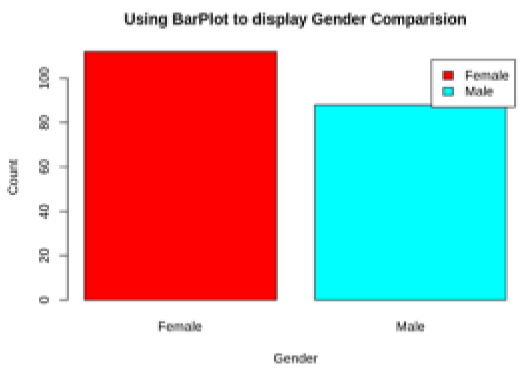
Algorithm: The k-implies calculation for partitioning, where each bunch's middle is addressed by the mean worth of the items in the group. Information: k: the quantity of groups, D: an informational collection containing n objects. Yield: A bunch of k groups. Strategy: (1) subjectively pick k items from D as the underlying group communities; (2) rehash (3) (re)assign each object to the bunch to which the article is the most comparative, in view of the mean worth of the articles in the group; (4) update the bunch implies, that is, ascertain the mean worth of the articles for each group; (5) until no change.

Methodology:

The informational index used to execute bunching and K Means calculation was gathered from a store in the shopping center.

The data set contains 5 attributes and has 200 tuples, addressing the information of 200 clients.The attributes in the dataset has CustomerId, gender, age, yearly income(k$), spending score on the size of (1-100).

Visualize the gender of customers:

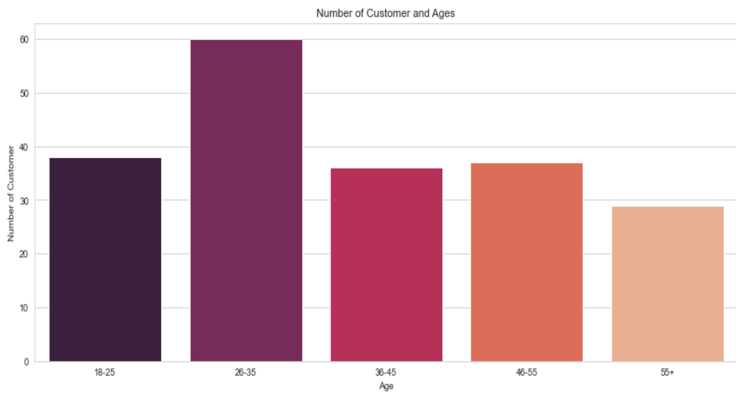


Visualize age of customers:

```
age18_25 = df.Age[(df.Age <= 25) & (df.Age >= 18)]
age26_35 = df.Age[(df.Age <= 35) & (df.Age >= 26)]
age36_45 = df.Age[(df.Age <= 45) & (df.Age >= 36)]
age46_55 = df.Age[(df.Age <= 55) & (df.Age >= 46)]
age55above = df.Age[df.Age >= 56]

x = ["18-25","26-35","36-45","46-55","55+"]
y = [len(age18_25.values),len(age26_35.values),len(age36_45.

plt.figure(figsize=(15,6))
sns.barplot(x=x, y=y, palette="rocket")
plt.title("Number of Customer and Ages")
plt.xlabel("Age")
plt.ylabel("Number of Customer")
plt.show()
```

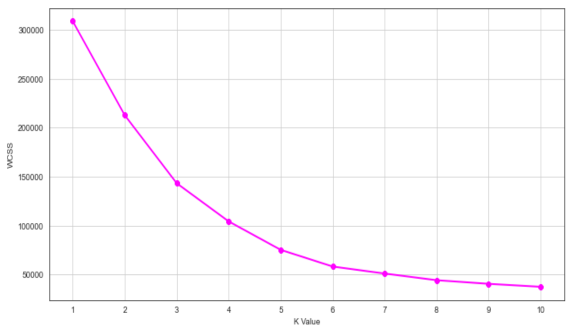


Elbow Method:

The elbow strategy depends on the perception that expanding the quantity of clusters can assist with lessening the amount of inside within-cluster variance of each cluster. This is on the grounds that having more clusters permits one to catch better gatherings of data objects that are more like one another.

To characterize the optimal clusters, Right off the bat, we utilize the clustering algorithm for different values of k. This is done by going k from 1 to 10 clusters. Then, at that point, we compute the complete intra-cluster sum of squares. Then, at that point, we continue to plot the intra-cluster amount of square based on the quantity of clusters. The plot means the estimated number of clusters needed in our model.

The ideal clusters can be found from the diagram where there is a bend in the chart



Visualize the clusters:

```
plt.figure(figsize=(10,6))
sns.scatterplot(x = 'Annual Income (k$)',y = 'Spending Score (1-100)',hue="label",
                palette=['green','orange','brown','dodgerblue','red'], legend='full',data = df  ,s = 60 )
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.title('Spending Score (1-100) vs Annual Income (k$)')
plt.show()
```

Conclusion

From the above perception it very well may be seen that Cluster 1 means the client who has high yearly pay just as high yearly spend. Bunch 2 addresses the Cluster having high yearly pay and low yearly spend. Cluster 3 addresses clients with low yearly pay and low yearly spend. Cluster 5 means the low yearly pay however high yearly spend.

Cluster 4 and group 6 means the client with medium pay and medium spending score.

References :

[1] I. S. Dhillon and D. M. Modha, "Concept decompositions for large sparse text data using clustering," Machine Learning, vol. 42, issue 1, pp. 143-175, 2001.
[2] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, "An efficient K-means clustering algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 24, pp. 881-892, 2002.
[3] MacKay and David, "An Example Inference Task: Clustering," Information Theory, Inference and Learning Algorithms, Cambridge University Press, pp. 284-292, 2003.
[4] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining Concepts and Techniques", Third Edition. [5] D. Aloise, A. Deshpande, P. Hansen, and P. Popat, "The Basis Of Market Segmentation" Euclidean sum-of-squares clustering," Machine Learning, vol. 75, pp. 245-249, 2009.
[6] S. Dasgupta and Y. Freund, "Random Trees for Vector Quantization," IEEE Trans. on Information Theory, vol. 55, pp. 3229-3242, 2009. [7]Puwanenthiren Premkanth, ―Market Segmentation and Its Impact on Customer Satisfaction with Especial Reference to Commercial Bank of Ceylon PLC.‖ Global Journal of Management and Business Research Publisher: Global Journals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1