

KALĀ



**THE JOURNAL OF
INDIAN ART HISTORY CONGRESS
(VOL: XXVII, 2021-22)**

Kala

The Journal of Indian Art History Congress

(Vol: XXVII, 2021-22)

Advisor

R.D. CHOUDHURY

Editor

UMESH CHANDRA DWIVEDI

INDIAN ART HISTORY CONGRESS

Guwahati, Assam

India

kala

The Journal of Indian Art History Congress (Annual Journal Devoted to the Study of Indian Art)

A Referral Journal

I rights reserved No. part of this publication may be reproduced in any form or by any means, when permission of the publisher/author. The responsibility for the facts stated, opinions expressed exclusions reached is entirely that of the contributors and the Indian Art History Congress takes no responsibility.

ISSN 0975-7945

Advisor

Dr. R.D. Choudhury

Former Director General, National Museum, New Delhi

&

Former Vice Chancellor, National Museum Institute of History & Art, Conservation & Museology (Deem University).
New Delhi,

Editor

Umesh Chandra Dwivedi

Retired Director of Archaeology, Jharkhand/Bihar

Cover Photo (Front)

Buddha & Vajrapani Gandhara,
Berlin Museum

Cover Photo (Back)

Pictorial Representation of Siddha Cakra Yantra
Auckland Art museum

Publisher

INDIAN ART HISTORY CONGRESS

Directorate of Archaeology Campus

G N.S. Road. Ambari Guwahati-781001

Phone-0361-2630102, Mob4e-60017221211

Web: www.indianarthistorycongress.com

Printed at

Saraighat offset press, Bamunimaidam, Guwahati-21

CUSTOMER SEGMENTATION USING CLUSTERING ANALYSIS

Mr.Pasagadugula Veeresh ^{*1}, Mrs. K.G.Prasanthi ^{*2}

Abstract

It is essential to analyse the huge amounts of data that our environment often produces. In the present technological world, where everyone is competing to the corporate strategy is supposed to be superior to everyone else adapted to the circumstances by hand. By considering all the potential customers today struggled with knowing what to buy and what to avoid. New ideas are very essential to business companies are unable to assess their primary customers by themselves. Where the Machine Learning occurs, and different techniques are used to discovered the hidden patterns in data to help in decision-making. The machine learning method called clustering involves comparing data from several groups such as market research, image processing, pattern recognition, search engines optimization and medical data processing and others. Customer market research involves segmentation, which is the topic of our endeavor. The definition of customer segmentation is the grouping of consumers based on their similar characteristics. In the present environment, it's essential for businesses should categorize their customers according to their age, location, gender, and additional characteristics. This enables businesses should concentrate on particular customers who are most likely to buy their products. The use of machine learning comes to provide an advantage over their business competitors. We can successfully to improve their business strategies, the primary intention of this project using the K-means algorithm to divide customers groups based on their attributes. Finally, by taking the mean value as be the major indication, the data from the various clusters inform us which groups the new customers belongs to.

Keywords:

Machine Learning, Customer Market Research, Segmentation-Means Algorithm.

1. INTRODUCTION

Customer segmentation is the process of dividing individuals who have characteristics relevant to marketing, such as age, gender, interests, and spending habits, into groups. Customer segmentation is a method used by companies to target particular, smaller groups of consumers with relevant messages that would encourage them to make a purchase. This technique is based on the concept that each and every customer is unique. In order to more effectively focus their marketing efforts to each segment, business also want to have a deeper understanding of their customers' preferences and needs. In order to divide customers into specific targeting groups, it is necessary to identify significant differentiators that separate them.

^{*1} MCA Student, Department of Master of computer Applications ^{*1},

Assistant Professor ^{*2}, Department of Information Technology ^{*2},

Vignan's Institute of Information Technology (Autonomous), Beside VSEZ, Duvvada,
Vadlapudi Post, Gajuwaka, Visakhapatnam – 530049.

vignaniit.edu.in

When determining customer segmentation techniques, factors also with a customer's demographics (age, race, religion, gender, family size, ethnicity, income, and level of education), geography (where they live and work), psychographics (social class, lifestyle, and personal traits), and affective (spending, consumption, usage, and desired benefits) tendencies are taken into account. The ability to adjust marketing strategies so that they are suitable for each customer category and to support business goals is called customer segmentation. It helps in identifying the items related to each client segment, managing supply and demand for those products, identifying and focused on a potential customer base, and predicting customer problems. By target specific consumer groups with a customer segmentation strategy, business owners might use its marketing resources more efficiently and increase their chances of cross-selling. When companies give customized messages to a set of clients as part of a marketing mix suited to their needs, it is simpler for them to identify innovative offers to motivate them to spend more.

2. EXISTING SYSTEM

Customer data is currently mainly stored through documentation and computer software, which is growing daily. At the end of the day, they will examine their data to determine how many products were sold, the number of real customers, etc. They identified who was beneficial to their business and improved their sales by analysing data that was collected. More paperwork and time are needed. Moreover, it is not a very effective method of locating the desired customer data.

3. PROPOSED SYSTEM

This fresh approach will be essential in conquering the existing process, that focuses on paperwork and computer digital data. Everyday data collection involves a growing quantity of documentation, which takes a lot of time. In today's world, new technologies were emerging. Machine learning is a powerful innovation that employs a variety of algorithms to predict the outcome. Consequently, to answer our problem statement, we'll use K-Means Clustering, which classifies the data into groups based on similar characteristics. The data will subsequently be presented.

CUSTOMER SEGEMENTATAION:

Due to the excessive competition in the business world, businesses have had to expand throughout time in order to increase profitability and draw in new customers. Based on their preferences. It takes a lot of effort and time to determine each customer's demands and meet them. This is due, among other things, to the fact that clients have a diverse set of needs of objectives, passions, and tastes. Segmenting customers, divides clients rather than employing a "one-size-fits-all" technique based on similar traits or practices, into groups. A marketing tactic known as customer segmentation divides divide the market into discrete, similar groups. Information used in the technique of client segmentation, which separates customers dividing them into categories is based on several variables, including geographic factors, economic trends, and demographic trends, and behavioural patterns. A client segmentation technique can help a company's marketing resources be better utilized.

MACHINE LEARNING:

In a wide range of business, like Facebook, where it helps us identify ourselves and our friends, and YouTube, where it recommends films based on our interests, we have seen machine learning in action. Unsupervised learning and supervised learning are the two categories used to describe. Assuming that the data in this case is targetable and that we want

to forecast in the future, such as identifying a person's worth or the amount of monthly costs, a data analyst regularly uses supervised learning to address problems like classification and regression. On the other hand, unsupervised learning may or may not have a label or goal in mind. Clustering, for instance, seems to have a fixed goal because it is based on a mathematical model. For example, in the case, we might like to groups of students based on their learning preferences. or product purchases. In order to increase consumer numbers and so create massive profits, the marketing business, particularly malls, is subject to intense competition. Machine learning is already being used by many markets and businesses to do this. Shopping centres and malls use machine learning models that specifically target the right people using the data they gather from customers. This not only raises sales and the number of visitors, but it also improves the efficiency of the business.

4. PROPOSED METHODOLOGY

K-Means Clustering Algorithm:

K-Means Clustering, an unsupervised learning technique, is used to fix clustering problems in machine learning. The unlabelled dataset is divided into k different clusters using an iterative process, and each cluster only comprises the datasets that are a part of that group and have characteristics in common. K here determines the minimum number of pre-defined clusters that must be generated as part of the process; for instance, if $K=2$, there will be two clusters, if $K=3$, there will be three clusters, and so on. It allows us to categorise the data into different groups and offers a workable technique for quickly and accurately determining the groups in the unlabelled dataset without the need for any training. The technique is centroid-based, with a different centroid for each cluster. Reducing the overall distances between each data point and its matching clusters is the main objective of this technique. An unlabelled dataset is first used as the algorithm's input, after which it is divided into k clusters. The algorithm then repeats this process until it has no more clusters left to use. The value of k in this method ought to be predetermined.

How does the K-Means Algorithm Work?

The working of the K-Means algorithm is explained in the below steps:

Step-1: Select the number K to decide the number of clusters.

Step-2: Select random K points or centroids. (It can be other from the input dataset).

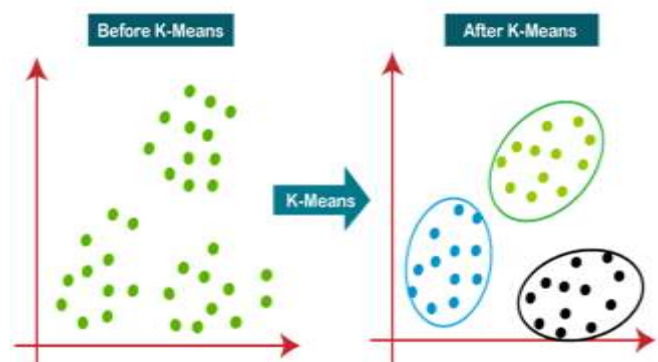
Step-3: Assign each data point to their closest centroid, which will form the predefined K clusters.

Step-4: Calculate the variance and place a new centroid of each cluster.

Step-5: Repeat the third steps, which means reassign each datapoint to the new closest centroid of each cluster.

Step-6: If any reassignment occurs, then go to step-4 else go to FINISH.

Step-7: The model is ready.



The Elbow method:

The Elbow approach applies k-means clustering on the dataset for a range of k values, such as from 1 to 10, and computes the average score for each value of k. By default, the distortion scores the sum of the square distances between each point and its designated center

is calculated. The optimal value for k can be seen graphically by plotting these overall characteristics for each model. The "elbow," or point of inflexion on the curve, is the optimal value of k if the line chart resembles an arm. The "arm" can point upward or downward, but if there is a sharp inflection point, it is likely that the underlying model works best at that point. To determine the ideal number of clusters, we employ the Elbow Method, which compares the Within Cluster Sum of Squares (WCSS) to the number of clusters (K Value). The WCSS calculates the sum of observations' distances from their cluster centroids using the formula.

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

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

Libraries Used:

Scikit-learn: It is a free Python machine learning software, sometimes known as sklearn. It is meant to interact with the Python numerical and scientific libraries NumPy and SciPy, and features support vector machines, random forests, gradient boosting, k-means, and DBSCAN, among other classification, regression, and clustering algorithms.

Seaborn: Seaborn is a matplotlib-based open-source Python library. It's used for exploratory data analysis and data visualization. Seaborn makes using data frames and the Pandas library a breeze. The graphs that are generated may also be readily changed.

NumPy (numerical Python): is a package that contains multidimensional array objects and tools for manipulating them. NumPy is a Python library that allows us to perform mathematical and logical operations on arrays. NumPy is widely used in combination with SciPy and Matplotlib (Scientific Python) (plotting library). This combination is frequently used as a substitute for MATLAB, a prominent technical computing platform. The Python counterpart to MATLAB, on the other hand, is today regarded as a more contemporary and comprehensive programming language.

Pandas: is a Python toolkit for data science, data analysis, and machine learning that is open-source. It is based on NumPy, a multi-dimensional arrays-supporting library. Pandas, being one of the most widely used data manipulation tools, works well with a variety of other Python data science modules.

Matplotlib: For 2D array charts, Matplotlib is a superb Python visualization library. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the entire SciPy stack. The ability to show vast volumes of data in simple images is one of the most essential advantages of visualization. Line, bar, scatter, histogram, and more graphs are available in Matplotlib.

5. EXPERIMENTAL RESULTS

The dataset for K-means algorithm clustering was provided by a retail outlet. The data set, which represents the data of 200 tuples, and have five attributes. In the main characteristics of the data collection customer ID, age, gender, yearly income (in k\$), and spending on a scale of 1 to 100. The value of a customer's purchases or expenditures at a mall is enhanced by a score for expenditures that goes from 1 to 100. Data cleaning must be done if a dataset has null values, duplicates, or other noisy data. Cleansing data makes sure it is accurate, useable, and available for analysis. Once the data is available, we can visualize it by contrasting the gender-specific annual income and spending score. There are five main types of plots, the study claims that provide examples of consumer groups participating in the following activities and consumer habits connected to yearly income and scores on expenditure.

1. High income low spending
2. Low-income high spending
3. A high spending low income
4. Average income average spending
5. High income high spending

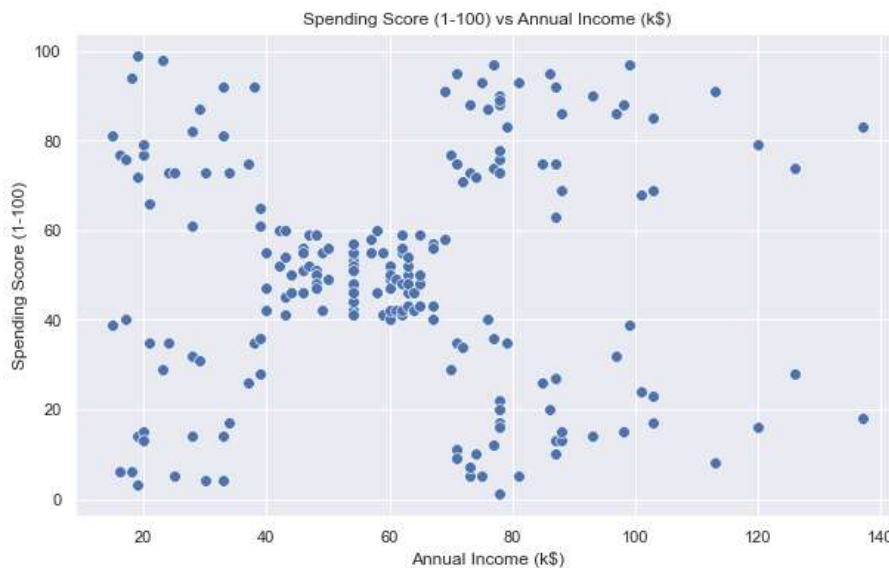


Fig: Annual Income vs Spending score

Now that there are lots of groups, we may build using a K-means model, but not great in detail. The Elbow method is efficient approach is used to forming clustering using K-means given a selection of k clusters (say, 1 to 10) and calculate the sum of the squares between each point and its centre in each case. Decide how many clusters there will be you the highest silhouette score possible. This explains the Elbow approach is determined. We observed that when K=5 is reached, WCSS is not moving quickly (Within Cluster Sum of Squares). Moreover, we get the number of clusters K=5 will be the ideal number of clusters based on the current data.

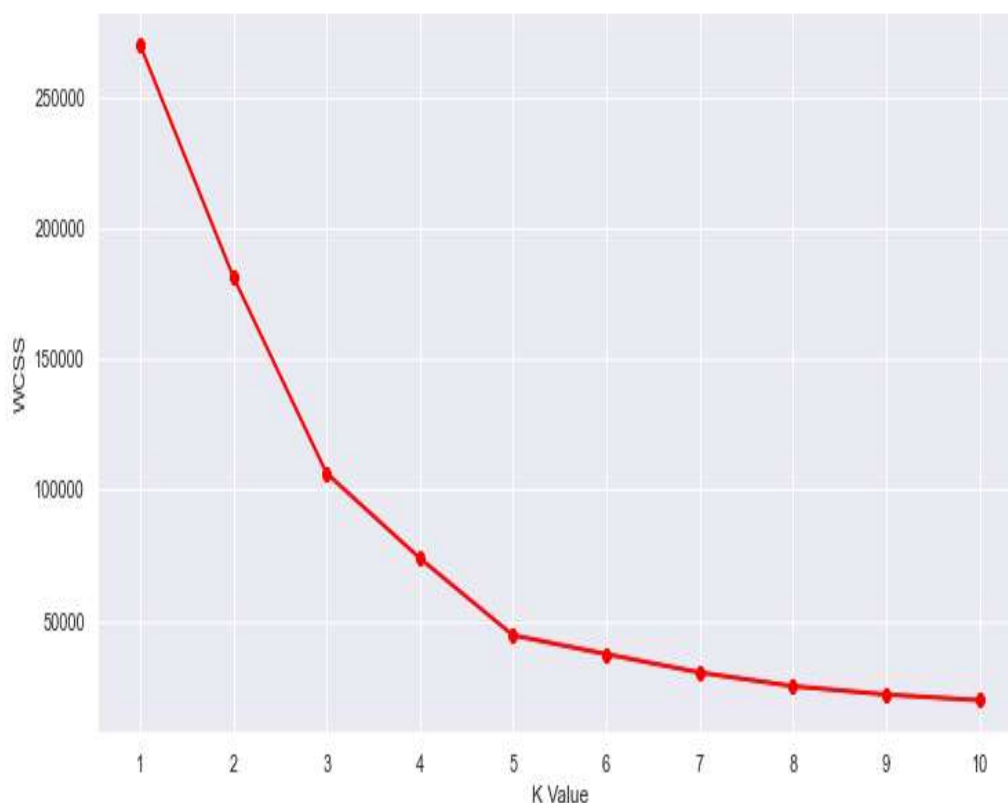


Fig: The Elbow method to find K value

Using the above strategy, we may categorise the plot into different groups, decide which clusters should be prioritized, and then give each group a label. K-means can be used to be employed to select which of the five clusters should targeted, specifically consumers with moderate incomes. High Income-High Spending Score, Spending Score, and Low income and high expenditures. The necessary customers been discovered.

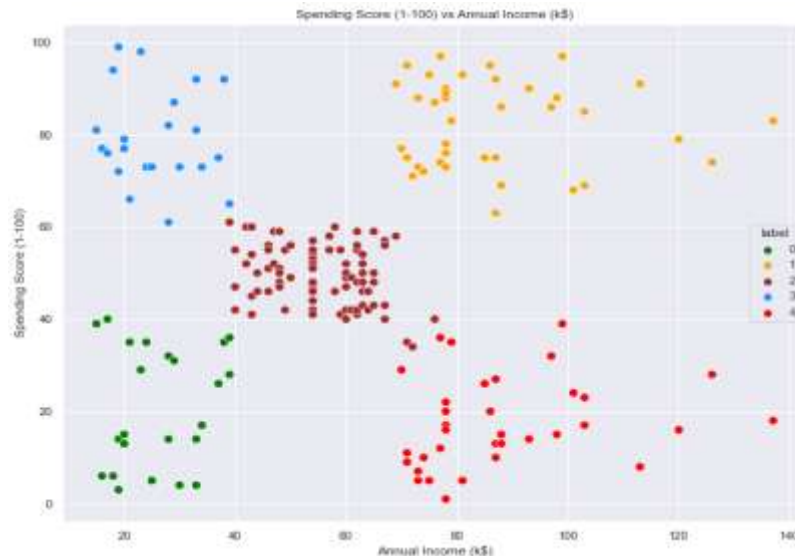


Fig: Segmenting Mall customers using K-means clustering

6. CONCLUSION

If you properly manage the best current customer segmentation process, however, the impact it can have on every part of your organization sales, marketing, product development, customer service, etc. is immense. Your business will possess stronger customer focus and market clarity, allowing it to scale in a far more predictable and efficient manner. Ultimately, that means no longer needing to take on every customer that is willing to pay for your product or service, which will allow you to instead hone in on a specific subset of customers that present the most profitable opportunities and efficient use of resources. That is critical for every business, of course, but at the expansion stage, it can often be the difference between incredible success and certain failure.

7. REFERENCES

- [1] Jiawei Han, Micheline Kamber, Jian Pei "Data Mining Concepts and Techniques", Third Edition.
- [2] 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.
- [3] S. Dasgupta and Y. Freund, "Random Trees for Vector Quantization," IEEE Trans. on Information Theory, vol. 55, pp. 3229-3242, 2009.
- [4] "Customer Segmentation Using K Means Clustering," Towards Data Science, Apr. 2019.
- [5] Ruhul Reddy, "Who's who: Understanding your business with customer segmentation," INTERCOM.
- [6] Kristen Baker, "The Ultimate Guide to Customer Segmentation: How to Organize Your Customers to Grow Better," Hunspot.
- [7] Tim Ehrens, "customer segmentation," TechTarget.

- [8] V.Vijilesh, "CUSTOMER SEGMENTATION USING MACHINE LEARNING," International Research Journal of Engineering and Technology (IRJET), vol. 08, no. 05, May 2021.
- [9] Expert Systems with Applications, vol. 100, Feb. 2018, "Retail Business Analytics: Customer Visit Segmentation Using Market Basket Data."
- [10] "CUSTOMER SEGMENTATION USING MACHINE LEARNING," IJCRT, AMAN BANDUNI and ILAVENDHAN A, vol. 05, 2018.
- [11] Tushar Kansal; Suraj Bahuguna; Vishal Singh; Tanupriya Choudhury, "Customer Segmentation using K-means Clustering," IEEE, Jul.
- [11] H.Krishna Sastri, South Indian Images of Gods and Goddess, Madras, Madras Government, Chennai, 1914
- [12] T.N.Subramanyam, South Indian Temple, Government Oriental Manuscript Library, Madras 1953-57
- [13] S.R.Balasubramanyam, Later Chola Temples, Mudgala Trust, Madras, 1979
- [14] C.Sivaramamurthi The Great Chola Temples, Archeological of India, New Delhi-2007.