Pratibodh
A Journal for Engineering
A free and Open Access Journal
Homepage: https://pratibodh.org
RATNA & BM RANVEER Foundation

Segmentation of customer groups using machine learning algorithms

Chaklan Suyash, Shorya, Tushar, Neelkmal
Department of Artificial Intelligence & Data Science
Jaipur Engineering College & Research Centre
suyashchaklan.ai25@jecrc.ac.in

Abstract

With the increased competition and saturation of marketplaces and consumers, doing well in the market has become extremely difficult especially for relatively smaller retailers and newly established organizations. The best way to do well in such a market is to understand your customers well and serve them accordingly. The manual labor to do that with extensive tables and manual sheets is an unnecessary task with lengthy procedures. With the advancements in technology and the smart innovation in computers has enabled us to do that task in a much easier and simplified way without much guesswork. Here is an analysis of such segmentation of customers to better understand their demographics and serve them better using modern day programming language, Python and the k-Means clustering algorithm of SKlearn library..

Article Status

Available online :

Keywords: *Machine Learning, Data, Customer Segmentation, Python, k-means algorithm*

1. Introduction

With the advancements in technology, the use of technologies that help make the manual work easier are being widely spread. With the insane amount of data present everywhere, data has become one of the most important assets of this century. With technologies like Machine Learning and Data Science, the use of data has been helpful to variety of people be it the government, consumers, sellers or different organizations[5].

Use of technologies like Data Mining has been very helpful in collecting and presenting raw data from deep areas on the internet and using it to obtain useful information[2]. Use of the K-Means algorithm is for clustering of the data using unsupervised learning. This algorithm is learning independent and can show varied results when used different datasets without training them each time.

The matplotlib library is used to represent the obtained data graphically making it easier to plot and demonstrate.

2. Literature Review:

The segmentation of customers is creating organized groups of the customers on a basis useful for the market to serve customers accordingly and make best profit[3].

Python brings the maximum versatility in programming combing concepts[1] of machine learning with code understandable easily. It is great with data refining scattering and using different libraries to get the job done.

Machine learning is a stem of Data Science that involves algorithms that help system learn, recognize patterns, improve performance and apply them in programs[4].

3. System Architecture:

The proposed program of customer segmentation is explained as follows:

Data Mining and Cleaning: Data is mined from different parts of the internet to be collected at a place and its cleaning is done to avoid discrepancies in the deployed model.

Different classifications of customers: Customers are classified into different ways and the model is applied to those classifications accordingly like based on gender, spending scores, age, annual income, etc[6][7].

K-Means Algorithm: The use of K-Means algorithm is done to sort the unlabeled data into multiple clusters for different types of classifications applied[9].

Centroids: Centroids are assigned in each cluster to calculate the average or the mean point distance of each point in the cluster from the centroid.

Clusters: Clusters are created from vector quantization while the original comes from signal processing. Clusters of different datasets are created and the segments of customers are noticed wherever the customer count goes high[8].

4. Observations:

Different observations are plotted and observed using clustering and graph plotting which is as follows in Fig.1:

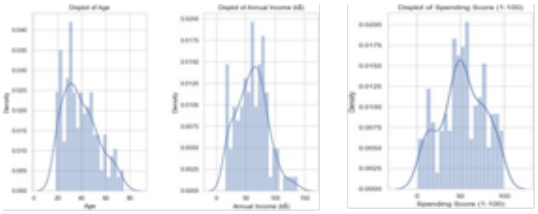

Figure 1. Displot of age/income/spending score vs density

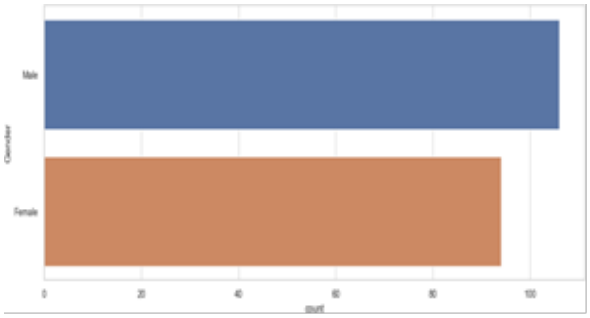

Figure 2 Graph of Count vs. Gender

The graphs of the age, annual income, spending score and the gender of customers as shown in Figure 3.

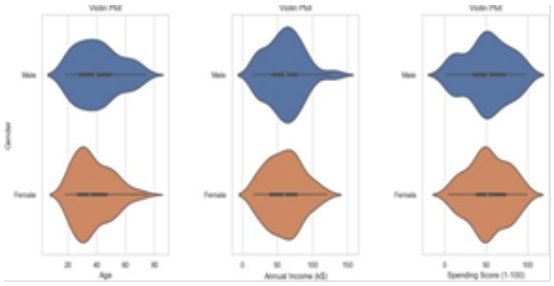

Figure 3. Violin plot of Gender vs Age, Income and Spending.

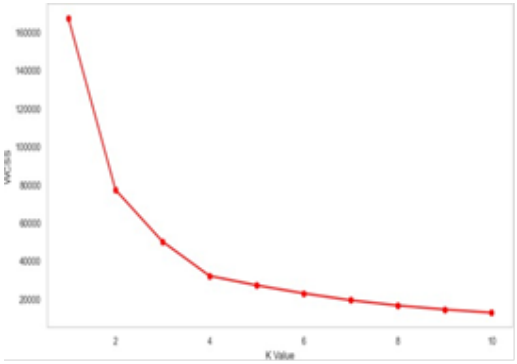

Figure 4. The graph of K Value vs WCCS(sum of squared distance between all points and the centroid) using elbow method.

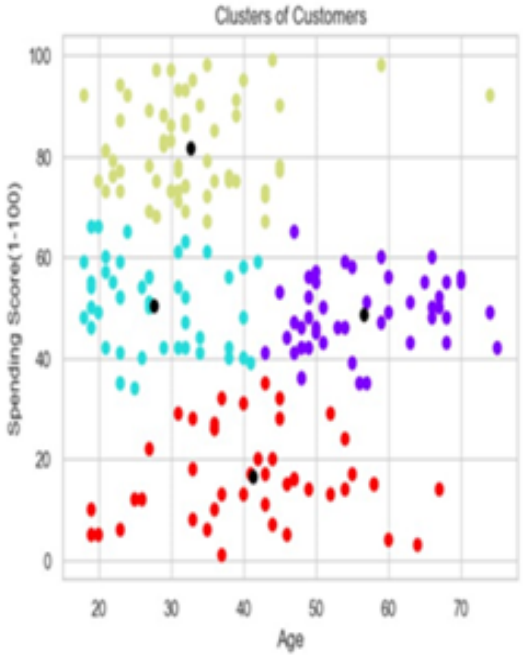

Figure 5. Colored Scatter Plot of age vs spending score of customers

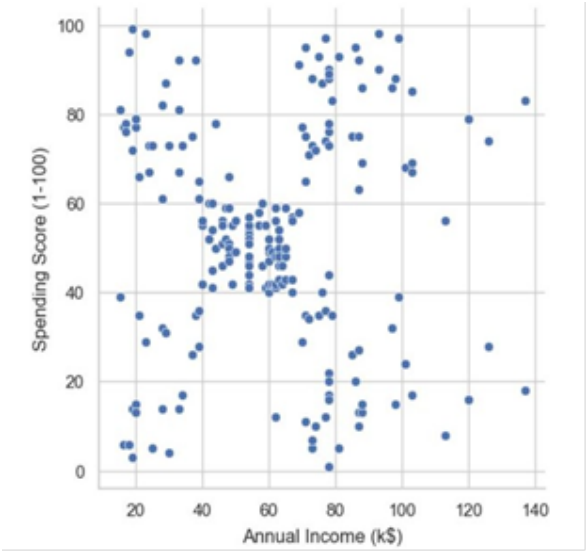

Figure 6. Scatter Plot of annual income vs spending score

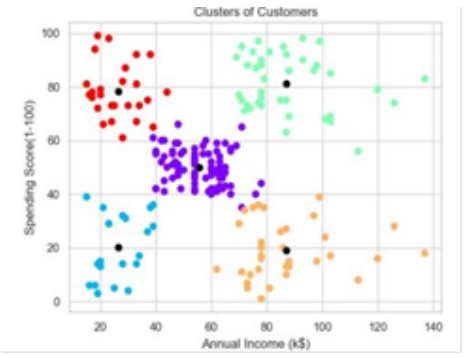

Figure 7. Scatter plot of annual income vs spending.

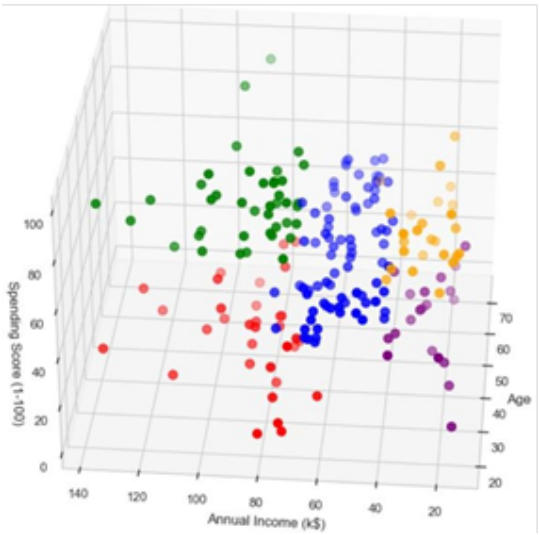

Fig. 8: 3D plot of income vs spending

5. Result:

The clusters and graphs suggest that in the gender demographics, male spend more in general than the female demographic. The demographic between annual income 40,000$ per annum and 60,000$ per annum have a spending score typically lying between 40-60. The clusters also suggest that the spending score of 60 and above is dominant in the age groups 20-40.

6. Conclusion:

The customer demographic after studied successfully leads to many advances on the use of right customer tactics which plays a very important role in increasing the sales. For example, the items in the spending score from 40-65 can have reduced prices for people of the age group 40-60 because the highest sales are observed in that group.

Such thoughtful uses of tactics would be a great booster in sales and help increase the profit[10].

7. References:

1.Blanchard, Tommy. Bhatnagar, Pranshu. Behera, Trash. (2019). Marketing Analytics Scientific Data: Achieve your marketing objectives with Python's data analytics capabilities. S.l: Packt printing is limited

2.Griva, A., Bardaki, C., Pramatari, K., Papakiriakopoulos, D. (2018). Sales business analysis: Customer categories use market basket data. Systems Expert Systems, 100, 1-16.

3.Hong, T., Kim, E. (2011). It separates consumers from online stores based on factors that affect the customer's intention to purchase.

4.Hwang, Y. H. (2019). Hands-on Advertising Science Data: Develop your machine learning marketing strategies… using python and r. S.l: Packt printing is limited

5.Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC.|| Global Journal of Management and Business Publisher Research: Global Magazenals Inc. (USA). 2012. Print ISSN: 0975-5853. Volume 12 Issue 1.

6.Puwanenthiren Premkanth, - Market Classification and Its Impact on Customer Satisfaction and Special Reference to the Commercial Bank of Ceylon PLC.|| Global Journal of