CUSTOMER SEGMENTATION USING K-MEANS CLUSTERING ALGORITHM

Submitted By

Shara Binte Osman (162-35-1676)

A thesis presented in partial satisfaction of the Bachelor of Science in Software Engineering degree requirement.

Department of Software Engineering

DAFFODIL INTERNATIONAL UNIVERSITY

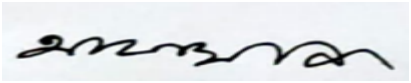
Spring-2022

APPROVAL

The thesis "CUSTOMER SEGMENTATION USING MACHINE LEARNING," submitted by Shara Binte Osman, 162-35-1676 to the Daffodil International University under the Department of Software Engineering, has been accepted as satisfactory for the completion of a portion of the prerequisites for a Bachelor of Science in Software Engineering, as well as approval of its style and contents.

Chairman

Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

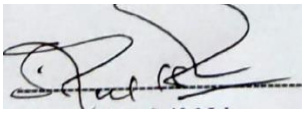


Internal Examiner 1

Afsana Begum
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Senior Lecturer
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

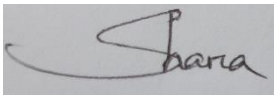


External Examiner

Prof. Dr. Md. Saiful Islam
Professor
Institute of Information and Communication Technology (IICT)
Bangladesh University of Engineering and Technology (BUET)

DECLARATION

It at this moment declares that I have done this thesis under the supervision of Dr. Imran Mahmud, Associate Professor, and Head Department of Software Engineering, Daffodil International University. It is stated that neither this thesis nor any part of it has been submitted to any other university to receive a degree.



_____

Shara Binte Osman
Student ID: 162-35-1676
Batch: 20 Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

Certified by:

_____

Dr.Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Faculty of Science & Information Technology
Daffodil International University

ACKNOWLEDGEMENT

By the grace of Allah, I've finished my thesis without interruption. I want to grant my heartfelt appreciation to my Advisor, Dr. Imran Mahmud sir, for his invaluable assistance and advice in my work. He helped me in resolving any issue that arose. Finally, without my parents' love and faithful backing, none of this would be possible. Thanks to their ongoing encouragement, support, and supplication, I am currently on the edge of completing my undergraduate degree.

Table of Contents

List of Figures

List of Tables

ABSTRACT

Customer segmentation by using data mining could assist organizations with leading client situated advertising and assembling differentiated systems focused on assorted clients. To enhance a business, it's very obligate to know about the client minutely, for this purpose customer segmentation is necessary. This exploration is tied in with customer segmentation into gatherings. We have collected our data from a supermarket through customers' membership cards. We have used unsupervised machine learning, a k-means algorithm for segmentation and clustering. By using unsupervised machine learning and k-means algorithm we get our targeted customers to increment the revenue of the company.

Keywords: Customer Segmentation, Supermarket, Unsupervised machine learning, K-means clustering

CHAPTER-1

INTRODUCTION

1.1 Background

Clients are viewed as the most significant resources to an ever-increasing number of supermarkets in the business sector. Instructions to acknowledge and additionally satisfy client needs are the key to supermarkets. However, there are many types of clients who come to the supermarket and it's very troublesome to know their needs. It's obscure to the companies that which types of shoppers want which types of products or services. In any case conscious about the insistence of a large number of clients, also it's an irregular interaction that what a shopper buy and which item. Over and above this's essential to provide an item according to the shopper demand and offer the support in like manner. For this reason, we have decided to conduct unsupervised machine learning to segment the clients and invent targeted customers for companies' revenue by the demeanor of k-means clustering. For targeted clients, it's very essential to be aware of their purchase product, annual income, spending score. Which group of shoppers buy more? Which type of products is a shopper buying often? What types of discounts do they want? What kind of shopper? Depending on this information the marketing team can make the proper scheme for companies. Generally speaking, for this segmentation marketing team can give assistance and shoppers can take needs from supermarkets. Which group the customers associated with that is unknown to us before implementing the clustering method but after implementing clustering it will be clear to us that the clients that data belongs to.

Moreover, this thesis will help to increase companies' revenue as well as demonstrate accurate order of customer segmentation.

1.2 Motivation of the Research

Nowadays numerous supermarkets are thriving. But these companies don't have proper knowledge of customer satisfaction, that's why supermarkets are sustaining with regards to their benefit.

So, our specific intention is to contrast and classify expected results with supermarket business and the dangerous component of business benefit. As per our solution, the marketing team can think about forwarding the business, and also, they can look for digital marketing for more publicity as well as for profit.

1.3 Problem Statement

The key problem of this thesis was gathering a proper dataset and ready up the dataset. We have tried to collect data from various supermarkets for exact datasets but because of the security issue of customers, no supermarket was not ready to give their clients data. So we have to excerpt data from "Kaggle" which was public data and it's an online platform of machine learning. After that, we have made and trained our model with the help of that dataset.

1.4 Research Question

- How accurately K-means clustering can be used for customer segmentation?

1.5 Research Objective

The fundamental objective of this thesis is to make customer segmentation based on a clustering algorithm for supermarkets. By doing this segmentation of customers we will understand which group of customers will play a vital to improve the company's

revenue. We can provide this targeted customer to the marketing team so that they can make plans for this group of customers like giving special discounts, occasionally giving offers, etc. Also, it mostly relies upon the product price, clients' age, clients' income, and clients' demand.

1.6 Research Scope

This work is for the developer of customer segmenting technologies. The objective of the study is to offer an optimized architecture for doing customer segmentation. If anybody wants to do customer segmentation tool will get a sharp ground for tool development. A scientific report has been created dependent on carrying out a K-means algorithm model in this study. An intelligence customer segmentation model developer will be benefited with the help of this study to make an enacting decision. Various segmentation techniques for customer segmentation have been mentioned by several researchers. The execution of the framework can be affected in light of the fact that the requirement for a legal algorithm areal. Various scholastic has researched customer segmentation methods with the help of other segmentation types like behavioral segmentation, demographic segmentation (Boyu Shen, 2021; Chih-Fong Tsai, Ya-Han Hu, and Yu-Hsin Lu, 2013; E.Y.L Nandapala and K.P.N Jyasena, 2020).

After doing a study on these reports all the aspects of the customer summation model will be known to an intellectual customer segmentation model developer.

1.7 Thesis Organization

In Chapter 2, we have extended on "Literature Review." In Chapter 3, we have recited our "Methodology." Like Dataset, Data Preprocessing and narrate about K-means clustering algorithm. Then in Chapter 4, we have shown our "Result" and "Discuss"

with regards to that. After that, in last Chapter 5, we highlighted our “Findings & Contributions” and showed the future scope of this work. Thusly, we have finished up our thesis paper.

CHAPTER-2

LITERATURE REVIEW

With the intention to exhibit a rile for customer segmentation for the business is focused on a real-world database from an online transaction platform, discuss behaviors or product preference and adopt corresponding marketing strategies of customers, used RFM model for identifying customer behavioral features (Boyu Shen, 2021). Analyzed use three algorithms on the purpose of the customers do the amount of shopping and average visit of customers into shop annually, for the careless, careful, standard, targeted and sensible customer cluster has been formed labeled by the application of clustering, also focused on higher buyers and frequent visitors vs higher buyers and occasional visitors with the help of the new clusters on flowed on applying mean shift cluster (Tusar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury, 2018). Compare results for correction with the help of two clustering techniques, to ameliorate the quality of services for feasible customer relationship management various customized marketing strategies have been aimed at each customer group are suggested based on the segmentation outcome (Chih-Fong Tsai, Ya-Han Hu, Yu-Hsin Lu, 2013). The methods of machine learning have been focused on the hazard in the application, "without a teacher" is thought in practice has been the practical method of machine learning (Natalya V. Razmochaeva, Dmitry M. Klionskiy, Vladimir V. Chernokulsky 2018). They had exhibited a soft clustering method to categorize online customers based on their purchasing data with the help of a mixed-class membership clustering approach for the purpose of attaining optimally segmentation (Roung-Shiunn Wu, Po-Hsuan Chou, 2011). For the customer, segmentation has used density-based algorithms as well as used centroid-based algorithms (A.S.M Shahadat Hossain, 2018). Used

CRM so that companies can identify customer behavior, innate, and by the help of CRM, companies can understand who are the profitable customer (E.Y.L Nadapala, K.P.N Jayasena, 2020). To definite congenial marketing strategies used hierarchical agglomerative algorithm in R programming language (Phan Duy Hung, Nguyen Thi Thuy Lien, Nguyen Duc Ngoc 2019). Increasing of the population has been discussed, the old aged patient who has chronic diseases for them the health cluster model has been used so that services can be provided in multiple platforms and can increase health status (Kim, Jong Tak, Hee-Jun Pan, Jonghum Kim, 2017). To pick out latent class analysis three cluster has been analyzed, on 377 participants the methodology of the clustering has been performed who attend (ACCHS) in Australia (Noble, Natasha 2015). School students whose age between 13-14 and 17-19 this thesis was for them to identify their health-related behavior, demographic characteristic. For this reason, two clusters have been optimized (Alzahrani, Saeed G, (2014). In the US the root of corporate default clustering of data mining has been researched and that was the fundamental purpose of this paper. To segment the problem and also complement each other K-means and CN2-SD have been used (Azizpour, 2014). For data analysis clustering is very obligate (Guha 2016). Into three groups K-means cluster has been divided these are poor, intermediate, good customers. Proxy value was the limitation of this paper. The study was basically focused on the demographic background (Grosskreutz, Henrik). The fundamental objective of this paper was to quest data mining access for customer segmentation. The company can redefine its product design (Brito, 2015). This paper is based on customer segmentation on telecommunication, they wanted to segment their client based on revenue and to distinguish high value from medium and low-value customers and handle each customer (Konstantinos T. Antonios, C. 2010). To map business opportunities and determine business strategies this cluster model has been used, for one promotion, remarketing regional marketing,

and customer loyalty programmers the business strategist has been mapped (Jinafu, L., Jianshuang L, Huaiqing H. 2011). In China, on telecom operators, this model has been applied to bully a customer value evaluation system (Hua S, Xiu S, Leung SCH, 2011). Based on generalized association rules and decision tree technology this paper has been made to target more services with more customers (Ma, H. 2015). To increase the share of wallet, market share, customer satisfaction this paper has been studied, business issue and available data has been covered hare (Baer D. 2012). In this paper they have used the K-means clustering algorithm to cluster customers into the group, to know customer purchase behaviors they made an e-commerce site for agro-business. They have just used only one algorithm for their paper (Anup Chandra Bepary, Zannatul Ferdous, Afsara Tasneem Misha, 2020). Their main target is doing machine learning implementation based on mall customers for this they have used customer income and their purchase product. They have used k-means and hierarchical algorithms (Sriramakrishnan Chandrasekaran, Abhishek Kumar 2019). From a vector characteristic of the cluster which is assumed to have a data density which is decreasing of a function have a present similarity of clusters, they have used density measurement, algorithm design, and analysis (David L. Davies). They have used K-means clustering and RFM model to categorize customers based on customers' value they also used POS customer transactions to improve the accuracy of prediction (Monireh Hosseini, Mostafa Shabani, 2015). They used three clustering algorithms which are k-means, fuzzy C-means, Hierarchical clustering algorithms they also used the RFM model to understand customer behavior, used KDD model (Surefunmi Idowu, Srivastav Kattukottai 2019). They do this research for understanding tourism-related behavior for tourism management, to segment the tourism-related market they have used a hierarchical algorithm, they have used an app for five months to collect data. Two major

clusters and four sub-clusters have been proposed (Jorge Rodriguez, Ivana Semanjski, Sidharta Gautam, Nico Van de Weghe, Daniel Ochoa, 2018). Depending on the hierarchical algorithm they have proposed a Q-criterion whose name is HACNJ, introducing of Q-criterion is the main contribution of HACNJ for clustering (Jianfu li, Jinashuang Li, Huaiquing He, 2011). Their main purpose was to determine the initial centroid of the K-means clustering algorithm, to find out the initial centroid of the K-means algorithm they have used SSE (Sum of Squared Error), depending on three data sets and number of clusters 2,3 and 4 the testing had been performed (Bernad Jumadi Dehotman Sitompul, Opim Salim Sitompul, Poltak Sihombing, 2019). A time series classification task is the patient risk significant. To evolve the approximate daily risk of a patient they had been begin by defining and extracting approximate risk processes. To identify patients at the risk of testing positive for the hospital they have applied the classification (J. Wiens, Jhon V Guttag, E.J. Horvitz 2012). They have focused on a decision model in mapping the classification of Indonesian telematics SMEs to conduct a hybrid data mining model (E T Tosida, F Andria, I Wahyudin, R Widianto, M Ganda· R R Lathif). To find out similarities between clients they have proses a segmentation methodology, they have found out the difference between transaction sequences then they segment the customers with help of customers' transaction data by the use of the hierarchical clustering method (Fang-Ming Hsu, Li-Pang Lu, Chun-Min Lin, 2012). Body of research into nine streams of literature they have focused enough review of past and current literature, relationship value management they have proposed a framework (Adrian Payne, Sue Holt, 2002).

CHAPTER-3

RESEARCH METHODOLOGY

3.1 Introduction

We used an unsupervised classification Machine Learning algorithm which is a K-means algorithm. This algorithm has been devoted based on some features. Our methodology started with assemblage data from a supermarket mall. The main part of the application of K-means clustering is the segmentation of customers and the customer segmentation is based on the customer data of the supermarket. Customer ID, Gender, Age, Annual income, Spending scores features helped us to mold our model.

3.2 Research Subject and Indication

Our research main goal is about clustering the customers of the supermarket mall. If we can perceive and cluster the customer into groups then we can perceive which group of customers will play a vital role to enhance the revenue of the company.

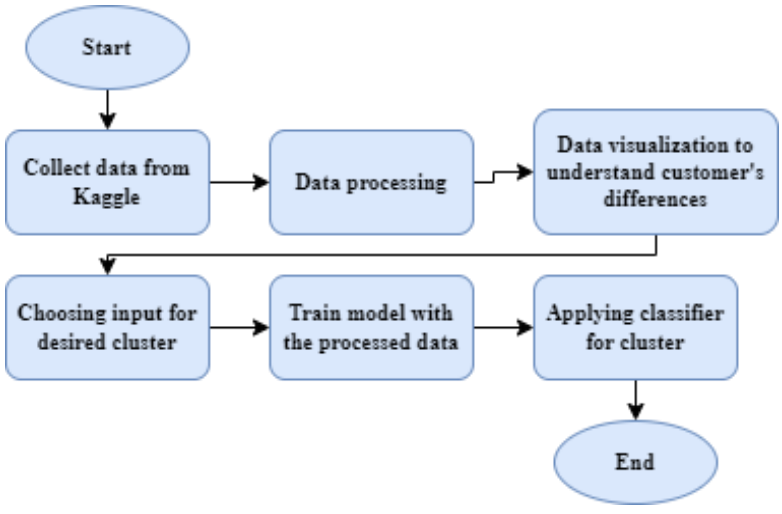


Figure 3.1: Research process flowchart

3.3 Data Collection and Dataset

We used data of a supermarket mall through customers' membership cards from Kaggle which has 200 datasets with 5 features. In this dataset ID is a unique value assigned to each customer, spending score based on the various measure as like annual income, how much money a customer exhausted in a year.

Table 3.1: Information of dataset

Customer ID	Gender	Age	Annual Income	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

3.4 Data Preprocessing

We collect our data from Kaggle, after that change our data into a CSV file. Before applying the K-means algorithm We use the panda's library in python to preprocess our dataset. First, we have applied default value handling because K-means clustering can't manage default values. As we can't reject to trim missing observations from segmentation that's why we imposed missing observations. We have counted the distance of K-means with categorical variables. When categorical variables are integrant variables then we replace them with the arithmetic sequence of correct dissimilation and when categorical variables are cardinal variables then we translate them into multiple binary numerical variables as the number of categorical classes. After that, we normalized our data to the unit of the rate that shouldn't pervert

correlative closeness of espial. We used the curse of dimensionality. We need to apply random initialization of centroids to constituent overlook frequent clustering process.

3.5 Implementation Requirement

Why it's known as the K-means? Since the letter addresses the number of clusters picked. To a specific perception which allots implies the closest bunch of some extraordinary capacity. Partake the k-means algorithm for clustering. To divide a data set into K distinct clusters a K-means clustering is facile and exoteric approach, to specify the desired number of clusters K, K-means has been performed then each observation to exactly one of the k clusters has been assigned (Boyu Shen, 2021).

Table 3.2: K-means Algorithm

Simplified simulation flow of the k-means algorithm
Start Inputs: X= (x1, x2, ..., xn) Determine: Clusters – k Initial Centroids – C1, C2, …., Ck Assign each input to the cluster with the closest centroids Determine: Update Centroids – C1, C2, …., Ck Repeat: Until Centroids don't change significantly (specified threshold value) Output: Final Stable Centroids – C1, C2, …., Ck End

In reality, the number of clusters is unknown to us. There are a few strategies to choose K that relies upon the space information and rule of thumbs.

With the help of K-means clustering, we can find out the cluster, where K is the free perimeter. K-means discover clusters in data and where the number of clusters is presented by the K value. First, we need to identify the K value before starting the algorithm then we need to find out some random points which are recked of as the center between clusters, this is addressed by centroids. After that own the distance of each data point from centroids.

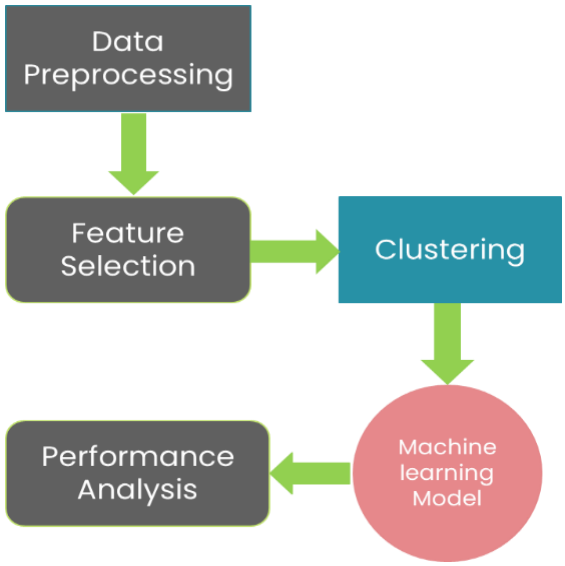


Figure 3.2: Methodology

The elbow method is one of the hearties used to discover the modal number of clusters. Here is the sum of distances of perceptions from their cluster centroids addresses Within-Cluster-Sum-of-Square (WCSS)

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

Here Y*i* is the centroid of observation Xi. It's equipped to augment the number of clusters and in restricted cases, every data point turns into its own cluster centroid. In figure 3.1 we can see WCSS is decreasing for 5 clusters, which means we are going to reach the true number of clusters because when WCSS is decreasing the clustering performance is increasing as well.

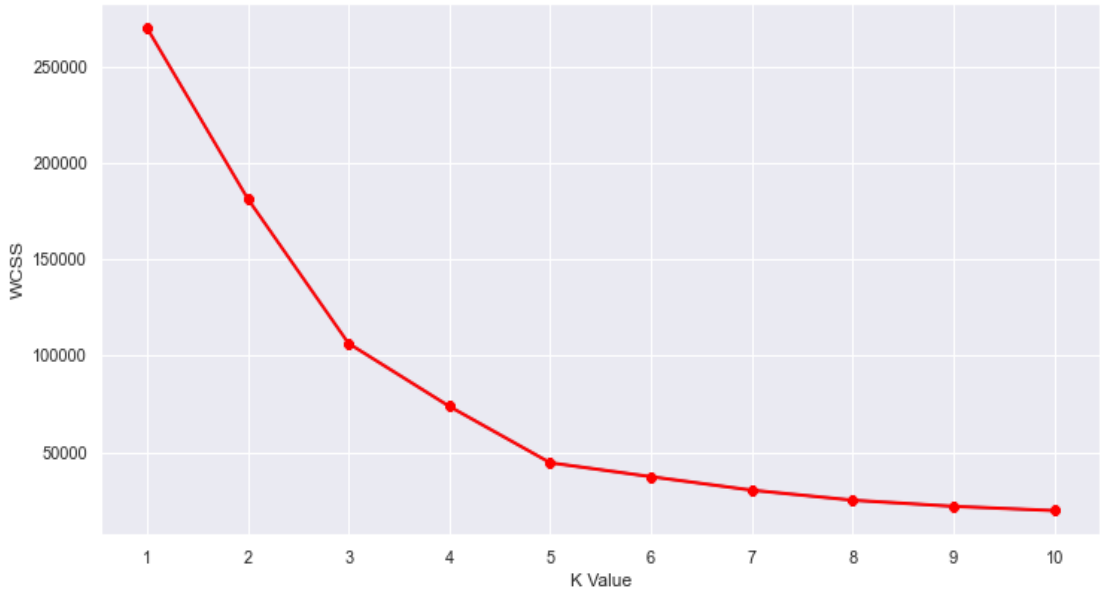


Figure 3.3: Elbow Method Graph

From Elbow Method we got our proper cluster number which is k-number. Depending on this K-value we will do the rest of the work. Here our K value is K=5.

CHAPTER-4

RESULT & DISCUSSION

4.1 Execution Parameters

The whole scheme was centered around finding out which group of customers will be used to increase the revenue of the company.

We have used unsupervised machine learning and K-means clustering algorithm for our model to segment supermarket customers.

4.2 Graphic Analysis

From those figures, we can find out the result of using machine learning algorithms after executing the model. All of the figures are based on customers' age, gender, annual income, spending score, and their clusters.

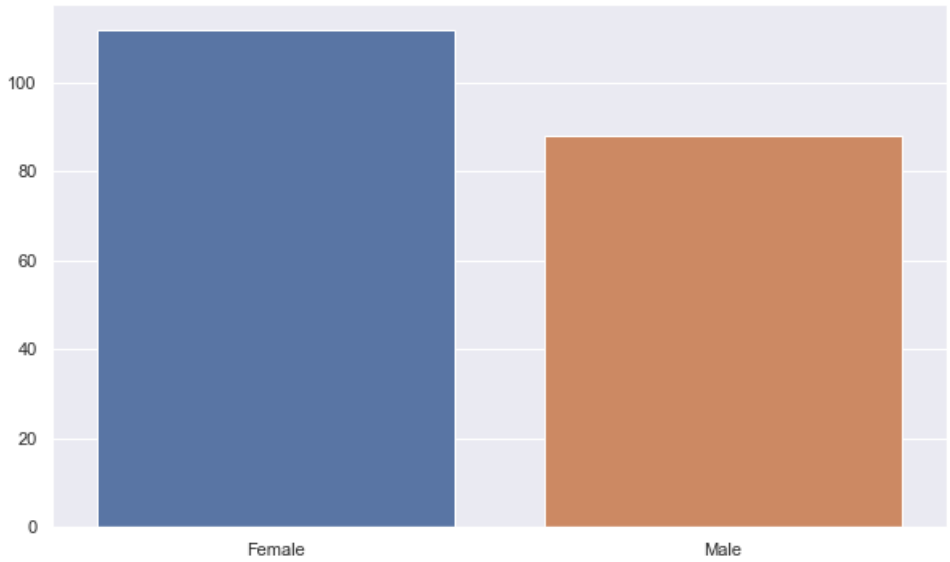


Figure 4.1: Distinct of Gender

From figure 4.1 we can see that female does more shopping from the supermarket.

Figure 4.2: Spending Score (1-100) vs Annual Income(k$) Cluster

In figure 4.2 we have applied the K-means clustering algorithm between spending score and annual income. From this, we have discovered a potential cluster to reach our objective from the clusters in general, that is whose annual income is 40-70 and spending score is 40-60.

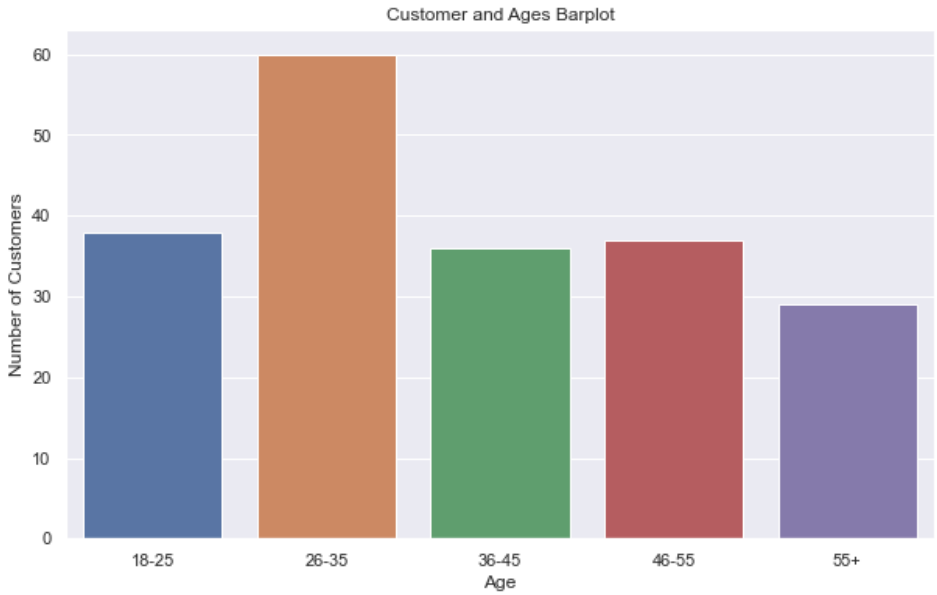


Figure 4.3: Customer and Ages Bar plot

We can view from figure 4.3 that our highest number of customer age is between 26-35.

Figure 4.4: Customer and Spending Score(1-100) Bar plot

We can observe from figure 4.4 that most spending score is 40-60 from all of the spending scores of the customers.

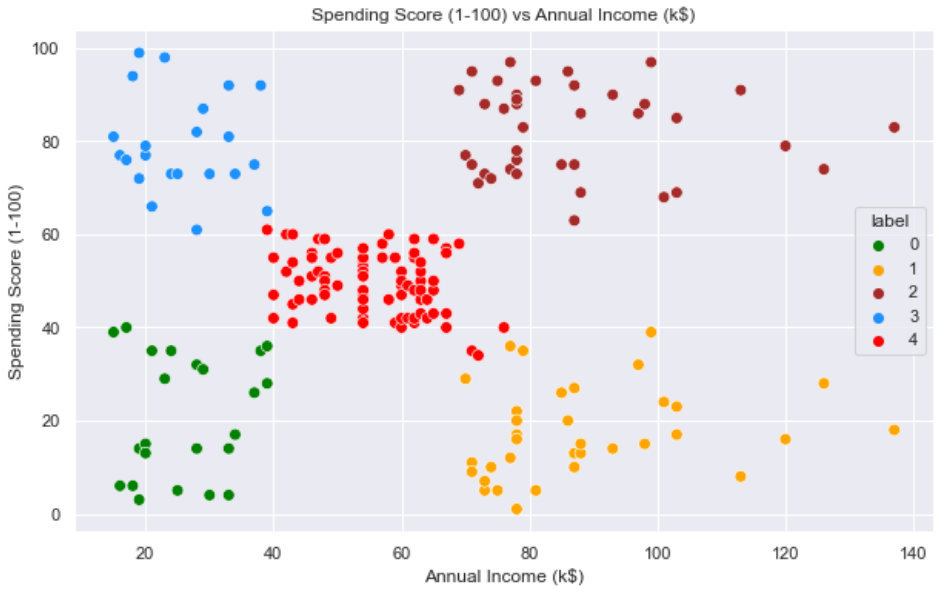


Figure 4.5: Annual Income (k$) and Spending Score (1-100)

We have labeled up all the clusters into label 0-4 in figure 4.5 and label 4 have our desired cluster.

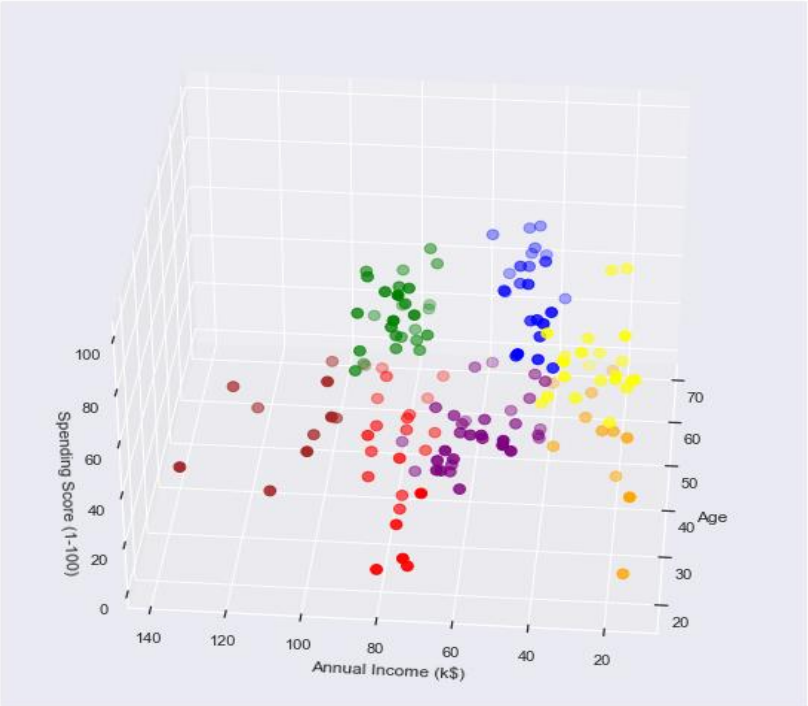


Figure 4.6: Annual Income(K$), Spending Score (1-100), and Age cluster

Ultimately, we have our most awaited result that we have got our targeted group of customers which was the fundamental requirement of our research.

We got our targeted customer based on the two most significant features which are annual income and spending score because as more a customer will acquire as much, he/she will do the shopping and we also compare these two features with the age feature so that we can figure out which age of customer we can do target.

CHAPTER-5

CONCLUSION & RECOMMENDATION

5.1 Discoveries, Contribution & Limitation

Nowadays so many supermarkets are growing. More and more people are also visiting these supermarkets. That`s why the owner of these companies should be aware of their customers. They should know customer demands, what type of products the customer wants, which kind of service from supermarkets they want. Because these customers are the main asset of their company to increase their revenue.

Any kind of public and private field will be benefited from the help of this segmentation and classification resolution. In this modern era, entrepreneurship and digital marketing are growing more and more, so with the help of this model, this sector will also be benefited. They can easily find out which group of customers are buying more products, which group of customers is purchasing repeatedly.

With help of this proposed model, the better achievement will be given. The optimistic outcome in this event is given by the affirmative set timing. By using K-means clustering we have been got our targeted customer group. Parallel to the different results, our model has achieved an excellent outcome.

As a new researcher, I don't have the experience of many years to conduct research and produce academic papers of such a large size individually, in my paper the scope and depth of discussion have been compromised on many levels compared to the works of

experienced scholars. I've pointed out some limitations from our paper in the future work part.

5.2 Recommendation for Future Works

This will work for the marketing system by the rule of this solution. This model is created by us to find out targeted customers. Collecting data from customers in the future also. With the help of predictable constructive output, companies can reduce their business risk.

In further for a more accurate result, we want to increase our dataset. For the purpose of the business, we want to affix additional features. The owners and buyers who conduct their life based on their income profit loss and who are very disquieted about their day-by-day shopping from them could be watchful. For any business, trickery is an immense issue.

We can make a website for the supermarket, by this, the popularity of supermarket will grow cause nowadays people are showing more interest on online shopping. By this customer can save their time and will be more satisfied.

If anyone wants to do customer segmentation they can implement other clustering algorithms like Hierarchical Clustering, Density-Based Clustering also can use other segmentation features like demographics. Also, can compare other clustering algorithms with the K-means algorithm to do customer segmentation and get more excellent output.

REFERENCES

[1] Boyu Shen, CONFCDS 2021, January 28-30, 2021, Stanford, CA, USA "E-Commerce Customer Segmentation via Unsupervised Machine Learning".

[2] Tusar Kansal, Suraj Bahuguna, Vishal Singh, Tanupriya Choudhury, 2018 IEEE "Customer Segmentation Using K-means Clustering".

[3] Chih-Fong Tsai, Ya-Han Hu, Yu-Hsin Lu, ©2013 Wily Publishing Ltd, "Customer Segmentation issue and strategies for an automobile dealership with two clustering techniques".

[4] Kim, Jong Tak, Hee-Jun Pan, and Jonghun Kim. "P2Pbased u-health cluster service model for silver generation in PBR platform." Peer-to-Peer Networking and Applications. 14 July 2015, SpringerScience, Business Media New York 2015.

[5] Nobel, Natasha E, Christine L, Paul, Nicole Turner, Stephen V. Blunden, Christopher Oldmeadow and Heidi E. Turon "A cross-sectional survey and latent class analysis of the prevalence and clustering of health risk factors among people attending an Aboriginal Community Controlled Health Service".

[6] Natalya V. Razmochaeva, Dmitry M. Klionskiy, Vladimir V. Chernokulsky, ©2018 IEEE "The Investigation of Machine Learning Methods in the problem of Automation of Sales Management Business-process".

[7] Alzahrani, Saeed G, Richard G, Watt, Aubrey Sheiham, Maria Aresu,, GeorgiosTsako "Patterns of clustering of six health-compromising behaviors in Saudi adolescents" volume14(5), Alzahran, BMC Public Health 2014

[8] Grosskreutz, Henrik, Mario Boley, and Maike KrauseTraudes, Subgroup discovery for election analysis: a case study in descriptive data mining."volume 6332, In International Conference on Discovery Science,Discovery Science pp 57-71.

[9] Brito, Pedro Quelhas, Carlos Soares, Sérgio Almeida, Ana Monte, and Michel Byvoet. "Customer segmentation in a large database of an online customized fashion business." Robotics and Computer-Integrated Manufacturing, Elsevier ,2015

[10] Azizpour, Shahriar, Kay Giesecke, and Gustavo Schwenkler. "Exploring the sources of default clustering."June 15, 2011; this draft February 24, 2014.

[11] Guha, Sudip, Nina Mishra "Clustering data streams: In Data Stream Management Springer Berlin Heidelberg, 2016

[12] Baer D. (2012) CSI: Customer Segmentation Intelligence for Increasing Profits. SAS Glob Forum. 2012:1-13.

[13] Ma, H. (2015) A Study on Customer Segmentation for E-Commerce Using the Generalized Association Rules and Decision Tree. 2015;(December):813-818.

[14] Hua S, Xiu S, Leung SCH. Expert Systems with Applications Segmentation of telecom customers based on customer value by decision tree model. Expert Syst Appl2012;39(4):3964-3973. doi: 10.1016/j.eswa.2011.09.034.

[15] Roung-Shiunn Wu, Po-Hsuan Chou, 1567-4223/$ - see front matter 2010 Elsevier B.V. All rights reserved "Customer Segmentations of multiple category data in e-commerce using a soft clustering approach"

[16] E.Y.L Nadapala, K.P.N Jayasena, 15th (IEEE) International Conference on Industrial and Information Systems (ICIIS) 2020 "The practical approach in Customer segmentation by using the k-means Algorithm".

[17] Phan Duy Hung, Nguyen Thi Thuy Lien, Nguyen Duc Ngoc, © 2019 Association for Computing Machinery. "Customer Segmentation using Hierarchical Agglomerative Clustering"
[18] A.S.M Shahadat Hossain, 2017 3rd IEEE "Customer Segmentation using centroid based and density-based clustering algorithms".
[19] Konstantinos T. Antonios, C. (2010). Data Mining Techniques in CRM: Inside Customer Segmentation".
[20] Jinafu, L., Jianshuang L, Huaiqing H. (2011) "A simple and Accurate Approach to Hierarchical Clustering Journal of Computer Information System".
[21] Anup Chandra Bepary, Zannatul Ferdous, Afsara Tasneem Misha (2020) "Customer Segmentation by using Machine Learning and E-commerce Solution".
[22] Sriramakrishnan Chandrasekaran, Abhishek Kumar (2019) "A clustering approach for customer Billing prediction in Mall, Machine learning".
[23] David L. Davies "Cluster separation measure".
[24] Monireh Hosseini, Mostafa Shabani, (2015) "New approach to customer Segmentation based on changes in customer value".
[25] Surefunmi Idowu, Srivastav Kattukottai (2019) "Customer Segmentation Based on RFM model using K-means, Hierarchical and Fuzzy C-means Clustering Algorithm".
[26] Jorge Rodriguez, Ivana Semanjski, Sidharta Gautam, Nico Van de Weghe, Daniel Ochoa, 2018) "Unsupervised Hierarchical Clustering Approach for tourism market segmentation based on Crowdsourced Mobile Phone data".
[27] Jianfu li, Jinashuang Li, Huaiquing He, (2011) "A simple and accurate approach to Hierarchical Algorithm".
[28] Bernad Jumadi Dehotman Sitompul, Opim Salim Sitompul, Poltak Sihombing, (2019) "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-means Algorithm".
[29] J. Wiens, Jhon V Guttag, E.J. Horvitz (2012) "Patient risk stratification for hospital-associated C. Diff as a time series classification task".
[30] E T Tosida, F Andria, I Wahyudin, R Widianto, M Ganda, R R Lathif "A hybrid data mining model for Indonesian telematics SMEs empowerment".
[31] Fang-Ming Hsu, Li-Pang Lu, Chun-Min Lin (2012) "Segmentation Customer by transaction data with concept hierarchy".
[32] Adrian Payne, Sue Holt, (2002) "Integrating the value process and Relationship Marketing".

PLAGIARISM REPORT

Turnitin Originality Report

Processed on: 22-Jan-2022 14:32 +06
ID: 1745898139
Word Count: 3816
Submitted: 1

162-35-1676 By Shara Binte Osman

Similarity Index	Similarity by Source
14%	Internet Sources: 11% Publications: 3% Student Papers: 5%

3% match (Internet from 02-Apr-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/5319/162-15-8057%20%2829_%29.pdf?isAllowed=y&sequence=1

2% match (student papers from 09-Nov-2021)
Submitted to Daffodil International University on 2021-11-09

1% match (Internet from 02-Apr-2021)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/123456789/4727/161-11-304%20%3d39%25.pdf?isAllowed=y&sequence=1

1% match (Internet from 07-Nov-2021)
https://www.researchgate.net/publication/271302240_Customer_Segmentation_Using_Clustering_and_Data_Mining_Techniqu

1% match (student papers from 15-Aug-2020)
Submitted to RMIT University on 2020-08-15

1% match (publications)
S. Sengole Merlin, Nisha Maria Arunkumar, Miriam A Angela. "Automated Intelligent Systems for Secure Live Migration", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018

1% match (student papers from 11-Dec-2017)
Submitted to Massey University on 2017-12-11

< 1% match (Internet from 07-Jul-2020)
https://dblp.dagstuhl.de/search/publ/bibtex/?q=stream%3Astreams%2Fjournals%2Fes%3A

< 1% match (Internet from 03-Apr-2021)
http://www.ieomsociety.org/harare2020/papers/431.pdf

< 1% match (student papers from 30-Jan-2020)
Submitted to Saint Anselm College on 2020-01-30

< 1% match (Internet from 16-Jul-2021)
https://www.analyticsvidhya.com/blog/2021/05/k-means-clustering-with-mall-customer-segmentation-data-full-detailed-code-and-explanation/

< 1% match (publications)
Liu, Qiuhong, Weiming Fu, Jiahu Qin, Wei Xing Zheng, and Huijun Gao. "Distributed k-means algorithm for sensor networks based on multi-agent consensus theory", 2016 IEEE International Conference on Industrial Technology (ICIT), 2016.

< 1% match (Internet from 15-Jul-2021)
https://sersc.org/journals/index.php/IJAST/article/download/17654/8921/

< 1% match (Internet from 02-Jan-2022)
https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning

< 1% match (publications)
Subhranil Koley, Aurpan Majumder. "Brain MRI segmentation for tumor detection using cohesion based self merging algorithm", 2011 IEEE 3rd International Conference on Communication Software and Networks, 2011

< 1% match (publications)
E T Tosida, F Andria, I Wahyudin, R Widianto, M Ganda, R R Lathif. "A hybrid data mining model for Indonesian telematics SMEs empowerment", IOP Conference Series: Materials Science and Engineering, 2019

< 1% match (Internet from 05-Nov-2020)
https://dblp.uni-trier.de/pid/w/NVdWeghe.html