Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 August; 6 (8s): 1150-1158

Customer Segmentation of Botanical Contact Printed Products Using K-Means Clustering

Adrita Biswas[1], Dr. Neeraj Rawat Sharma[2*]

[1]Ph.D. Scholar at Banasthali Vidyapith, India
[2*]Asstt. Professor, department of design, Banasthali Vidyapith, Rajasthan, India
Email: nirajdesigns47@gmail.com

Abstract

The growing demand for sustainable fashion has catalysed an interest in botanical contact printed products worldwide, which utilize nature to create unique imprints and textile designs. In this study, we have segmented the available data into distinct clusters using the k- means method, which is an unsupervised machine learning technique, to determine the data points that lie within the confines of this market. By analysing the data collected—including earnings and expenditures, preferences, and demographic information—the aim is to uncover meaningful patterns that can shape marketing strategies for evolving sustainability.

Recognizing customer groups is essential for businesses for several reasons. In the first place, it helps companies direct marketing more precisely, increasing the possibility that within the available resources a message would be heard by the targeted industry groups. Apart from that, it is possible to optimize the range of products as a defined customer segment would appreciate the product features. Thirdly, it helps to gain loyalty because people want to be recognized with brands that resonates their belief systems. Insightful data derived from K- means clustering helps businesses increase customer satisfaction, revenue, and growth in the sustainable textile sector. Having consistent actionable marketing strategies based on data is vital to the growth of companies in the textile sector that wish to tap into the more environmentally focused customers.

Keywords: *Clustering, K-Means, Data Mining, Customer Segmentation, Botanical Printing*

Abbreviations: RFM- Recency, Frequency, Monetary Value, WCSS- Within Cluster Sum of Square, EM- Elbow Method, ASM- Average Silhouette Method, GSM- Gap Statistic Method.

1. Introduction

1.1 Botanical Contact Printing

A shift is currently occurring in the textile industry as there is an increase in demand from consumers for sustainable and environmentally friendly products [1]. In this case, the botanical contact printing technique could be appealing to such customers as it produces beautiful textile prints paired with environmental sustainability. This method uses natural materials to imprint plant designs onto fabric, resulting in products that are both aesthetically pleasing and sustainable [2]. With the growth in demand for such textiles, it is vital for companies to understand the constituent customer segments in order to diversify marketing strategies accordingly. [3].

Botanical contact printing is not just a process; it is an integral part of further movement aimed at reducing the environmental footprint of the fashion industry. Traditional textile manufacturing often involves harmful chemicals and processes that can have detrimental effects on the environment. On the contrary, botanical contact printing technique that uses natural fabrics, natural dyes, and other materials wins the rest of the population that is more concerned with the material transparency and ethics of textile production in their consumer choices.

1.2 Customer Segmentation

Customer segmentation is key in marketing strategies and involves breaking down a wide customer base in smaller but more manageable groups with common traits and characteristics [4]. This approach gives the opportunity to assess different behaviours and needs of the audience, helping in the development of focused marketing strategies and initiatives that connect with the audience on different levels [5]. This particular study seeks to apply the resilient unsupervised machine learning algorithm K-Means, K-Means [6], as a means to determine potential different customer segments in the customer market for botanical contact printed merchandise [7]. By studying customer data on buying behaviours, preferences, and demographics, marketers look to provide customer insights that can improve customer satisfaction and increase sales in this emerging market for green textiles [8].

There is a need of contact printed products for business to print contact and there is ever- increasing market for printed

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 August; 6 (8s): 1150-1158

products. Companies need to study different consumer attitudes and behaviours which is a complex dome [9]. Understanding who their customers are—what drives their purchasing decisions and how they interact with brands—becomes crucial [10]. This research seeks to fill that gap by employing K-means clustering to analyse customer data and identify distinct segments with similar characteristics.

1.3 K-means Clustering

Segmentation, in particular, is an approach in machine learning that aims to cluster data into smaller groups based on particular attributes which may include colour, size, and shape. These attributes have in the past few years been studied in different domains [11],[12],[13], [14], [15].

Out of these, K-Means is the most popular and simplest as well, most efficient method of clustering since it reduces the within-cluster sum of squares (WCSS), thereby producing compact and well-separated clusters. It focuses on and minimizes absolute distances to each of the data points, giving compact and well-separated cluster results disproving the within- cluster sum of squares (WCSS).

However, the most evident challenge is finding the most optimal cluster number (K) needed for accurate results [16], [17]. Other such approaches include the Elbow Method (EM) [18][19][20], Average Silhouette Method (ASM)[21] and Gap Statistic Method (GSM) [22].The focus of this specific study is customer segmentation with the application of the K- Means algorithm.

Research Objectives

The objectives of this research are:

1. **Identification of Customer Segments**: To identify how many distinct customer segments are there within the market for botanical contact printed products.
2. **Analysis of Segment Characteristics**: To explore and identify unique characteristics, preferences, and behaviours of those identified segments.

In order to contribute valuable knowledge to the field of sustainable fashion marketing especially with a relatively new area such a botanical contact printing with the necessary tools to engage the target customer base effectively, the study was conducted.

2. Research Methodology

A structured methodology was implemented to identify the customer segments for botanical contact printed products using K-means clustering. The methodology has several critical steps: a strategy for sampling, data collection, data pre-processing, clustering implementation, and result validation.

2.1 Sampling Strategy

The first step was executed with having a clearly stated direction on sampling. A set of pop- up stores or experience centres are stationed strategically in foot traffic friendly ecosystems like eco stores, fairs of art & clothing, or community celebrations of sustainability. The selection was made to acquire a varied set of prospective buyers who were likely to be interested in buying something related to contact printed botanical stuff, e.g. Art Galleries, Craft Council Exhibition, Sustainable fairs, Handloom & Handicraft Exhibitions, Design fairs etc. Only participants who made at least one purchase were allowed, maximizing the effectiveness of purchasing intent in capturing the sample ensuring the dataset's relevance was clear. Within a few months' time, 305 customers were profiled at the sites and the profile served a good set of values in terms of behavioural analysis.\

2.2 Data Collection

Surveys were conducted during scheduled times at the pop-up stalls in the form of snapshots using survey interlaces. The 'Responding' portion was made in a way to capture important behavioural and demographic data. Customers stated their ages and their monthly wage which made it easy to see to see trends for income. In addition, the survey asked the customers the about their purchase frequency and liking which helped to find their position and behaviour in the Recency, Frequency and Monetary score or in other words, the RFM score was ascertained. This explanation paints a full picture of buyers on the customer spectrum while capturing critical demographic data through the different shapes of data fields.

2.3 Data Pre-processing

The data collected went through a number of sequential steps so that the data would be of the right quality for the intended analysis. The first step focused on allocated handling missing data. Age and income questions that had no responses were regarded and either substituted, or omitted from the data set as a means in keeping the integrity of the data. Omitting sets was important in validating the accuracy of the data in question. In the next step, the age and income metrics were put through normalization. This provided that the two variables clustered together, lending to a better relationship depiction. Specifically, the values were scaled to a range of 0 to 1, which is particularly important in clustering algorithms where distances between points dictate cluster formation. Additionally, while the primary focus was on numerical data, any categorical responses, such as gender or background, can also be encoded appropriately to facilitate further analysis.

2.4 K-Means Clustering Implementation

The K-means algorithm is used next to hone in on different customer profiles based on age as well as monthly income. This in turns splits data into different segments. The first step in the analysis involves deciding how many clusters in the data set there are. This is achieved through the Elbow method. The method involves graphing K vs WCSS and determining what k value has the lowest WCSS.

When optimal K has been computed, K means should be run on the clean dataset. Each customer should be assigned, by their monthly income and RFM score, closest to the cluster centroid. The algorithm repeated the calculations for the cluster centroids resulted in stable clusters that accurately represented customer segments. After clustering, the two resulting segments were analysed to identify their characteristics. The first segment was characterized by younger customers with lower monthly incomes, likely prioritizing affordability and value in their purchases. In contrast, the second segment comprised older customers with higher monthly incomes, who were expected to place a greater emphasis on quality and sustainability in their purchasing decisions.

2.5 Validation of Clusters

The clustering results were to be validated next, for which additional analyses were performed. One of the key metrics used for validation was the Elbow method. This was calculated for each data point in R Studio, which helped us to understand what the similar data points have in common in their own cluster compared to others. Furthermore, for better perusal, visual analysis was conducted by generating a scatter plot to represent the clusters visually. This plot clearly depicted the distinct buyer segments based on their monthly income and RFM score, which contributed greatly in identification some behavioural facts for marketing strategies.

Figure 1Flow chart for the K-means clustering algorithm

The steps of data analysis cannot be initiated until the collection is systematized and processed to make it uniform in representation and entail assigning values to absences, redefining values to fit global standards rationalization of dominant languages, etc.The processed data is fed into the K-means clustering algorithm to delineate customer groupings. These customer groupings can contain segments which can be further subdivided or the total number of clusters. The total number of clusters are determined by the Elbow method. The Elbow method is used to find the number of clusters such that any further addition to the cluster does not yield any increase in 'value'. All possible segments are cluster analysed to find the groups unique features and preferences.

3 Implementation & Analysis

Before applying k-means, the collected dataset was loaded in the RStudio software and the necessary packages are installed.

```
library(corrplot)
library(cluster)
library(factoextra)
library(ggplot2)
```

The tabulated dataset was then loaded and checked for missing values using the following code.

```
data <- read.csv("C:/Users/Arnab Das/Downloads/adrita.csv")
data <- na.omit(data)
str(data)
```

This resulted in the following output where the system 'handles' the missing value and displays the final table.

```
'data.frame':   300 obs. of  9 variables:
 $ Date                         : chr  "17-Dec-22" "17-Dec-22" "17-Dec-22" "17-Dec-22" ...
 $ Gender                       : chr  "Female" "Male" "Female" "Female" ...
 $ Age                          : int  28 26 39 55 29 44 51 48 50 49 ...
 $ Current.Status.              : chr  "Job" "Job" "Job" "Job" ...
 $ Monthly.income               : int  20000 20000 60000 36000 32000 62000 61000 55000 62000 72000 ...
 $ Purchase.amount              : int  2200 2600 2200 15000 3300 1500 1800 2200 300 550 ...
 $ No.of.items                  : int  1 2 1 5 3 1 1 1 1 1 ...
 $ No.of.days.since.last.purchase: int  980 980 980 980 979 979 979 979 979 979 ...
 $ RFM.Score                    : num  4.33 7.67 4.33 9 8 ...
 - attr(*, "na.action")= 'omit' Named int 301
  ..- attr(*, "names")= chr "301"
```

The dataset was investigated more on the relationship between the variables by taking a look at the correlation matrix.

```
                 Monthly.income   RFM.Score   Purchase.amount
Monthly.income             1.00        0.00             -0.05
RFM.Score                  0.00        1.00              0.63
Purchase.amount           -0.05        0.63              1.00
```

In accordance with the output, there appears to be a weak inverse correlation between monthly income and the purchase amount. Given that neither the purchase amount nor the purchase frequency significantly affects the monthly income, there was a need to further clarify the connection between monthly income and the RFM Score.

There is a primary consideration that is needed to be addressed prior to fitting the K- means clustering model. K-means clustering, like many other methods, does suffer from the phenomenon known as clustering bias, which, in effect, is a distortion in the clustering outcome that arises from variables that are measured in different units. More tailoring variables and/or clustering distorters to a common measurement is a necessary step singled out as the distance metric the clustering algorithm employs is only reliable if the variables are of a similar scale.

There are many ways of addressing the problem of metric space distorters to scale, and the Problem of Clustering with Multi-Scale Criteria is a particularly egregious example dedicated to the powerful tool known as variance reduction. Such processes involve a variance-reducing mechanism in which the distance—which as a signature of the feature comparison—is in absolute terms, and is meaningless, suffers from the syndrome known as high dimensionality. Within this framework of this analysis, to be specific, and the R is the environment undergoing the analysis, the scale () function is the particular standardized function of choice that was utilized. Since this function supplies a matrix, the Base-R style is streamlined and a clearer and less complicated option for coding as opposed to the tidy verse style or framework. This standardization step is to ensure that all variables are equally relevant to the clustering process, leading to more accurate results.

```
df <- data[-c(1,2,3,4,6,7,8)]
df[, c("Monthly.income", "RFM.score")] = scale(df[, c("Monthly.income", "RFM.score")])
```

3.1 Fitting and evaluating the model

The identification of the relevant clusters is carried out using the K-means clustering. This is done in few steps, firstly, the number of clusters is taken as three, with the parameter *nstart* configured to 20. This setting helps the algorithm to perform 20 separate runs with different random initializations, allowing R to automatically select the optimal clustering solution based on the minimization of the total within-cluster sum of squares.
In addition, a random seed is chosen in order to preserve the persistent nature of the outcome over different runs of the code.

```
set.seed(123)
km.out <- kmeans(df, centers = 3, nstart = 20)
km.out
```

Output:

```
K-means clustering with 3 clusters of sizes 177, 13, 110

Cluster means:
  Monthly.income RFM.Score
1       32418.93  5.141243
2      130000.69  4.846154
3       61745.45  5.151515

Clustering vector:
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28  29  30 
  1   1   3   1   1   3   3   3   3   3   3   3   3   1   1   1   3   1   1   3   3   3   3   3   3   1   3   1   1   1 
 31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58  59  60 
  3   1   1   3   3   3   3   3   3   1   1   1   1   1   3   1   1   3   3   3   3   2   3   1   1   1   1   1   1   1 
 61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90 
  1   2   1   3   3   3   1   3   2   1   1   3   3   3   3   1   1   3   3   1   3   1   3   3   3   1   1   3   3   1 
 91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 
  1   1   3   1   1   2   1   1   3   3   1   1   3   1   1   3   1   1   2   3   3   1   3   1   1   1   1   1   1   3 
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 
  3   1   2   3   3   1   1   1   1   1   3   1   1   3   1   1   3   1   2   3   2   1   1   1   1   3   1   3   3   3 
151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 
  1   3   3   1   3   1   1   1   3   1   1   1   1   1   1   1   1   3   1   1   1   1   3   1   1   1   1   1   1   1 
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 
  3   1   3   3   3   3   1   1   3   1   1   1   1   1   1   3   3   1   2   1   1   1   1   1   1   3   3   3   1   3 
211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 
  1   1   3   1   1   1   1   1   1   2   1   1   1   1   1   1   1   1   1   1   1   1   1   1   3   3   3   3   3   1 
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 
  1   1   1   1   3   1   1   3   1   3   2   2   1   3   1   1   1   1   3   1   1   3   3   3   2   1   3   1   3   1 
271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 
  3   3   1   1   1   1   1   1   1   1   3   1   3   1   3   1   3   1   1   1   1   1   3   1   1   3   1   1   3   3 

Within cluster sum of squares by cluster:
[1] 15365439648 21601260095 12786873100
 (between_SS / total_SS =  75.2 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"    "size"        
[8] "iter"         "ifault"      
```

From the output, it is evident that the analysis identifies three distinct clusters, and the number of observations in each of them are 177, 13, and 110, respectively. For each cluster, the squared distances of each observation to the cluster centroids are calculated, and the clusters are formed based on the minimum distance association to the centroids. These initial results appear valid; however, further exploration of alternative cluster counts is to be carried out to ensure optimal segmentation.
To do this, the number of clusters is methodically altered from one to ten. This iterative process is evaluated using function called a scree plot, which displays the number of clusters along the x-axis and the total within-cluster sum of squares (WCSS) along the y- axis. In the context of this case study, ten K-means models are generated using only the *price* and *number of reviews* as input features. The WCSS from each model is stored in a variable for use in constructing the scree plot. This method is in line with the Elbow Method, which is a well-known way to figure out how many clusters are best for a particular dataset.

```
wss <- numeric(n_clusters)
set.seed(123)
for (i in 1:20) {
km.out <- kmeans(df, centers = i, nstart = 20)
wss[i] <- km.out$tot.withinss
}
wss_df <- tibble(clusters = 1:n, wss = wss)

library(tidyverse)
wss_df <- tibble(clusters = 1:20, wss = wss)
scree_plot <- ggplot(wss_df, aes(x = clusters, y = wss, group = 1)) +
geom_point(size = 4)+
geom_line() +
scale_x_continuous(breaks = c(2, 4, 6, 8, 10)) +
xlab('Number of clusters')
scree_plot
```

This generates the graph for the WSS Curve for the optimal number of clusters.

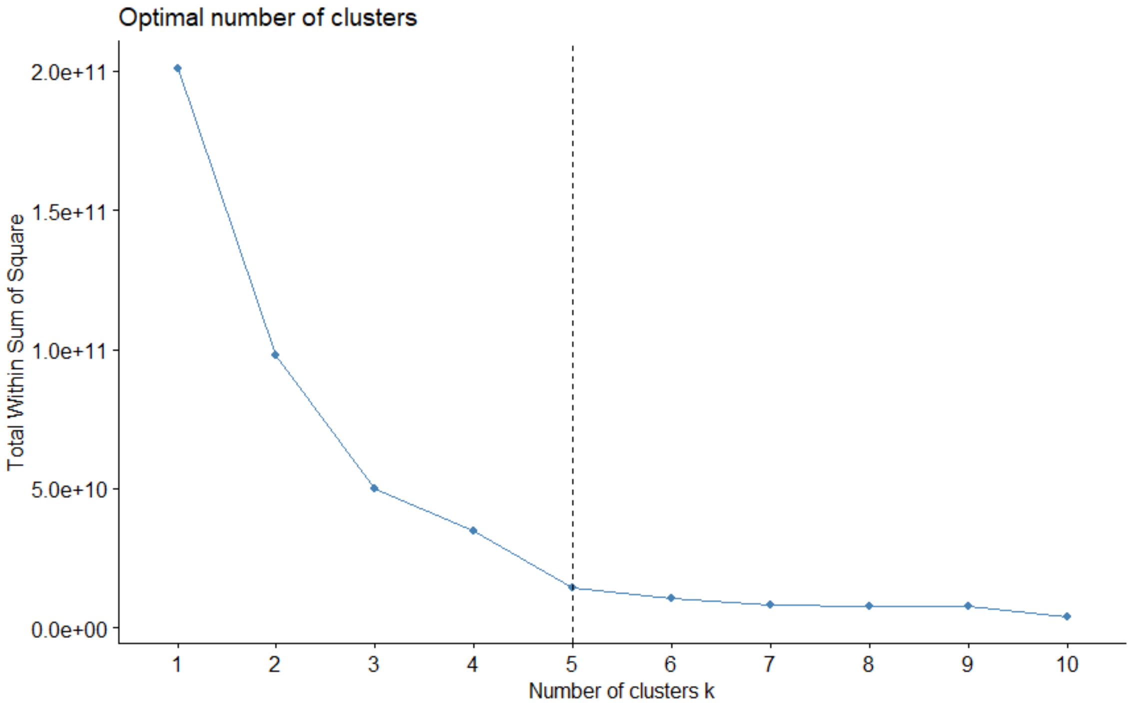

Figure 2 Optimal number of Clusters (Elbow Method)

Studying the scree plot, it is observed that the WCSS total does increase, but only up to a certain number of clusters and then decreases. This method is used the most to find clusters that 'drop'. In this case where the drop in WCSS is relatively flat. After looking closely at the data, the answer becomes more obvious. The image does indicate that five clusters is the best number of clusters since any additional clusters does not significantly increase the score for the model.
Thus, to determine the number of clusters, the value of \(k \) is set to 5, with a seed again applied for reproducibility. In the R programming, clustering can be carried out using the k means () function in built with the relevant package library. It requires specifying the number of desired clusters (*k*) which was obtained by the Elbow method and, optionally, parameters such as the number of random initializations (*nstart*) to improve the clustering outcome.

```
set.seed(123)
km.out <- kmeans(df, centers = 5, nstart = 20)
km.out
```

Subsequently, a scatterplot can be generated to visually represent the relationship between Monthly income and the RFM Score, with points can be color-coded in the programme according to their respective cluster IDs. This is depicted in Figure 3.

Journal for Re Attach Therapy and Developmental Diversities
eISSN: 2589-7799
2023 August; 6 (8s): 1150-1158

```
df$cluster_id <- factor(km.out$cluster)
ggplot(df, aes(Monthly.income, RFM.Score, color = cluster_id)) +
geom_point(alpha = 0.25) +
xlab("Monthly.income") +
ylab("RFM.Score")
```

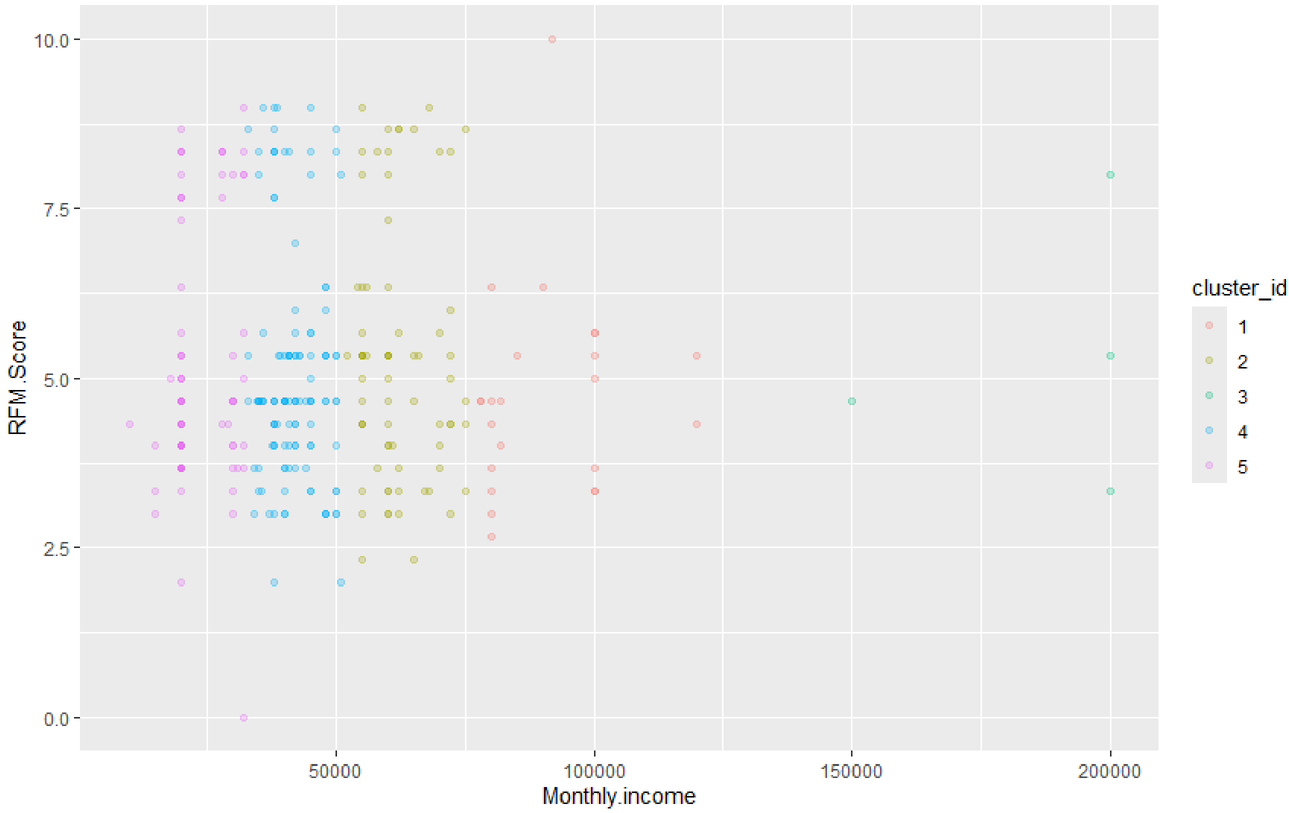

Figure 3 Generated Scatterplot for the clustering

3.2 Actual Clustering

To showcase the results of a clustering process, a cluster plot is usually created. This plot gives a spatial description of the clusters in two dimensions, using methods such as Principal Component Analysis (PCA) whenever the dataset exceeds two dimensions. In R, the fviz cluster () function from the facto extra package is most effective for these types of visual representations. This function shows points with colours indicating the different clusters they belong to and cluster centroids, making it easy to visualize the clustering.

```
km.final <- kmeans(df, 5)
View(df)
km.final$tot.withinss
km.final$size
df$cluster <- km.final$cluster
head(data, 5)
```

```
clusplot(df,
df$cluster,
color = TRUE,
shade = TRUE,
labels = 4,
lines = 1)
```

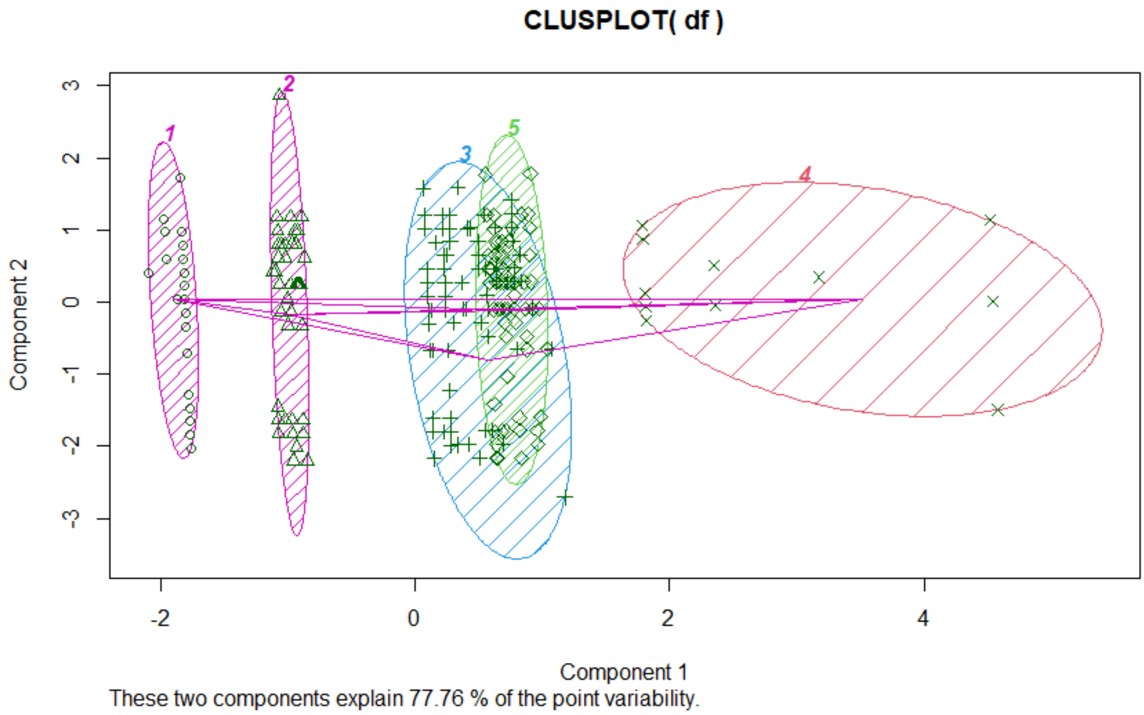

Figure 4 Final Clusterplot

3.3 Analysis

In the cluster plot where Component 1 is the salary earned per month as analysed using RFM and Component 2 is the spending score; it is possible to draw important conclusions especially on the behaviours of the customers in buying the botanical contact printed products.

Occasional Buyers (Cluster 1) and Cautious Consumers (Cluster 2) exhibit some overlap, suggesting that individuals in these groups may share similar spending behaviours despite differing income levels. For instance, lower-income customers in the Cautious Consumers segment may still demonstrate high spending scores, indicating a willingness to prioritize and invest in botanical contact printed products that resonate with their interests. In contrast to the Budget-Conscious Shoppers and Value Seekers clusters which show greater dispersion, Cluster 3 is more compact. It is likely that the Affluent Enthusiasts are individuals earning a high income and therefore capable of spending a great deal on premium botanical contact printed products. This is an example of a high-tier prospective customer. As for the Budget-Conscious Shoppers, this may also include lower-income groups who are reflective of spending scores and budget limitations on what such products and services they are able to purchase. Finally, the Value Seekers segment might also portray moderate-income customers who do not earn high salaries, but somehow, they spend more and this high spending score suggests that they are ready to spend on some appealing botanical products.

4 Conclusion

The ever-growing market of botanical contact printed products makes it crucial to comprehend customer segmentation for businesses aiming to succeed within the eco-friendly textile industry. This study employs K-means clustering to analyse purchase data, aiming to uncover for tailored marketing within the purchase behaviour, preferences, and demographic of the eco-friendly textile industry. These clusters can inform targeted marketing strategies. For example, the Affluent Enthusiasts could be approached with premium botanical contact printed products, whereas those in the Budget-Conscious Shoppers segment could be offered low-cost direct mail pieces or promotions designed to boost engagement. This improved customer insight enables more effective marketing that maximizes sales while building loyalty to this segment of botanical contact printed products.

As we move forward, it is crucial for businesses to embrace data-driven decision-making in their marketing strategies, ensuring that they remain responsive to the evolving preferences of their customer base. Through effective customer segmentation, companies can not only meet the demands of the market but also contribute to a more conscious future in fashion.

References

1. Md. T. Rahaman, A. D. Pranta, Md. R. Repon, Md. S. Ahmed, and T. Islam, 'Green production and consumption of textiles and apparel: Importance, fabrication, challenges and future prospects', *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 10, no. 2, p. 100280, Jun. 2024, doi: 10.1016/j.joitmc.2024.100280.
2. Özlenen. Erdem İşmal, 'Patterns from Nature: Contact Printing', *Journal of the Textile Association.*, vol. . 77, pp. 81–91, 2016.
3. O. Pricopoaia, N. Cristache, A. Lupaşc, R. C. Dobrea, M.-V. Tureatca, and L. G. Dinulescu, 'The Impact of Marketing Strategies on Promoting Sustainability in the Fashion Sector', *Sustainability*, vol. 17, no. 12, p. 5546, Jun. 2025, doi: 10.3390/su17125546.
4. T. Tavor, L. D. Gonen, and U. Spiegel, 'Customer Segmentation as a Revenue Generator for Profit Purposes', *Mathematics*, vol. 11, no. 21, p. 4425, Oct. 2023, doi: 10.3390/math11214425.
5. C. J. Yu, L. Wu, Y. Chiao, and H. Tai, 'Perceived quality, customer satisfaction, and customer loyalty: the case of lexus in Taiwan', *Total Quality Management & Business Excellence*, vol. 16, no. 6, pp. 707–719, Aug. 2005, doi: 10.1080/14783360500077393.
6. H.-H. Bock, 'Clustering Methods: A History of k-Means Algorithms', 2007, pp. 161– 172. doi: 10.1007/978-3-540-73560-1_15.
7. A. K. Jain, 'Data clustering: 50 years beyond K-means', *Pattern Recognit Lett*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/j.patrec.2009.09.011.
8. Y. Li, X. Chu, D. Tian, J. Feng, and W. Mu, 'Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm', *Appl Soft Comput*, vol. 113, p. 107924, Dec. 2021, doi: 10.1016/j.asoc.2021.107924.
9. F. Şahin and C. Söylemez, *Navigating the Shifting Landscape of Consumer Behavior*. IGI Global, 2024. doi: 10.4018/979-8-3693-1594-1.
10. L. G. . Schiffman and L. Lazar. Kanuk, *Consumer behavior*. Prentice Hall, 1991.
11. H. F. Kaiser, 'On Cliff's Formula, the Kaiser-Guttman Rule, and the Number of Factors', *Percept Mot Skills*, vol. 74, no. 2, pp. 595–598, Apr. 1992, doi: 10.2466/pms.1992.74.2.595.
12. J. M. John, O. Shobayo, and B. Ogunleye, 'An Exploration of Clustering Algorithms for Customer Segmentation in the UK Retail Market', *Analytics*, vol. 2, no. 4, pp. 809– 823, Oct. 2023, doi: 10.3390/analytics2040042.
13. T. Banu, '{Customer Segmentation With Machine Learning for Online Retail Industry', *The European Journal of Social and Behavioural Sciences*, vol. 31, no. 2, pp. 111–136, 2022.
14. R. E. Bawack, S. Fosso Wamba, and K. D. A. Carillo, 'A framework for understanding artificial intelligence research: insights from practice', *Journal of Enterprise Information Management*, vol. 34, no. 2, pp. 645–678, Feb. 2021, doi: 10.1108/JEIM- 07-2020-0284.
15. R. Gelbard, O. Goldman, and I. Spiegler, 'Investigating diversity of clustering methods: An empirical comparison', *Data Knowl Eng*, vol. 63, no. 1, pp. 155–166, Oct. 2007, doi: 10.1016/j.datak.2007.01.002.
16. I. K. khan *et al.*, 'Addressing limitations of the K-means clustering algorithm: outliers, non-spherical data, and optimal cluster selection', *AIMS Mathematics*, vol. 9, no. 9, pp. 25070–25097, 2024, doi: 10.3934/math.20241222.
17. K. Tabak Kızgın and S. Alp, 'Studying the Impact of Changing Consumer Behavior During Crisis Periods Through Store Classification', *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 288, Nov. 2024, doi: 10.1007/s44196-024-00694-3.
18. J. E. Gentle, L. Kaufman, and P. J. Rousseuw, 'Finding Groups in Data: An Introduction to Cluster Analysis.', *Biometrics*, vol. 47, no. 2, p. 788, Jun. 1991, doi: 10.2307/2532178.
19. R. L. Thorndike, 'Who Belongs in the Family?', *Psychometrika*, vol. 18, no. 4, pp. 267–276, Dec. 1953, doi: 10.1007/BF02289263.
20. D. J. Ketchen and C. L. Shook, 'The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique. ', *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996.
21. A. Rasid Mamat, F. Susilawati Mohamed, M. Afendee Mohamed, N. Mohd Rawi, and M. Isa Awang, 'Silhouette index for determining optimal k-means clustering on images in different color models', *International Journal of Engineering & Technology*, vol. 7, no. 2.14, p. 105, Apr. 2018, doi: 10.14419/ijet.v7i2.14.11464.
22. I. K. Khan *et al.*, 'Determining the optimal number of clusters by Enhanced Gap Statistic in K-mean algorithm', *Egyptian Informatics Journal*, vol. 27, p. 100504, Sep. 2024, doi: 10.1016/j.eij.2024.100504.