**Nanyang Technological University**

**School of Mechanical and Aerospace Engineering**



**MA4829**

**Analysis of Oral-B Survey Data Using Data Mining**

28 Feb 2021

1. Abstract

Data mining is evidently important in our information rich society as an analytical tool. Many businesses and industries utilise data mining to aid them in making informed decisions by determining the trends in the marketplace and gaining insights on consumer preference as well as product positioning. Data mining is able to do so by employing software which will search for patterns in a large set of data.

The goal of this project is to analyse Oral-B’s survey data using data mining techniques which comprise of Supervised and Unsupervised learning. In Supervised learning, the dataset contains labeled training data and examples of the method are Regression and Classification. In Unsupervised learning, the dataset contains unlabelled training data and examples of the method are Agglomerative Hierarchical Clustering, Principal Component Analysis (PCA) and Association Rule Mining. With the information and knowledge gained through the two learning processes, we will be able to come up with optimal strategies of product design and development to cater the product to the consumers and boost sales.

1. Analysis

Principal Component Analysis (PCA)

Principal Component Analysis was done on the Characteristic Rating and the Features that Customers are willing to pay a Premium for. After data cleaning, preprocessing was done. The rows of data are transposed into vertical columns. The processed data was then saved as an excel file, labelled as “FeaturesPremiumCRating.xlxs”. RStudio is the primary software used to perform the PCA.

The cleaned data was first imported into RStudio, then categorised by Characteristics, Gender, Types and Features Premium. The function prcomp was used to perform the PCA. The Features that customers are willing to pay a premium for and how highly they rated the list of characteristics were analysed.

After PCA was done, the top two principal components are plotted against each other and clustered in terms of gender and toothbrush types in order to identify if different gender groups have any particular features and characteristics that customers are looking for.

The code used in PCA can be found in part (a) of the appendix.

Association Rule Mining and Apriori Algorithm

Association Rule Mining is a procedure which aims to observe frequently occurring patterns among the itemsets, where support, confidence and lift are used as attributes of association between items. Support shows the proportion of transactions in the database in which an item appears, this signifies the popularity of an interest. Secondly, Confidence tells us the number of times the relationships are found to be true. Lastly, Lift is the ratio of Confidence to Support, this signifies the likelihood of the itemset A to be True when itemset B is True as well.

Apriori Algorithm is used for finding frequent itemsets in a dataset for Association Rule mining. RStudio provides open source software and has a built-in library function called ‘arules’ which implements the Apriori Algorithm. Using RStudio, we have taken inspiration from Market Based Analysis to compare items in each itemsets and finding correlations between them, by computing strong rules through Association Rule Mining, specifying the minimum support and confidence levels based on our needs.

Using information from the Survey we have collated data from Question 8. We categorised the data on an excel based on whether a person would be **willing** to pay extra for any of the 15 features mentioned, hoping to find a correlation between features which people are willing to pay for. We imported the excel to RStudios to run the Apriori Algorithm.

The code used in Apriori Algorithm analysis can be found in part (b) of appendix.

Agglomerative Hierarchical Clustering (AHC)

Using the excel file used for PCA “FeaturesPremiumCRating.xlxs”, the Characteristic Rating components were transposed and renamed as “CratingCSV” to do Agglomerative Hierarchical Clustering. This method groups the different characteristics that each customer is willing to pay for into multiple clusters based on how similar the data were. Rstudio is the primary software where AHC was done.

The excel file “CratingCSV” was first imported into Rstudio, then a check was done on the data set to make sure that the row which labeled the characteristics are considered as characters by using the head() and tail() function. The row names to be used for the plot later on were then generated with rownames(). Using the require(stats) and as.matrix() function, the distance matrices were defined and rounded off to 3 decimals. With the hclust() function, the hierarchical tree (dendrogram) can then be plotted from the distance matrices generated beforehand. A tidy up was then done on the dendrogram by utilizing the factoextra package available in Rstudio to make it neater.

The code used in Agglomerative Hierarchical Clustering can be found in part (c) of appendix.

1. Assumptions

Within the data collected, there were many unfilled data sets. Data cleaning was done prior to summary analysis. Some assumptions were made to fill the gaps and ambiguity of the raw data. Assumptions that were made are listed as such:

1. Unfilled numerical cells will assume 0 in value
2. Numerical cells that were filled with description was given an assumed value

In particular the rating of the looks of product the following descriptions were discovered

|  |  |
| --- | --- |
| Description | Assumed Rating |
| Indifferent | 1 |
| Other (Too plain) | 1 |
| Other (OK, but more to be desired) | 2 |
| Other (OK, but ordinary) | 2 |
| It’s ugly | 1 |

The summary analysis is tabulated below:

Gender and Type Analysis

|  |  |
| --- | --- |
| Analysis | Number of Entries |
| Total of Survey Entries | 21 |
| Male Users | 13 |
| Female Users | 8 |
| Manual Toothbrush Users | 17 |
| Rechargeable Toothbrush users | 3 |
| Both Manual and Rechargeable | 1 |
| Average Current Price | $3 |
| Average Price Customer willing to pay | $5.30 |
| Battery Life Rating | 2.167/3 |
| How much more customer willing to pay for rechargeable | $13.13 |
| Average Rating of Product Looks | 1.88/3 |

Table 3.1: Breakdown of Survey Demographics

The demographics of customers are categorised into 6 different groups. In particular, the gender and the 3 toothbrush types. This immediately shows the majority of the target customers. It can be shown in Figure 3.1 that the majority of the customers use manual toothbrushes. Manual toothbrush users account for approximately 81% of total survey data.

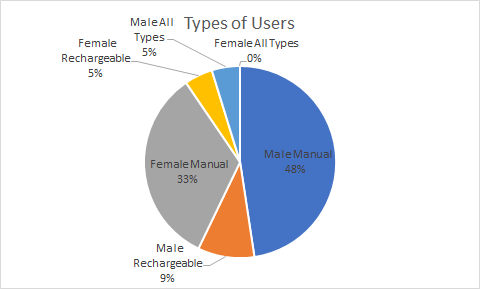


Figure 3.1: Pie Chart of Survey Demographics

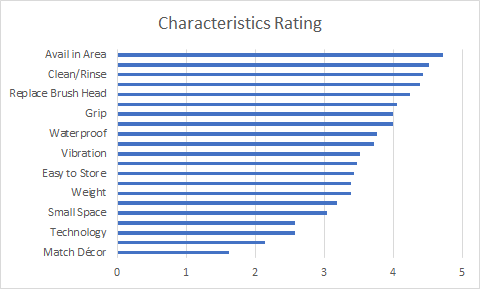


Figure 3.2: Characteristics Rating Ranking from Highest to Lowest Rating

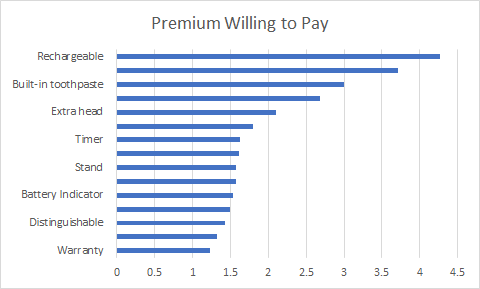


Figure 3.3: Features that Customers are willing to pay a premium

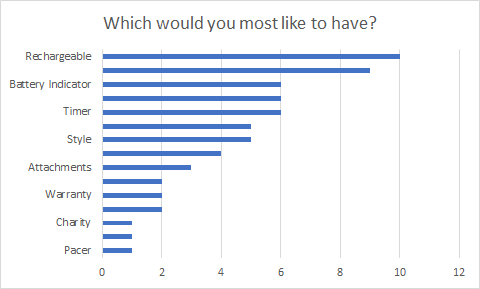


Figure 3.4: Most Wanted Feature that Customers Want

\

1. Results

Principal Component Analysis

We will look at the analysis of the Characteristic Rating then move on to the analysis of the Features Premium.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| PC No. | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
| Standard Deviation | 2.2281 | 1.8568 | 1.6235 | 1.39457 | 1.26646 | 1.11561 | 1.04042 |
| Proportion of Variance | 0.2364 | 0.1642 | 0.1255 | 0.09261 | 0.07638 | 0.05927 | 0.05155 |
| Cumulative Proportion | 0.2364 | 0.4006 | 0.5261 | 0.61869 | 0.69507 | 0.75433 | 0.80588 |
| PC No. | PC8 | PC9 | PC10 | PC11 | PC12 | PC13 | PC14 |
| Standard Deviation | 0.9995 | 0.91166 | 0.737 | 0.69822 | 0.59024 | 0.54048 | 0.48197 |
| Proportion of Variance | 0.04757 | 0.03958 | 0.02587 | 0.02321 | 0.01659 | 0.01391 | 0.01106 |
| Cumulative Proportion | 0.85345 | 0.89303 | 0.91889 | 0.94211 | 0.95869 | 0.97261 | 0.98367 |
| PC No. | PC15 | PC16 | PC17 | PC18 | PC19 | PC20 | PC21 |
| Standard Deviation | 0.36457 | 0.3108 | 0.22085 | 0.19657 | 0.1374 | 0.08478 | 6.04E-17 |
| Proportion of Variance | 0.00633 | 0.0046 | 0.00232 | 0.00184 | 0.0009 | 0.00034 | 0.00E+00 |
| Cumulative Proportion | 0.99 | 0.9946 | 0.99692 | 0.99876 | 0.9997 | 1 | 1.00E+00 |

Table 4.1 Importance of components for Characteristic Rating

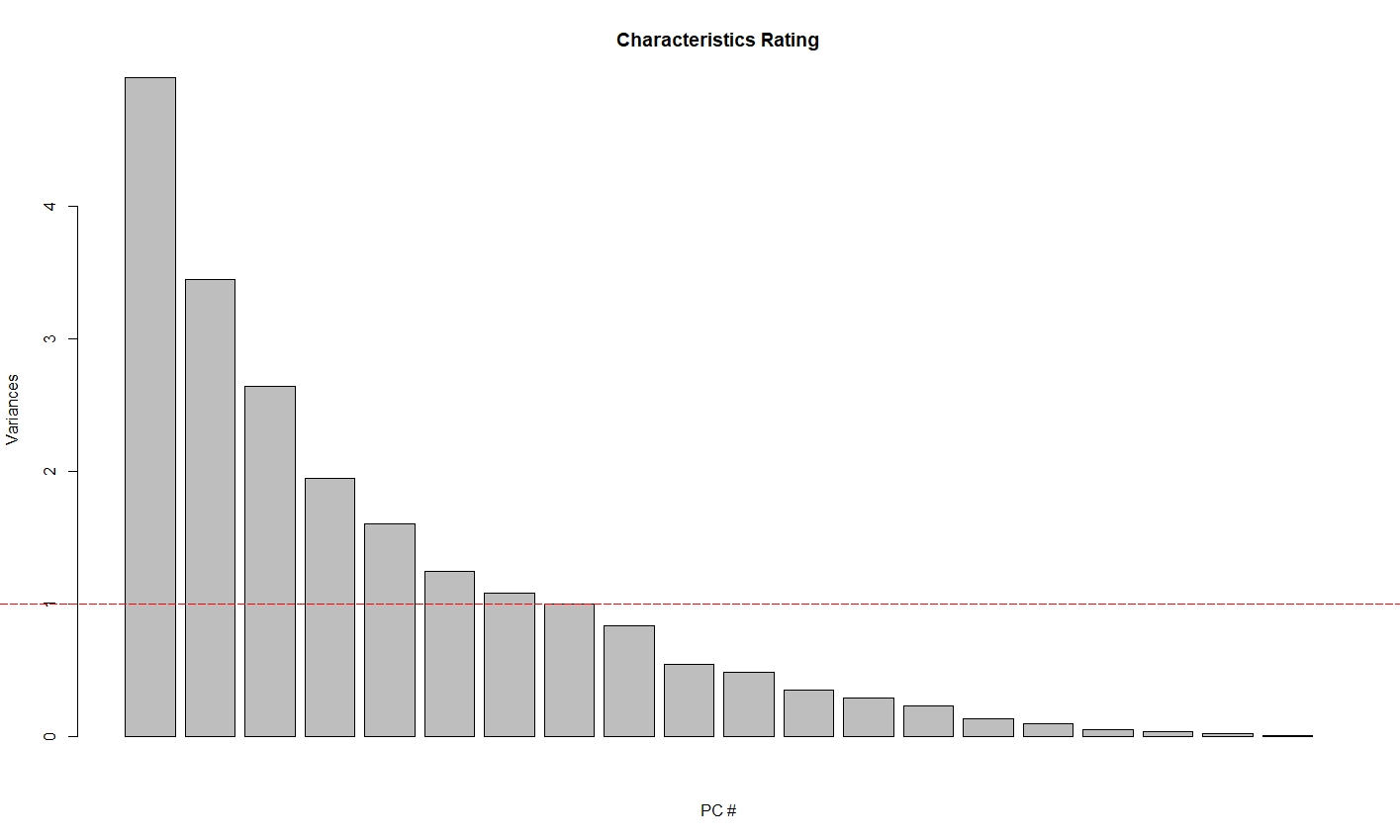
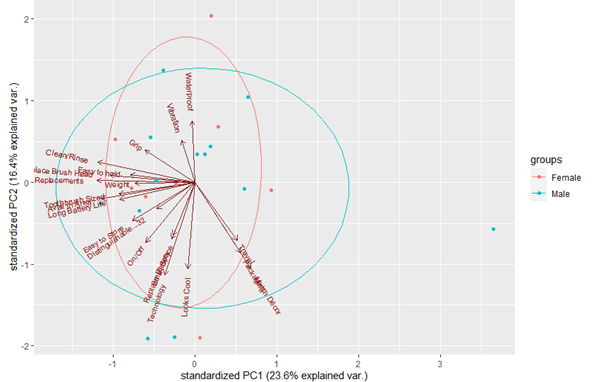


Fig.4.1 This figure is the screeplot for the PCs of the Characteristic Ratings. It shows the variances or eigenvalues of all the PCs. The red dotted line represents the eigenvalue of 1.

In Figure 4.1, a horizontal line is drawn where the value of variance or the eigenvalue is 1. This will show us which principal component we should consider up till. An eigenvalue lower than 1 indicates that the component will explain less than a single variable is able to and as such, can be neglected without losing too much representation of the original data. Based on Figure 4.1, we can narrow down our analysis to PCs 1 to 8. Based on the cumulative variance shown in Table 4.1, the first eight PCs represent up to approximately 85.3% of the variance of our original dataset which is acceptable.



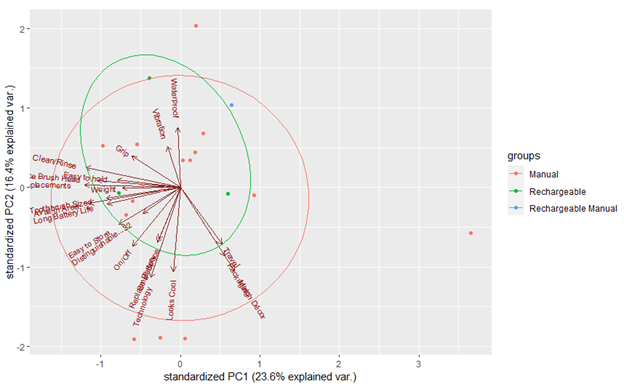


Fig. 4.2 This figure shows the Characteristic Rating PCA clustered by gender and type of toothbrush used.

From Figure 4.2, the clusters largely overlap each other. Therefore, it can be seen that no meaningful relation can be drawn from gender or type of toothbrush used. That is to say that the desirability of features are approximately identical across genders and type of toothbrush used.

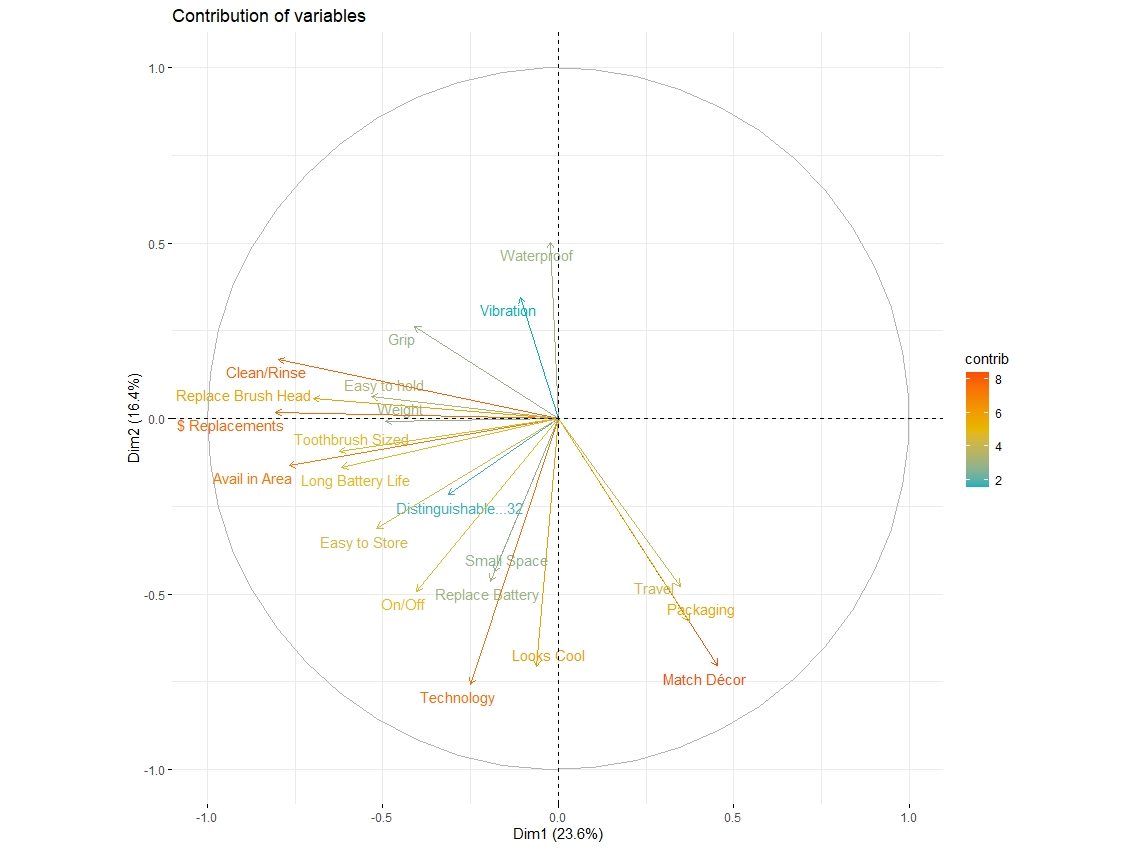


Fig. 4.3 This figure shows the contributions of all the variables towards PCs 1 (x axis) and 2 (y axis) of the Characteristic Rating.

From Figure 4.3, it can be observed that the degrees of contributions can be broken down into 4 main groups in decreasing order:

Group 1: Match Decor, Clean/Rinse, $ Replacements, Availability in Area, Technology.

Group 2: Packaging, Travel, Looks Cool, On/Off, Easy to Store, Toothbrush Size, Replaceable Brush Head.

Group 3: Small Space, Replaceable Battery, Weight, Easy to Hold, Grip, Waterproof.

Group 4: Vibration, Distinguishable.

This shows us the features that we should focus on to produce a more attractive product. The variables’ colour indicates the magnitude of their contribution according to the legend on the right. While their relative directions indicate their correlation; same direction indicating positive correlation and opposite directions indicating negative correlation.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| PC No. | PC1 | PC2 | PC3 | PC4 | PC5 |
| Standard Deviation | 1.9843 | 1.5919 | 1.3584 | 1.226 | 1.10264 |
| Proportion of Variance | 0.2625 | 0.1689 | 0.123 | 0.1002 | 0.08105 |
| Cumulative Proportion | 0.2625 | 0.4314 | 0.5545 | 0.6547 | 0.73572 |
| PC No. | PC6 | PC7 | PC8 | PC9 | PC10 |
| Standard Deviation | 1.007 | 0.9246 | 0.87136 | 0.64404 | 0.5491 |
| Proportion of Variance | 0.0676 | 0.057 | 0.05062 | 0.02765 | 0.0201 |
| Cumulative Proportion | 0.8033 | 0.8603 | 0.91093 | 0.93859 | 0.9587 |
| PC No. | PC11 | PC12 | PC13 | PC14 | PC15 |
| Standard Deviation | 0.4517 | 0.44256 | 0.3535 | 0.25803 | 0.16833 |
| Proportion of Variance | 0.0136 | 0.01306 | 0.00833 | 0.00444 | 0.00189 |
| Cumulative Proportion | 0.9723 | 0.98534 | 0.99367 | 0.99811 | 1 |

Table 4.2 Importance of components for Features Premium

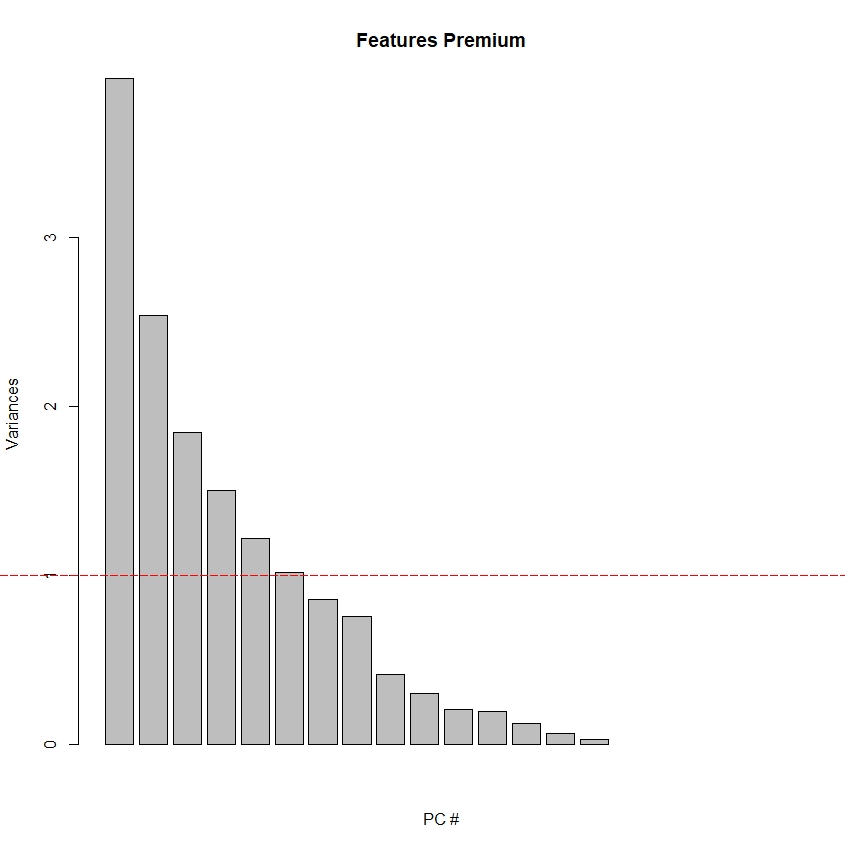


Fig. 4.4 This figure is the screeplot for the PCs of the Features Premium. It shows the variances or eigenvalues of all the PCs. The red dotted line represents the eigenvalue of 1.

Similar to the analysis done on the PCs of the Characteristic Ratings, the same screeplot is made to find the PC till which we can consider. The Features Premium can be reduced to just 6 PCs, which is still able to represent 80.33% of the original data.

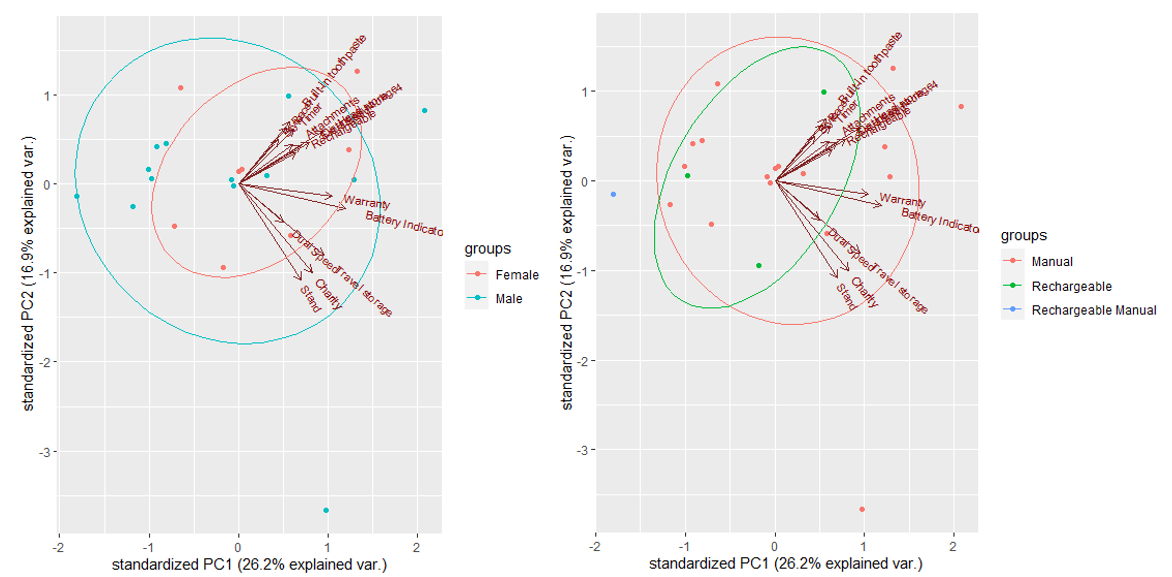


Fig. 4.5 This figure shows the Features Premium PCA clustered by gender and type of toothbrush used.

Much like the PCA for the Characteristic Ratings, the Features Premium PCA also possesses largely overlapping clusters which show little to no correlation towards gender or types of toothbrush used.

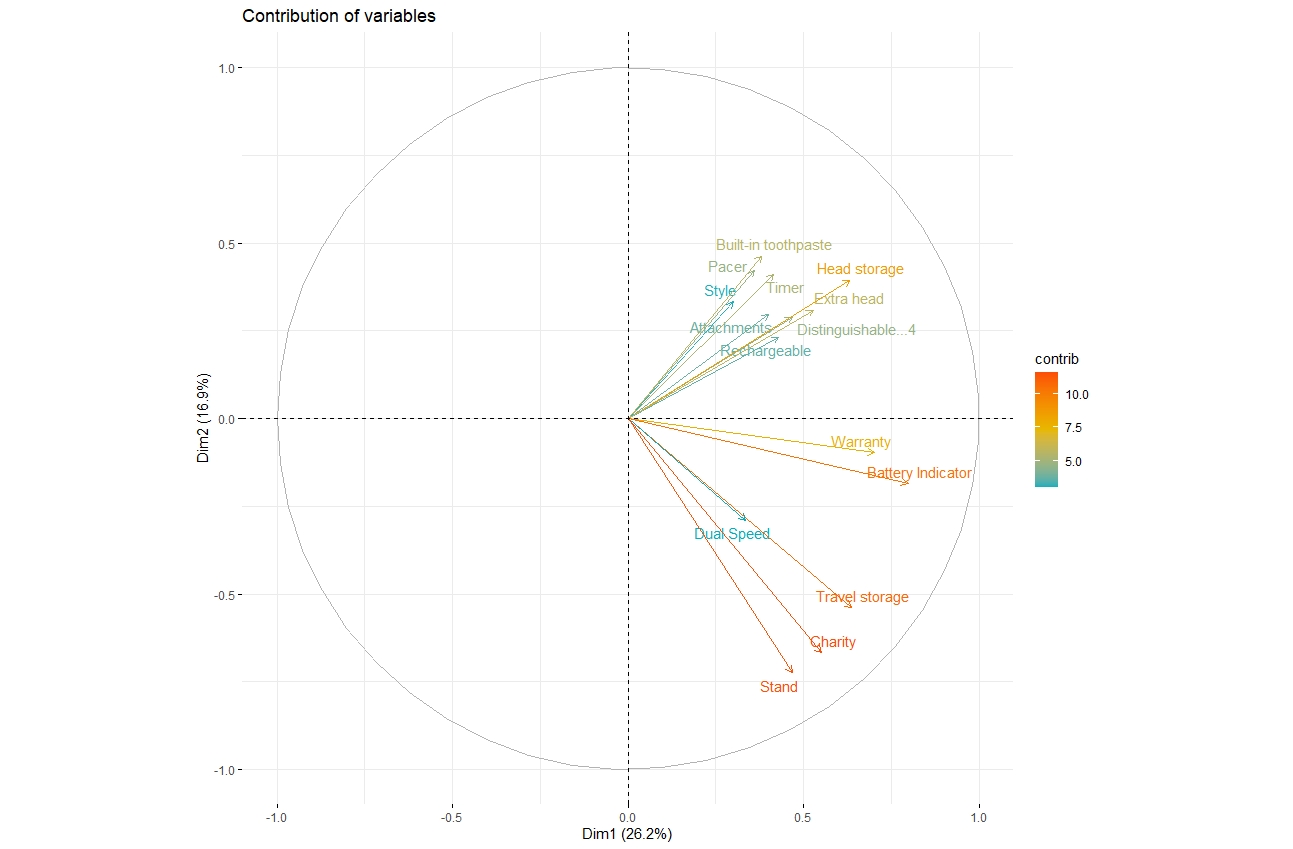


Fig. 4.6 This figure shows the contributions of all the variables towards PCs 1 (x axis) and 2 (y axis) of the Features Premium.

From Figure 4.6, it can be seen that most of the variables are positively correlated to each other as all of them point to the same side of the graph. Once again, the variables’ degrees of contribution can be broken down into 4 main groups in decreasing order:  
 Group 1: Stand, Charity, Travel Storage, Battery Indicator.

Group 2: Head Storage, Warranty.

Group 3: Built-in toothpaste, Pacer, Timer, Extra Head, Distinguishable.

Group 4: Dual Speed, Rechargeable, Attachments, Style.

Association Rule Learning

The algorithm ended up with a dataset of 15 columns with 21 rows. Following up, we chose a specific level of support to only include itemsets which were moderately popular and a high confidence level such that our results only includes higher conditional probability of occurrence of both LHS and RHS.

Performing the Apriori Algorithm at our chosen levels of support, confidence. 14 rules are generated, sorted by descending lift ratios and organised in the table below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Rules | LHS | RHS | Support | Confidence | Lift | Count |
| 1 | Warranty | Extra head | 0.09 | 1.00 | 4.2 | 2 |
| 2 | Distinguishable | Timer | 0.09 | 1.00 | 3.5 | 2 |
| 3 | Attachments | Dual Speed | 0.09 | 0.66 | 2.3 | 2 |
| 4 | Style | Stand | 0.19 | 0.80 | 1.86 | 4 |
| 5 | Rechargeable,Style | Stand | 0.09 | 0.66 | 1.55 | 2 |
| 6 | Rechargeable,Battery Indicator | Stand | 0.09 | 0.66 | 1.55 | 2 |
| 7 | Stand | Rechargeable | 0.28 | 0.66 | 1.40 | 6 |
| 8 | Rechargeable | Stand | 0.28 | 0.60 | 1.40 | 6 |
| 9 | Stand,Battery Indicator | Rechargeable | 0.09 | 0.66 | 1.40 | 2 |
| 10 | Style | Rechargeable | 0.14 | 0.60 | 1.26 | 3 |
| 11 | Travel storage | Stand | 0.09 | 0.50 | 1.16 | 2 |
| 12 | Battery Indicator | Stand | 0.14 | 0.50 | 1.16 | 3 |
| 13 | Battery Indicator | Rechargeable | 0.14 | 0.50 | 1.05 | 3 |
| 14 | Style,Stand | Rechargeable | 0.09 | 0.50 | 1.05 | 2 |

Table 4.1 : Rules generated by apriori algorithm

Explanation

Referencing Rule 1, at a Lift ratio of 4.2. Given that a person is willing to pay for a ‘Warranty’ feature, the likelihood of that person willing to pay for a ‘Extra’ feature goes up by 4.2 times. Also, a higher value for Confidence ratio shows a higher likelihood to pay for the feature ‘Extra Head’ when willing to pay for ‘Warranty’.

Apriori Algorithm Plot

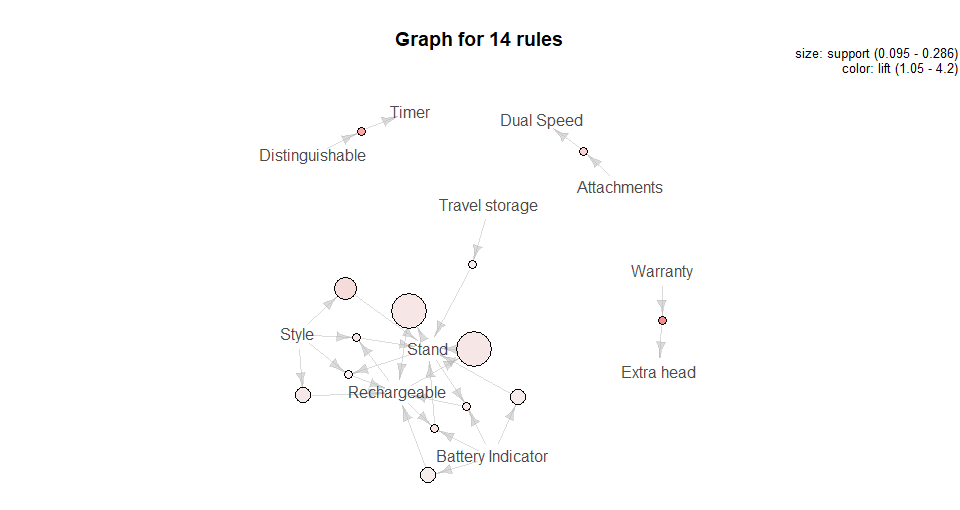
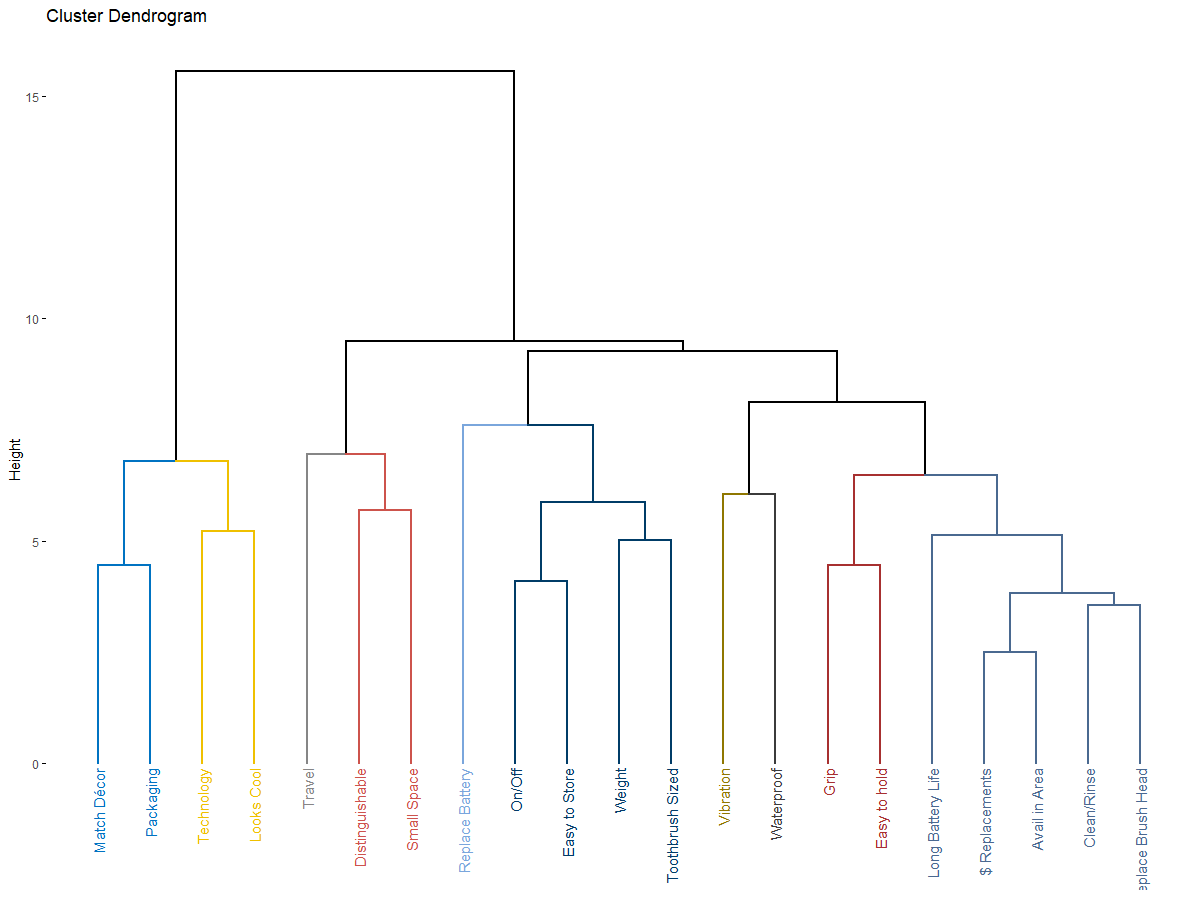


Fig 4.4: Plot of Itemsets

Agglomerative Hierarchical Clustering (AHC)

Fig 4.5: Dendrogram for Characteristic Rating

The horizontal axis of the Dendrogram represents the characteristics of the product that were rated by survey respondents. By drawing a cut-off line in Fig 4.5, the clusters for characteristic can be obtained.

Cluster 1: “Match Décor”, “Packaging”, “Technology”, “Looks Cool”

Cluster 2: “Travel”, “Distinguishable”, “Small Space”

Cluster 3: “Replace Battery”, “On/Off”, “Easy to Store”, “Weight”, “Toothbrush Sized”

Cluster 4: “Vibration”, “Waterproof”

Cluster 5: “Long Battery Life”, “$ Replacements”, “Avail in Area”, “Clean/Rinse”,

“Replace Brush Head”

1. Discussion

Principal Component Analysis (PCA)

It is observed from Table 4.1 that the proportion of data represented in PC1 and PC2 are 26.25% and 16.89% respectively for features that customers are willing to pay a premium for. While in characteristics rating PC1 and PC2 have 23.64% and 16.42% respectively. This indicates that PCA is unable to effectively represent the variance within our dataset.

Attempts were made to cluster the dimensionally reduced data in terms of the respondent’s gender and their current type of toothbrush. However, as can be seen in Figures 4.2 and 4.5, no clear relations can be drawn from either categories.

Association Rule Learning (Apriori Algorithm)

Based on this Itemset Plot, several patterns and observations can be made:

- The most popular feature was ‘Stand’ and ‘Rechargeable’, with the next being ‘Style’.   
- People willing to pay for the ‘Stand’ feature are likely to pay for the ‘Rechargeable’ feature as well.

- People willing to pay for ‘Attachments’ are likely to pay for the ‘Dual Speed’ as well.

- There are some whom are willing to pay for a ‘Battery Indicator’ and are likely to pay for the ‘Rechargeable’ feature as well as a ‘Stand’

Based on this information, we can see a huge correlation between some features. Features including ‘Style’, ‘Stand’, ‘Rechargeable’, ‘Battery Indicator’ and ‘Travel Storage’. Although there are some features which are highly correlated with each other, such as ‘Attachments’ and ‘Dual Speed’, focusing our design based on multiple features which are correlated with one another is a smarter choice.

Analysis of Table and Plot

Based on the Table 4.1, Rule 5 has a Lift value of 1.55 which suggests positive correlation between the two itemset.Also, the data shows that people that are willing to pay for a Rechargeable & Style feature are willing to pay for a Stand feature as well. Moreover, Rule 9,with a Lift value of 1.4 shows that people who are willing to pay for a Stand & Battery Indicator feature are likely to pay for a Rechargeable feature as well. This analysis shows that focusing our design based on ‘Style’,’Rechargeable’,’Stand’ and ’Battery Indicator’ will be beneficial to our product, this is in line with the analysis shown in regard to Fig 4.4.

Agglomerative Hierarchical Clustering (AHC)

By analysing the different clusters observed from AHC, the following relationship can be made. From cluster 1, it can be seen that the product's aesthetics is prioritized here. From cluster 2, it can be seen that design for ease of packing during travel is the key factor for this cluster. From cluster 3, it is certain that the design for convenience of usage is of importance here. From cluster 4, the main trend here is the toothbrush having additional functions. Lastly from cluster 5, cost and benefit is of the majority factor for customers. These data sets show us which characteristics should be implemented together to bring the most optimal results.

Proposed Design and Pricing

Now that our analysis on the given data and the mined data is done, in this section, we will propose a next generation design for the Oral-B toothbrush that can capture a larger market and potentially generate more revenue for the company.

Proposed Design

Starting from the pre-analysis point of view from the survey results. It is clear that the current context for toothbrushes used is predominantly manual toothbrushes instead of electronic or rechargeable toothbrushes. However, when asked which feature they would like the most and which feature they would pay the highest for, it was the rechargeable option. Thus, at a narrowed down level, the proposed design will be a rechargeable electric toothbrush that would fit in with the new and relevant demands of today.

For the current electric toothbrushes available in the market, there are common trends that are noteworthy. Examples inferred include being bulky, heavy and the need to replace the batteries.

The **characteristic** we propose to focus on is the ability to be waterproof. Although the results do not specifically point out waterproof as being a characteristic as the most demanded, we choose this as a high number of customers would like this characteristic and the other factors such as replaceable head and longevity of battery life could be protected with the ability for it to resist water damages. This is supported in the dendrogram for characteristics when vibration is associated with waterproofing. This shows that the likelihood of customers expecting this characteristic when vibration(electric/automatic) toothbrushes are true is highly correlated.

Now we will elaborate on the **features** focused on the new design. From the analysis above, the 3 most important features would be rechargeability, having a stand and the style of the toothbrush. It is also good to note that people who wanted features like rechargeability also wanted features like stand, and a battery indicator. Since the travel storage feature is also associated with these features, it should also be noted. Thus, these features specifically should be emphasized on the most in the new design for the toothbrush.

In summary, listed below are the key features and characteristics that the new design should focus on:

1. Waterproof
2. Rechargeability
3. Stand to be included
4. Focus on style of the toothbrush
5. Replaceable Head
6. Longer Battery Life

Pricing

Referring to table 3.1 summary Data, it can be observed that the mean cost of a toothbrush is approximately $3, and the price that customers are willing to pay has a mean value of $5.30. In addition, customers are also willing to top up $13.13 for the additional rechargeable feature. As seen in figure 3.1. Pie chart of demographics. This indicates that approximately 81% of consumers are still using manual toothbrushes, and rechargeable features are the most sought after feature as extracted from the survey.

Referencing association rule learning from the Apriori Algorithm, the rechargeable feature is associated with “stand” and “style”. Therefore, in proposing the cost of the new design, additional cost of $1.32 for “stand” and “$1.57” for style must be accounted for. The breakdown of summary data can be found in appendix part (d)

Therefore, the final price to be proposed is calculated as such:

Estimated Selling Price = Current Mean Price + Cost of “style” feature + Cost of “stand” feature + Rechargeable Feature Premium

Estimated Selling Price = $3 + $1.32 + $1.57 + $13.13 = **$19.02**

1. Conclusion

Processing large volumes of data as demonstrated in this assignment, has exposed the intricacy of data processing. The original raw data had ambiguity, and data gaps that were manually filtered and given an assumed value. Preprocessing of the raw data set also allows for software to process more efficiently. Through the means of various supervised and unsupervised learning techniques, key clusters of key features and characteristics within the large dataset were able to be extracted and then utilised for designing the next generation product. Data mining allows design goals to be clearly defined and identified. As such cost and design proposals were made to focus on key characteristics that consumers have highlighted. This sharp focus allows resources to be directed and assigned efficiently.

References

[1]L. Hayden, "Principal Component Analysis in R", *https://www.datacamp.com/*, 2018. [Online]. Available: https://www.datacamp.com/community/tutorials/pca-analysis-r. [Accessed: 27- Feb- 2021].

[2]H. Jabeen, "Market Basket Analysis using R", *https://www.datacamp.com/*, 2018. [Online]. Available: https://www.datacamp.com/community/tutorials/market-basket-analysis-r. [Accessed: 27- Feb- 2021].

[3]"Agglomerative Hierarchical Clustering - Datanovia", Datanovia. [Online]. Available: https://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering/. [Accessed: 27- Feb- 2021].

Appendix

* 1. RStudio Code for PCA Analysis

#import excel before this

FeaturesPremiumCRating

#categorising

fp<-(FeaturesPremiumCRating[, 1:15])

gender <- (FeaturesPremiumCRating[, 16])

type <- (FeaturesPremiumCRating[, 17])

gendertype <- (FeaturesPremiumCRating[, 18])

crating<-(FeaturesPremiumCRating[, 19:29])

#pca of features premium

fp.pca <- prcomp(FeaturesPremiumCRating[, 1:15], center=TRUE, scale = TRUE)

print(fp.pca)

plot(fp.pca, type ="l")

summary(fp.pca)

#pca of characteristics rating

crating.pca <- prcomp(FeaturesPremiumCRating[, 19:29], center=TRUE, scale = TRUE)

print(crating.pca)

plot(crating.pca, type ="l")

summary(crating.pca)

library(devtools)

#I manually install ggbiplot here

#go here: https://github.com/vqv/ggbiplot

#$download the zip file ("clone or download"). In the zip file you'll

#find a folder called R. In this folder are two R scripts.

#Open these two and just run them.

#This is the manual way of creating the function ggbiplot.

#clustering process

ggbiplot(fp.pca)

ggbiplot(crating.pca)

genderfp <- c(gender)

#FP Clustered Gender

Fp.gender <- c(rep("Female",2),"Male","Female",rep("Male",6),"Female","Male","Female",rep("Male",2),"Female","Male",rep("Female",2),rep("Male",2))

ggbiplot(fp.pca, ellipse = TRUE, groups = fp.gender)

ggbiplot(crating.pca, ellipse = TRUE, groups = fp.gender)

#FP Cluster Type

fp.type <- c("Rechargeable",rep("Manual",6),"Rechargeable Manual",rep("Manual",5),"Rechargeable",rep("Manual",5),"Rechargeable","Manual")

ggbiplot(fp.pca,ellipse=TRUE, groups=fp.type)

ggbiplot(crating.pca, ellipse = TRUE, groups = fp.type)

#FP Cluster Gender+Type

fp.gendertype <- c("Female Recharge", "Female Manual", "Male Manual", "Female Manual", "Male Manual", "Male Manual", "Male Manual", "Male All", "Male Manual", "Male Manual", "Female Manual", "Male Manual", "Female Manual", "Male Recharge", "Male Manual", "Female Manual", "Male Manual", "Female Manual", "Female Manual", "Male Recharge", "Male Manual")

ggbiplot(fp.pca ,groups=fp.gendertype)

ggbiplot(crating.pca ,groups=fp.gendertype)

#Plotting Screeplot for Features Premium and Characteristic Rating

screeplot(fp.pca,npcs=20,xlab="PC #",main="Features Premium")

abline(h = 1, col="red", lty=5)

screeplot(crating.pca,npcs=20,xlab="PC #",main="Characteristic Rating")

abline(h = 1, col="red", lty=5)

#Plotting the contribution of variables of Features Premium

fviz\_pca\_var(fp.pca, col.var = "contrib", # Color by contributions to the PC

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, # Avoid text overlapping

title="Contribution of variables")

#Plotting the contribution of variables of Characteristic Rating

fviz\_pca\_var(crating.pca, col.var = "contrib", # Color by contributions to the PC

gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),

repel = TRUE, # Avoid text overlapping

title="Contribution of variables")

* 1. RStudio Code for Apriori Algorithm

install.packages("arules")

install.packages("arulesViz") cf

library(arules)

library(arulesViz)

Import excel (file name as "ASR")

View(ASR)

ASR[is.na(ASR)] = 0

ASR=as.matrix(ASR)

ASR=as(ASR,"transactions")

rules=apriori(ASR,parameter=list(supp=0.05,conf=0.5,minlen=2))

rules=sort(rules, by="lift") inspect(rules)

plot(rules,method="graph",control=list(type="itemsets"))

* 1. RStudio Code for Agglomerative Hierarchical Clustering

#Check dataset

head(CratingCSV)

tail(CratingCSV)

str(CratingCSV)

head(CratingCSV)

#define rowname as character type if not yet defined

CratingCSV$rowname <- as.character(x = CratingCSV$rowname)

str(CratingCSV)

head(CratingCSV)

#generate rownames

rownames(CratingCSV) <- c(CratingCSV$rowname)

head(CratingCSV)

#rename data with edited rowname

CHdendro <- CratingCSV

#generating the matrix for dendro plot

require(stats)

res.dist <- dist( x = CHdendro, method = "euclidean")

x <- as.matrix(res.dist)[1:21, 1:21]

round(x, digits = 3)

require(stats)

res.hc <- hclust(d = res.dist,

method = "complete")

plot(x = res.hc)

#tidy the dendro

require(factoextra)

fviz\_dend(x = res.hc, ex = 0.7, lwd = 0.7)

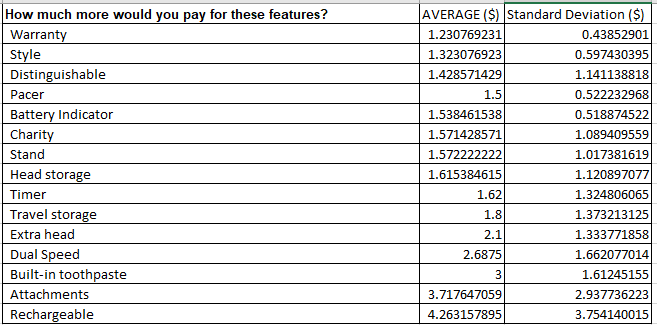
#apply colour and change color settings, thickness etc

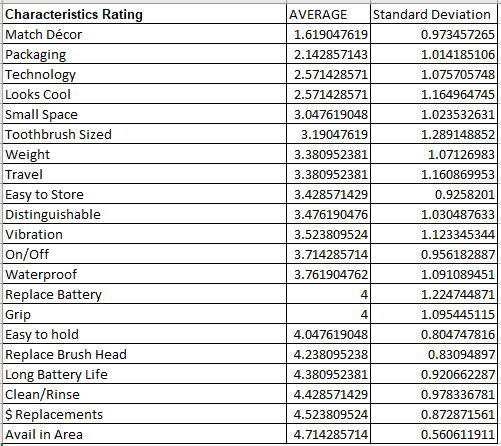
require(grDevices)

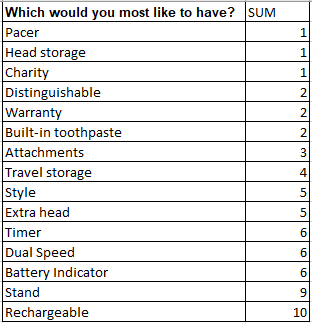
colors()

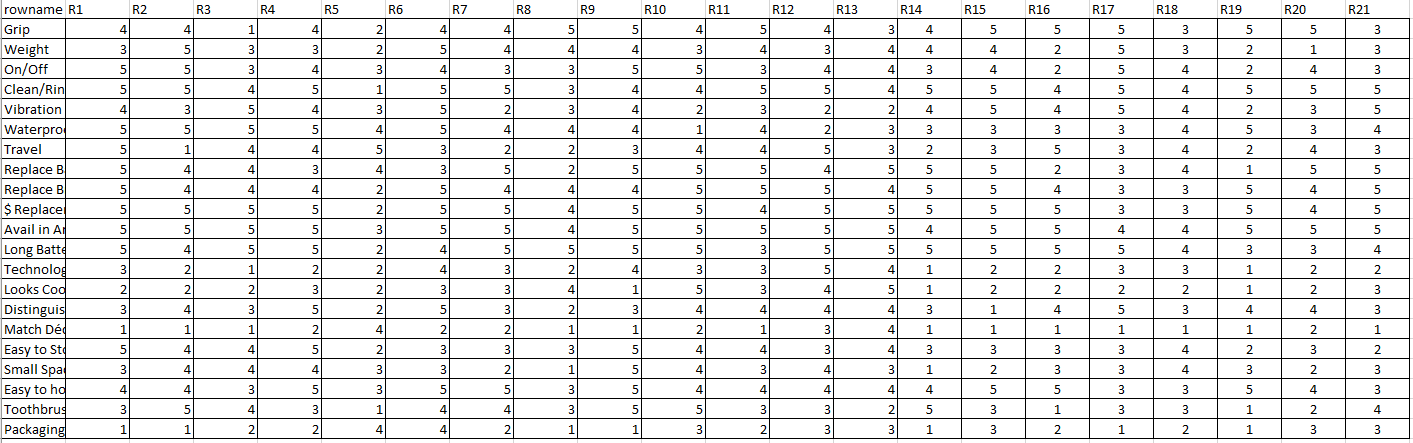
fviz\_dend(x = res.hc, ces = 0.8, lwd= 0.8, k=10, k\_colors = "jco")

* 1. Summary Data







* 1. CratingC
  2. Distance matrices generated from Rstudio

Grip Weight On/Off Clean/Rinse Vibration Waterproof

Grip 0.000 7.165 6.473 5.512 7.928 7.862

Weight 7.165 0.000 5.318 7.521 6.226 6.789

On/Off 6.473 5.318 0.000 6.554 6.473 7.165

Clean/Rinse 5.512 7.521 6.554 0.000 7.591 7.237

Vibration 7.928 6.226 6.473 7.591 0.000 6.055

Waterproof 7.862 6.789 7.165 7.237 6.055 0.000

Travel 8.624 9.040 7.017 9.268 7.165 8.315

Replace Battery 8.315 7.591 5.606 7.017 7.928 8.624

Replace Brush Head 5.318 7.091 6.392 3.546 7.309 6.942

$ Replacements 6.392 7.928 7.017 3.830 8.124 7.237

Avail in Area 6.055 8.315 6.712 3.546 8.124 7.091

Long Battery Life 6.473 6.866 5.968 5.118 7.237 7.862

Technology 9.381 6.866 7.091 10.784 9.040 9.764

Looks Cool 10.235 7.728 8.315 11.352 9.924 10.388

Distinguishable 6.712 6.473 5.880 6.633 7.309 6.942

Match Décor 13.656 10.976 11.580 15.216 12.197 12.577

Easy to Store 7.091 5.699 4.094 7.309 6.473 6.226

Small Space 8.923 7.017 5.790 9.212 7.237 7.309

Easy to hold 4.461 6.789 6.055 4.801 7.165 6.141

Toothbrush Sized 8.864 5.014 5.880 8.440 7.165 8.315

Packaging 11.804 9.492 9.871 12.947 9.547 10.832

Travel Replace Battery Replace Brush Head $ Replacements

Grip 8.624 8.315 5.318 6.392

Weight 9.040 7.591 7.091 7.928

On/Off 7.017 5.606 6.392 7.017

Clean/Rinse 9.268 7.017 3.546 3.830

Vibration 7.165 7.928 7.309 8.124

Waterproof 8.315 8.624 6.942 7.237

Travel 0.000 7.862 8.059 9.155

Replace Battery 7.862 0.000 6.392 7.017

Replace Brush Head 8.059 6.392 0.000 2.895

$ Replacements 9.155 7.017 2.895 0.000

Avail in Area 8.685 6.712 3.546 2.507

Long Battery Life 8.502 6.942 4.909 3.964

Technology 7.309 9.268 9.871 11.071

Looks Cool 7.994 9.817 9.871 11.071

Distinguishable 6.942 8.624 6.789 7.521

Match Décor 10.081 12.947 13.997 15.147

Easy to Store 6.226 6.473 6.712 7.165

Small Space 6.055 8.440 8.502 9.097

Easy to hold 7.521 7.017 4.094 4.094

Toothbrush Sized 9.492 7.017 7.659 8.440

Packaging 7.928 10.784 11.580 12.947

Avail in Area Long Battery Life Technology Looks Cool

Grip 6.055 6.473 9.381 10.235

Weight 8.315 6.866 6.866 7.728

On/Off 6.712 5.968 7.091 8.315

Clean/Rinse 3.546 5.118 10.784 11.352

Vibration 8.124 7.237 9.040 9.924

Waterproof 7.091 7.862 9.764 10.388

Travel 8.685 8.502 7.309 7.994

Replace Battery 6.712 6.942 9.268 9.817

Replace Brush Head 3.546 4.909 9.871 9.871

$ Replacements 2.507 3.964 11.071 11.071

Avail in Area 0.000 4.461 11.259 11.443

Long Battery Life 4.461 0.000 10.235 10.337

Technology 11.259 10.235 0.000 5.219

Looks Cool 11.443 10.337 5.219 0.000

Distinguishable 7.381 7.591 7.165 7.728

Match Décor 15.556 14.619 6.789 6.633

Easy to Store 7.309 6.309 6.789 7.659

Small Space 9.097 8.805 5.790 7.381

Easy to hold 4.342 5.318 8.981 9.764

Toothbrush Sized 9.268 7.862 7.728 8.252

Packaging 13.187 12.743 6.055 5.699

Distinguishable Match Décor Easy to Store Small Space

Grip 6.712 13.656 7.091 8.923

Weight 6.473 10.976 5.699 7.017

On/Off 5.880 11.580 4.094 5.790

Clean/Rinse 6.633 15.216 7.309 9.212

Vibration 7.309 12.197 6.473 7.237

Waterproof 6.942 12.577 6.226 7.309

Travel 6.942 10.081 6.226 6.055

Replace Battery 8.624 12.947 6.473 8.440

Replace Brush Head 6.789 13.997 6.712 8.502

$ Replacements 7.521 15.147 7.165 9.097

Avail in Area 7.381 15.556 7.309 9.097

Long Battery Life 7.591 14.619 6.309 8.805

Technology 7.165 6.789 6.789 5.790

Looks Cool 7.728 6.633 7.659 7.381

Distinguishable 0.000 10.686 5.699 5.699

Match Décor 10.686 0.000 10.735 9.155

Easy to Store 5.699 10.735 0.000 4.801

Small Space 5.699 9.155 4.801 0.000

Easy to hold 5.968 12.825 5.880 7.451

Toothbrush Sized 7.659 11.259 5.880 6.866

Packaging 9.155 4.461 9.437 7.728

Easy to hold Toothbrush Sized Packaging

Grip 4.461 8.864 11.804

Weight 6.789 5.014 9.492

On/Off 6.055 5.880 9.871

Clean/Rinse 4.801 8.440 12.947

Vibration 7.165 7.165 9.547

Waterproof 6.141 8.315 10.832

Travel 7.521 9.492 7.928

Replace Battery 7.017 7.017 10.784

Replace Brush Head 4.094 7.659 11.580

$ Replacements 4.094 8.440 12.947

Avail in Area 4.342 9.268 13.187

Long Battery Life 5.318 7.862 12.743

Technology 8.981 7.728 6.055

Looks Cool 9.764 8.252 5.699

Distinguishable 5.968 7.659 9.155

Match Décor 12.825 11.259 4.461

Easy to Store 5.880 5.880 9.437

Small Space 7.451 6.866 7.728

Easy to hold 0.000 8.059 10.735

Toothbrush Sized 8.059 0.000 9.602

Packaging 10.735 9.602 0.000