QUEEN'S UNIVERSITY BELFAST

Detailed Report on Finding Key Factors affecting Customer Decisions using Conjoint Analysis

40386497
Marketing Analytics (MGT7215)
MSc. In Business Analytics
Word Count: 1848
2022

Introduction

Market segmentation is an important aspect in marketing strategy because it enables companies to determine and target client groups with customised marketing messages and product advertisements. Conjoint analysis has been a well-liked method for market segmentation in recent years, enabling researchers to pinpoint the most important qualities of a good or service and gauge how much customers value them. However, designing the survey and deciding which attributes and levels to test can be difficult and time-consuming when performing a conjoint analysis. Using fractional factorial design, a statistical method that enables researchers to significantly reduce the number of experimental runs necessary to analyse the various combinations of qualities, is one way to deal with this problem.

Researchers in a product development study found that a fractional factorial design gave them the ability to evaluate numerous product features while still achieving statistically significant findings (O. Toubia, 2005). Additionally, they were able to cut the number of experiments runs needed by up to 75% by using fractional factorial design, which decreased the time and expense associated with data collection.

Another study found that by employing fractional factorial design, they were able to pinpoint the crucial characteristics and levels that varied market segments valued most, allowing them to adjust their marketing strategies and product offerings accordingly (Kim, 2011). Additionally, they were able to estimate consumer demand for various product features and assess the potential effects of various pricing strategies.

A study in the effect of conjoint analysis in customer choice modelling discovered that by controlling for interactions between distinct qualities and levels, they could evaluate the significance of each while employing a fractional factorial design (Green, 1978). Additionally, they discovered that by combining conjoint analysis with fractional factorial design, they could determine the trade-offs that customers were willing to make between various product attributes, allowing companies to create more successful marketing campaigns.

In this report, we implement conjoint analysis using Fractional Factorial design to find the key factors that affect customer decisions. This study also aims to understand how market simulations are predicted using Principal Component Analysis. The visualisations are mapped in Tableau.

Methodology

First, the factors and their respective levels are loaded. As we can see in fig. 1 that there are five factors that are considered to be analysed among all laptop profiles. Factor ‘Brand’ has four levels, ‘Hard Drive’ has three levels, ‘RAM’ has three levels, ‘Screen Size’ has three levels and ‘Price’ has four levels.

1. Brand (Apple, Lenovo, Dell, Acer)
2. Hard Drive (128, 256, 512)
3. RAM (2, 4, 8, 16)
4. Screen Size (12.1, 15.4, 17.3)
5. Price (900, 1200, 1500, 2000)

Fig. 1

Now, we calculate the number of possible product profiles by multiplying the degree of levels from each attribute which gives an output of 576 profiles. Since, it is time consuming and not feasible to analyse all the profiles, we use fractional factorial design to come up with a smaller subset of optimal profiles that will yield similar results without performing as many experiment runs. We choose the subset size to be 20 so we can rank each profile from 1-20 based on customer preferences. The subset obtained has a similar distribution across all brands and includes all individual level values thus making it ideal for our analysis.

Now, we calculate the correlation between each variable to make sure the profiles we have chosen are optimal. Ideally, correlation values between variables should be closer to zero (indicating that the variables are orthogonal to each other) in order to have an optimal subset of profiles.

```
> print(cor(caEncodedDesign(design)))
                Profile       Hard         RAM Screen.Size      Price
Profile      1.00000000 -0.4382478 -0.04646616  -0.2806271 -0.1131734
Hard        -0.43824776  1.0000000  0.10317569   0.2858176 -0.1905429
RAM         -0.04646616  0.1031757  1.00000000  -0.4894689 -0.1249576
Screen.Size -0.28062709  0.2858176 -0.48946886   1.0000000 -0.3354075
Price       -0.11317343 -0.1905429 -0.12495756  -0.3354075  1.0000000
```

Now, we read in the data for customer preferences. By conducting a survey amongst 132 customers who were going to purchase a laptop in the next 3 months, we have recorded their preferences by asking them to rank profiles on a scale of 1-20 (where, 1 is the lowest ranking profile and 20 being the highest ranked profile.

Conjoint Analysis

After reading the customer preferences, we calculate the part worths for conjoint analysis. Part worths are simply numerical values that measure how much each attribute level affects the customer decisions. The baseline case chosen:

Brand	Hard Drive	RAM	Screen Size	Price
Acer	128 GB	2 GB	12.1 in	$900

The analysis model excludes or simply sets the baseline attribute values to zero and compares all other levels in that attribute with the respective baseline level.

```
> part.worths
    intercept  Apple Lenovo   Dell Acer 128 GB 256 GB 512 GB 2 GB   4 GB   8 GB  16 GB 12.1 in 15.4 in 17.3 in $900 $1,200 $1,500 $2,000
V1     14.044 -6.909 -6.639 -0.584    0      0 -0.066 -5.202    0  0.589 12.470 10.977       0  -0.220  16.175    0 -7.856  0.370 -3.936
V2     14.138  3.517  0.052  0.064    0      0  1.103 -5.219    0 -1.030  3.047  8.750       0   2.120   7.258    0 -2.651 -3.248 -0.464
V3     11.318  6.124  3.061 -0.537    0      0  2.338 -1.582    0 -1.108 -3.490  2.678       0   1.523  -2.507    0 -0.114 -4.080  0.850
V4     11.542  1.584 -1.567  1.215    0      0 -1.598 -4.712    0  0.239  3.261  4.592       0   0.363   6.469    0 -8.070  4.276 -1.756
V5     11.318  6.124  3.061 -0.537    0      0  2.338 -1.582    0 -1.108 -3.490  2.678       0   1.523  -2.507    0 -0.114 -4.080  0.850
```

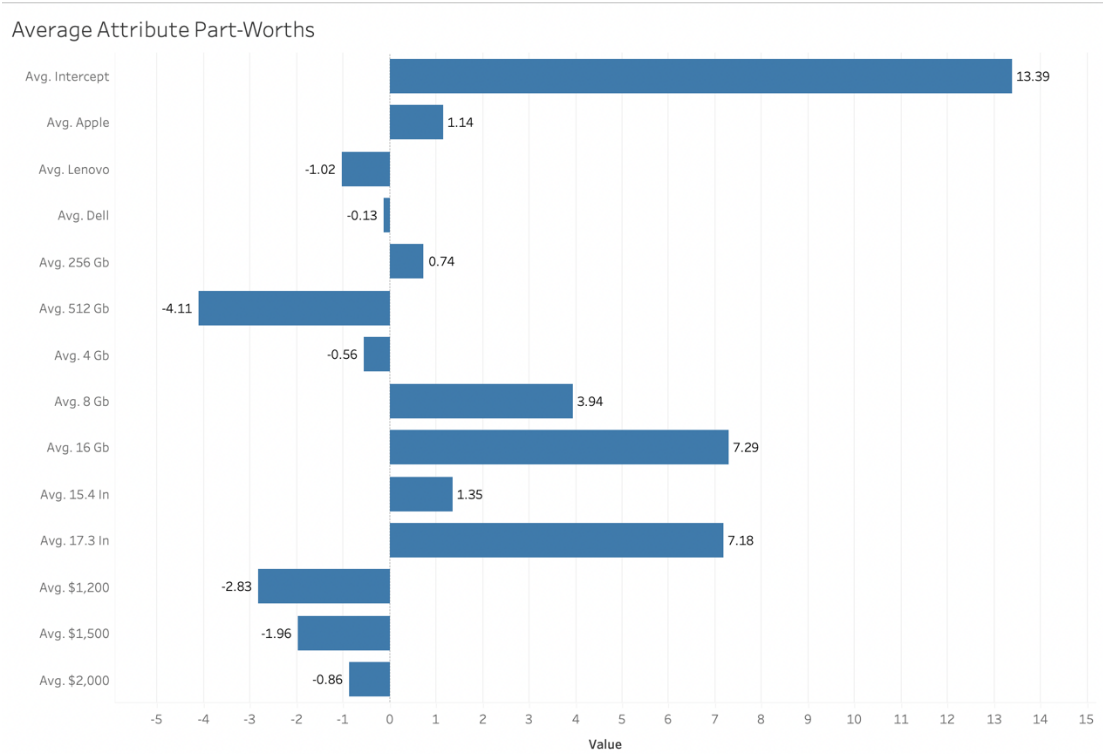
As observed in the above figure, in profile 1, the brand value Acer is set to zero while Apple has a value of -6.909, which means that customer 1 is 6.9 times more likely to choose an Acer laptop over Apple. Similarly, we can analyse other attribute values and determine the most important features for every respective customer. We visualise these values in Tableau to understand the important predictors in the dataset.

Perceptual and Preference Mapping

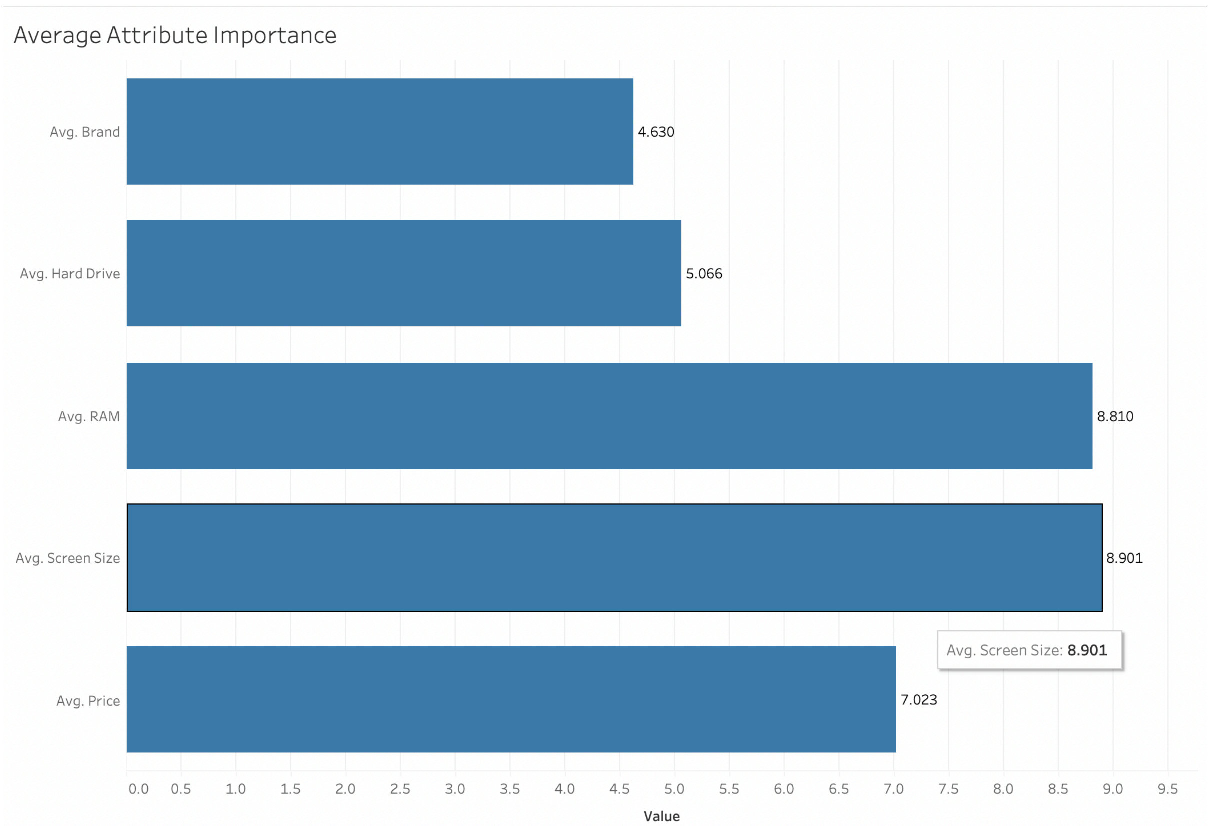
There are 8 product profile currently present in the market from which we need to determine the important factors the customers might find and subsequently the choice they would make by using Principal Component Analysis. Principle Component Analysis (PCA) is a multivariate technique used to analyse large datasets without losing interpretability while preserving the maximum amount of information (Jolliffe, 2002). It creates Principle Components generally of the same degree as attributes chosen. These Principle Components are orthogonal with no overlap on each other. We run PCA analysis in R on the perception profiles provided by referencing them to the attribute and levels vector. We plot the PVE graph to better understand the variance in the dataset. The Perceptual and Preference Mapping is visualised in Tableau using the perception scores obtained from PCA analysis.

Results and Discussion

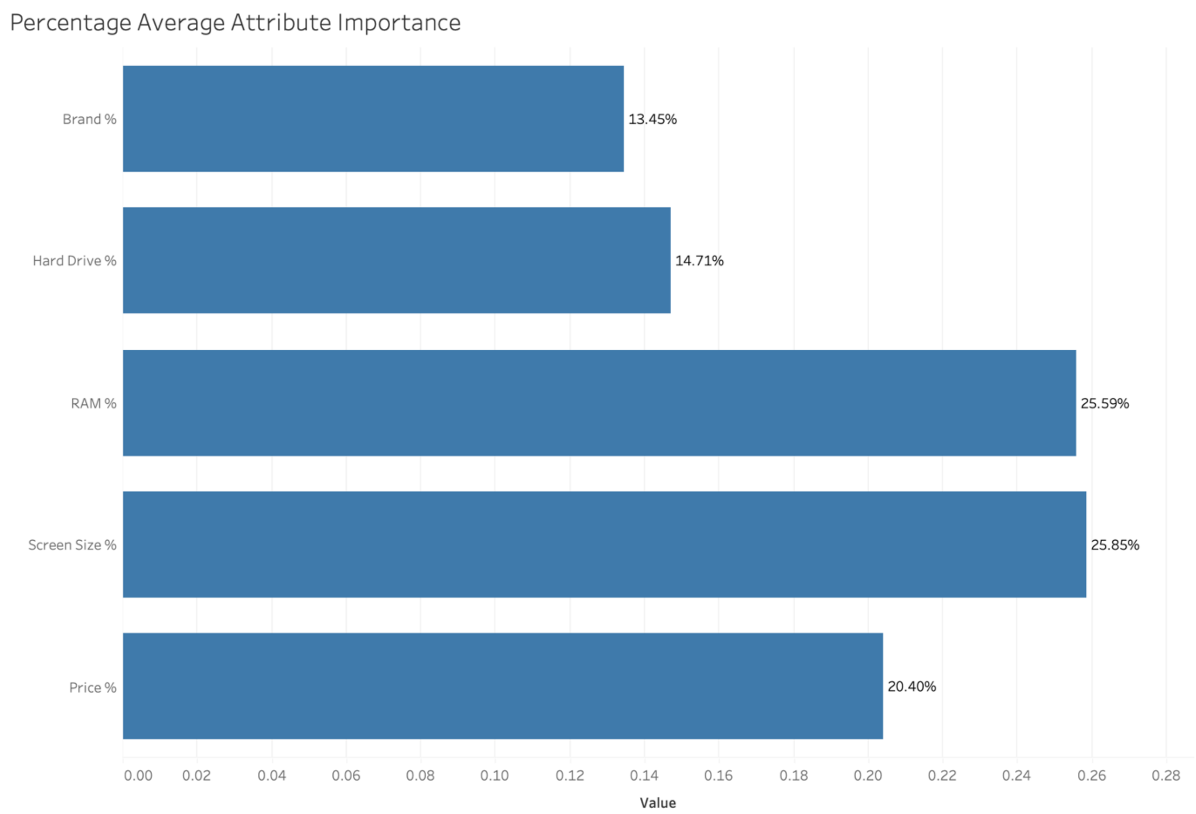
To visualise our results from Conjoint Analysis, we load the partworths scores into tableu. First, we visualise the average scores of every attribute. Level partworths enable us to dive deeper and understand what specific attribute levels affect customer decisions. Level partworths are created using average preference scores. Levels that are strongly preferred by the customer get higher scores (www.conjointly.com, 2023).



As seen above, the average score of brand 'Apple' is 1.14 times more likely to be chosen than our baseline brand, 'Acer'. Attribute scores of 'RAM' and 'Screen Size' have the highest average scores when compared with other factor levels, indicating that customers gave the highest scores for these factors. In general, customers are strongly inclined towards purchasing laptops with a RAM of 16Gb and a bigger screen size of 17.3 inches.

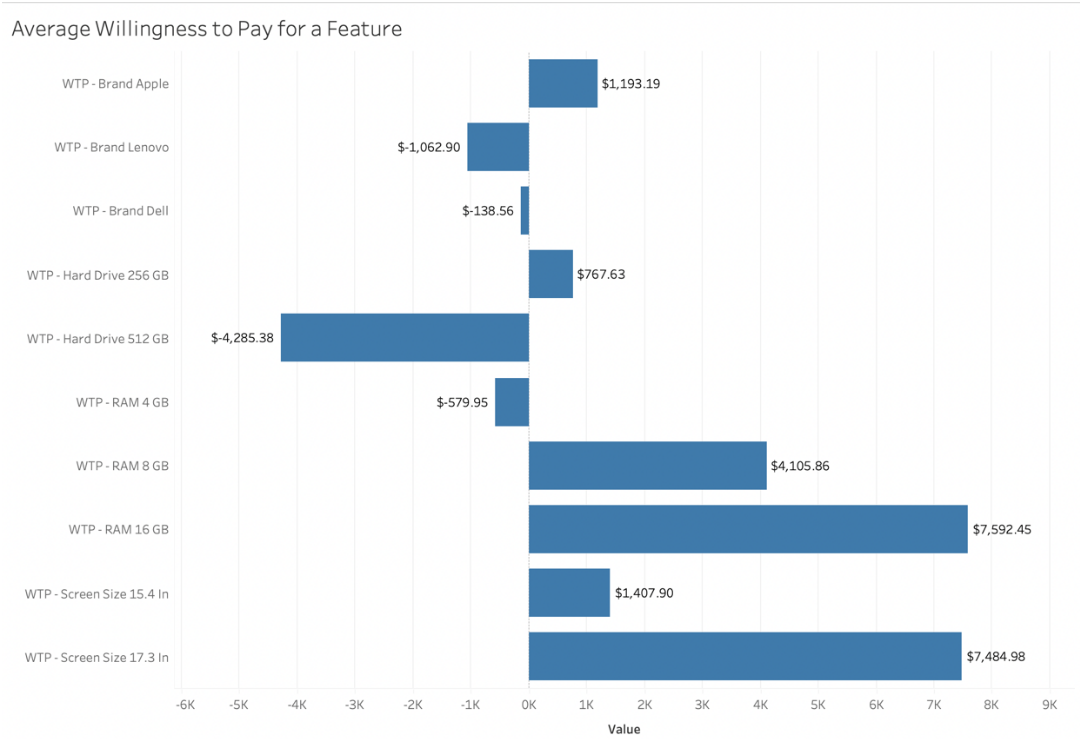


The above graph displays the relative importance of each attribute group with each other. We create calculated fields to find the maximum value of every attribute level. Attributes 'RAM' are the most preferred fields while hard drive storage and brand name has the least effect on customer decisions.

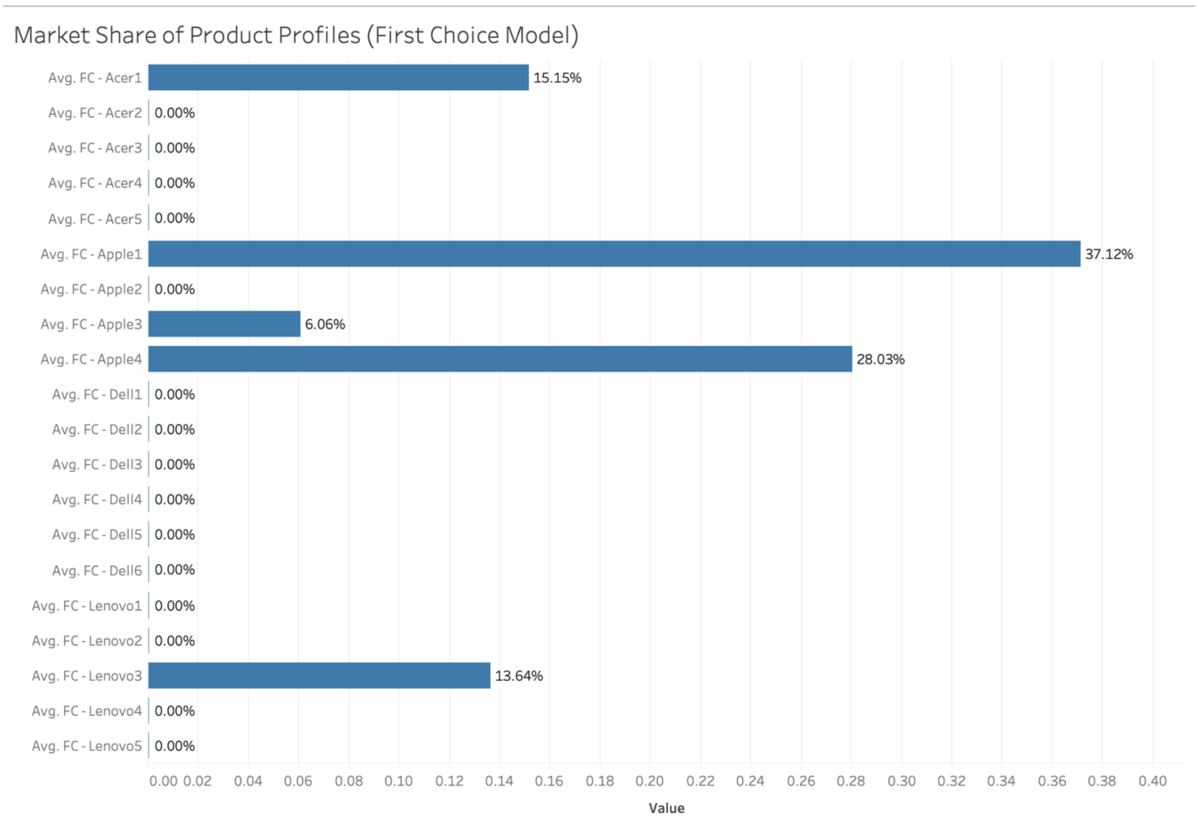


The above graph shows percentage comparison of the important features for customers. This is done by adding every attribute level by the sum of all attribute levels. 13.5% of the customers consider the brand while purchasing a laptop. Similarly, 14.71% consider hard drive, 25% consider RAM, 25.85% consider Screen Size and 20.40% consider Price.

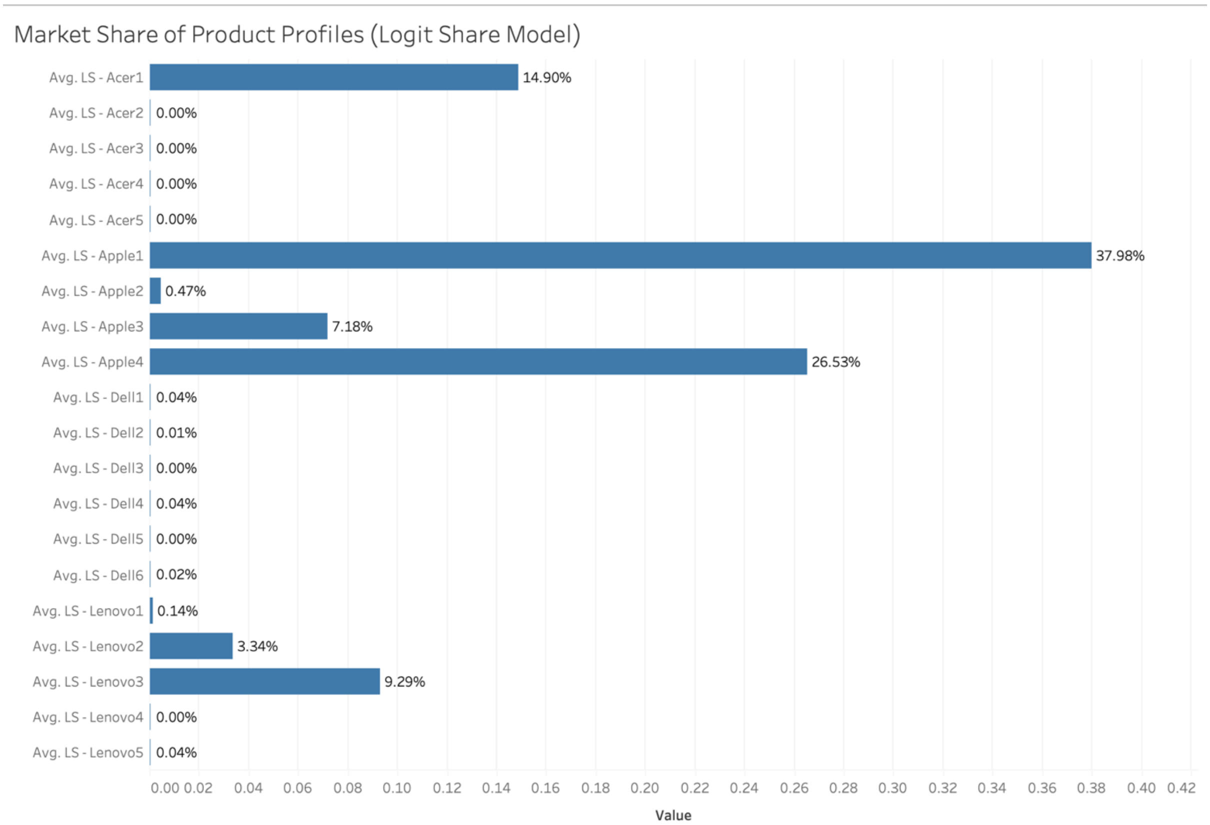
Now we calculate how many customers are likely to pay more or less for a specific feature. As seen below, customers are willing to pay $1193.19 more for a laptop from Apple rather than Acer. They are more likely to pay more for a 256 GB hard drive but not for a higher feature model. Interestingly, customers are likely to pay the much more for a laptop with the highest RAM or screen size.



Now we create a market simulation in order to calculate the market share of each product profile if they were to be sold by our company. Rather than We implement two methods, namely, First Choice and Logit share, to determine the market share.



First Choice model is based on the rule that the highest scored attribute will be chosen. The model assigns the entire share to the attribute level with the highest utility score and assigns a zero share to the remaining levels. From the above graph we can infer, that 37% of the customers are more likely to choose the first apple product profile (256 Gb, 15 Gb, 17.3in, $1500) and 28.03% are more likely to choose the fourth apple product file (128 Gb, 16 Gb, 17.3 in, $1200)



Unlike First Choice, Logit Share uses weighted averages to assign higher values to important features and lower values to less important features. This model gives a slightly more accurate and descriptive result than its counterpart. As seen in the Logit Share model graph for market share, we observe that the same products profiles have the highest utility values but the

Principal Component Analysis

```
> pca
Standard deviations (1, .., p=4):
[1] 1.3449609 1.1062161 0.7201919 0.6698431

Rotation (n x k) = (4 x 4):
                   PC1        PC2         PC3         PC4
Hard        -0.4016417  0.6537233 -0.09809668 -0.63380349
RAM          0.5253598  0.4214415  0.73907022 -0.01262323
Screen.Size  0.4842366  0.5135965 -0.63160698  0.32063396
Price        0.5728838 -0.3622866 -0.21266142 -0.70379518
```

We calculate the values for every loading factor across the four principal components that are formed. The principal components are orthogonal to each other where the maximum variance of the dataset occurs in the first principal component. Each principal component defines a different direction in spacePC1 has a standard deviation (SD) of 1.344, PC2 has a SD of 1.106, PC3 has a SD of 0.72 and PC4 has a SD of 0.67.

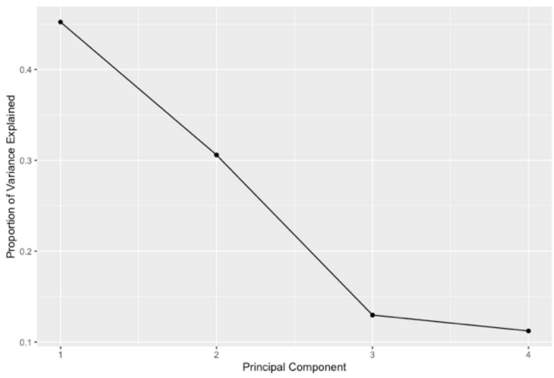
```
> pca$x
             PC1        PC2        PC3         PC4
[1,] -0.55046864  1.7255816 -0.5959191 -0.18921244
[2,]  2.11936283  0.5120236  0.1358411 -0.83595043
[3,] -0.01623207 -1.0619074  0.6174900  0.02447888
[4,] -0.84362735 -1.3527313 -0.0843035  0.35745539
[5,] -2.35457834  0.7900277  0.2465051 -0.14129017
[6,] -0.06304918 -0.7270029  0.4460739 -0.77893839
[7,]  1.12641272  0.8346585  0.6932314  1.24483427
[8,]  0.58218004 -0.7206500 -1.4589190  0.31862288
```

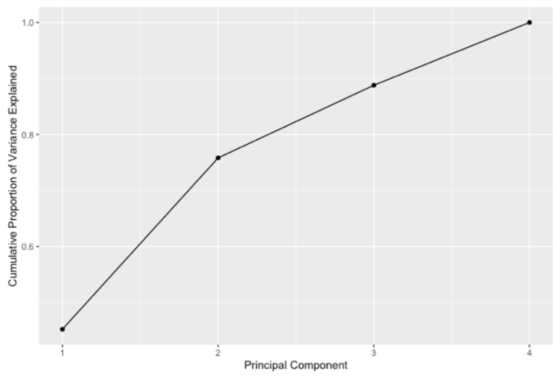
The above figure shows the singular values of every perception in every principal component.

Although PCA reduces the complexity and dimensionality, there is an even better method to determine the level of variance captured in these principal components. We calculate the Proportion of Variance explained (PVE) by each principal component.

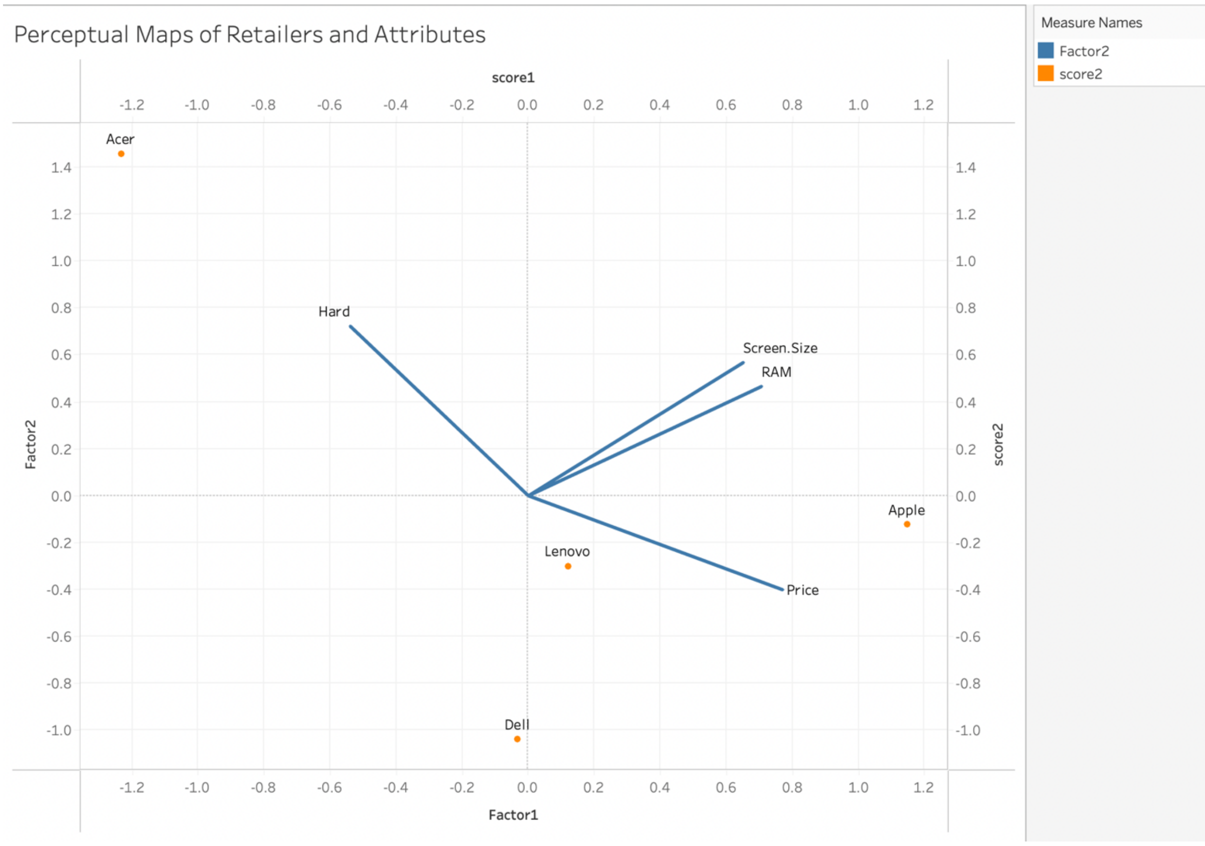
	varExp	pve
1	1.8089198	0.4522300
2	1.2237140	0.3059285
3	0.5186764	0.1296691
4	0.4486898	0.1121725

It was found that, PC1 composed of 45% of total variance explained, PC2 is composed of 30% PVE while PC4 is composed of 11% PVE.





After graphing the PVE percentages, it is observed that there is a break at point 3 in the elbow-plot. This indicates that the majority variance is explained in the first three principal components. In order to understand how different market brands perform on the attributes we used, we graph the visualisation in tableau. To achieve this, we join the perception attributes and perception scores. Since there are no similar columns, we have to join them using Union.



As observed in the above graph, the attribute vectors for Screen Size and RAM are very close to each other indicating high positive correlation. Hard Drive and Price are negatively correlated, while Screen Size and RAM are somewhat correlated to Price.

We can also understand how the different brands perform in different attribute categories. This is achieved by drawing perpendicular vectors from the brand data point to the attribute vector. The brand Apple does exceptionally well in RAM, Screen Size and Price category which also happen to be the most important predictors. Apart from ACER, all other brands do poorly in the Hard Drive section.

Conclusion

In conclusion, the most important features according to the customers in the survey were Screen Size and Ram, while Price also had a relatively higher importance than Brand and Hard drive. Approximately, 64% of the consumers preferred an apple, while dell got a market share of zero in both models. In order to successfully expand the product offerings of the company, they should focus more on features like Screen Size, Ram. They can also provide competitive prices to divert the market to their favour. Since Apple is the best performing brand in all the important predictors they should try to include more Apple products to increase their overall profit.

Bibliography

Anon., n.d. *www.conjointly.com.* [Online]
Available at: https://conjointly.com/guides/how-to-interpret-partworth-utilities/
Green, P. E. &. S. V., 1978. Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research,* 2(3), pp. 103-123.
Jolliffe, I., 2002. *Principal Component Analysis.* s.l.:s.n.
Kim, J. P. J. &. L. H., 2011. Product positioning using conjoint analysis with fractional factorial design.. *Journal of Product Innovation Management,* 28(3), pp. 380-397.
O. Toubia, E. J. J., 2005. Reducing choice overload in iterative choice settings: The role of anticipated regret.. *Journal of Consumer Research,* 32(2), pp. 199-209.
www.conjointly.com, 2023. [Online]
Available at: https://conjointly.com/guides/how-to-interpret-partworth-utilities/

Appendix

R code

```
#####################
# Conjoint Analysis #
#####################

setwd("/Users/kartik/Downloads")
## Install Packages (if needed)
install.packages("conjoint")
install.packages("readxl")
install.packages("dplyr")
install.packages("ggplot2")


## Load Packages and Set Seed
library(conjoint)
library(readxl)
library(data.table)
library(dplyr)
library(ggplot2)

#Set seed to make results reproducible
set.seed(1)

## Set up attributes and levels as a list
attrib.level <- list(Profile = c("Apple", "Lenovo", "Dell", "Acer"),
        Hard = c("128 GB", "256 GB", "512 GB"),
        RAM = c("2 GB", "4 GB", "8 GB", "16 GB"),
        ScreenSize = c("12.1 in", "15.4 in", "17.3 in"),
        Price = c("$900", "$1,200", "$1,500", "$2,000"))
```

```
## Create the fractional factorial design
experiment <- expand.grid(attrib.level)
#design <- caFactorialDesign(data=experiment, type="fractional", cards=20, seed=1)
design <- read.csv("product_profiles.csv")

#Rectify data typo
design$Hard[design$Hard == "356 GB"] <- "256 GB"
print(design)

## Check for correlation in fractional factorial design
print(cor(caEncodedDesign(design)))

## Export design for survey
write.csv(design, file.choose(new=TRUE), row.names = FALSE) ## Name the file
conjoint_profiles.csv

## Run the conjoint analysis study
## Read in the survey preference results
df <-  read_excel("conjoint_profiles.xlsx",
                 sheet = "Conjoint Profiles", range = "l11:ae143")
pref<-as.data.frame(t(df))

## Set up attributes and levels as a vector and Estimate the part-worths for each respondent
attrib.vector <- data.frame(unlist(attrib.level,use.names=FALSE))
colnames(attrib.vector) <- c("levels")
part.worths <- NULL
for (i in 1:ncol(pref)){
  temp <- caPartUtilities(pref[,i], design, attrib.vector)
  ## Pick the baseline case
  ## Adjust coding as needed based on number of attributes and levels
  ## Base Case: Brand CR, Shipping $0, Restock 0%, Retdays 7 days, Price $150
  Base_Profile <- temp[,"Acer"]; Base_Hard <- temp[,"128 GB"]; Base_RAM <- temp[,"2 GB"]
  Base_ScreenSize <- temp[,"12.1 in"]; Base_Price <- temp[,"$900"]
  ## Adjust Intercept
  temp[,"intercept"] <- temp[,"intercept"] - Base_Profile - Base_Hard - Base_RAM -
    Base_ScreenSize - Base_Price
  ## Adjust Coefficients
  ## Brand
  L1 <- length(attrib.level$Profile) + 1 ## Add 1 for the intercept
  for (j in 2:L1){temp[,j] <- temp[,j] - Base_Profile}
  ## Shipping
  L2 <- length(attrib.level$Hard) + L1
  for (k in (L1+1):L2){temp[,k] <- temp[,k] - Base_Hard}
  ## Restock
  L3 <- length(attrib.level$RAM) + L2
  for (l in (L2+1):L3){temp[,l] <- temp[,l] - Base_RAM}
  ## Retdays
```

```
  L4 <- length(attrib.level$ScreenSize) + L3
  for (m in (L3+1):L4){temp[,m] <- temp[,m] - Base_ScreenSize}
  ## Price
  L5 <- length(attrib.level$Price) + L4
  for (n in (L4+1):L5){temp[,n] <- temp[,n] - Base_Price}
  part.worths <- rbind(part.worths, temp)
}
rownames(part.worths) <- colnames(pref)

#Export Partworths for visualisation
write.csv(part.worths, file.choose(new=TRUE), row.names = FALSE)

## Read in perception and preference data
#per <-read_excel("conjoint_profiles.xlsx", sheet = "Perceptions.csv")## Choose
perceptions.csv file
per <- read.csv("perceptions.csv")

## Run Princple Components Analysis on Perceptions
pca <- prcomp(per[,2:length(per)], retx=TRUE, scale=TRUE)

## Perceptual Map Data - Attribute Factors and CSV File
attribute <- as.data.table(colnames(per[,2:length(per)])); setnames(attribute, 1, "Attribute")
factor1 <- pca$rotation[,1]*pca$sdev[1]; factor2 <- pca$rotation[,2]*pca$sdev[2]; path <-
rep(1, nrow(attribute))
pca_factors <- subset(cbind(attribute, factor1, factor2, path), select = c(Attribute, factor1,
factor2, path))
pca_origin <- cbind(attribute, factor1 = rep(0,nrow(attribute)), factor2 =
rep(0,nrow(attribute)), path = rep(0,nrow(attribute)))
pca_attributes <- rbind(pca_factors, pca_origin)
write.csv(pca_attributes, file = file.choose(new=TRUE), row.names = FALSE) ## Name file
perceptions_attributes.csv

## Perceptual Map Data - Brand Factors and CSV File
score1 <- (pca$x[,1]/apply(abs(pca$x),2,max)[1])
score2 <- (pca$x[,2]/apply(abs(pca$x),2,max)[2])
pca_scores <- subset(cbind(per, score1, score2), select = c(Profile, score1, score2))
write.csv(pca_scores, file = file.choose(new=TRUE), row.names = FALSE) ## Name file
perceptions_scores.csv

pca$x

#Calculate PVE values
pca_var = data.frame(varExp = pca$sdev^2)
pca_var = pca_var %>%
  mutate(pve = varExp / sum(varExp))

#Plot PVE
```

```
ggplot(pca_var, aes(as.numeric(row.names(pca_var)), pve)) +
  geom_line() +
  geom_point()+
  xlab("Principal Component") +
  ylab("Proportion of Variance Explained")

#Plot cumulative PVE
ggplot(pca_var, aes(as.numeric(row.names(pca_var)), cumsum(pve))) +
  geom_line() +
  geom_point() +
  xlab("Principal Component") +
  ylab("Cumulative Proportion of Variance Explained")
```