Special Work Performance 33

Group 15 (Giorgi Modebadze (602191) – Radoslav Evtimov (570341) – Ecenaz Bal (429775))

**Summary**

The purpose of this document is to describe the survey data gathered from 593 people regarding different Bluetooth speakers. The main interest of survey is to assess people`s overall awareness of main players on the market, whether respondents own or indent to buy new speaker and which attributes play key role during decision making. 5 item Subjective Knowledge Scale was used for analyzing *consumer's perception of the amount of information they have stored in their memory*. The first part of this documents provides overview of the survey and reports about interesting findings, while the second part tries to create homogenous segments by different using clustering techniques.

**Demographics**

Out of 593 participants 44% were female, 53% male and 3% did not provided their gender. 28 nationality were represented in sample. The majority of people 56% were residents of Germany, followed by Turkey - 7%, Belgium – 4%, France – 3%, US - 3%. Others were less than 3%. Most of the people were Students with 56%, followed by Employed - 34%. Self-employed, Unemployed and Retired totaled 10%. The larger part of respondents – 82% were between 18 and 29 years old. The income distribution by occupation was following:



**Own and intention to buy**

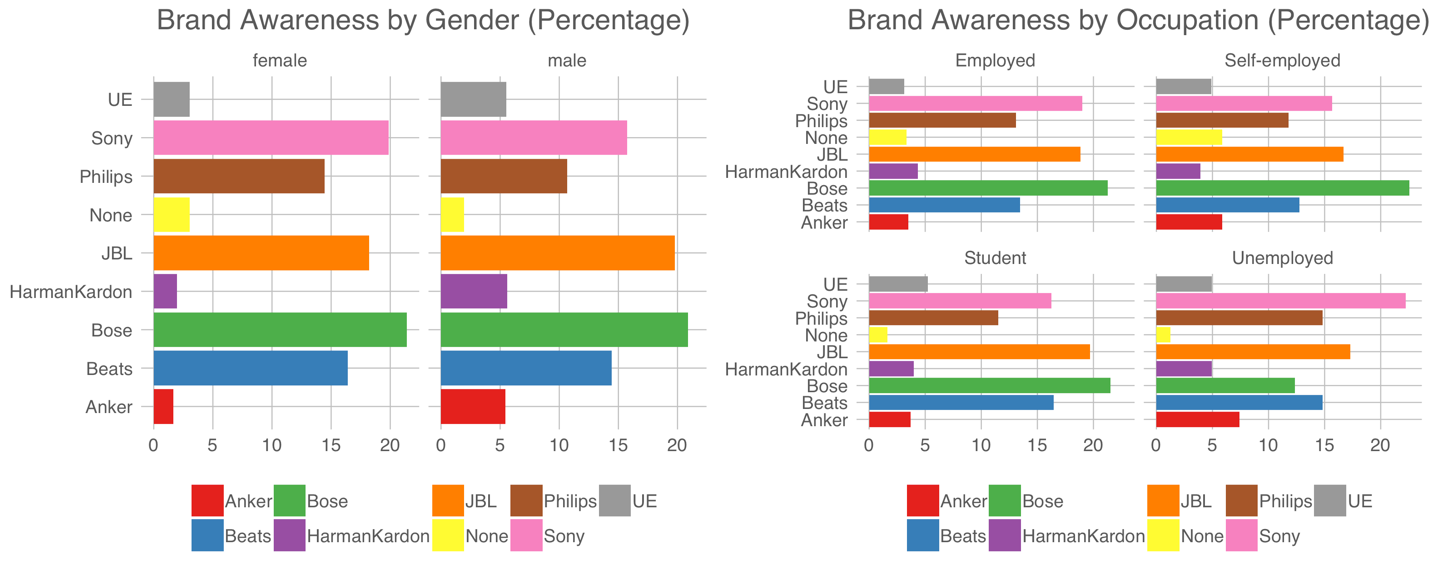
When it comes to owning the speaker, most people 55% did not own speaker, while 45% did. It is interesting to note that this proportion still exists if we spread participants by Occupation. But when it comes to gender, 51% of males own the speakers, while for the females the number is only 39%. It can indicate that interest for the speakers do not differ by age or occupation. We cannot say that Students are more likely to buy them than Employed people. But its clear that males are more interested in owning one than females.

67% of participants are not planning to purchase new speaker, while 33% do. This proportion is maintained within the gender and occupation. Only self-employed people tend to be lower on scale to purchasing new one - 24%. Income wise the lowest proportion goes to the people who earn most. Only 26% plans to buy one.

**Brand Awareness**

The most well know brand between participants is Bose. Out of 593 people, 391 know about it.

Close competitors are: JBL – 354 and Sony – 327. Least popular brands are – UE – 84 HarmanKardon – 78 and Anker with 74 votes. There is a difference between proportion for gender as well as occupation.

****

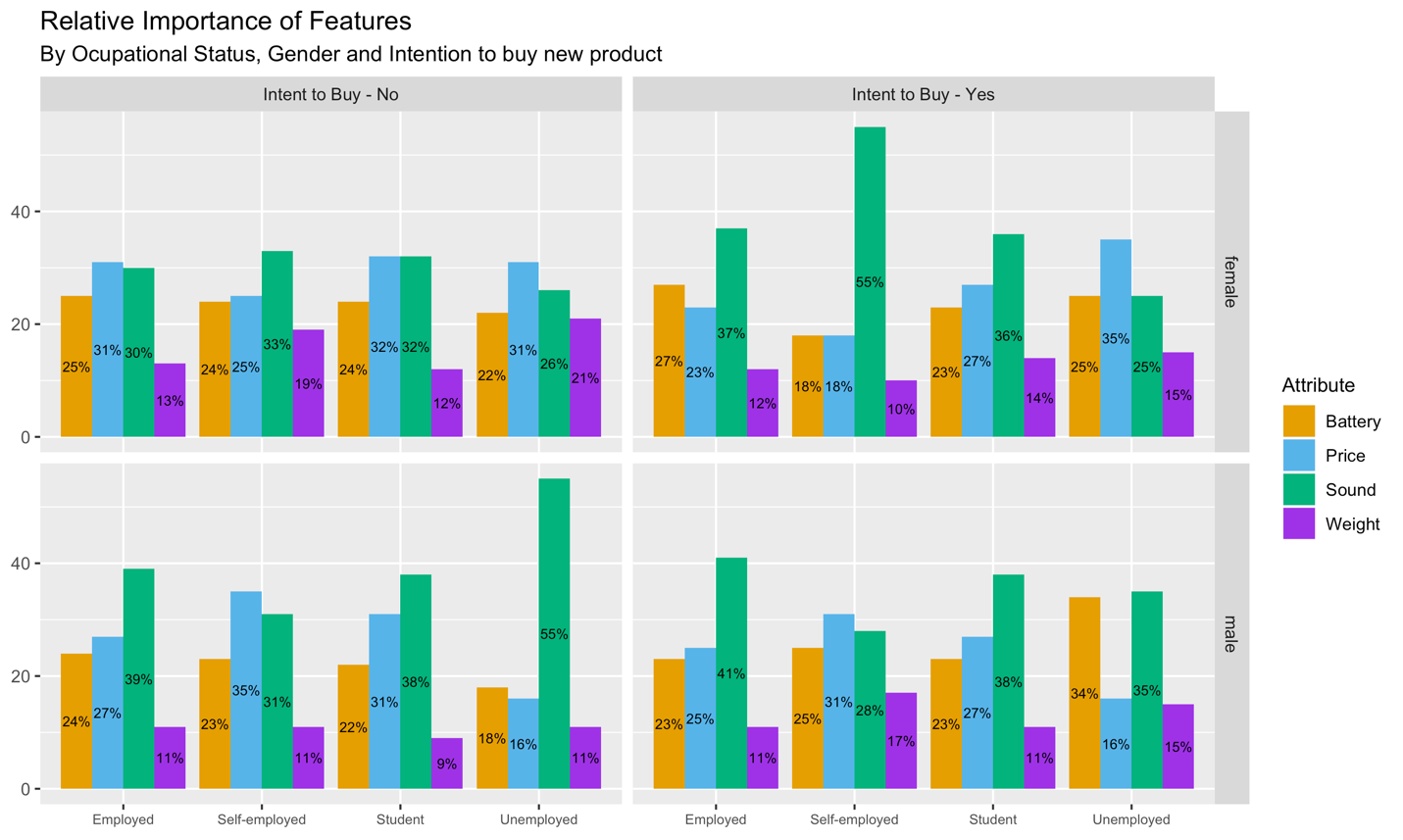
It is worth noting that for people who are willing to purchase new speakers, Bose is still number one with 138 votes out of 198 with closest to it JBL – 133. Anker, UE and HarmanKardon still remain on the bottom with , 27, 26 and 29 votes.

**Subject Knowledge**

The survey contained 5 item Subject Knowledge Scale to determine participants perception regarding the information they think they know about speakers. Using factor analysis, two factors were enough to explain 74% of the variation. Third item – “*Among my circle of friends, I'm one of the "experts" on portable Bluetooth speakers* “seem to be most different of other items and respondents tend to have twice the lower points for this item, than on average. When it came to gender, Females on average put lower points than men, meaning females do not believe they are experts on Bluetooth speakers. It is important to mention, that people who answered that, they were going to buy product, also tend to give themselves higher scores. It can indicate that, buying Bluetooth speaker is less of impulsive decision and before actual purchase people actually do their research. There is quite difference also by occupation. Self-employed people tend to score themselves highest, followed by students and employed people.

**Relative Importance of Features**

Survey provides the data about relative importance of features. People were asked to assign 100% of weights to the four features: Battery, Price, Sound and Weight. Taking overall average: The sound is most important feature – 36%, followed by price – 29%, battery – 23% and weight – 12%. This pattern continues within the more relevant group, people with intention to buy a new product, especially in two largest groups employed and students. For employed females Battery life is more important than price. The graph below provides full detailed information.



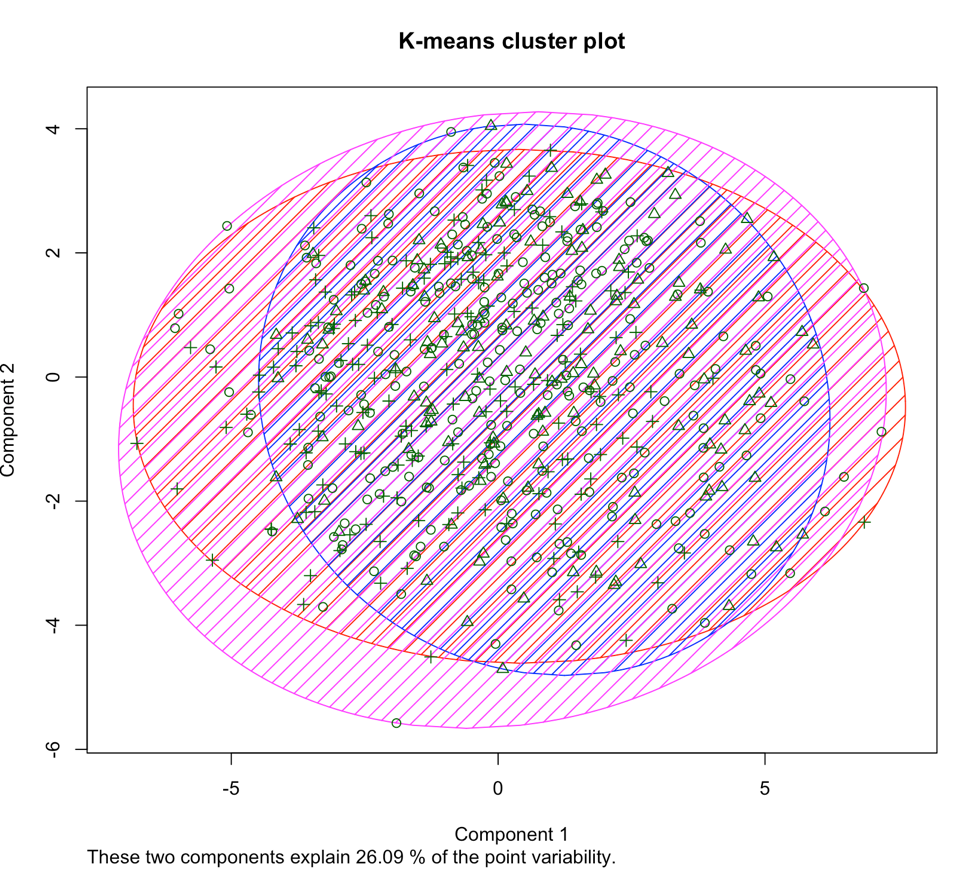
**Clustering**

After general overview of the survey data and seeing some patterns in it, clustering techniques are used to discover new customer segments. The main goal of this process is to create segments that are similar within the group and different from other groups, so it will be easy for marketers to create specific marketing campaigns for each of them and target them better.

As the first step specific variables from the data are chosen, that to our belief can have biggest impact in the clustering process. Variables are chosen based on the exploratory data analysis done in the first part. The variables are: Own, Intent to Buy, all of Subject Knowledge, PII and Relative importance, Gender, Age, Residence, Occupation and Income. Also new variable – Brand Knowledge is created – which is sum of all brand awareness questions. It can help to adjust Subject Knowledge. During the process, several pairs of variables were selected, so we could find best pair of variables, that could give best segments.

The first clustering technique we will be using is K-means, even though it rises a problem. K-means relies on Euclidean distance, while our data contains some binary data, like Own and Intent to buy, as well as non-metric data, like Gender, Age, Residence and so on. The good thing is that binary factors can be coerced to numeric with no alteration of meaning. And non-metric data can be converted to binary variables by dummying them. The mlr package`s create Dummy Features were used for this purpose. As the goal we chose 3 cluster, which could be extended maximum to five, as we believe that this is the maximum number of segments, marketers can use to target and create special campaigns for.

Using K-means did not show much success, as it created groups that were very similar to each other, make it impossible to interpret them and allowing business decision making. The main difference occurred in the relative importance of features, but the features, like income level, gender, occupation were very similar. This result was expected, as it is not optimal to cluster binary values with k-means, but as it was mix of binary and non-binary data it is always worth to try, especially if we take *no free lunch theorem* into consideration.



As the second approach into clustering survey data hierarchical clustering was used. As K-means, hierarchical clustering also starts with proper selection of variables. Other than that, it has its own constraints that needs to be taken into considerations before moving forward. As it is distance-based algorithm, where primary information is the distance between observations, it only accepts numeric values. But most of the variables of survey data are factor variables. To convert them into Euclidean distance, daisy function from cluster library was used. After getting distance matrix, hclust function was used for creating hierarchical cluster. For the first attempt complete linkage method was used. For checking goodness-of-fit cophenetic correlation coefficient was used which gave 0.38 for “complete” and 0.31 for “single” linkage. Meaning, its still highly ambiguous groups, but results are much better than K-means. The graph below represents the groups that were created after choosing four clusters from the dendrogram.

The main differentiator between groups were Occupation and Income Level. Also for this people different features of Bluetooth speakers played role.

Just Interpret results I sent on FB

