**SWP 3 - Describing Data, Segmentation & Clustering**

***3a. Data Structure and Key Findings***

The given dataset consists of the results derived from an exploratory market research questionnaire, regarding the respondent’s personal information, as well as involvement, preferences and opinions regarding Bluetooth speakers. Some answers had reversed answer scales and patterns in order to check for respondent’s attention and prevent unusable answer sets.

The resulting data basis indivData consists of 593 respondents answering 36 answers each. 3% of the participants did not indicate their gender and therefore will be dropped from the subsequent analysis data basis due to inconclusive answers. Furthermore, we will exclude the 13% of respondents that did not state their monthly income, due to the implications that uncertainty holds for marketing research purposes. The cleaned data set consist of 506 entries. The two largest economic subsets consist of 57.90% students and 31.82% employed people. Roughly half of respondents, 57.90% stated a monthly household income below 1000€. 82.60% of respondents are aged between 18 and 29, indicating a rather young segment. 57.11% of respondents stated their country of residence as Germany, while the rest were widely spread to over 30 different countries. Due to the nature of Bluetooth speakers, there are no special implications derived here, due to technical similarities across markets and the lack of local distribution. 54.55% of respondents are male, 45.45% are female.

The questionnaire comprised multiple Bluetooth speaker properties the participants had to value. They were asked to distribute 100% across the four attributes “Battery”, “Price”, “Sound” and “Weight”. The respondents valued the importance of the attributes as shown in Figure 1.



Figure 1. Relative Importance Rating of Properties (in %)

Multiple implications can be derived from that result. The sound is valued by the participants as the most important property of Bluetooth speakers, whereas males tend to value it slightly higher than females. The second most important factor is the price of the speaker. Since 57.9% having an income lower than 1000€ this is not a surprising but an important outcome. Figure 1 clearly shows where the focus of a marketer has to be, namely on the sound and price. In combination with the fact that 82.6% of the respondents are aged between 18 and 29 years, we can state that when targeting the participants of the presented study, marketers or sellers should focus to present the Bluetooth speakers as a low-priced but still high-quality good, as well as presenting it as a young and cool device. A light weight scarcely increases the value of a Bluetooth speaker. When taking gender into account withing the four properties, it can be stated that females are willing to forego sound quality in order to get a lower-priced product. Thus, from the marketer’s point of view it can be advantageous to differentiate the product in that regard. When differentiating the four properties shown in Figure 1 by the two most important age groups (18-24 years and 25-29 years) we can state that with increasing age, the preferences slightly shift. The respondents above an age of 24 years valued the sound and the battery capacity slightly higher than the 18–24-year-old group. For that increase in product quality, they accept a corresponding increase in price. A discrimination in terms of this findings might be considered when selling Bluetooth speakers to the considered age groups. The aforementioned implications are derived by the results shown in Figure 2. The presented numbers are means across the age groups in percent.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Age | Battery | Price | Sound | Weight |
| 18-24 years | 23.50 | 29.59 | 35.13 | 11.77 |
| 25-29 years | 24.39 | 27.62 | 36.60 | 11.40 |

Figure 2. Relative Importance Rating in % (by age group)

In marketing, knowing your potential consumer is essential if not even the most important aspect. The more detail a marketer or business has about the consumers, the more specific can the groups be targeted, which ideally results in higher sales. With respect to understand the consumers of Bluetooth speakers in the given data set, we decided to first filter the data by people who already own a Bluetooth speaker. In the given data set, 234 out of 506 participants already own a Bluetooth speaker, that corresponds to a share of 46.25%. From these 234 people, 191 are between 18 and 29 years old, consisting of 79 females (41.36%) and 112 males (58.64%). This suggests that the main consumers are young men between 18 and 29 years. The consequences of defining this demographic group as the main consumer would be as follows when we take into account the results from property analysis. The sound of the Bluetooth speaker would be the most important attribute. Furthermore, the consumer group would be willing pay a higher price when they perceive the sound as being high-quality. In the context of consumer targeting, it is furthermore even more important to take a deeper look at the variable “IntentToBuy” because it may occur that that group varies from the group already owning the good. In the given data set 174 respondents intent to buy a Bluetooth speaker. 151 of them are in the age between 18 and 29 years, of which 45.70% are female and 54.30% are male. We can state that that results regarding the age distribution corresponds to the findings about participants who already own a Bluetooth speaker.

Through further differentiation of the group of 151 people who intends to buy a Bluetooth speaker we were able to derive the following results. Among the people who already own a speaker, 51 intent to buy a new one. From these 51 people 19 people are female and 32 people are male. This is not a surprising result when considering the previous findings. As stated before, males value the sound of the speakers as the most important attribute. With technical progress the sound quality gets better with every new generation of Bluetooth speaker. Thus, males buy Bluetooth speaker more frequently in order to get better sound quality. Females in the given data set, however, are more price sensitive in the context of Bluetooth speakers and do not value the sound as important as males do. Therefore, they may use a Bluetooth speaker over a longer period of time. An interesting finding is that among the people in the data set who do not own a Bluetooth speaker but intent to buy one the share of females and males is 50:50. That implies, that when looking at consumers between 18 and 29 years who do not own a Bluetooth speaker, males and females should be targeted in the same proportion, although young males previously seemed to be the main consumer group. When creating follow up marketing campaigns, however, it might be more efficient to target males rather than female, as shown above.

In order, to further specify the marketing strategy, we decided to group the participants aged between 18 and 29 years who intent to buy a Bluetooth speaker based on their subjective knowledge regarding the product. The variable “Subjective Knowledge” is built by the mean of 5 different subjective knowledge measures in the questionnaire, calculated for every participant. It can take values from 1 to 7. Group 1 is defined by respondents who stated their subjective knowledge between 1 and 3 (44 people), Group 2 consists of the people who stated their subjective knowledge between 3 and 5 (71 people) and Group 3 contains every person who stated its subjective knowledge above 5 (36 people). This split across leads to the following results (Figure 3).



Figure 3. Intention to buy among Subjective Knowledge Groups (age 18-29)

Most of the participants who intent to buy a Bluetooth speaker have a subjective product knowledge between 3 and 5 (Group 2). This is neither very high nor very low, they have a rather average subjective knowledge about the product. In Group 2 (medium subjective knowledge) as well as in Group 3 (high subjective knowledge), males are overrepresented. The females in the given data set rather perceive their product knowledge as low, therefore most of them show up in Group 1. These findings are consistent with the previous analysis. As already stated, males have a higher focus on the sound of Bluetooth speaker than females do (Figure 1). That can now be explained by Figure 2. Since males perceive their product knowledge as higher than females do, it consequently influences their buying decision in that direction. The same effect holds for females, who base their buying decision mainly on the product price.

The presented findings suggest that males might have a higher involvement in the product than females do. The higher the consumer’s product involvement, the higher the probability they are going to buy the product. In order to evaluate the participant’s involvement levels, we built groups in the same manner we did with the subjective knowledge (age 18-29, “IntentToBuy” = 1). Again, most respondent’s contain in Group 2 with a medium product involvement between 3 and 5. Marketers should therefore prepare for that involvement level when creating advertisements for Bluetooth speakers. Another challenge for the businesses is to increase the product involvement in the long term in order to increase sales. In contrast to the subjective knowledge, there are more males than females in every Group.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Group 1 | | Group 2 | | Group 3 | |
| Males | Females | Males | Females | Males | Females |
| 10 | 7 | 62 | 57 | 10 | 5 |

Figure 4. Personal Product Involvement Groups (age 18-29)

When taking the results from the analysis about who does not own a Bluetooth speaker but intents to buy one into account, we can conclude that one task of businesses is to increase the product involvement of females. They are as important as male consumers, with the difference that they have a lower perceived subjective knowledge (Figure 3) and therefore may focus on slightly other things than males (Figure 2). When increasing the product involvement, subjective knowledge might increase and that can lead to more customers. Therefore, in the long term it is important to increase the female as well as the male consumers product involvement to maximize sales.

***3b. Clustering***

In the first step three columns were added to the dataset, each containing the mean of the brand awareness, knowledge and involvement values. This will be useful for the evaluation and interpretation of the clusters.

For the clustering it was first necessary to find a subset of the data that contained a sufficient amount of meaningful variables to yield good clusters but didn’t contain too many variables, so the clusters would be sufficiently spaced for the algorithm to correctly indentify them. In the graph below one can see the MDS of three subsets.

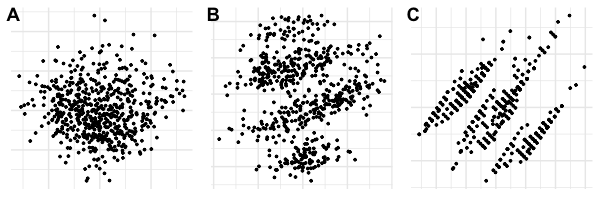


Figure 5. Multidimensional Scaling

A is the MDS of the entire dataset. No clusters are visible. C is the MDS of a subset containing the variables Own, IntentToBuy, RelImp\_Price and Income. Three distinct and well spaced clusters are visible. The subset used for B contains the same variables as C but additionally the above mentioned averages of brand awareness, knowledge and involvement. One can see the outlines of four different clusters, which is promising. Unfortunately the clusters are not sufficiently distinct for any algorithm to yield a good result. Therefore we settled on the subset whose MDS is displayed in graph C.

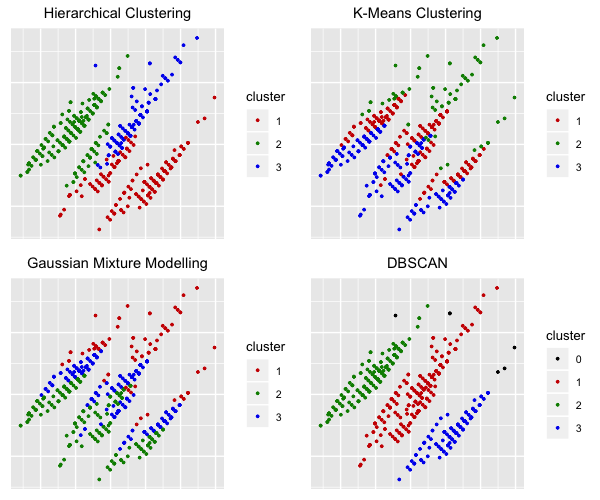


Figure 6. Clustering results

In the next step we tried out several different clustering algorithms to find the one which achieves the best result. The graphical results of the cluster analysis can be seen in the graph above. While Hierarchical Clustering, K-Means Clustering and Gaussian Mixture Modelling (GMM) have achieved decent results, only Density-Based Spatial Clustering of Applications with Noise (DBSCAN) was able to correctly indentify the three clusters. This makes sense as the clusters are well spaced and have roughly the same density. A great thing about DBSCAN is also that it marks as outliers points that lie alone in low-density regions (i.e. whose nearest neighbors are outside of the set epsilon neighborhood) and adds them to another cluster. Below one can see the mean values for each cluster and each variable, including the averages of brand awareness, knowledge and involvement at the end.

Ein Bild, das Text, Zeitung enthält.

Automatisch generierte Beschreibung

Figure 7. Results of the DBSCAN

Cluster 0 consists of six outliers. All of them have a rather low subjective knowledge about Bluetooth speakers. However, Cluster 0 is not representative since it only contains six people, we therefore excluded it from the analysis. Cluster 1 are the „Scrimpers“ with 251 elements. They mostly don’t own and don’t want to buy a bluetooth speaker. Overall they are quite average, but have little knowledge and generally care about the price. Cluster 2 are the „Owners“ with 205 elements. They all own and don’t want to buy a speaker, have high knowledge, high involvement and care about sound. Furthermore they have the highest age and the highest income out of the three main clusters. Cluster 3 are the „Poor Graduates“ with 131 elements. They all don’t own a bluetooth speaker, but want to buy one and show rather high involvement. They are the youngest and have the lowest income, but also the highest education level.