Consumer Preferences Analysis Using MFA

Angélica Portocarrero and Atenea Rojas

August 3, 2024

Universidad AUTÓNOMA de Occidente

1 Study Context

In this case study, a consumer goods company aims to better understand the preferences and perceptions of its customers regarding its products. To achieve this, information has been collected on product characteristics, consumer preferences, and demographic and qualitative data of the consumers. Analyzing this data will help the company identify patterns and effectively segment consumers.

2 Data Dictionary

Variable	Description
product_quality	Perceived product quality on a scale of 1 to 10, with higher values indicating better quality.
product_durability	Perceived product durability on a scale of 1 to 10, with higher values indicating greater durability.
preference_taste	Preference for the taste of the product on a scale of 1 to 10, with higher values indicating greater preference.
preference_price	Preference for the price of the product on a scale of 1 to 10, with higher values indicating greater preference.
age	Age of the consumer in years, providing demographic information.
income	Annual income of the consumer in dollars, providing information on socioeconomic status.
gender	Gender of the consumer (Male or Female).
education_level	Education level of the consumer (High School, Bachelor's, Master's, PhD).

Table 1: Descriptions of variables used in the analysis.

3 Analysis Objective

The objective of the analysis is to use Multiple Factor Analysis (MFA) to identify patterns in consumer perceptions of product characteristics, preferences, and how these relate to demographic and qualitative data. This will enable the company to segment consumers and design more effective marketing strategies.

4 Discussion Questions

1) What are the main dimensions identified in the MFA analysis, and how much variance do they explain?

2) How are the product characteristics related to consumer preferences according to the MFA results?

3) Can you identify any distinct consumer segments based on the MFA individual factor map? Describe them.

4) How do demographic variables (age and income) influence consumer preferences and product perceptions?

5) What insights can be drawn from the qualitative variables (gender and education level) in relation to product quality and durability?

6) Discuss the contributions of quantitative and qualitative variables to the main dimensions. Which variables are most influential?

7) How can the company use these insights to improve their marketing strategies and product development?

8) What are some potential limitations of this study and the MFA method in analyzing consumer preferences?

9) Propose additional variables or data that could be included in future analyses to gain a more comprehensive understanding of consumer behavior.

10) How could the company integrate these findings into their customer relationship management (CRM) systems for personalized marketing?

5 Exploratory Data Analysis

Before starting any kind of analysis, it is important to perform a previous process in order to explore the data initially. This Exploratory data analysis aims to gather insights and hints of the patterns and relationships of the data. The information collected here will support conclusions obtained by the MFA, demonstrating it's importance in any kind of data analytics process.

The first operations performed with the data were directed towards discovering the shape of the dataset. In this case, as shown in **Table 2** it was discovered that the dataset had a shape of a hundred rows and eight columns, without null or duplicate values.

Rows	**Columns**	**Null values**	**Duplicate values**
100	8	0	0

Table 2: Dataset shape

Next, **Table 3** shows the descriptive statistics for each variable providing valuable insights into consumer perceptions, preferences, and demographic characteristics, aiding the company in understanding its customers better.

	product_q	**product_d**	**pref._taste**	**pref._price**	**age**	**income**
Min.	4.079	1.870	5.809	-0.4038	18.0	27226
1st Qu.	6.429	5.139	7.251	3.6942	30.0	42921
Median	7.115	5.959	7.918	5.6152	39.0	49607
Mean	7.073	6.077	7.912	5.2425	40.8	49894
3rd Qu.	7.950	7.205	8.530	6.7595	52.0	55791
Max.	8.905	9.636	10.293	9.5282	65.0	76494
SD	1.109	1.545	0.9077	2.0981	13.84	9103.3
CV	15.678	25.430	11.472	40.022	33.940	18.245

Table 3: Descriptive statistics of variables

For product quality, the perceived scores range from 4.079 to 8.905. The median score of 7.115 and mean of 7.073 indicate that, on average, consumers rate the product quality relatively high. The standard deviation (1.109) and coefficient of variation (15.678) suggest moderate variability in perceived product quality among consumers. This suggests that while most consumers perceive the product quality positively, there is some variability that could be explored further to understand different consumer segments' views on quality.

For product durability, the scores range from 1.870 to 9.636, with a median value of 5.959 and a mean of 6.077 that indicate that consumer perceptions of product durability are more varied, with an average leaning toward moderate durability. The higher standard deviation (1.545) and coefficient of variation (25.430) indicate significant variability in durability perceptions. This indicates that the company may need to investigate why perceptions of durability vary widely and address any concerns to improve overall consumer satisfaction.

Preference for taste has scores ranging from 5.809 to 10.293, with a median value of 7.918 and a mean of 7.912 that show that consumers generally have a strong preference for the product's taste. The low standard deviation (0.977) and coefficient of variation (11.472) suggest that taste preferences are relatively consistent among consumers. This means that the product's taste is a strong selling point with consistently high preferences, which can be emphasized in marketing strategies.

Preference for price displays scores from -0.4038 to 9.5282, with a median value of 5.6152 and a mean of 5.2425. The high standard deviation (2.0981) and coefficient of variation (40.022) reflect a substantial variability in price preferences. This leads to assume that price sensitivity varies greatly among consumers, suggesting the need for segmented pricing strategies or promotions to cater to different consumer groups.

For age, the range is from 18.0 to 65.0 years. The median age is 39.0 years, and the mean is 40.8 years, indicating a diverse age distribution with a slight skew towards older consumers. The standard deviation is 13.84, and the coefficient of variation is 33.91, reflecting considerable age diversity. This indicates that marketing strategies should account for a wide age range, potentially tailoring messages to different age groups to maximize appeal.

Finally, income ranges from \$27,266 to \$76,494, with a median income of \$49,607 and a mean of \$49,894, suggesting that the consumer base has a moderate income level. The standard deviation is \$9,103.3, and the coefficient of variation is 18.245, indicating moderate income variability among the consumers. This implicates that the company should consider income diversity when designing marketing campaigns and product pricing, ensuring affordability and value for different income segments.

Next, **Figure 1** illustrates the gender distribution of the consumer sample, showing the proportion of male and female participants. It can be seen that female participants represent 53% of the sample, meaning that female participants slightly outnumber males. This indicates a slight skew towards female consumers in the data collection process.

The near-equal representation of male and female consumers suggests that the company's products appeal broadly across genders. Leveraging the insights gained from this balanced representation, the company can design marketing strategies to target both male and female consumers without significant bias.

Figure 1: Gender distribution

Figure 2 presents the distribution of education levels among the consumer sample, categorized into Bachelor's, High School, Master's, and PhD levels.

It can be seen that 22% of the participants hold a Bachelor's degree, indicating a substantial portion of the sample has completed undergraduate education. then, 17% of the participants have a High School education, representing the smallest group in the sample. This suggests that a smaller segment of the consumers have only completed high school. With 35% of the participants holding a Master's degree, this group is the largest segment, indicating a significant portion of the consumer base has pursued graduate education. Lastly, 26% of the participants hold a PhD, representing a notable segment of highly educated consumers.

The distribution shows a diverse range of education levels, with a significant proportion of highly educated consumers (Master's and PhD). This suggests that the company's products may appeal more to individuals with higher education levels, who might have specific preferences and higher purchasing power. Marketing strategies can be tailored to highlight aspects of the product that appeal to well-educated consumers, such as quality, innovation, and value for money. Additionally, understanding the preferences of the smaller High School-educated segment could provide opportunities to broaden the product's appeal.

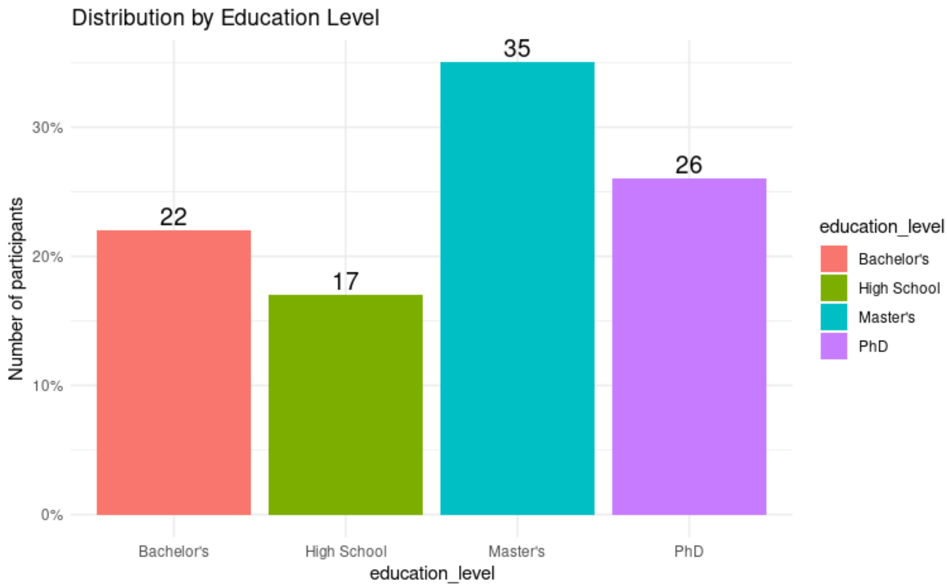

Figure 2: Education Level Distribution

Table 4 presents the distribution of education levels segmented by gender, providing insights into the educational background of male and female participants.

The data shows that a higher percentage of female participants have obtained a Master's degree compared to their male counterparts, while males have a slightly higher representation at the Bachelor's and High School education levels. Understanding these differences in educational attainment can help the company tailor its marketing and communication strategies to resonate better with each gender group, taking into account their educational backgrounds.

	Bachelor's	High School	Master's	PhD	Sum
Female	10.00	6.00	22.00	15.00	53.00
Male	12.00	11.00	13.00	11.00	47.00
Sum	22.00	17.00	35.00	26.00	100.00

Table 4: Education level per gender

Figure 3 displays the distribution of education levels among female and male participants in pie charts, providing a visual representation of the data from **Table 4**.

The pie charts highlight the higher concentration of female participants with Master's degrees (41.5%) compared to males (27.7%). Conversely, males have a more balanced distribution across all education levels, with a notable 25.5% holding Bachelor's degrees and 23.4% each holding High School and PhD levels. This visualization reinforces the insights from Table 4, indicating that female consumers in the sample are more likely to have advanced degrees. This could imply that female consumers may value educational content and detailed product information, which can be utilized in marketing campaigns to better appeal to this demographic. For male consumers, with their more evenly spread educational backgrounds, a varied approach that includes both practical and detailed information may be effective.

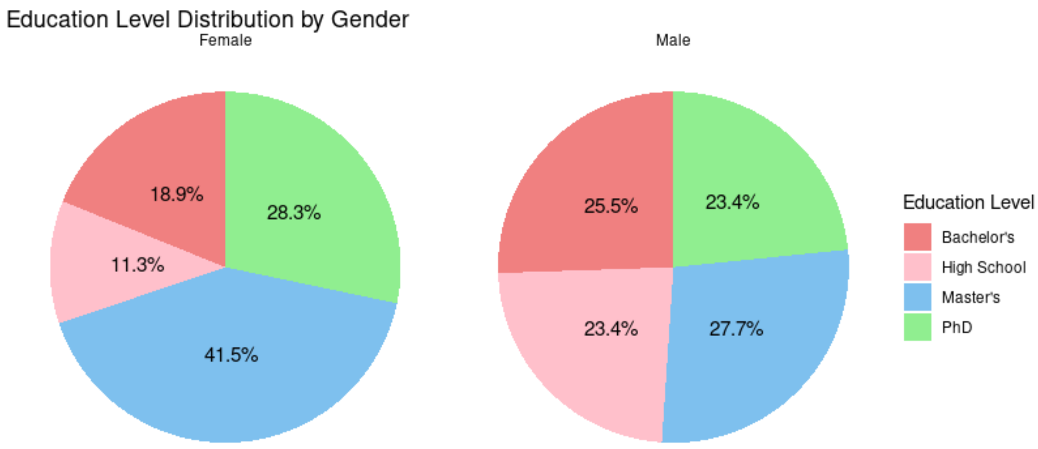

Figure 3: Education level distribution by Gender

Lastly, for this exploratory data analysis phase, boxplots were created to visualize consumer perceptions and demographic data based on gender and education level.

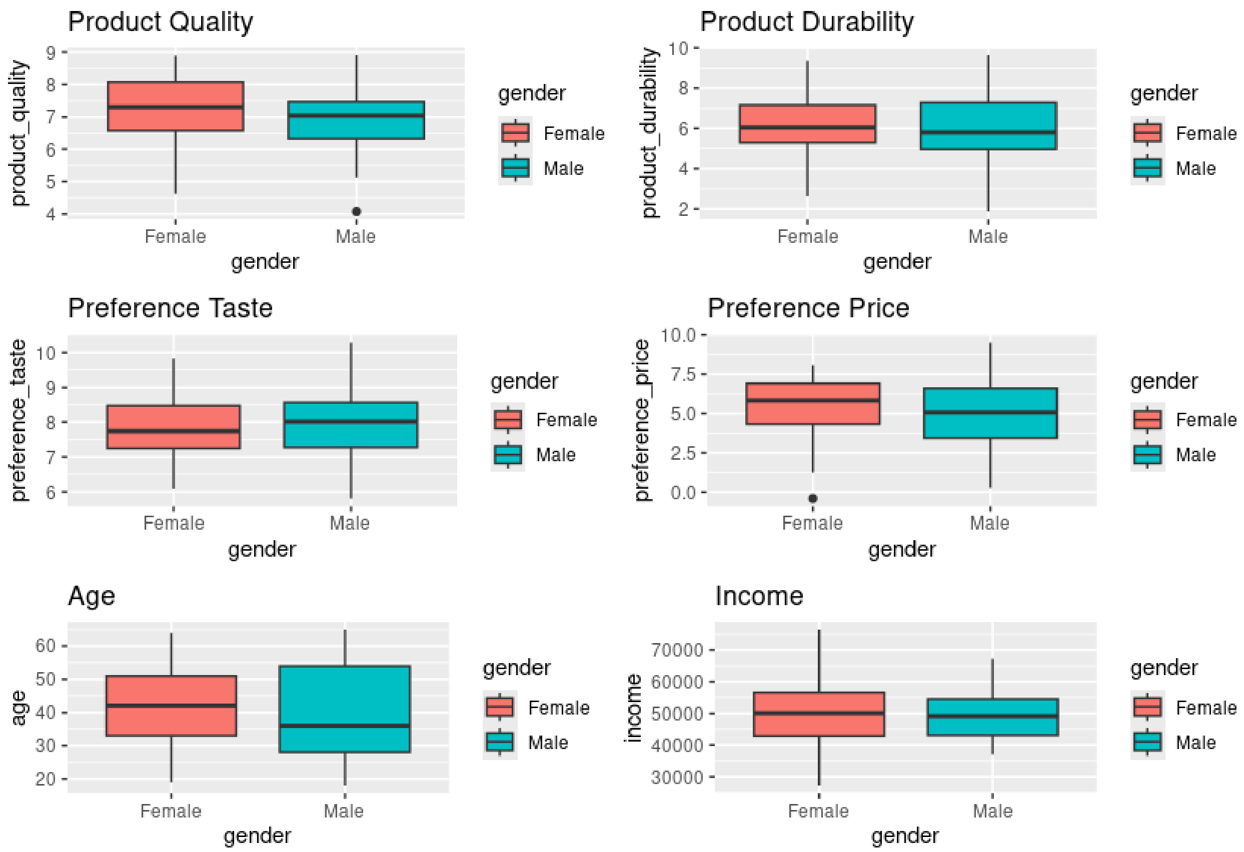

Figure 4: Variables by Gender

As seen on **Figure 4**, for gender, the boxplots revealed that male and female customers have similar median values for perceived product quality, with slightly higher median ratings for males, and females exhibiting greater variability. Both genders showed comparable medians for product durability, with males slightly higher, and consistent variability in responses.

Taste preference medians were nearly identical, with males slightly higher and females showing more variation. Preference for price had similar median values for both genders, but females displayed a wider range of responses. The median age was close for both genders, with females slightly older and males showing a wider age range. Median income values were similar for both genders, with females slightly higher and exhibiting greater variability.

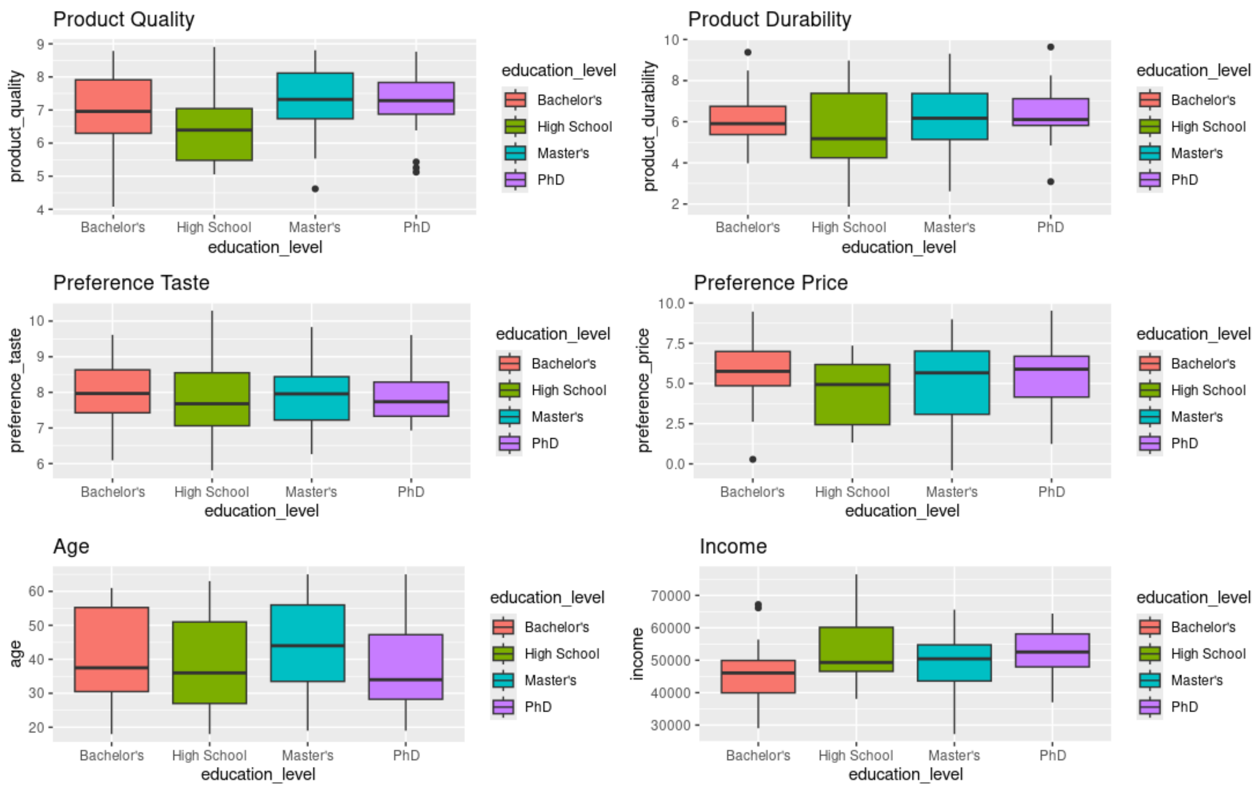

Figure 5: Variables by Education Level

In **Figure 5** boxplots are segmented by education level. In this case, the boxplots indicated that perceived product quality ratings were relatively similar across all education levels, with greater variability among those with Master's degrees and Bachelor's degrees, and showing a few very distanced outliers for PhD degree holders. Perceived durability ratings were comparable, with High School graduates showing a lower median a bigger variability.

Taste preference medians were quite similar, with Bachelor's degree holders slightly higher and High School graduates showing more variation. Price preference medians were similar, with Master's degree holders exhibiting a broader range of responses, and Bachelor's degree holders a lower variability with an outlier that indicates a case of a strong dislike for the product's price.

The age distribution showed that PhD holders were generally younger, while Master's degree holders were older, with varying age spreads across education levels. Median income values almost increased with higher education levels, with PhD holders having the highest median income however High School graduates seemed to have slightly higher income that Bachelor's degree holders.

These boxplots provided valuable insights into how consumer perceptions and demographic factors varied based on gender and education level, guiding further analysis in the case study.

6 Analysis Results

1) What are the main dimensions identified in the MFA analysis, and how much variance do they explain?

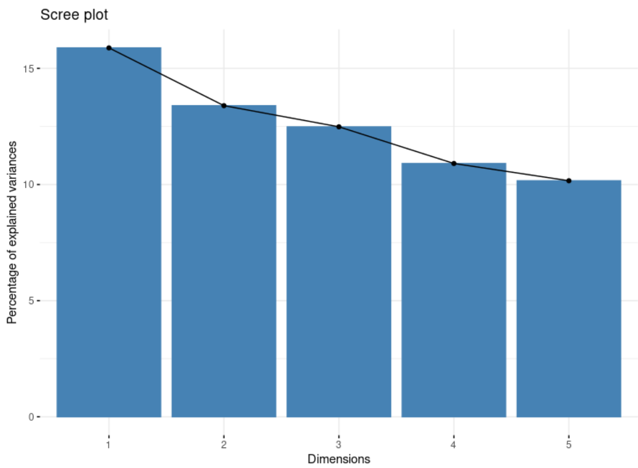

Figure 6: FAMD Screeplot

According to the screeplot, it can be identified that the main dimensions explain approximately 29% of the data, taking into account that we took the first two dimensions as the main ones. It is advisable to take the number of dimensions that explain at least 70% of the data, however, in this case there are dimensions that are not very relevant, this can also be seen in the correlation matrix of **Table 5** where very weak correlations are evident.

	prod_quality	prod_durability	pref_taste	pref_price	age	income
prod_quality	1	0.09	-0.07	0.21	-0.11	-0.04
prod_durability	0.09	1	-0.04	-0.02	0.08	-0.02
pref_taste	-0.07	-0.04	1	-0.12	0.10	0.09
pref_price	0.21	-0.02	-0.12	1	-0.12	0.07
age	-0.11	0.08	0.10	-0.12	1	0.12
income	-0.04	-0.02	0.09	0.07	0.12	1

Table 5: Correlation Matrix

2) How are the product characteristics related to consumer preferences according to the MFA results?

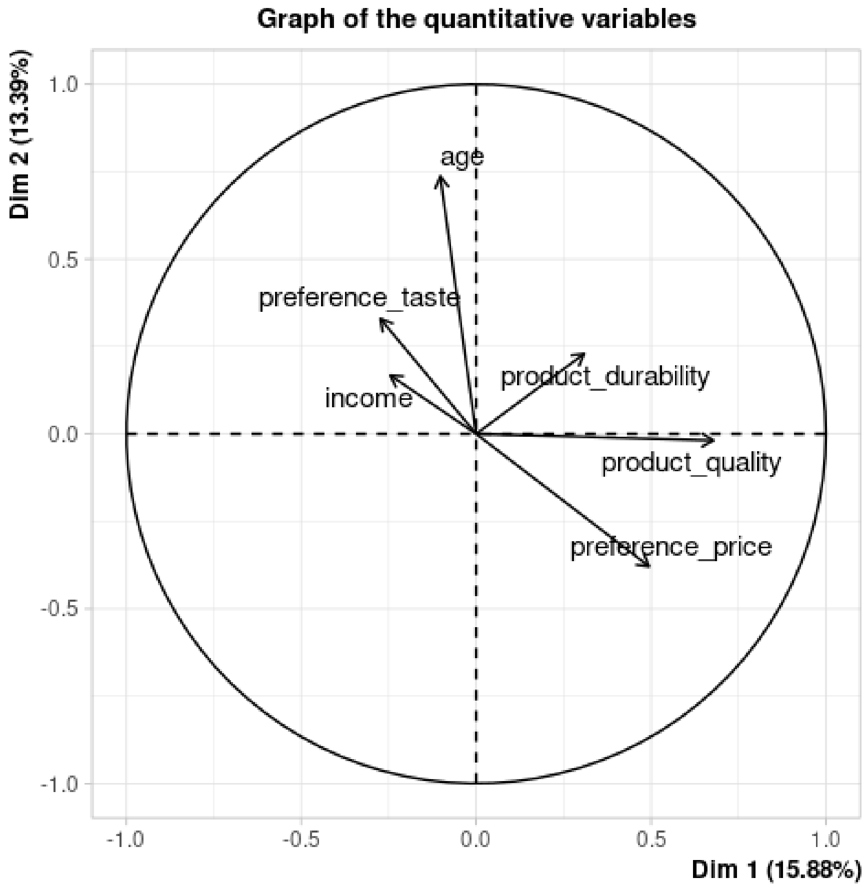

Figure 7: FAMD Biplot

First, product_quality and preference_price have a positive, though not very strong, relationship. This suggests that consumers who value product quality may also consider price similarly, indicating that those seeking high quality products may be willing to pay higher prices. However, the non-exact proximity of their positions also suggests that price is not the most critical factor for all consumers who value quality, but it is still an element to consider. This positive correlation reflects a market segment where both quality and price are important factors in decision making.

In addition, product_durability is positively correlated with product_quality, as both vectors are oriented in a similar direction. This indicates that consumers who value durability also tend to value quality. These attributes tend to be important to those consumers seeking reliable and durable products, suggesting that durability is perceived as an essential component of overall product quality.

On the other hand, preference_taste and income show a positive correlation, indicating that consumers with higher income tend to value product taste more highly. However, since income is closer to the center of the graph, this relationship is less pronounced, suggesting that income has a moderate influence on taste preference, and that this attribute may be valued by consumers of different income levels.

As for age, its vector is almost perpendicular to that of product_quality, indicating a lack of direct correlation with this product characteristic. In addition, age is close to the outer edge of the biplot, indicating significant variability. Product_durability, on the other hand, is closer to the center, suggesting lower variability. Although age and product_durability do not show a strong relationship, there could be a slight association. However, age seems to have a more notable relationship with preference_taste, as their vectors are closer, suggesting that taste preferences may vary more significantly with consumer age than other product characteristics.

3) Can you identify any distinct consumer segments based on the MFA individual factor map? Describe them.

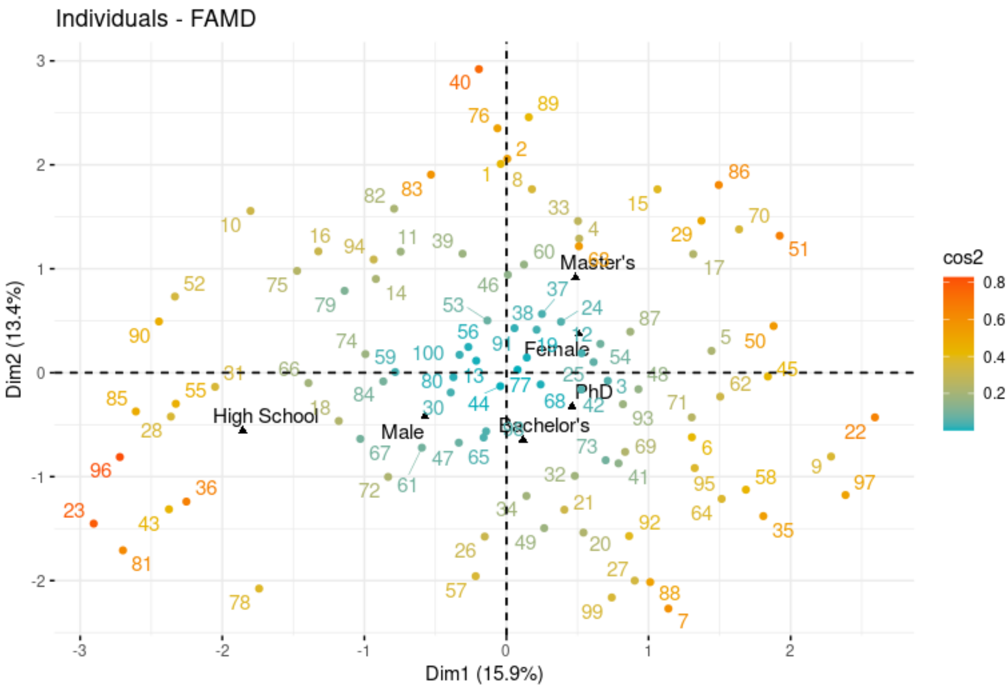

Figure 8: Individual factor map

As displayed on **Figure 8**, the segments of individuals are non-existent, since the individuals are almost completely mixed, which can be seen later in the clustering analysis.

In the clustering analysis process, a series of methods were performed in order to select the optimal amount of clusters for the data. The silhouette, elbow and gap statistic methods returned the results displayed on **Figure 9**, **Figure 10**, and **Figure 11** respectively.

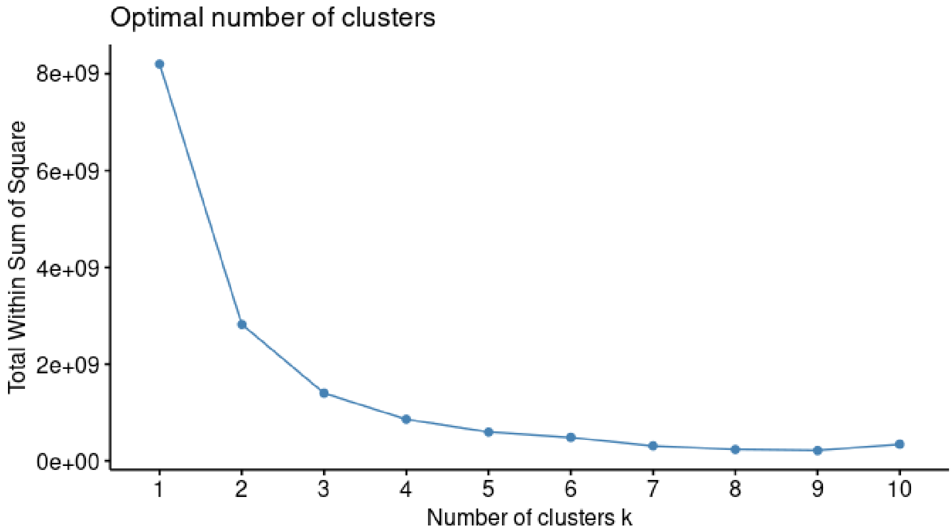

Figure 9: Elbow Method

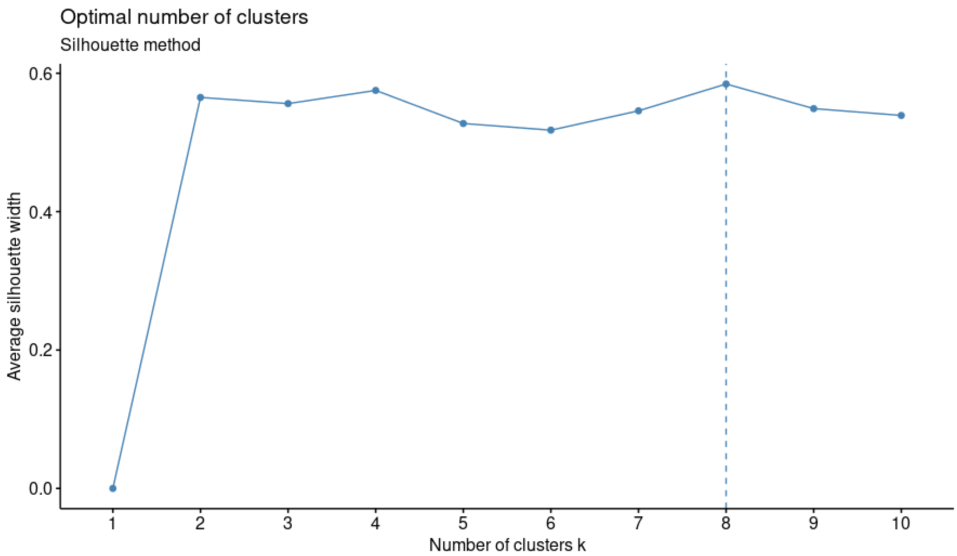

Figure 10: Silhouette Method

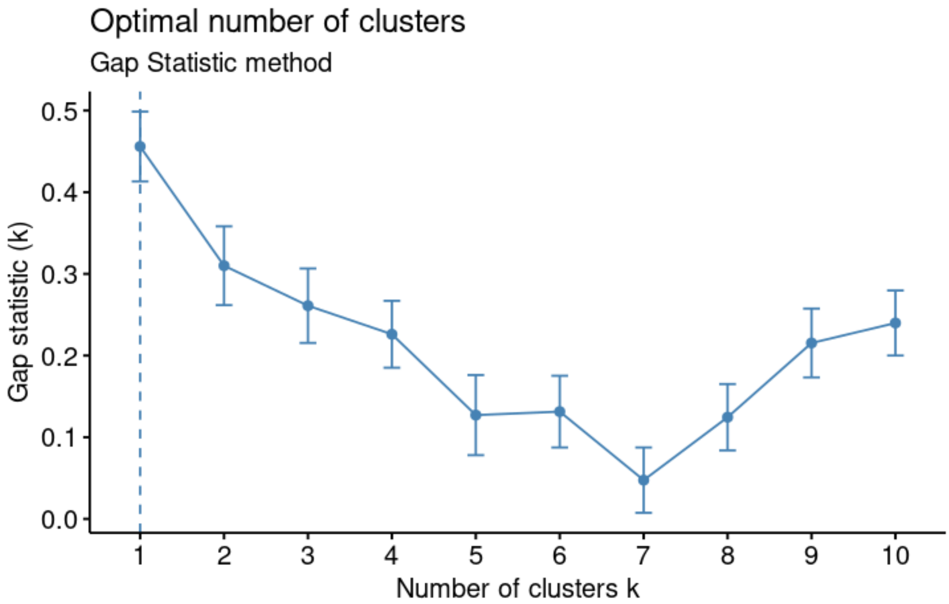

Figure 11: Gap Statistics

The most convenient result was returned by the elbow method, hence why, two clusters were implemented. This method was considered the most appropriate because it uses the within-cluster sum of squares (wss) to determine the optimal number of clusters. The elbow point, where the rate of decrease sharply slows, indicates a suitable number of clusters. Comparing it with other methods like silhouette and gap statistics, the elbow method provided the most coherent and interpretable segmentation of the data, showing a clear distinction in the reduction of wss, which aligns well with the observed data distribution and variance.

In **Figure 12**, it can be seen that the two clusters are heavily intersected, showing a minimal distinction between individuals.

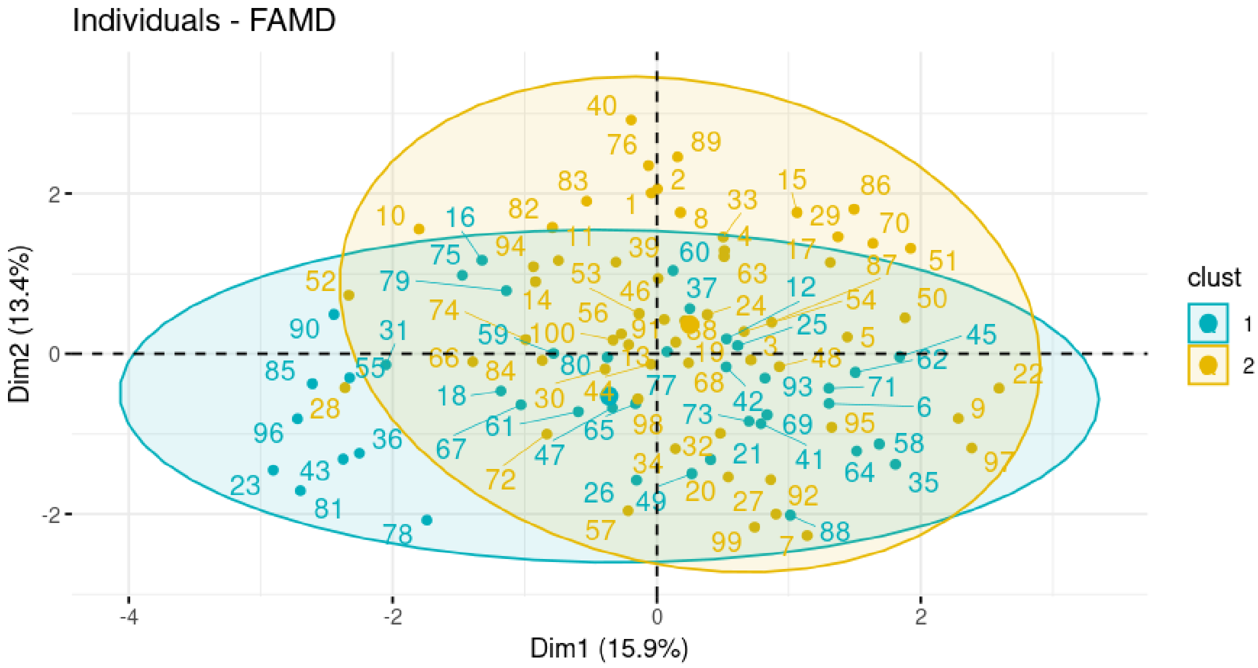

Figure 12: Cluster Individuals FAMD

Figure 13 displays multiple boxplots through which we can see the differences between each cluster. It can be seen that clusters don't have strong differences, however the most prominent are seen in the variables preference_taste, age, and income.

In cluster 2, the median of age is around 10 years higher than cluster 1, however the individuals from cluster 1 seem to have a higher income than those who belong to cluster 2. The individuals in cluster 2 tend to have a slightly higher preference for the product's taste in comparison to those from cluster 1.

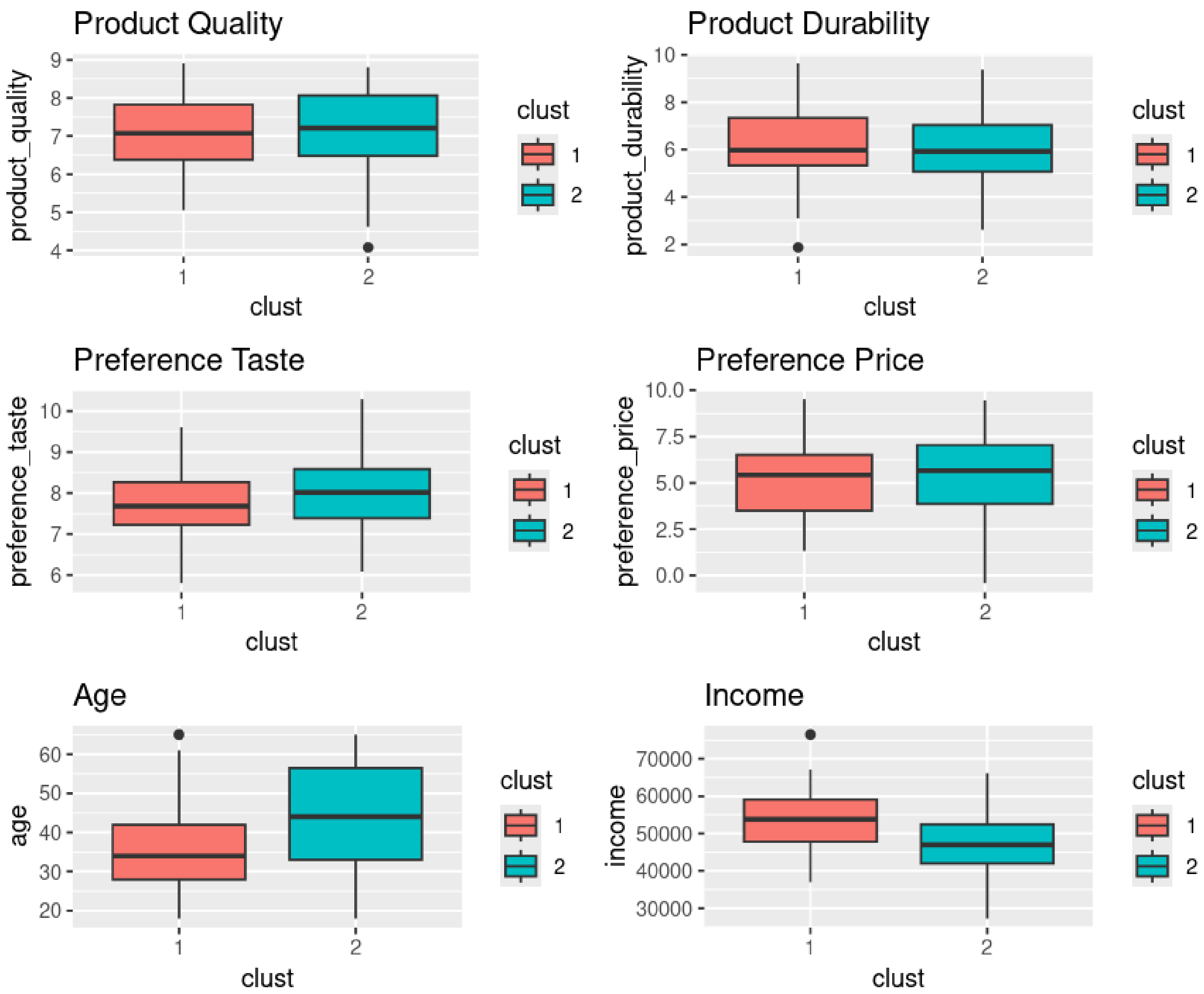

Figure 13: Variables by cluster

Attribute	Group 1 Mean	Group 2 Mean	p-value	95% CI
Product Quality	6.975	7.140	0.4701	[-0.617, 0.287]
Preference Taste	7.757	8.020	0.1465	[-0.621, 0.094]
Preference Price	5.105	5.338	0.5798	[-1.070, 0.602]
Product Durability	6.158	6.020	0.6663	[-0.496, 0.772]
Income	53158	47625	0.002958	[1941, 9125]
Age	36.2	44.0	0.004307	[-13.100, -2.510]

Table 6: Results of Welch Two Sample t-test for various attributes between clusters

In addition to this, a t-test analysis was performed, where the following results were obtained for various consumer attributes across the identified clusters. As shown in **Table 6**, the analysis revealed that for product quality, the t-test resulted in a p-value of 0.4701, indicating no significant difference between the clusters, with mean values of 6.975 for group 1 and 7.140 for group 2. The 95% confidence interval (CI) for the difference in means was [-0.617, 0.287], which includes zero, further suggesting no significant difference.

Similarly, for preference taste, the p-value was 0.1465, also showing no significant difference, with mean values of 7.757 for group 1 and 8.020 for group 2. The 95% CI was [-0.621, 0.094], again including zero, indicating a lack of significant distinction. The preference price attribute yielded a p-value of 0.5798, with means of 5.105 and 5.338 for groups 1 and 2, respectively, and a 95% CI of [-1.070, 0.602], supporting the absence of distinct segmentation.

For product durability, the p-value was 0.6663, with nearly identical means of 6.158 and 6.020, and the 95% CI for the difference was [-0.496, 0.772], indicating no significant distinction between the clusters. However, significant differences were found in income and age, with p-values of 0.002958 and 0.004307, respectively. The mean income for group 1 was 53,158 compared to 47,625 for group 2, with a 95% CI of [1941, 9125]. The mean age was 36.2 years for group 1 and 44.0 years for group 2, with a 95% CI of [-13.100, -2.510].

Despite these differences in income and age, the lack of significant variation in the primary product-related attributes across clusters, as indicated by the p-values and confidence intervals, suggests that there are no strongly defined consumer segments in terms of product preferences. This uniformity across key variables implies that the identified clusters do not represent distinct consumer groups with unique preferences, thereby supporting the conclusion that the segments are not clearly differentiated.

4) How do demographic variables (age and income) influence consumer preferences and product perceptions?

From the biplot displayed on **Figure 7**, it can be seen that the Age variable has a stronger influence on Dimension 2, as it points upwards and strays away from the origin of the plot. It is somewhat distant from other variables, suggesting it might not have a strong direct correlation with product quality, durability, or preference price, which are more aligned with Dimension 1.

Age is in the general direction of preference_taste, but not closely aligned, this might suggest that older consumers have a slight preference for the taste of the product, but this relationship is not very strong. When it comes to product characteristics, the weak association with product quality and durability implies that age might not be a significant factor in how consumers perceive these characteristics.

In the same figure, it can be seen that the income variable has a rather weak influence on the other variables as it is positioned near the origin. Its direction is close to preference taste, which may indicate an association between these two variables, indicating that higher income individuals might have a slightly stronger preference for the taste of the product. In the same way, its direction also indicates that there is a negative correlation between income and preference_price, which would suggest that people with higher income wouldn't really care about the product's price.

However, given the small influence of income, it's not strong enough to confirm any of these suggestions, leading to affirm that the proximity to the origin indicates that income does not have a strong influence on perceptions of product quality, durability, or preference price.

This lack of influence of income on product quality and preference price is supported by the low correlations between these variables displayed on the correlation matrix of **Table 5**

5) What insights can be drawn from the qualitative variables (gender and education level) in relation to product quality and durability?

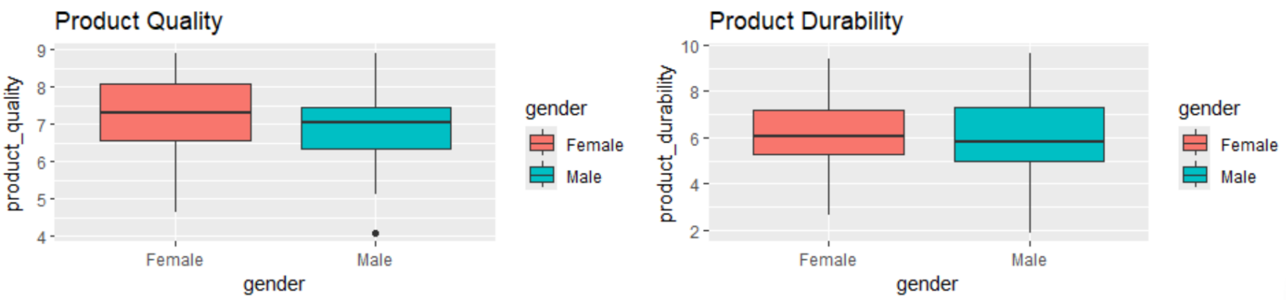

Figure 14: Perceived Product Characteristics by Gender

For Product Quality, the data indicates that women perceive product quality as slightly superior compared to men. The median rating for women is higher, although the dispersion of their data is slightly greater. In contrast, men show less variability in their ratings, with a lower median. In addition, there is a visible outlier for men, which could be due to a significantly negative experience with the product or expectations that are very different from the common ones. This outlier could indicate a case where a male rated the product quality much lower than the rest of his group, which deserves further exploration to understand the reasons behind this extreme perception in order to enhance customer satisfaction.

In terms of Product Durability, women also tend to give higher ratings compared to men, although less noticeable in this case. The median for women is slightly higher, and the distribution of ratings shows a tendency to rate product durability higher. Men's perception of durability is a little more varied and with a slightly lower median, but even so, the difference between genders is less pronounced than for Product Quality.

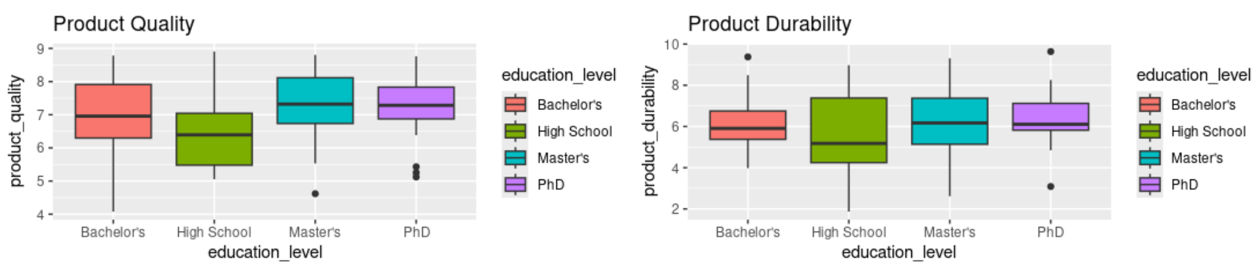

Figure 15: Perceived Product Characteristics by Education Level

In relation to Product Quality, the analysis reveals several notable differences across education levels, as shown in **Figure 15**.

Individuals with a Bachelor's degree perceive product quality as relatively high, indicated by a high median and low dispersion in their ratings. This suggests a consistent perception of product quality within this group. In contrast, individuals with a High School education have the lowest median rating and exhibit greater variability in their assessments, reflecting diverse expectations and experiences.

For those with a Master's degree, the median rating is high, similar to Bachelor's, but with a slightly wider range of opinions, as indicated by the presence of an outlier. This suggests that while the overall perception is positive, there are occasional significant deviations in experience. Individuals holding a PhD degree also show a high median with minimal dispersion, indicating a uniform and positive perception of product quality.

Regarding Product Durability, individuals with a Bachelor's degree tend to provide moderate median ratings, accompanied by slight variability. The High School group exhibits the most variability in durability ratings, along with a lower median, indicating a wide range of opinions on the product's durability. This might point to a diverse set of expectations or varying levels of product satisfaction.

Those with a Master's degree show a relatively high median rating and moderate dispersion, similar to their assessments of product quality. Meanwhile, PhD holders demonstrate a high median and low dispersion in their ratings, suggesting a consistent perception of product durability among this highly educated group, excluding some isolated cases that are represented by outliers for both very high and low ratings.

These suggest that individuals with higher education levels (Master's and PhD) tend to have more uniform and positive perceptions of both product quality and durability. In contrast, those with a High School education show greater variability in their perceptions, indicating a broader range of expectations and experiences with the products. These findings can guide marketing and product development strategies, particularly in targeting and improving perceptions among groups with more varied opinions.

6) Discuss the contributions of quantitative and qualitative variables to the main dimensions. Which variables are most influential?

In Dimension 1, **product_quality**, **education_level**, and **preference_price** are the variables that contribute the most, with 29.02%, 28.81%, and 15.39%, respectively (as shown in **Table 7**). This indicates that the perception of product quality, education level, and price play crucial roles in the variability observed in this dimension. The high contribution of **product_quality** suggests that this variable is particularly important in the first dimension, highlighting its relevance in the overall evaluation of products.

On the other hand, in Dimension 2, the most influential variable is **age**, with a contribution of 40.54%. This suggests that the age of individuals is a determining factor in this dimension, possibly reflecting generational differences in product preferences and perceptions. In addition, **education_level** also contributes significantly with 25.91%, as shown in **Figure 16**, indicating that educational background plays a notable role in shaping preferences, although to a lesser extent compared to age.

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
product_quality	29.02	0.03	3.53	1.26	0.45
product_durability	6.01	3.92	0.27	28.18	26.24
preference_taste	4.71	8.08	0.00	15.88	15.82
preference_price	15.39	10.72	2.06	0.69	30.21
age	0.66	40.54	1.35	1.32	4.12
income	3.79	2.09	43.26	2.76	7.57
education_level	28.81	25.91	48.06	39.76	15.54
gender	11.62	8.70	1.47	10.15	0.04

Table 7: Variable contributions

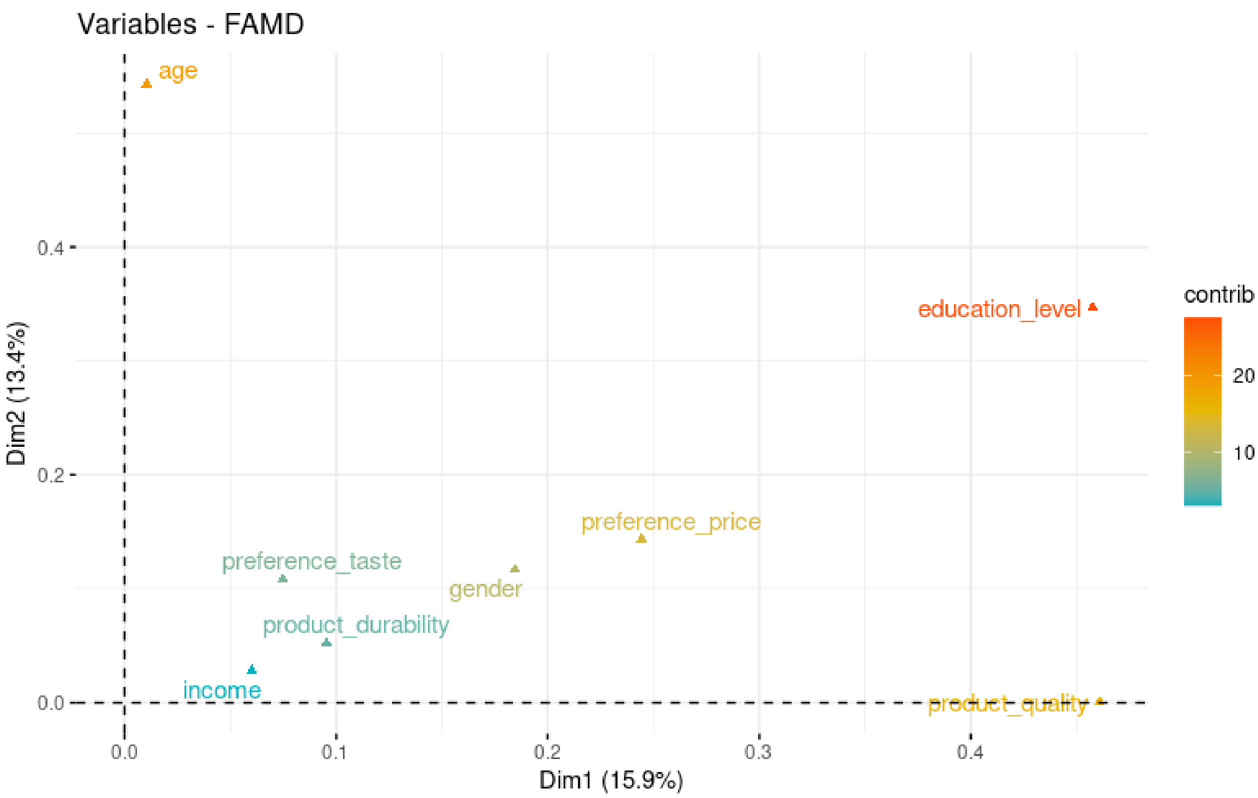

Figure 16: Variable contributions graph

7) How can the company use these insights to improve their marketing strategies and product development?

The company can use these insights to optimize its marketing and product development strategies by focusing on the most influential factors in consumer perception. Since product_quality and preference_price are the most prominent variables in Dimension 1, it is suggested that the company should emphasize these aspects in its advertising campaigns. For example, they could highlight the high quality of their products and offer promotions that emphasize value for money, targeting consumers who prioritize both quality and cost when making purchasing decisions.

In Dimension 2, the age variable emerges as the most influential factor, followed by preference_price. This indicates that the age of consumers plays an important role in determining their preferences and perceptions. The company can leverage this information to segment its market more effectively by creating customized products and messages for different age groups. For example, products that emphasize advanced technology or modern design might appeal to younger consumers, while those that emphasize durability and value for money might appeal to older consumers.

In addition, the consistent impact of preference_price on both dimensions suggests that price is a cross-cutting factor that influences diverse demographics. The company could consider strategies such as introducing different product lines at different prices to meet the needs of a broader spectrum of consumers. This would not only improve customer satisfaction, but also allow the company to capture a larger market share by catering to a variety of preferences and price sensitivities.

The company should also keep in mind that income is not a crucial factor in the perception of the products, acknowledging that a campaign that segments clients by their income would not be the most effective strategy.

8) What are some potential limitations of this study and the MFA method in analyzing consumer preferences?

Despite the valuable insights provided by Multiple Factor Analysis (MFA), there are several limitations to consider in both the study and the method used. First, MFA assumes that linear relationships between variables adequately capture the characteristics of the data, which may not be true if there are nonlinear or complex relationships between variables. This may lead to an incomplete or biased interpretation of the underlying factors influencing consumer preferences.

In addition, the method relies heavily on the variance of the variables, which means that variables with higher variance may dominate the main dimensions, as was observed with the significant contribution of variables such as product_quality and age. This may obscure the importance of other variables that could be relevant but have lower variance in the data. Also, the use of qualitative variables in MFA may result in difficulties in interpreting the dimensions, as the quantification of these variables may not fully capture the subtleties of consumer opinions or attitudes.

Another limitation is that the study is based on consumer self-reported data, which may introduce biases such as social desirability bias or recall errors. Participants might have provided answers that they consider socially acceptable or not accurately recall their experiences and perceptions. This could influence the results and thus the conclusions derived from the analysis.

9) Propose additional variables or data that could be included in future analyses to gain a more comprehensive understanding of consumer behavior.

The variables proposed for future analysis in the **Table 8** have been created considering the nature of the initial data set, which appears to be composed of customer survey responses. These variables are derived from questions directed to consumers about their perceptions and experiences with products, suggesting that they are self-reported data. Given this context, the new variables were designed to capture additional aspects that might not have been addressed in the original dataset, but are relevant to better understand consumer behavior.

Variable	**Description**
country	Country or region of residence of the customer.
purchase_site	A qualitative variable describing where the customer purchased the product (website, catalog, store, etc.).
marital_status	Customer's marital status.
sustainability_importance	Importance placed on sustainability, with higher values indicating a greater concern for sustainable practices.
purchase_frequency	Frequency of product purchases, indicating how often a consumer buys products.
average_spending	Average spending per purchase in dollars, indicating the consumer's spending level.
customer_satisfaction	Overall satisfaction with the product, rated on a scale from 1 to 10.
net_promoter_score	Likelihood of recommending the product to others, rated from 1 to 10.
advertising_influence	Degree to which advertising influences the consumer's purchasing decisions.
peer_recommendations	Impact of recommendations from friends and family on purchasing decisions.

Table 8: Proposed Variables

10) How could the company integrate these findings into their customer relationship management (CRM) systems for personalized marketing?

Integrating the findings from the analysis into the company's Customer Relationship Management (CRM) systems can significantly enhance the effectiveness of personalized marketing strategies. By utilizing the detailed insights obtained from variables such as product quality, durability, taste preferences, price sensitivity, age, and income, the company can create more targeted and relevant marketing campaigns.

Firstly, the CRM system can segment customers based on their preferences for product quality and durability. For example, customers who prioritize high-quality products can be targeted with promotions that highlight the premium features and superior quality of new or existing products. Conversely, those who are more price-sensitive can receive tailored offers that emphasize value-for-money and special discounts. This segmentation allows the company to communicate more effectively with different customer groups, increasing the likelihood of engagement and conversion.

Additionally, demographic data such as age and income can be leveraged to further personalize marketing efforts. Younger customers, who may be more inclined to try new and innovative products, can be targeted with campaigns that showcase the latest product lines or technological advancements. Meanwhile, older customers, who might prefer reliability and value, can be approached with messages emphasizing product durability and long-term savings. Moreover, income data can help in crafting offers that are financially appealing to specific customer segments, such as installment plans or bundled discounts.

Furthermore, incorporating psychographic data, such as sustainability, can refine marketing strategies. Customers who value sustainability can be targeted with eco-friendly product lines and messaging that aligns with their values.

Overall, by integrating these findings into their CRM systems, the company can deliver highly personalized marketing messages that resonate with individual customers preferences and characteristics. This personalized approach enhances customer satisfaction and also fosters loyalty and increases sales by addressing the specific needs and desires of each customer segment.

7 Conclusion

In conclusion, the analysis conducted throughout this work has allowed us to explore consumer preferences and product characteristics, using techniques such as Multiple Factor Analysis (MFA) and cluster analysis. Through MFA, key relationships between variables were identified, as well as demographic differences in income and age between clusters. However, the absence of significant differences in other main attributes indicates that there are no clearly differentiated consumer segments.

Importantly, during the analysis, it was noted that the data appeared to be randomly generated, evidenced by the presence of unusual values and some inconsistencies that did not make practical sense. This random data presented certain challenges and hindered the analytical process, affecting the depth and accuracy of the conclusions that could be drawn.

Nevertheless, a quality analysis was achieved, demonstrating the capability of the methodologies employed to reveal relevant patterns and provide valuable insight into the data, even under less than ideal conditions. This work highlights the importance of reliable and accurate data for more robust and actionable insights in the field of consumer analytics.

References

[1] B. Escofier and J. Pages. Multiple factor analysis (afmult package). *Computational Statistics & Data Analysis*, 18(1):121–140, 1994.

[2] A. KASSAMBARA. *Practical Guide To Principal Component Methods in R: PCA, M(CA), FAMD, MFA, HCPC, factoextra.* Multivariate Analysis. STHDA, 2017.