For cited: *Yıl, O. V., Kadıoglu, O., Tuncer, Ö. & Cebecı, U. (2021). Applying Data Mining Techniques to Analyze Different Generation Customers' Preferences for Buying Gold Necklaces. Journal of Current Research on Engineering, Science and Technology, 7 (1), 93-104.*

Journal of Current Research on Engineering, Science and Technology (JoCREST)
ISSN: 2651-2521

Year: 2021
Volume: 7
Issue: 1

Crossref doi: 10.26579/jocrest.75

Research Article/Araştırma Makalesi

Applying Data Mining Techniques to Analyze Different Generation Customers' Preferences for Buying Gold Necklaces

Osman Volkan YIL[1], Oguzhan KADIOGLU[2], Özgenur TUNCER[3] & Ufuk CEBECI[4]

Keywords
Generation, Data Mining, Gold Jewelry, Gold Necklaces.

Abstract
Increasing competition in the jewelry market, expensive raw materials, and high cost of labor bring the demands and requests of the customers to a significant level. At the same time, meeting these expectations ensures that the companies are more sustainable and increase their profits. Companies should determine their vision and strategies by analyzing preferences and design products suitable for users.
This study aims to determine the relationship and correlation between the factors such as brightness, size, color, karat, design, and weight that affect consumer preferences and behaviors of Y and Z generations in the Turkish gold jewelry market using data mining techniques. This research leads to a better understanding of how various factors affect the purchasing behavior of generations and enables companies in the jewelry industry to determine their vision and brand strategy or other strategies. The data of this study which consists of gold necklaces preferences were obtained from a survey conducted with 243 participants from Turkey in 2021. Results and findings are explained in this study.

Article History
Received
15 June, 2021
Accepted
30 July, 2021

1. Introduction

Throughout history, precious metals have had an important place for societies. Humankind used these precious metals while shopping as trading materials, and sometimes as jewelery. With the acceleration of technology and the world's economic wealth, buying jewelry started to be a part of people's lives more and more. This participation affects mostly gold items due to their high value, opportunities of investment, and the perception of the beauty of gold jewelry formed in the society. Just like all products that have been trading in the market, different consumers have developed preferences for gold jewelry products, also.

In order to develop a better understanding of consumers and how they behave while making a purchasing decision, companies consider customer preferences.

[1] Corresponding Author. ORCID: 0000-0001-5724-7678. Istanbul Technical University, yil15@itu.edu.tr
[2] ORCID: 0000-0001-6101-1481. Istanbul Technical University, kadioglu15@itu.edu.tr
[3] ORCID: 0000-0003-3440-2561. Goldenline Bogazici Hediyelik Esya ve El Sanatları San. Ve Tic. A.S, ozgenurtuncer@goldenlinejew.com
[4] ORCID: 0000-0003-4367-6206. Istanbul Technical University, cebeciu@itu.edu.tr

For the purpose of setting business strategies and visions, businesses are developing considerable efforts to keep up-to-date knowledge on their customer preferences, which is critical to overcoming problems (Boccia and Sarnacchiaro, 2018). Analyzing and understanding customer preferences arising from different circumstances will help companies get a better market position against their competitors. The generational approach is an effective way to analyze the impact of changes on purchases. Variations of economic and social differences such as technology activities, economic conditions, life perceptions, opportunities, expectations, and social norms are witnessed by different generations (Heaney, 2007). To analyze the effects of these differences in the modern world, the main focus should be on Generation Y and Generation Z because their population and consumption percentage in the world is greatly high. Also, this will help companies to get a grasp on changing trends by population groups. With the help of an increasing number of data mining techniques, which can be useful to make relevant comments on differences as well as similarities, analysis can be done effectively.

2. Material and Method

2.1 Material

There are many different models in gold necklace models for consumer preferences and designers' creativity. These models emerge due to the different qualities and quantities of the product. Many models available provide a choice for customers. In order to determine the purchasing preferences of gold necklaces, the gold necklace models produced by the "Goldenline Jewelry and Giftware Industry and Trade Inc." company were examined and the factors that would affect the purchasing decision were determined in line with these models and the opinions of the employees. Accordingly, the factors to be evaluated in the study were determined as size, color, carat, design, and weight.

2.2. Method

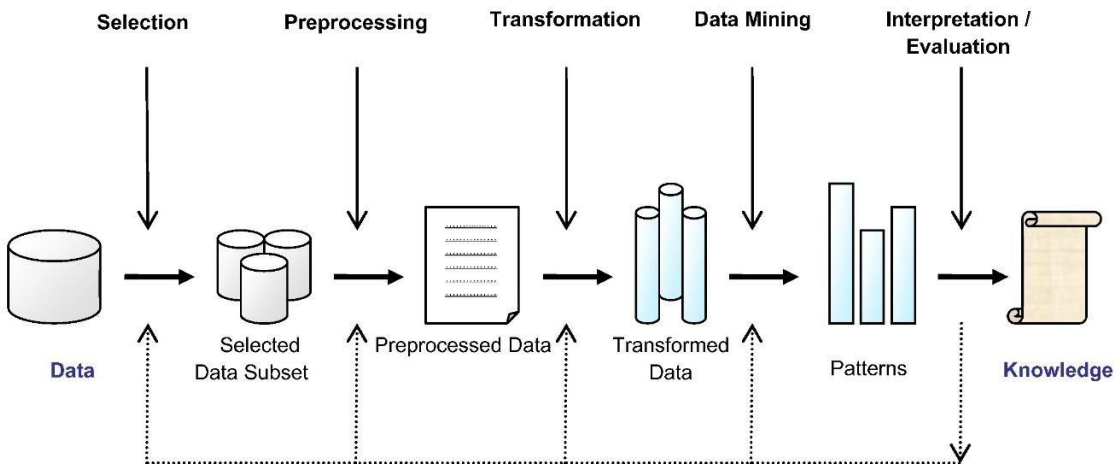
There is an abundant flow of information in every aspect of life. To begin with, data means collected information. Data is usually stored in electronic devices and it is mostly including facts and numbers to make precise decisions, comments, and developments in related fields. Data is a unit of information. After evaluating the data in a meaningful context, it transforms into information. Data can be measured, analyzed, and visualized using programming tools to get a grasp on what sort of favorable information lies beneath it. Unprocessed data is called raw data. After analyzers remove undesired data to minimize error-causing outliers, they make clear-cut judgments on the remaining data.

As information technology develops, analyzing data becomes more and more important. One remarkable technique of doing this analysis is data mining. Data mining is the instrument for detecting relationships and correlations in vast data sets (Hand and Adams, 2014). Vast data sets are also called databases. Researchers deal with larger datasets every day as a result of the development in information technology.

Information and technology, and database systems have developed from basic file processing to more complex database systems starting from the 1960s. After the 1970s, affiliative database systems have been developed to make meaningful comments on raw data. Starting in the late 1980s, huge progress in computer technology has led to establishing new methods of gathering data as well as storing data. Data analysis now can be used by applying a huge number of methods (Han et.al., 2011).

To get knowledge from databases, the term "Knowledge discovery in databases (KDD)" arose in the literature (Fayyad et. al., 1996). It is used by analyzers to make meaningful discussions on data by applying different methods as clearly seen in Figure 1. First of all, raw data is collected. Then, considering the problem and need, a selection is made from this data. After preprocessing the selected data subset and making a transformation process, which is transforming the dataset into a form where data mining can be applied, selected Data mining techniques are applied. Patterns are found to get knowledge from the data.

Figure 1: KDD process (Kavakiotis et. al., 2017).



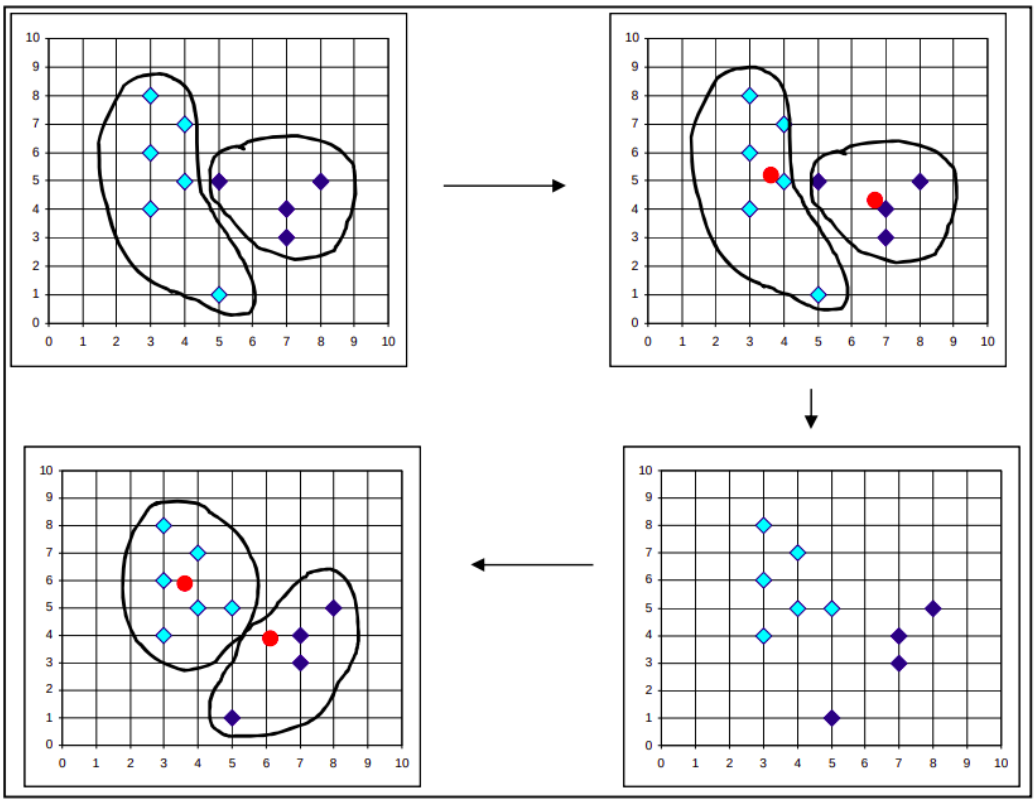
Data mining is also highly used by marketers to get a grasp on consumers' thoughts (Wang, 2013). Data mining is used for analyzing correlations on shopping habits of consumers, especially in discount seasons, trend changing times. It helps marketers to detect relations between different products and how consumers are approaching different trends and develop habits that can be detected as well as specifying stock and inventory numbers based on these relations (Raorane and Kulkarni, 2011).

Clustering is also known as data segmentation is the process of grouping or segmenting objects into subset or "clusters". These objects are similar between them and are dissimilar to the objects that belong to other clusters. (Hastie et. al., 2009). The purpose of clustering is based on the principle of "maximizing the intraclass similarity and minimizing the interclass similarity" (Han et. al., 2011). Clustering is a data mining technique that groups project objects. As opposed to classification, groups are not predetermined in this method. The best form of clustering that can be created occurs at the end of the analysis. Several clustering types have been applied in the literature, such as k-means clustering, k-medoids

clustering, hierarchical clustering, and spectral clustering. Each method can be distinguished from others by its weaknesses and strengths.

K-means, one of the oldest and most widely used clustering algorithms, was developed by J.B. MacQueen in 1967 (MacQueen, 1967). The K-Means algorithm is an unsupported learning and clustering algorithm. The value k in K-Means determines the number of sets and must take this value as a parameter. K-means are called because K forms the original set and the center of each set is the average of the values in the set. The algorithm puts records of a statistically similar nature into the same group. An element is only allowed to belong to a set. The value in the cluster center represents the cluster. Figure 2 shows the working principle of K-means clustering.

Figure 2: Working principle of K-means clustering algorithm (Shukla and Naganna, 2014)

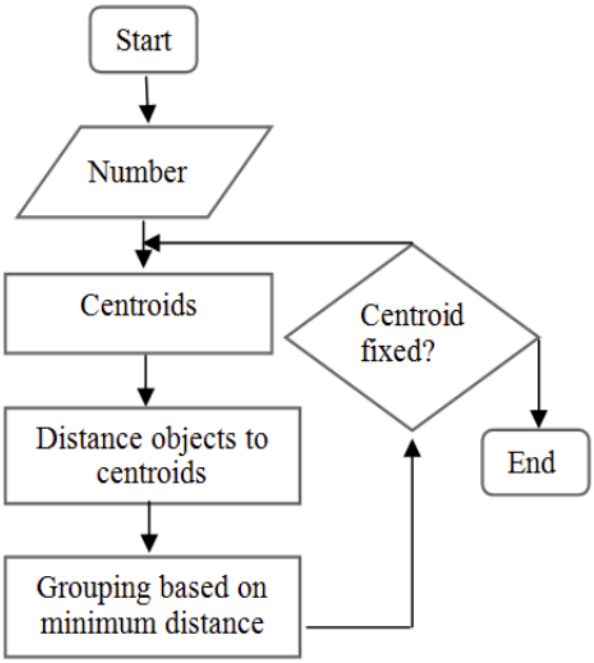


Stage of K-Means methods are:

- Select 'k' number of clusters #1
- Calculate centroid of k clusters #2
- • Compute Euclidean distance of each object in the data-set from each of the centroids #3
- Assign each object to the cluster that it is nearest to based on the distances calculated in the previous step. #4
- Calculate the centroid of each cluster again #5

- Reassign the objects to clusters based on the least distance if no change then go to 7 or else go to 5 #6
- Stop #7 (Saraiya and Ganage, 2018)

Figure 3: Flowchart of K-means Algorithm (Sarker et.al, 2018)



3. Application

There are so many features to consider when designing gold necklaces. Customers need to think about these features before making a purchasing decision, as well as designers. Six of the features have been chosen as a part of this research, which are brightness, size, color, karat, design, and weight. A questionnaire was conducted with 243 people across Turkey. 189 of survey responders belong to Generation Y, whilst 54 of them belong to Generation Z. Questionnaire has been built using Google Forms. Mainly, social media platforms have been used to reach people. They were asked to evaluate the importance of brightness in regarding other factors, with coefficients of "9, 7, 5, 3, 1, 1/3, 1/5, 1/7, 1/9".

- 9: Factor 1 is 9 times more important according to Factor 2.
- 7: Factor 1 is 7 times more important according to Factor 2.
- 5: Factor 1 is 5 times more important according to Factor 2.
- 3: Factor 1 is 3 times more important according to Factor 2.
- 1: Factor 1 is equally important according to Factor 2.
- 1/3: Factor 2 is 3 times more important according to Factor 1.
- 1/5: Factor 2 is 5 times more important according to Factor 1.
- 1/7: Factor 2 is 7 times more important according to Factor 1.

- 1/9: Factor 2 is 9 times more important according to Factor 1.

To prevent getting different coefficients from the same factors, brightness had been chosen as the constant factor. The questionnaire and sample answers can be seen in below Table 1.

Table 1: Conducted questionnaire with sample answers

Date of Birth	Evaluate the importance of the brightness factor according to the size factor.	Evaluate the importance of the brightness factor according to the color factor.	Evaluate the importance of the brightness factor according to the karat factor.	Evaluate the importance of the brightness factor according to the design factor.	Evaluate the importance of the brightness factor according to the weight factor.
1980-1999	3	1/3	1/5	1/9	9
1980-1999	5	3	1/5	1/9	9
2000 and latter	7	1	3	1/3	3
1980-1999	1	3	5	1/9	1
2000 and latter	1	5	1/3	9	1

After getting the results, a weighted score is calculated for each survey responder. Every factor is ratioed according to its coefficient. By doing that, every responder's coefficient sum was made equal to 1 as seen in Table 2.

Table 2: Weighted score of each factor

Generation	Weighted Brightness	Weighted Size	Weighted Color	Weighted Karat	Weighted Design	Weighted Weight
Gen Y	0.05	0.02	0.16	0.27	0.49	0.01
Gen Y	0.06	0.01	0.02	0.32	0.58	0.01
Gen Z	0.17	0.02	0.17	0.06	0.52	0.06
Gen Y	0.08	0.08	0.03	0.02	0.72	0.08
Gen Z	0.16	0.16	0.03	0.48	0.02	0.16

Then, using SPSS 26.0 Statistical Software, descriptive statistics are calculated for Generation Y (Table 3), and Generation Z (Table 4). Means of Y data lie between 0.1 and 0.28, while standard deviations are between 0.11 and 0.18. In addition, means of Z data lie between 0.11 and 0.28, while standard deviations are between 0.11 and 0.19. Z-values are stored as variables for further calculations.

Table 3: Descriptive statistics for Generation Y Data

	N	Minimum	Maximum	Mean	Std. Deviation
Weighted Brightness	189	0.0217	0.6429	0.1365	0.1650
Weighted Size	189	0.0039	0.5068	0.1022	0.1106
Weighted Color	189	0.0030	0.8591	0.1248	0.1366
Weighted Karat	189	0.0046	0.8412	0.2187	0.1735
Weighted Design	189	0.0039	0.8217	0.2876	0.1894
Weighted Weight	189	0.0039	0.7616	0.1302	0.1283
Valid N (listwise)	189				

Table 4: Descriptive statistics for Generation Z Data

	N	Minimum	Maximum	Mean	Std. Deviation
Weighted Brightness	54	0.0283	0.5833	0.1503	0.1622
Weighted Size	54	0.0046	0.6402	0.1174	0.1421
Weighted Color	54	0.0055	0.4655	0.1367	0.1264
Weighted Karat	54	0.0039	0.7836	0.1825	0.1783
Weighted Design	54	0.0116	0.8232	0.2887	0.1982
Weighted Weight	54	0.0036	0.4451	0.1245	0.1171
Valid N (listwise)	54				

K-means clustering method was applied to Z-values of each factor using SPSS. Cluster numbers are determined as two for both generations' data. Initial cluster centers are shown in Table 5 and Table 6.

Table 5: Initial cluster centers for Generation Y

Gen Y	Cluster	
	1	2
Zscore: Weighted Brightness	-0.24861	0.09594
Zscore: Weighted Size	-0.82791	-0.77079
Zscore: Weighted Color	5.37631	-0.79016
Zscore: Weighted Karat	-1.18191	-1.13509
Zscore: Weighted Design	-1.46251	-1.35766
Zscore: Weighted Weight	-0.93219	4.92157

Table 6: Initial cluster centers for Generation Z

Gen Z	Cluster	
	1	2
Zscore: Weighted Brightness	-0.38975	-0.13710
Zscore: Weighted Size	-0.70370	3.68071
Zscore: Weighted Color	-0.98276	-0.74356
Zscore: Weighted Karat	3.37200	-0.92089
Zscore: Weighted Design	-1.01706	-1.24093
Zscore: Weighted Weight	-0.95682	0.03042

After 7 iterations for Generation Y (Table 7), and 4 iterations for Generation Z (Table 8), final cluster centers are reached.

Table 7: Iteration history for Generation Y

Iteration	Change in Cluster Centers	
	1	2
1	4.726	5.097
2	0.319	0.195
3	0.196	0.141
4	0.094	0.077
5	0.081	0.068
6	0.039	0.035
7	0.000	0.000

Table 8: Iteration history for Generation Z

Iteration	Change in Cluster Centers	
	1	2
1	3.401	3.395
2	0.212	0.246
3	0.109	0.152
4	0.000	0.000

4. Results and Discussion

Final cluster centers are reached after final iterations (Table 9 and Table 10).

Table 9: Final cluster centers for Generation Y

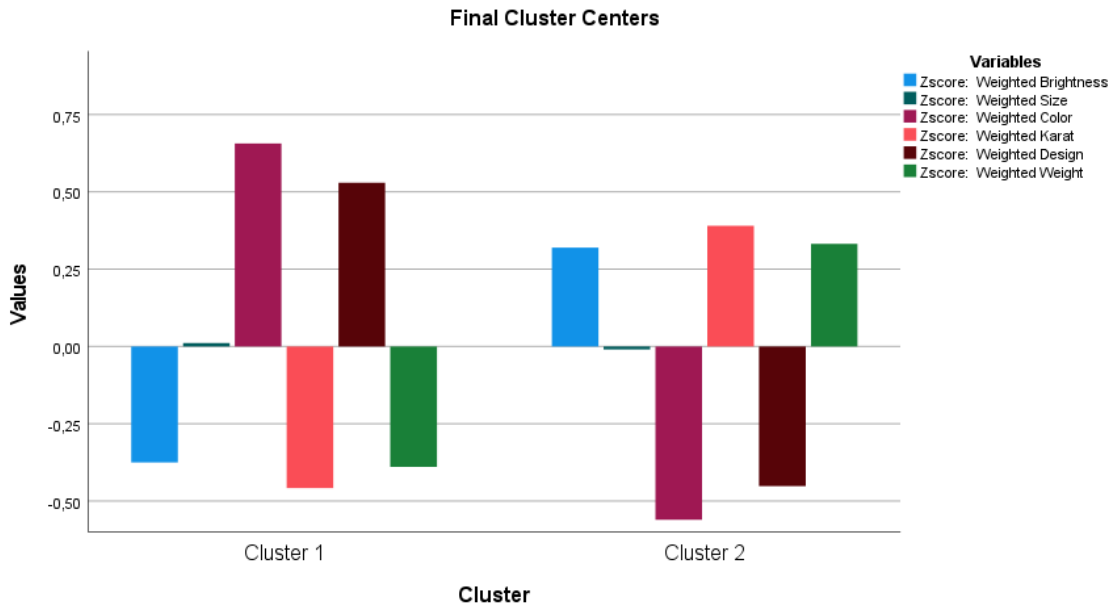
Gen Y	Cluster	
	1	2
Zscore: Weighted Brightness	-0.37505	0.31989
Zscore: Weighted Size	0.01125	-0.00959
Zscore: Weighted Color	0.65686	-0.56026
Zscore: Weighted Karat	-0.45768	0.39037
Zscore: Weighted Design	0.52953	-0.45166
Zscore: Weighted Weight	-0.38919	0.33195

Table 10: Final cluster centers for Generation Z

Gen Z	Cluster	
	1	2
Zscore: Weighted Brightness	-0.52351	0.70560
Zscore: Weighted Size	-0.44050	0.59372
Zscore: Weighted Color	0.11003	-0.14831
Zscore: Weighted Karat	0.43297	-0.58357
Zscore: Weighted Design	0.29604	-0.39901
Zscore: Weighted Weight	-0.01957	0.02637

For the purpose of seeing results more accurately, cluster centers are graphed using SPSS (Figure 4 and Figure 5).

Figure 4: Visualized final cluster centers for Generation Y



According to two clusters:

- For cluster 1, design and color are the most important features with the highest significance.
- For cluster 2, brightness, karat, and weight are the most important features but with less significance compared to cluster 1 results.
- Therefore, combining color and design is more important to address Generation Y consumers in terms of gold necklaces.

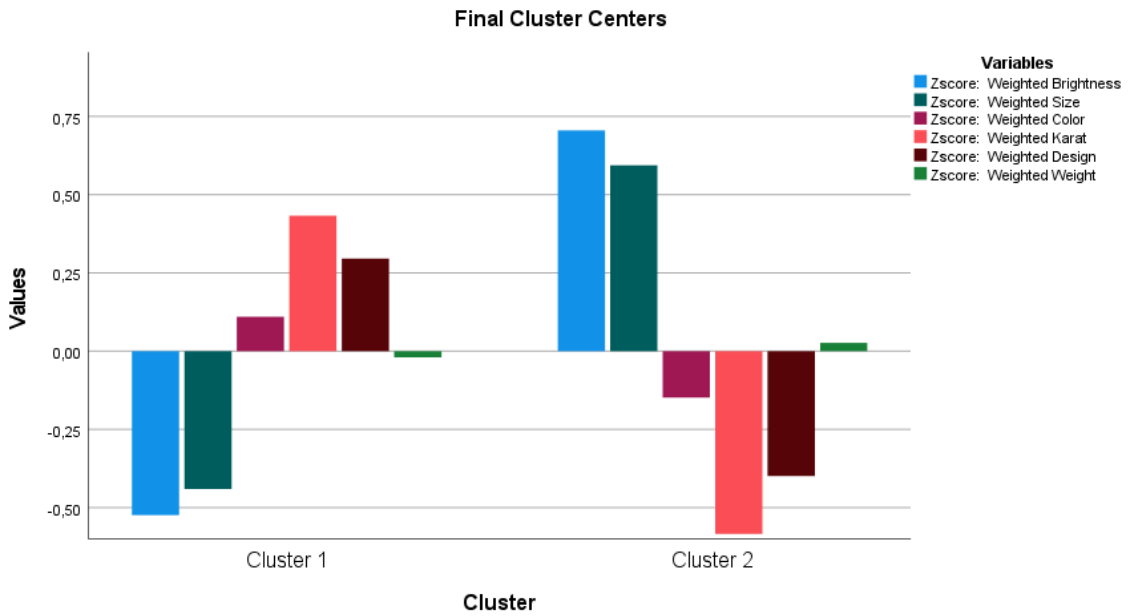
The ANOVA table for Generation Y results is below in Table 11.

Table 11: ANOVA table for Generation Y

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: Weighted Brightness	22.675	1	0.884	187	25.648	0.000
Zscore: Weighted Size	0.020	1	1.005	187	0.020	0.887
Zscore: Weighted Color	69.554	1	0.633	187	109.810	0.000
Zscore: Weighted Karat	33.768	1	0.825	187	40.942	0.000
Zscore: Weighted Design	45.203	1	0.764	187	59.196	0.000
Zscore: Weighted Weight	24.417	1	0.875	187	27.913	0.000

Brightness, color, carat, design, and weight factors are significant, while size factor is not. Color has the highest F-value, which shows that it is the most meaningful factor in terms of Generation Y customer preferences.

Figure 5: Visualized final cluster centers for Generation Z



According to two clusters:

- For cluster 2, brightness and size are the most important features with the highest significance.
- For cluster 1, karat and design are the most important features but with less significance compared to cluster 2 results.
- Therefore, combining brigthness and size is more important to address Generation Z consumers in terms of gold necklaces.

ANOVA table for Generation Y results are below in Table 12.

Table 12: ANOVA table for Generation Z

Gen Z	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Zscore: Weighted Brightness	19.947	1	0.636	52	31.381	0.000
Zscore: Weighted Size	14.123	1	0.748	52	18.890	0.000
Zscore: Weighted Color	0.881	1	1.002	52	0.879	0.353
Zscore: Weighted Karat	13.644	1	0.757	52	18.028	0.000
Zscore: Weighted Design	6.379	1	0.897	52	7.115	0.010
Zscore: Weighted Weight	0.028	1	1.019	52	0.027	0.869

Brightness, size, karat, and design factors are significant, while color and weight factors are not. Brightness has the highest F-value, which shows that it is the most meaningful factor in terms of Generation Z customer preferences. F-values of Karat and Size are relatively high compared to the remaining factors.

5. Conclusion

There is an abundant flow of information in every aspect of life. As information technology develops, analyzing data becomes more and more important. Data

mining provides the processing and analyzing the data, and the processed data allows us to access knowledge.

There are several data mining methods. While this research was being conducted, many of these methods were applied and it was observed that for the questionnaire data, k-means clustering was the most suitable, efficient, and accurate technique. K-means clustering is an unsupervised learning data mining technique that is widely used to cluster the elements of a set.

This study demonstrates the clusters between the factors such as brightness, color, karat, design, size, and weight that affect consumer preferences of Y and Z generations in golden necklaces by using data mining techniques. At the same time, this research helps to develop vision and strategies according to the customer spectrum. It ensures the production and marketing of optimum necklaces suitable for customers' preferences.

For further research, this study can be repeated for different industries and different product groups. Alongside with, the study focused on the effect of the generation factor on preferences, which can be changed or combined with different factors such as age, gender, education level, and occupation.

6. Acknowledge

We would like to thank the Goldenline Bogazici Hediyelik Esya ve El Sanatları San. Ve Tic. A.S. company for their support during the work.

References

Boccia, F., & Sarnacchiaro, P. (2018). The impact of corporate social responsibility on consumer preference: A structural equation analysis. **Corporate Social Responsibility and Environmental Management,** 25(2), 151-163.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. **AI magazine,** 17(3), 37-37.

Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. **The Morgan Kaufmann Series in Data Management Systems,** 5(4), 83-124.

Hand, D. J., & Adams, N. M. (2014). Data Mining. **Wiley StatsRef: Statistics Reference Online,** 1-7.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. **Springer Science & Business Media.**

Heaney, J. G. (2007). Generations X and Y's internet banking usage in Australia**.** **Journal of Financial Services Marketing,** 11(3), 196-210.

Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. **Computational and structural biotechnology journal,** 15, 104-116.

MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations**. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability** (Vol. 1, No. 14, pp. 281-297).

Raorane, A., & Kulkarni, R. V. (2011). Data mining techniques: A source for consumer behavior analysis. arXiv preprint arXiv:1109.1202.

Saraiya, P. R., & Ganage, Y. (2018). Study of clustering techniques in the data mining domain. **International Journal of Computer Science and Mobile Computing,** 7(11), 31-37.

Sarker, A., Shamim, S. M., Zama, M. S., & Rahman, M. M. (2018). Employee's performance analysis and prediction using K-means clustering & decision tree algorithm. **Global Journal of Computer Science and Technology.**

Shukla, S., & Naganna, S. (2014). A review on K-means data clustering approach. **International Journal of Information and Computation Technology,** 4(17), 1847-1860.

Wang, Y. (2013). Clustering, Classification, and Factor Analysis in High Dimensional Data Analysis.