Clustering of Customer Attitudes Towards Eco-Innovations - Evidence from Bulgaria

Georgi Marinov[1], Teodoro Gallucci[2], Vesselina Dimitrova[3] and Petyo Boshnakov[4]
[1) 3) 4)] *University of Economics, Varna, Bulgaria*
[2)] *University of Bari Aldo Moro, Bari, Italy*
E-mail: gmarinov@ue-varna.bg; E-mail: teodoro.gallucci@uniba.it
E-mail: vesselina.dimitrova@ue-varna.bg; E-mail: pboshnakov@ue-varna.bg

Please cite this paper as:

Marinov, G., Gallucci, T., Dimitrova, V. and Boshnakov, P., 2023. Clustering of Customer Attitudes Towards Eco-Innovations - Evidence from Bulgaria. In: R. Pamfilie, V. Dinu, C. Vasiliu, D. Pleşea, L. Tăchiciu eds. 2023. *9th BASIQ International Conference on New Trends in Sustainable Business and Consumption.* Constanța, Romania, 8-10 June 2023. Bucharest: ASE, pp. 274-279

DOI: 10.24818/BASIQ/2023/09/050

Abstract

The research objective of the paper is twofold: firstly, to investigate how the adoption of innovative practices in the wine industry from the environmental, social, and conventional points of view influences consumers' choices in Bulgaria, and secondly, what is the connection between eco-innovation and consumer attitudes. Based on the results from the previous research study, we calculate index values for different types of innovations and we use the k-means clustering procedure to explore consumers' attitudes towards eco-innovations in Bulgaria by determining an optimal number of clusters. Bulgarian young consumers (<26 age) are environmentally friendly and orientated towards eco-innovations in the wine industry with special emphasis on recycling (water, energy) and replacing materials. Our findings confirm other recent studies that preferences for eco-innovations in the wine industry are correlated with a willingness to pay more for organic wines. In the present paper, eco-innovation in the wine industry is studied from the demand side, which is traditionally neglected. To our knowledge, clustering analysis for customer attitudes regarding eco-innovations in the wine industry is applied for the first time in Bulgaria. The results of the study can be helpful for wine managers, technical specialists, and wine-producing companies to prioritize their green efforts for the youth generation in Bulgaria.

Keywords

Wine sector, eco-innovations, customers, hierarchical clustering, k-means, Bulgaria.

DOI: 10.24818/BASIQ/2023/09/050

Introduction

Eco-innovation is related to the aim to use fewer resources (saving water, energy etc.) and process impact on the environment (waste recycling, prevention of production pollution). This innovation is strongly related to environmental awareness and understanding from buyers and sellers. Eco-innovation can influence consumers' purchasing decisions and can improve the market position of enterprises because it motivates their economic and social performances. According to Egri and Herman (2000) and Sumrin et al. (2021), eco-innovation is keeping knowledge about consumers' behavior and other stakeholders regarding green trends in the economy.

The research interest is provoked by two aspects: firstly, to consider the increasing importance of environmental concerns for consumers in industries, closely related to sustainability as the wine industry, and secondly, to consider the increasing role of wine eco-innovations as a source of competitiveness. (Rabadan and Bernabeu, 2021) The research objective of the paper is twofold: firstly, to investigate how the adoption of innovative practices in the wine industry from an environmental, social, and conventional point of view influences consumers' choices in Bulgaria, and secondly, what is the connection between eco-innovation and consumer attitudes.

Literature review

Cluster analysis has been successfully adopted to address data clustering problems in different domains such as medical science, manufacturing, the financial sector, urban development, industries, sales, and marketing because of its flexibility (Ikotun et al., 2023). In literature, there are different variants of k-means algorithms such as batch k-means, incremental batch k-means (Forgy, 1965), online k-means (Linde, Buzo and Gray, 1980), and incremental online k-means (MacQueen, 1965).

For example, it is possible to highlight the use of k-means algorithm for color quantization in computer graphics (Abernathy and Celebi, 2022), or the application of k-means machine learning algorithm to create a solution by the synthetic data to enhance the detection of the geologic potential field-generated bodies (Eshimiakhe and Lawal, 2022). Other important studies deal with consumer analysis for instance, Tabianan, Velu and Ravi (2022), analyze the clusters such as event type, products, and categories to support vendors to identify the groups that share similar criteria focusing on the highly profitable segment to the least profitable segment. Higuchi and Maehara (2021) analyzed non-hierarchical clusters (NHCA) with the *k*-means method in the motivational profile of quinoa consumers in Modern Metropolitan Lima for highlighting any differences between consumers. Other research to mention is the use of k-means clustering algorithm based on the adaptive learning particle swarm optimization (ALPSO) algorithm for “developing a customer segmentation method to achieve the division of customer groups in the grape market in China”. (Li et al., 2021).

Considering literature in the wine industry, the k-means algorithm has been used for analyzing both quality and preferences. For instance, McCune et al. (2021) proposed a modified k-means algorithm to cluster the Bordeaux wine dataset based on the original k-means clustering. The analysis has the goal to allow the selection of a specific number of wines that vendors would like to propose to consumers that are more representative of their offerings. This approach provides insight ”*by grouping similar wines so that a vendor can make more informed decisions through the unsupervised learning*”. Katarya and Saini, (2022) analyzed how to improve wine tasting using primary component analysis (PCA) and k-means clustering algorithms to recommend wines. The recommendation approach has been useful to offer insights to all types of users, beginners, and regular wine drinkers to enhance their current preferences. From this non-exhaustive review, it emerges that the k-means algorithm is a very flexible and suitable tool for analyzing correlations between consumers and their choices to better understand how to anticipate consumer choices.

Methodology

The starting point of this study is based on the results of previous research on the innovations in the wine industry regarding an empirical study that used a questionnaire for managers and technical specialists in Bulgarian wine-producing companies and fuzzy analytical hierarchical process (fuzzy AHP) assessment (Boshnakov, Dimitrova and Marinov, 2022), further in the text the BDM index. The results are presented in three dimensions - conventional, ecological, and social, highlighting the assessment of each item. The results of the study showed that priority is given to conventional and eco-innovations, compared to social innovations in the wine sector, according to managers and technical specialists in Bulgarian wine-producing companies. Table no.1 presents the BDM results for all wine innovations.

In the present paper, we used the values of the BDM fuzzy analytical hierarchical process (fuzzy AHP) assessment to calculate the indexes, on which the cluster analysis for Bulgarian consumers is made. As an instrument, we use a survey, with core questions in Likert form, with a 5 degree scale (1 to 5 or -2 to +2, "strongly disagree" to "strongly agree").

At the same time, our research interest is more open to evaluating only consumers’ eco-innovations in the wine industry, because we believe that these findings especially can encourage wine managers and technical specialists in the development of more sustainable practices for recycling and reduction of raw materials.

Table no. 1. Innovations in the wine industry according to managers and technological specialists in Bulgarian wine-producing companies - a decision hierarchy

CONVENTIONAL INNOVATIONS	ECO-INNOVATIONS	SOCIAL INNOVATIONS
Product innovation (0.30): Significant improved products onto the market (0.1), QR code (0.13)/website (0.18)/newsletter (0.19), wine club (0.18), training course (0.19), green activities promotion (0.21), IT technologies (0.16)	Reduction of material use (0.38): Resource efficiency per unit of output (0.44), organic certification (0.56)	Recognizing wine innovations of indigenous people and local community (0.36)
Grape-growing techniques and technologies (0.27): Use of organic (0.18), chemical (0.42) and innovative substances (0.39)	Replacing material (0.28): less greenhouse gas intensive alternatives (0.42), emission monitoring (0.58)	Recognizing locally developed wine innovations and experimentation (0.32)
Grape-transformation techniques and technologies (0.44): Selective cryoextraction (0.45), wine bio-informational research (0.55)	Recycling (0.34): Reduction of consumption through recycling water (0.32), waste (0.33), materials (0.35)	Piloting and testing local policy wine innovations (0.32)

Source: Boshnakov, Dimitrova and Marinov, 2022 (with their values from the managers assessments), based on Frigon et al., 2020.

After the starting framework of this study is mentioned above, we now explain the steps of clustering used in the present study in Bulgaria. The adopted k-means clustering procedure generates clusters using the cluster's object mean value. K-means divides the dataset into non-overlapping and independent k numbers, without internal structures or labels, such that the observations in one cluster are similar to each other and dissimilar to those in the remaining sets (Niu et al., 2021). The k-means then maximize the inter-cluster distance between samples and minimize the intra-cluster distance.

The goal of the k-means algorithm is to find locally optimal solutions, accounting for a clustering error. We make the calculations in R (R Core Team, 2018), and use the Hartigan and Wong (1979) algorithm. To solve the clustering problem with k clusters, the process starts with one cluster and it finds its optimal position, corresponding to the centroid of the dataset. Clustering is performed on a data matrix so that initially the data is coerced to a matrix of numeric values and the number of clusters and the maximum number of iterations are entered. Next, a random set of rows is chosen as the initial centers. Within the following optimization procedure, points are allotted into k groups, so that the sum of squares to the cluster centers is minimized.

Unlike other clustering algorithms, the k-means clustering requires the number of clusters, k, to be set by the researcher. Since many variants are available, setting k can be a difficult task. One of the core problems with the algorithm is to find the optimum number of clusters, because most realizations do not contain explanations for the selection of particular values for k (see Pham, Dimov **and** Nguyen (2005) for a more detailed explanation). Iterative experiments with different values of k can be helpful, and in our study, we proceed in this way. We rely on approaches that are used in the factor analysis, such as the scree plot and factor loadings to assess the optimum number of clusters.

From the methodological approach, first, after cleaning the data, we compute the index values, for all three dimensions. Secondly, we use the k-means clustering procedure and determine the optimal number of clusters using the scree method and comparing its suggestions to other alternatives. Third, clusters are analyzed according to the sample characteristics. We also check the clusters by the average silhouette approach to assess the quality of clustering. In the next step, clusters are visualized, and some descriptive statistics are added.

Our main choice is the k-means algorithm, because it is a very simple and fast approach, able to deliver robust and consistently interpretable outcomes. We deliberately choose this rather mechanistic approach, to not impose any prior suggestions to the clusters.

Descriptive statistics of the sample are organized as follows. The survey for our study was conducted in March 2023, our respondents are Bulgarian citizens, aged 18-63, the average age is 28.7 years, and the median age is 22 years, all with high school education or university degree, living in North-eastern Bulgaria. Since our intent is to explore presumably the views of the young people, two thirds of our sample are aged below 27 years. After cleaning the data, we proceed with 246 valid responses, 153 women and 93 men.

The initial part of our empirical analysis consists in calculating the values for the indexes, for conventional, organic, and social innovations. We use the weights from the managers' version of the BDM index since we consider it to be more likely to reflect or coincide with the views of the general public - technical specialists tend to put additional attention to specific details, which are unfamiliar or unpopular to the general public. We used the estimates of global variables as the basis for the clustering, and applied an iterative process, to assess the optimum number of clusters.

As recommended in the literature, we first normalize the data, using the mean values of the indexes as centers and the variances as scales. Further, the calculations are based on normalized data. The scree plot suggests that the optimum number of clusters is 5, with 4 or 3 still acceptable (see Figure no. 1).

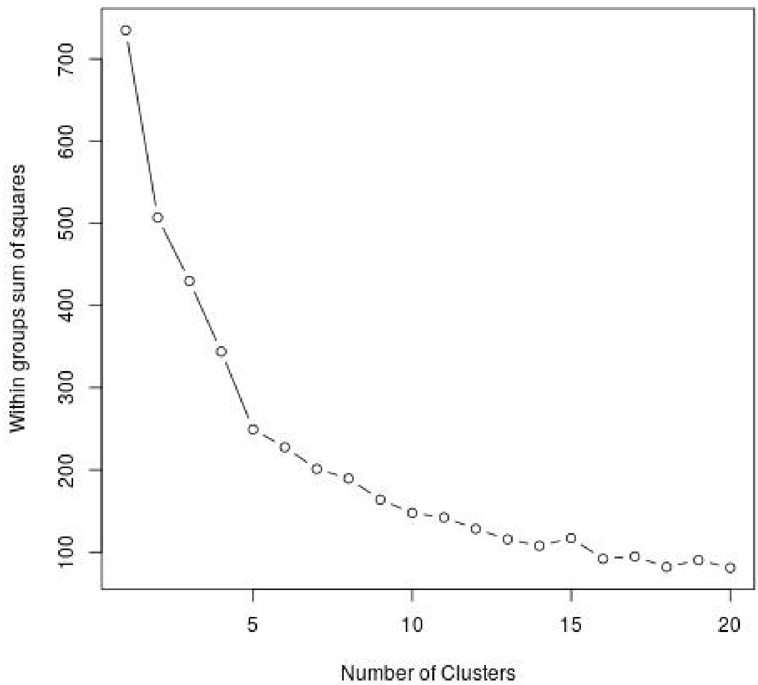

Figure no. 1. Scree plot on normalized data
Source: The authors

Proceeding further with 5 clusters, we compare their main characteristics using several statistical tests. The resulting 5 clusters are of sizes 55, 92, 52, 38, 9, resp. between_SS / total_SS = 66.1 %. According to the results, we can decide to continue the analysis with either 5 or 4 clusters.

To verify our findings about the number of clusters, we also make cluster mapping (Figure no. 2).

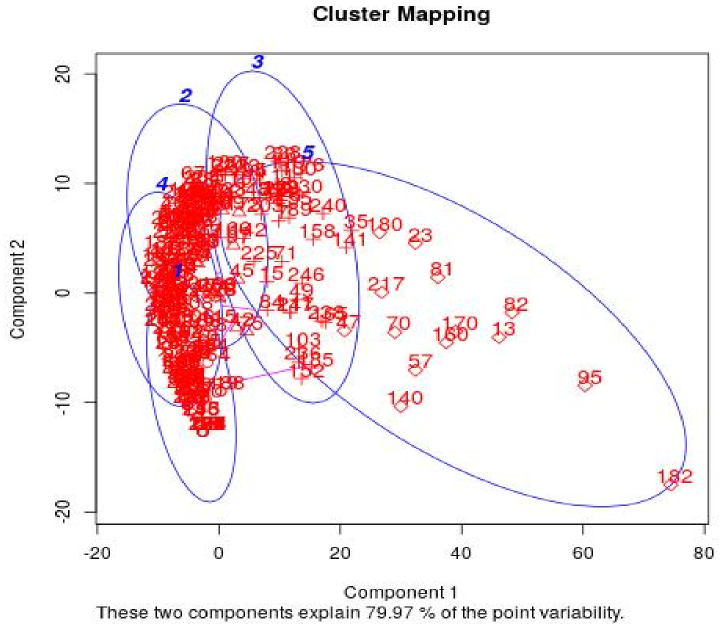

Figure no. 2. Cluster mapping, 5 clusters
Source: The authors

Cluster mapping (Figure no. 2) shows the first two components comprise nearly 4/5 of the variability in the sample, therefore it can be fairly well used as a basis for choosing the appropriate number of clusters. We proceed further with 5 clusters.

Results and discussion

We represent the main results from the cluster analysis in Table no. 2, which contains the basic characteristics for each cluster, such as the average values for the variables used for classification, as well as the main features of the clusters.

Table no. 2. Characteristics of respondents by cluster

	1	2	3	4	5
Variable 1 (conventional, mean value)	4.71	4.91	3.87	4.59	2.52
Variable 2 (eco, mean value)	4.14	4.48	4.27	2.66	2.53
Variable 3 (social, mean value)	3.67	4.76	4.44	4.18	2.69
Respondents	55	92	52	38	9
F	30	59	39	21	4
M	25	33	13	17	5
Age mean	25.6	28.6	30.4	31.8	26.9
Age median	22	23	22.5	23.5	23
Edu high	30	52	30	15	6
Edu university	25	40	22	23	3
Organic wines (are better)	0.56	0.40	0.15	-0.13	0.34

Source: The authors

Due to its small size (9 respondents), we can exclude the data of the last cluster from our further analysis, considering them as outliers, and concentrate on the remaining 4 clusters. From the remaining, clusters 1 and 2 comprise respondents giving more attention to eco- innovations, and people who are also more enthusiastic about organic wines, while clusters 3 and 4 contain respondents who relatively give less attention to eco-innovations, are slightly older and are more skeptical about organic wines (see Table 2).

To assess the views of young people, we conducted a more detailed analysis, assessing differences in the answers to questions about the elemental variables (composing the global ones) given by younger people (< 26 years) and the rest of the sample. We discovered significant differences for the global variable "eco-innovations", with young people giving more importance to it (4.00 vs. 3.90) than the rest of the sample.

Sub-indicators also reveal the importance of reducing resource use and recycling for young people. More detailed screening in comparison with elder respondents (age 26 and more) reveals existing differences in several sub-questions. T-tests show existing statistical differences between young people and the rest of the sample for questions "recycling", with scores of 4.60 for young people vs. 4.31 for the rest, "less greenhouse gas intensive alternatives" (4.57 vs. 4.30), and "emission monitoring" (4.57 vs. 4.28). However, since the last two questions are sub-questions, and they are the constituent parts of "replacing material", we accept that differences in this variable also exist, with younger people stronger recognizing the importance of recycling and replacing materials for greener production.

Conclusion

In the present paper, eco-innovation in the wine industry is studied from the demand side, which is traditionally neglected. To our knowledge, clustering analysis for customer attitudes regarding eco-innovations in the wine industry is applied for the first time in Bulgaria. Our findings show that Bulgarian young consumers (< 26 age) are environmentally friendly and oriented towards eco-innovations in the wine industry with a special emphasis on recycling (water, energy) and replacing materials. Our findings confirm the ones of other recent studies, that preferences for eco-innovations in the wine industry are correlated with willingness to pay more for organic wines (Rabadan and Bernabeu, 2021). The results of the study can be helpful to wine managers, technical specialists, and wine-producing companies to prioritize their green efforts for the youth generation in Bulgaria. The main extensions of the research are related to the comparison between eco-innovations and conventional and social innovations in the Bulgarian wine industry.

References

Abernathy, A. and Celebi, M.E., 2022. The incremental online k-means clustering algorithm and its application to color quantization. *Expert Systems with Applications*, 207, p.117927. https://doi.org/10.1016/j.eswa.2022.117927.

Boshnakov, P., Dimitrova, V. and Marinov, G., 2022. Innovations in Wine Sector: Evidence from Bulgaria Based on Fuzzy AHP Approach, New Trends in Sustainable Business and Consumption: *BASIQ 2022, 8th International Conference : Conference Proceedings, 25-27 May 2022, Graz, Austria, Bucharest* : Bucharest Univ. of Economic Studies Publ. House, pp. 583-590, http:// doi.org/10.24818/BASIQ/2022/08/077.

Egri, C.P. and Herman, S., 2000. Leadership in the NorthAmerican environmental sector: values, leadership styles, and contexts of environmental leaders and their organization. *Academy of Management Journal,* 43(4), pp.571-604, http:// doi.org/ 10.2307/1556356.

Eshimiakhe D. and Lawal K., 2022. Application of K-means algorithm to Werner deconvolution solutions for depth and image estimations. *Helyon*, 8(11), e11665.

Forgy E., 1965. Cluster analysis of multivariate data: Efficiency vs. interpretability of classification. *Biometrics*, 21, pp.768-780.

Frigon, A., Doloreux, D. and Shearmur, R., 2020. Drivers of eco-innovation and conventional innovation in the Canadian wine industry. *Journal of Cleaner Production*, 275, p.124115. https://doi.org/10.1016/j.jclepro.2020.124115.

Hartigan, J.A. and Wong, M.A., 1979. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1), pp.100-108.

Higuchi, A. and Maehara, R. 2021. A factor-cluster analysis profile of consumers. *Journal of Business Research*, 123, pp.70-78.