Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 24, Issue 1, 2024
PRINT ISSN 2284-7995, E-ISSN 2285-3952

A STATISTICAL STUDY ANALYSIS ON EXPLORING CONSUMPTION PATTERNS REGARDING FOOD LOSS AND WASTE

Daniel NIJLOVEANU*, Victor TIȚA*, Nicolae BOLD*, Dragoș SMEDESCU, Alexandru FÎNTÎNERU**,Valentina TUDOR**, Cosmina SMEDESCU**, Emanuela JERCA*****

University of Agronomic Sciences and Veterinary Medicine of Bucharest, *Faculty of Management and Rural Development – Slatina Branch, **Faculty of Management and Rural Development - Bucharest, ***Faculty of Horticulture, 59 Marasti Boulevard, District 1, 011464, Bucharest, Romania, E-mails: nijloveanu_daniel@yahoo.com, victortita@yahoo.com, bold_nicolae@yahoo.com, dragos.smedescu@managusamv.ro, fintineru.alexandru@managusamv.ro, tudor.valentina@managusamv.ro, cosminasmedescu@gmail.com, emanuela.jerca@doctorat.usamv.ro

Corresponding author: bold_nicolae@yahoo.com

Abstract

Food loss ans waste, referred next as FLW, has a great extent on the economic and behavioural patterns of consumption. In this paper, we aim to present statistical univariate and multivariate analyses based on a statistical study run during a period of several months in the Romanian territory. In this matter, the main purpose of the analysis developed in this paper is to determine several patterns of FLW phenomenon. This is made based on a methodology which comprises two main sections: the univariate analysis, consisting in statistical determinations of indicators and direct observation, and the multivariate analysis, related to the clustering analysis. The clustering analysis is based on the usage of hierarchical method, which has as final result the determination of clusters describing patterns of consumption related to food loss and waste. After the analysis was run, five main clusters related to food waste behaviour patterns were determined, with slight differences between food consumption habits, but with various combinations between the parameters taken into consideration. The obtained clusters offer important information for policymakers and stakeholders aiming to customize interventions and programs to target specific demographic groups or segments of the population.

Key words: *food waste and loss, clustering, consumption pattern, Romania*

INTRODUCTION

Food loss and waste (FLW) represent a significant issue with far-reaching implications on both economic and behavioral aspects of consumption patterns [9]. FLW not only impacts the financial bottom line but also reflects broader societal trends [12] in how food is produced, distributed, and consumed [1]. Understanding the dynamics of FLW is crucial for addressing sustainability challenges and improving resource efficiency in the food supply chain [8, 10].

In this paper, we delve into a detailed analysis of FLW, focusing particularly on its prevalence and patterns within the context of Romania. Over the course of several months, we conducted a comprehensive statistical study covering various regions of the country. Our objective is to provide insights into the nature of FLW phenomenon and identify key trends and patterns.

MATERIALS AND METHODS

The characteristics of the statistical study

The main data used in the analysis which will be presented further in detail was obtained based on a statistical study. The study has several specific characteristics, such as the purpose, the methodology, the target group and the measurement instruments.

The purpose of the study was related to the investigation of the consumer behaviour related to food loss ans waste. The final aspects of the study were also related to suggesting countermeasures and prevention actions for FLW.

The main measurement instrument was a statistical questionnaire, which included as core features open and choice-based questions related to food consumption behavioural aspects.

The target group was considered to be formed of adults of various age groups and socio-economic environments. The online distribution of the questionnaire to the sample group also led to a relatively even geographical distribution of the responses.

The methodology of the study comprised of several phases:

(1)The questionnaire design: this phase consisted in the creation of the development of an online questionnaire composed of structured questions to assess food shopping frequency, storage methods, awareness of food waste, and other relevant factors.

(2)The sampling process: the sample group was selected through a convenience random sampling method, resulting in a sample volume of 364 respondents.

(3)The questionnaire implementation: The questionnaire was distributed online through a number of social media platforms and via email to a diverse sampling of participants from various geographic regions, delimited by the development regions of Romania (NUTS 2-level divisions).

The study also took into account ethical procedures; thus, the anonymity and the confidentiality of the respondents was ensured. In the end, several limitations were also delimited, including possible problems of sample representativeness and subjective reporting of consumer behavior related to food waste.

The analysis methodology

For this analysis, the determination of patterns was considered to be completed as several clusters which describe consumption patterns of respondents related to FLW were aimed to be generated. In this matter, we will establish the main methodology based on two main components: the univariate analysis and the multivariate analysis.

Regarding the univariate analysis, the main purpose is related to the description of the main indicators taken into account in the analysis in order to draw several preliminary conclusions of the study. The analysis starts by describing the independent variables in the questionnaire, such as the frequency of responses and their distribution. In this phase, tables, graphs, and measures of central tendency (such as median, mean) and dispersion (such as standard deviation) are determined to illustrate and interpret this data.

As for the multivariate analysis, the main purpose is to identify and establish correlations between several indicators, in order to define patterns within the responses and to extrapolate based on the sample. The method used for the multivariate analysis is the Hierarchical Cluster Analysis (HCA) method [4, 13]. The phases used for the multivariate analysis methodology are related to:

(1)The selection of the variables: the variables that were taken into consideration were:

(a)the food purchase frequency (FPF), determined as the number of food purchase events per month;

(b)the amount of money used for food purchase (FPM), expressed in lei per month;

(c)the cause of food waste (FWC), as grouped by categories;

(d)the perceived percent of food thrown away monthly (FWP);

(e)the responses to specific affirmations: „At the end of a meal, I throw away leftovers." (A1); „I freeze the leftovers from the meal and will eat them later or prepare other products from them." (A2);„The leftovers from the table are used in pet food." (A3);„I check the expiry date of the food at the time of purchase and will buy products with a longer shelf life." (A4);„I do selective recycling." (A5), with three possible answers: „Frequently", „Rare" or „Never".

(2)The encoding of data: the variables are encoded for integers. Thus, all the variables were transformed in numerical data using the Ordinal Encoding (OE) method (map values to integers).

(3)The choice of clustering method: the specific methods taken into consideration were the HCA method and the DBSCAN approach. The HCA method was chosen for the specific set of data.

Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 24, Issue 1, 2024
PRINT ISSN 2284-7995, E-ISSN 2285-3952

(4)The determination of the optimal number of clusters: this will be made based on the specific method of HCA called dendrogram cut.
(5)The run of clustering algorithm: the steps of the algorithm will be presented further.
(6)The interpretation of the results: the generated clusters and their main characteristics will be presented.
(7)The validation of the clusters: this will be made using dendrogram analysis.
The HCA approach consists in the development of several steps:
(1)The determination of the distance matrix: a matrix of distances between all pairs of observations in the data set is computed. Distance will be calculated using a metric such as Euclidean distance.
(2)The determination of the dendrogram: the distance matrix is used to construct the dendrogram, which is a graphical representation of the similarity relationships between observations.
(3)The cutting of the dendrogram: cutting dendrograms at a certain height allows the determination of a specific number of clusters. This can be done by selecting a cut height that separates the dendrogram into an appropriate number of clusters, based on criteria such as maximizing the inter-cluster distance or analyzing the structure of the dendrograms.
(4)The final clustering: based on the selected cut height, each observation will be classified into a corresponding cluster.
Using the specified approach, the final result will consist in the clusters obtained, which will be defined by specific characteristics of the variables taken into account.

RESULTS AND DISCUSSIONS

The univariate analysis

For the univariate analysis, we will determine specific observations related to several indicators extracted from the responses of the sample participants. Firstly, we will present them.
The presentation starts with the demographic data of the respondents. Fig. 1 presents the data related to age group.

We can observe that the majority of the respondents (approximately 73%) are part of the 18 – 35 years age group.

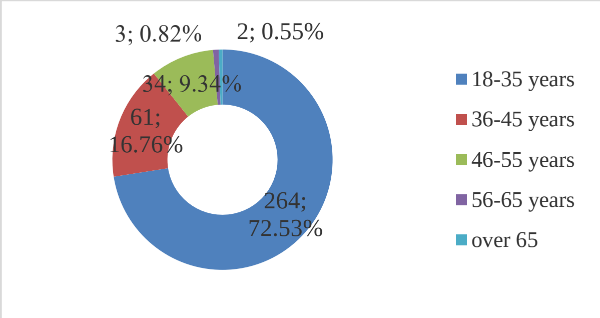


Fig. 1. The classification of respondents by age group
Source: data processing from online questionnaire.

Next, Fig. 2 presents the data related to gender.

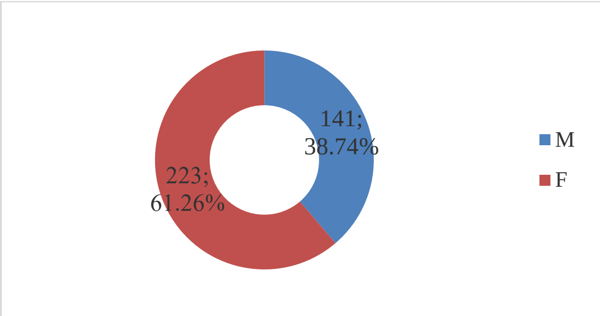


Fig. 2. The classification of respondents by gender
Source: data processing from online questionnaire.

The structure of the respondents related to gender shows a balanced proportion of responses, slightly higher for the feminine respondents.
Next, Fig. 3 presents the data related to the latest form of education of the respondents.

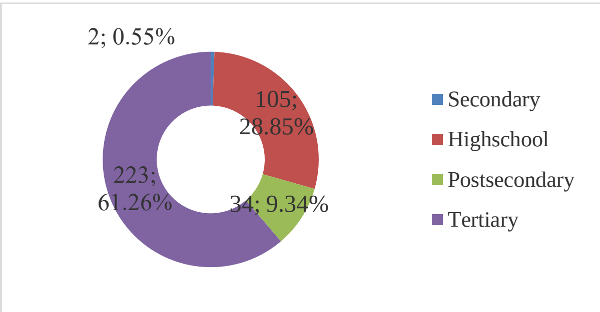
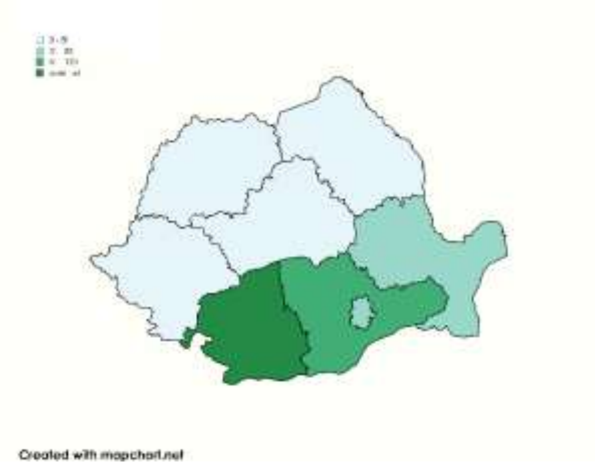


Fig. 3. The classification of respondents by latest form of education
Source: data processing from online questionnaire.
We can observe that the majority of the respondents has as latest form of completed education tertiary levels, related to university

Scientific Papers Series Management, Economic Engineering in Agriculture and Rural Development
Vol. 24, Issue 1, 2024
PRINT ISSN 2284-7995, E-ISSN 2285-3952

education (e.g., bachelor's degree, master's degree).

Next, Map 1 presents the data related to the geographical distribution of the respondents.



Map 1..The classification of respondents by NUTS-2 regions
Source: data processing from online questionnaire.

We can observe that the majority of the respondents, approximately 50% of the respondents (171 responses), is established in RO41 region (South-West Oltenia Region).

Next, Fig. 4 presents the data related to the socio-professional status of the respondents.

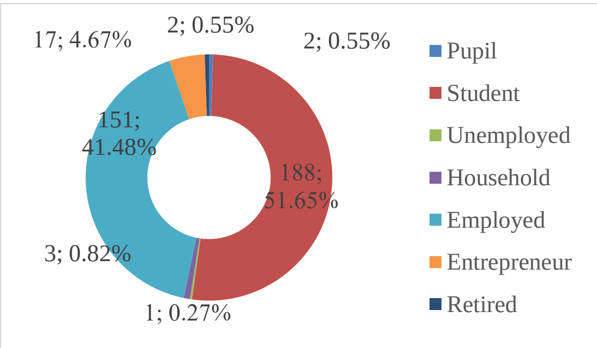


Fig. 4. The classification of respondents by socio-professional status
Source: data processing from online questionnaire.

We can observe that the main socio-professional statuses of the respondents are related to studentship (in a higher form of education) and employed.

Next, 5 presents the data related to the income of the respondents.

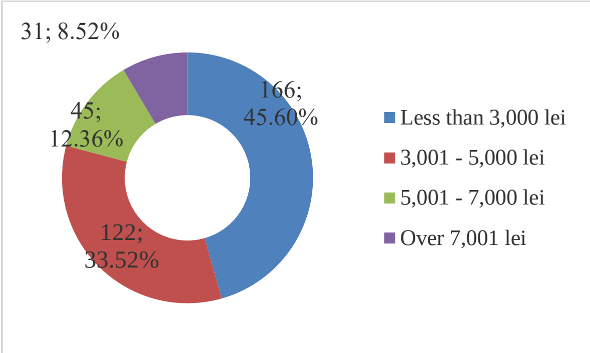


Fig. 5. The classification of respondents by income
Source: data processing from online questionnaire.

We can observe that the main two categories of income of the respondents are related to a medium level, between 3,000 and 7,000 lei.

Next, 6 presents the data related to the residence of the respondents.

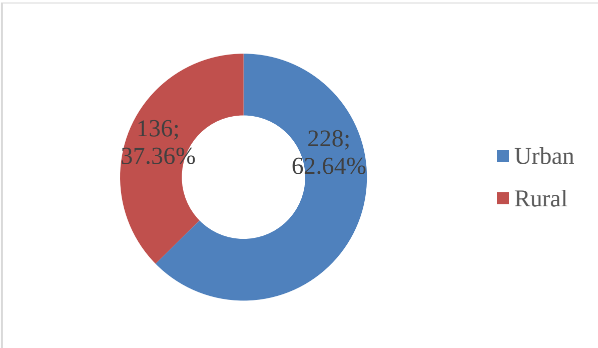


Fig. 6. The classification of respondents by residence
Source: data processing from online questionnaire.

We can observe that a slightly majority of the respondents is established in the urban environment (approximately 65%).

Next, Fig. 7 presents the data related to the frequency of food purchase events of the respondents.

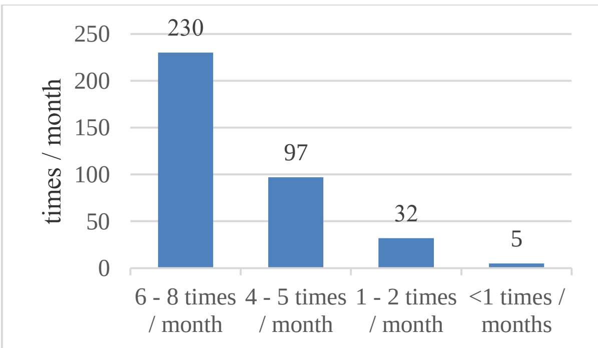


Fig. 7. The frequency of food purchase events
Source: data processing from online questionnaire.

We can observe that many respondents have a frequent habit of buying food, resulting a predominant frequency of 6 to 8 food purchasing events per month.
Next, 8 presents the data related to the amount of financial resources spent monthly on food.

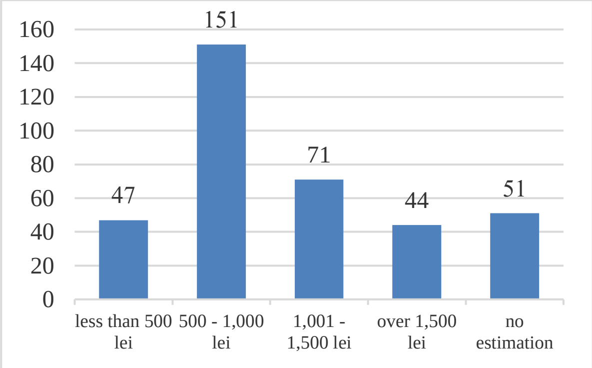


Fig. 8. The amount of money spent on food monthly
Source: data processing from online questionnaire

The majority of the respondents spend an amount of money between 500 and 1,000 lei per month on food, with sensibly equal amount of money for other categories, including the respondents that could not estimate the sum spend monthly on food. Regarding the causes of food waste, Fig. 9 presents the data related to the main causes of FLW identified among the respondents.

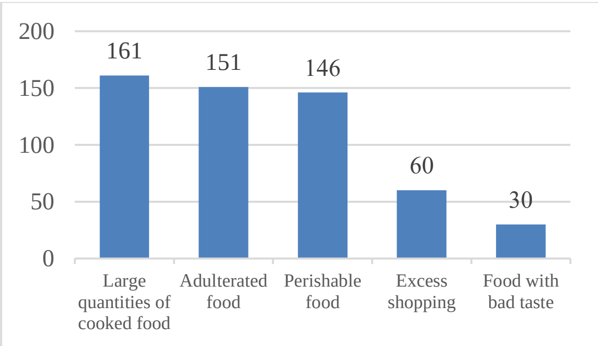


Fig. 9. The main causes of the food waste identified among the respondents
Source: data processing from online questionnaire.

The three main causes chosen by the respondents were related to the large quantities of cooked food, the perishability of the food and the natural processes of the degradation of the food that is not consumed. Next, 10 presents the data related to the perceived amount of food waste rationed by the total amount of purchased food.

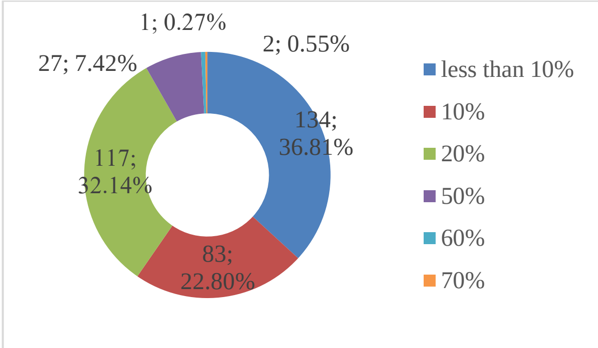


Fig. 4. The perceived percent of food waste monthly
Source: data processing from online questionnaire.

The majority of the respondents (over 90%) perceived the amount of wasted food between none and 20% of the total amount of purchased food, with two thirds of this majority perceiving the amount below or equal to 10%.
Finally, 11 shows several responses given to specific affirmations regarding food waste.

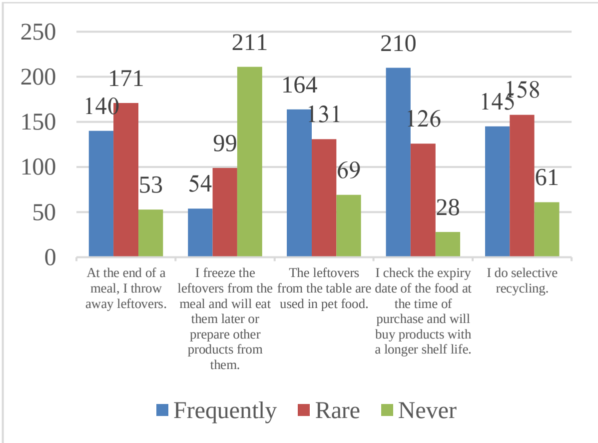


Fig. 5. Responses given to specific affirmations related to food waste
Source: data processing from online questionnaire.

As we can see, the respondents seldom throw away leftovers at the end of the meal. Regarding later usage of the leftovers, a large majority never freezes them in order to reuse them. A large part of the respondents uses the leftovers for the pets in the household and many respondents try to avoid food waste by purchasing food with a larger shelf time. Finally, the respondents do selective recycling of the food rarely to frequently.
As a preliminary conclusion, the profile of the respondent comprises the next demographic and behavioural patterns: the main age group

is between 18 and 35 years, mainly with studies up to tertiary education. Also, the majority of the respondents were concentrated geographically in the NUTS-2 regions South-West Oltenia and South. The main socio-professional statuses are related to employed people and students, with predominant incomes between 3,000 and 5,000 lei or under 3,000 lei. Also, a slight majority of the respondents have the residence in the urban area, with a frequent habit of purchasing food with a frequency of 6-8 times per month. The main amount of money used to purchase food is situated between 500 and 1,000 lei. The main self-identified causes of food waste are large quantities of purchased food and the perishability character of food, with a perceived amount of food waste up to 20%. Finally, respondents have a frequent habit of reusing leftovers, either by preservation or consumption by pets. Also, the respondents have a frequent habit of purchasing food with a higher expiry date and do selective recycling of food.

The multivariate analysis

The methodology presented in the previous sections regarding the multivariate analysis was conducted on the resulted data. The enumerated variables were encoded using two methods: categorical encoding, in case of questions with unique answers, and One-Hot Encoding (OHE) [6], related to multiple-choice answers. For the latter case, each choice was encoded with binary data (digits 0 and 1), 0 having the significance of not being chosen and 1 otherwise. Regarding the variables denoted by A#, the encoding was made by the response given by the user (0 – Never, 1 – Rare, 2 – Frequently). The encoding results are shown in **Error! Reference source not found.**.

Table 1. The variable encoding chart

No.	Variable		Abbreviation	Possible values set
1.	The food purchase frequency		FPF	{1..4}
2.	The amount of money used for food purchase		FPM	{1..5}
3.	The cause of food waste – choice	Excess shopping	FWC - 1	{0,1}
4.		Perishable food	FWC - 2	{0,1}
5.		Food with bad taste	FWC - 3	{0,1}
6.		Adulterated food	FWC - 4	{0,1}
7.		Large quantities of cooked food	FWC - 5	{0,1}
8.	The perceived percent of food thrown away monthly		FWP	{1..6}
9.	„At the end of a meal, I throw away leftovers.”		A1	{0,1,2}
10.	„I freeze the leftovers from the meal and will eat them later or prepare other products from them.”		A2	{0,1,2}
11.	„The leftovers from the table are used in pet food.”		A3	{0,1,2}
12.	„I check the expiry date of the food at the time of purchase and will buy products with a longer shelf life.”		A4	{0,1,2}
13.	„I do selective recycling.”		A5	{0,1,2}

Source: own processed data.

After the encoding, the values were used in the HCA approach of clustering.

A specific Machine-Learning based software was used (Orange [3]).

Various results were obtained, with the main results being the dendrogram chart, which is presented in **Error! Reference source not found.** and shows the main clusters obtained after the data analysis.

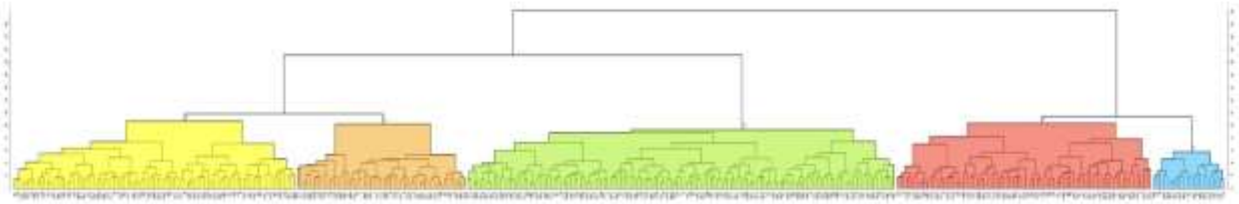


Fig. 12. The obtained dendrogram for the variables taken into account

Source: Own determination.

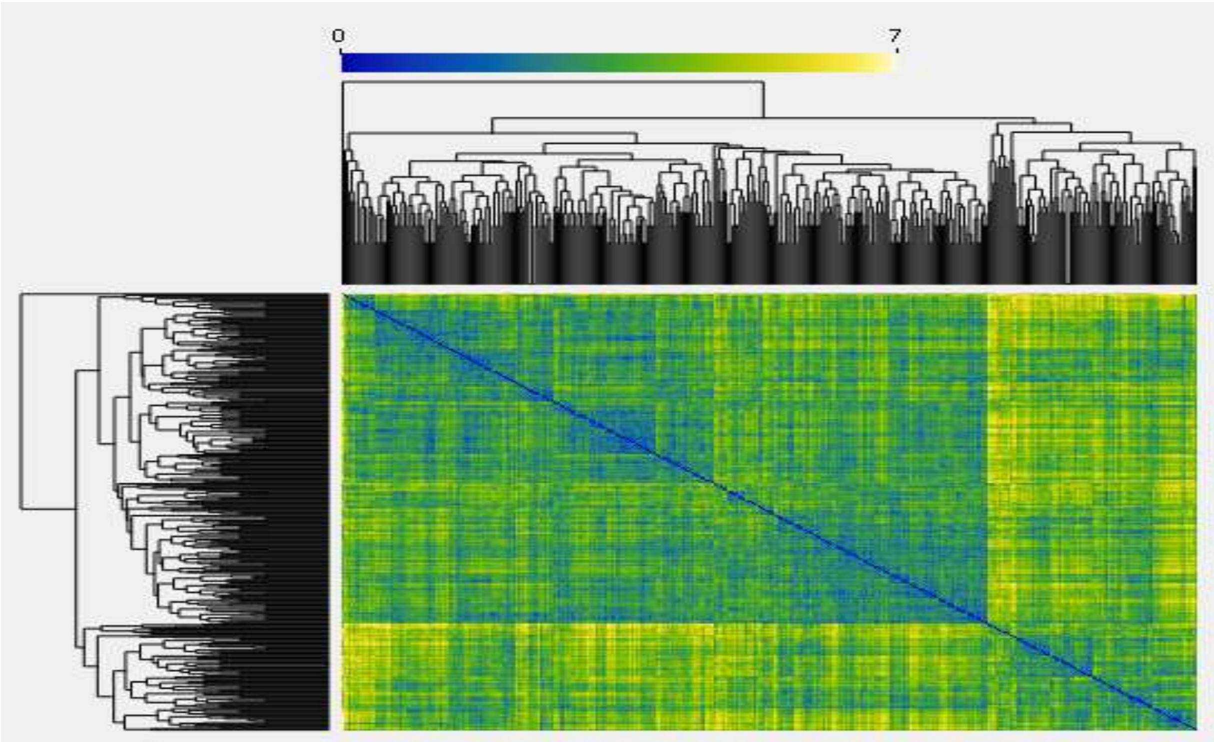


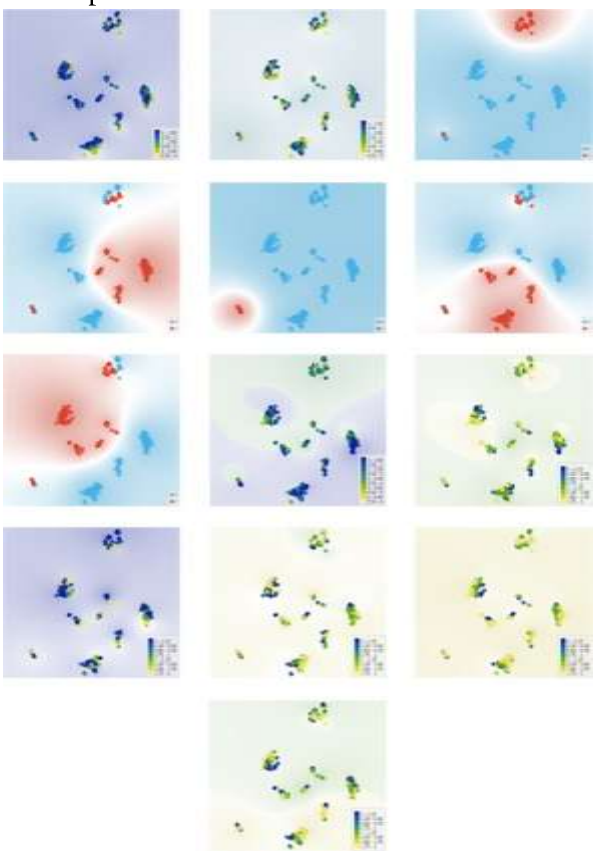
Fig. 13. The distance map correspondent to the obtained distance matrix
Source: own determination.

The distance matrix, whose corresponding distance map is presented in the next Figure 13 presents the corresponding distance map which determines the differences between each response.

Also, a Louvain clustering algorithm was applied and 8 clusters were obtained automatically. A Louvain clustering algorithm is a graph clustering method used to identify communities or groups of nodes with greater connectivity to each other than to nodes outside the group in a graph [2].

Further, an empirical DBSCAN analysis (i.e., an alternative method of clustering) [5] was run. A detailed analysis of the appliance of this method will be detailed in future papers. After the run of the analysis, two main clusters were obtained.

For a more detailed and visually-aid analysis of data, a t-SNE (t-Distributed Stochastic Neighbor Embedding, i.e., a dimensionality reduction and data visualization technique [14]) was used (perplexity of 30, exaggeration of 2 and 13 PCA components). Usually, a t-SNE analysis is made in order to reduce the multi-dimensional character of the data in a 2D space. The results are shown in



.

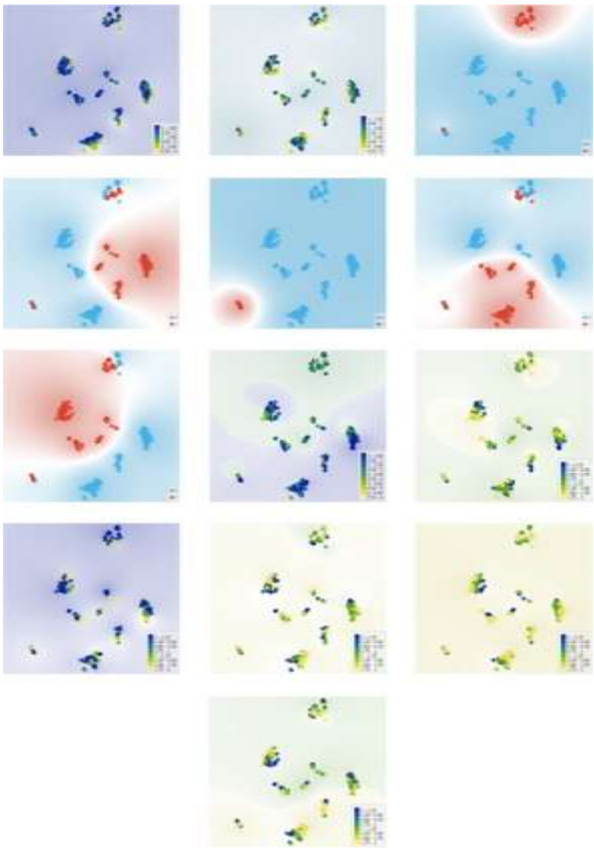


Fig. 14. The t-SNE data visualisations by variables (in order, FPF, FPM, FWC1-FWC5, FWP, A1-A5, clusters)
Source: data processing from online questionnaire.

The data was obtained after the appliance of the distance matrix calculated using Euclidean distance [9] and the options of the dendrogram were considered as shown in **Error! Reference source not found.**, with no pruning appliance (i.e., taking into account all input data, with no filtering process) and the manual selection of the number 5 for the cluster. The number was chosen after a process of direct observation, with a purpose of a balanced set of obtained clusters. Finally, the method chosen for the linkage was the Linkage Ward [7], which is a linkage method used in agglomerative clustering, a type of hierarchical clustering. This method focuses on minimizing the variation within the clusters formed at each step of the clustering process.

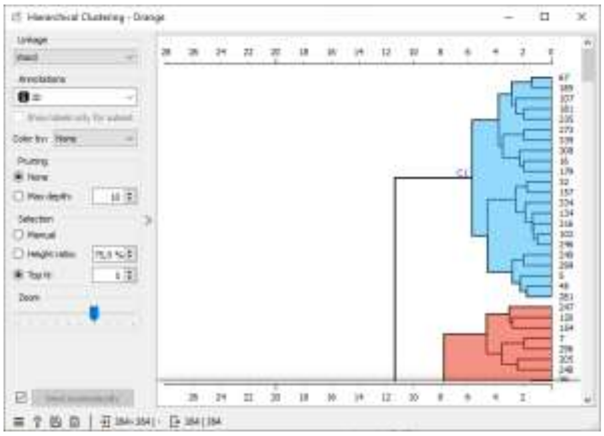


Fig. 15. The HCA parameters in the software
Source: data processing from online questionnaire.

Based on the results, five clusters of responses were obtained. The description of the clusters is made in the next paragraphs.

The C1 cluster (the hungry caring spenders), coloured in blue in the dendrogram, contains 22 respondents. This cluster is characterized by respondents which buy very frequently food (6-8 times per month), who cannot estimate the amount of money used for food purchasing, with a perceived amount of food waste under 10% of purchased food, with an antagonist behaviour related to food waste reuse, as they rarely throw away leftovers and use them predominantly for pets. They are more inclined to purchase food with longer shelf life and do more often selective recycling. As for the main cause of food waste identified by this cluster, the process of food adulterating after being cooked (FWC – 4) was one of the most predominant.

The C2 cluster (the unpredictive frequent food buyers), coloured with red in the dendrogram, is formed of 77 respondents. These respondents are frequent buyers of food (6-8 times per month) and the amount of money spent on food is either above 1,500 lei or cannot be estimated. The perceived amount of food waste is estimated to be between 10% and 20% of the purchased food. Also, they are throwing away leftovers at a moderate frequency (rare or frequently), but the ones that are not thrown away are preserved and used with a large probability in future meals or less probably for pets. They are also inclined to purchase products with longer shelf life and have an unpredictive behaviour related to selective recycling. The main identified causes for the food waste for this cluster were the perishability of the food or large quantities of cooked food which were not consumed.

The C3 cluster (the balanced no-left over purchasers), coloured with green in the dendrogram, is formed of 129 respondents. They also buy very frequently food (6-8 times per month), with a monthly food budget mainly situated between 500 and 1,000 lei. They perceive to waste approximately 20% of the purchased food. They rarely of frequently throw away leftovers and they almost never preserve them in order to reuse. The food waste is rarely or frequently used for pets. Also, they are extremely inclined to buy food with longer expiry dates and their tendency to do selective recycling is moderate. The main causes selected for food waste were determined as large quantities of cooked food which is then altered.

The C4 cluster (the less-hungry non-waster wasters), coloured in orange in the dendrogram, contains 51 respondents. The frequency of their purchase events is unpredictively situated between 4 and 8 times per month, with a food budget below 1,000 or even 500 lei per month. They perceive to waste food in a proportion of less then 10% of the purchased food products. They rarely or frequently throw leftovers and they almost never preserve them for reuse, the ones that eventually are not thrown are used as pet food. In order to limit the food waste, they are also inclined to buy food which lasts longer, as they have unpredictive behaviour related to selective recycling, made rarely or frequently. The main cause of food waste is considered to be the food alteration after large quantities of it were cooked.

Finally, the C5 cluster, (the uncertain food habit consumers), coloured in yellow in the dendrogram, contains 85 respondents. The majority of the respondents also buy food very frequently (6 to 8 times per month), with a budget situated with a greater probability between 500 and 1,000 lei, rarely up to 1,500 lei. They claim to waste food to a proportion closer to 10%, with tendencies up to 20%. They tend to keep leftovers, rarely throwing them away, with a totally unpredictive behaviour related to their preservation. There may be a higher chance of using them as pet food. Most commonly, food with longer availability periods is chosen, with an uncertain behaviour related to selective recycling, but more inclined to have habits related to this process. The main causes related to food waste are selected to be food categories that are easily perishable.

In conclusion, food waste is a generalized phenomenon in Romania, but with very specific aspects related to food waste patterns among the consumers.

CONCLUSIONS

The findings presented in this paper shed light on the intricate dynamics of food loss and waste (FLW) within the Romanian context, offering valuable insights into its prevalence, determinants, and potential interventions. Through a comprehensive statistical analysis spanning several months, we have identified several key patterns and trends that characterize FLW behavior in the country.

One of the key contributions of our analysis is the identification of distinct clusters representing different consumption patterns related to FLW. Through hierarchical clustering analysis, we have delineated several clusters that capture the diverse range of behaviors and attitudes towards waste within the Romanian population. These clusters provide valuable insights for policymakers and stakeholders seeking to tailor interventions and initiatives to specific demographic groups or segments of the population.

ACKNOWLEDGEMENTS

This work was supported by a grant of the University of Agronomic Sciences and Veterinary Medicine of Bucharest, project number 2023-007 acronym **ReWaFA**, within IPC 2023.

REFERENCES

[1]Cattaneo, A., Sánchez, M. V., Torero, M., Vos, R., 2021, Reducing food loss and waste: Five challenges for policy and research. Food Policy, 98, 101974.
[2]Combe, D., Largeron, C., Géry, M., Egyed-Zsigmond, E., 2015, I-louvain: An attributed graph clustering method. In Advances in Intelligent Data Analysis XIV: 14th International Symposium, IDA 2015, Saint Etienne. France, October 22-24, 2015. Proceedings 14 (pp. 181-192). Springer International Publishing.
[3]Demsar, J., Curk, T., Erjavec, A., Gorup, C., Hocevar, T., Milutinovic, M., Mozina, M., Polajnar, M., Toplak, M., Staric, A., Stajdohar, M., Umek, L., Zagar, L., Zbontar, J., Zitnik, M., Zupan, B., 2013, Orange: Data Mining Toolbox in Python, Journal of Machine Learning Research 14(Aug): 2349−2353.
[4]Dugad, R., Ahuja, N., 1998, Unsupervised multidimensional hierarchical clustering. In Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'98 (Cat. No. 98CH36181) (Vol. 5, pp. 2761-2764). IEEE.
[5]Khan, K., Rehman, S. U., Aziz, K., Fong, S., Sarasvady, S., 2014, February, DBSCAN: Past, present and future. In The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014) (pp. 232-238). IEEE.
[6]Mougan, C., Álvarez, J. M., Ruggieri, S., Staab, S., 2023, August. Fairnes simplications of encoding protected categorical attributes. In Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society (pp. 454-465).
[7]Nielsen, F., Nielsen, F., 2016, Hierarchical clustering. Introduction to HPC with MPI for Data Science, 195-211.
[8]Nijloveanu, D., Tita, V., Bold, N., Fintineru, G., Smedescu, D., Chiurciu, I.A., Smedescu, C., Patrachioiu, G., 2023, Dynamics of food loss and waste causes along the agrifood chain in Romania. Scientific Papers. Series "Management, Economic Engineering in Agriculture and rural development", Vol. 23(4), 569-580.
[9]Pakravan-Charvadeh, M. R., Flora, C., 2022, Sustainable food consumption pattern with emphasis on socioeconomic factors to reduce food waste. International Journal of Environmental Science and Technology, Vol. 19(10), 9929-9944.
[10]Paraschivu, M., Cotuna, O., Matei, G., Sarateanu, V. 2022, Are food waste and food loss a real threat for food security?. Scientific Papers. Series "Management, Economic Engineering in Agriculture and rural development", Vol. 22(1), 479-484.
[11]Ran, X., Xi, Y., Lu, Y., Wang, X., Lu, Z., 2023, Comprehensive survey on hierarchical clustering algorithms and the recent developments. Artificial Intelligence Review, Vol.56(8), 8219-8264.
[12]Schanes, K., Dobernig, K., Gözet, B., 2018, ood waste matters-A systematic review of household food waste practices and their policy implications. Journal of cleaner production, 182, 978-991.
[13]Sreedhar Kumar, S., Madheswaran, M., Vinutha, B. A., Manjunatha Singh, H., Charan, K. V., 2019, A brief survey of unsupervised agglomerative hierarchical clustering schemes. Int J Eng Technol (UAE), Vol. 8(1), 29-37.
[14]Van der Maaten, L., Hinton, G., 2008, Visualizing data using t-SNE. Journal of machine learning research, 9(11).