

Article

Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation

Emilia Herman ^{1,*}, Kinga-Emese Zsido ¹ and Veronika Fenyves ²

¹ Faculty of Economics and Law, George Emil Palade University of Medicine, Pharmacy, Science and Technology of Targu Mures, 540142 Targu Mures, Romania

² Faculty of Economics and Business, University of Debrecen, 4032 Debrecen, Hungary

* Correspondence: emilia.herman@umfst.ro

Abstract: Nowadays there is a large amount of information at our disposal, which is increasing day by day, and right now the question is not whether we have a method to process it, but which method is most effective, faster and best. When processing large databases, with different data, the formation of homogeneous groups is recommended. This paper presents the financial performance of Hungarian and Romanian food retail companies by using two well-known cluster analyzing methods (K-Mean and K-Medoid) based on ROS (Return on Sales), ROA (Return on Assets) and ROE (Return on Equity) financial ratios. The research is based on two complete databases, including the financial statements for five years of all retail food companies from one Hungarian and one Romanian county. The hypothesis of the research is: in the case of large databases with variable quantitative data, cluster analysis is necessary in order to obtain accurate results and the method chosen can bring different results. It is justified to think carefully about choosing a method depending on the available data and the research aim. The aim of this study is to highlight the differences between the results of these two grouping procedures. Using the two methods we reached different results, which means a different evaluation of financial performance. The results demonstrate that the method chosen for grouping may influence the assessment of the financial performance of companies: the K-Mean method produces a greater variety of groups and the range of results obtained after grouping is larger; whereas, the group distribution and the results obtained by the K-Medoid method are more balanced.

Keywords: database; cluster analysis; K-Mean method; K-Medoid method; financial performance; financial indicators; financial statements; food retail companies; Hungary; Romania



Citation: Herman, E.; Zsido, K.-E.; Fenyves, V. Cluster Analysis with K-Mean versus K-Medoid in Financial Performance Evaluation. *Appl. Sci.* **2022**, *12*, 7985. <https://doi.org/10.3390/app12167985>

Academic Editors: Małgorzata Charytanowicz and Piotr A. Kowalski

Received: 20 June 2022

Accepted: 5 August 2022

Published: 10 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Nowadays we are working with growing and more complex databases, we have more and more information, and the data and information to be processed and analyzed are very diverse, comprehensive and heterogeneous. Large databases and the available information must be systematized and sorted; thus, the most appropriate solution, option or method is to break the data down into homogeneous groups. The essence of this is that the elements of each group are similar, close to each other, but different from the elements of the other groups [1]. With the help of homogeneous groups, we will be able to evaluate and analyze each group better and more correctly, and finally the evaluation of the entire database will give a more realistic picture.

In the case of financial analysis, it is very important to process large databases correctly and to evaluate the results. Evaluating the financial performance of companies requires comparing very different companies in terms of size and performance based on the processing and analysis of the multi-annual reports of these companies.

The average values of a database, containing variable and significantly different data, cannot be representative, thus cluster analysis is the best solution for database analysis.

There are many techniques and different methods for homogeneous grouping. Clustering methods are presented by Madhulatha [2], who offers great insight into clustering

algorithms and their applicability, characteristics and limitations. The most commonly used methods are K-Mean and K-Medoid algorithms [3].

The K-Mean method uses the group mean (centroid) for grouping the data. The limitation of this method is that very different, distinct values can distort the results. For the definition of reference points, the K-Medoid method uses some representative values (called medoids) instead of mean values [4]. The K-Medoid method is considered less sensitive to outliers in comparison with K-Mean clustering [5].

Prior to grouping the databases, identifying and eliminating the outliers is recommended for more accurate results. Due to the negative effects of the outliers on the results, it is necessary to eliminate them [6]. More ways to identify the outliers are recognized such as the distance-based method, density-based method and graph methods [7].

Hypotheses and Purpose of the Paper

From the very beginning of the present research, in the phase of creating the database, but also further on, when the first calculations were made, many research questions appeared based on which the research hypotheses were formulated. The hypotheses of this research are:

1. For large databases, any type of analysis (social, financial, economic, statistical, etc.) should start firstly with an examination of the variability of the data (how different or similar the elements in the database are). If the elements of the database differ greatly from each other, the formation of homogeneous groups is necessary for effective and correct analysis.
2. The selection and determination of the indicators or characteristics based on which the groups will be formed are very important, as they can influence the obtained results. These indicators or characteristics must reflect the aim of the research.
3. The method chosen for the formation of groups influences the obtained results. The different sizes and contents of groups may influence the results.

Based on these assumptions, *the purpose of this paper* is to perform a comparative analysis between the results of two well-known and frequently used clustering methods (K-Mean and K-Medoid) in the field of financial performance. The processing of databases with large and variable data is possible and recommended only with homogeneous group resolution, based on cluster analysis. The chosen method and grouping procedure will directly influence the results. Cluster analysis is frequently used in all scientific fields such as pharmacy, medicine, energy, economy, biological issues, etc. Our research presents the differences between the results obtained by using two well-known clustering methods for the evaluation of financial performance taking into account three main financial ratios (Return on Sales—ROS, Return on Assets—ROA and Return on Equity—ROE).

This study answers the question of whether it is worthwhile to process databases with large and variable data using cluster analysis with K-Mean or with K-Medoid in the field of financial performance. Moreover, it compares the results obtained using these two clustering methods.

2. Brief Literature Review

There is a great deal of literature dealing with the processing of databases containing large and varied elements and the breakdown of data into homogeneous groups. Clustering methods are used in many fields such as biology, disease classification, archeology, image segmentation, social classifications and market segmentation, as well as database analysis in many fields [8], e.g., entrepreneurship [9], transportation [10], medical research [11], engineering [12] and industrial clusters performance [13].

In the field of economics and finance, large and complex data and databases are needed more and more. As a result, there is a growing need for the processing of economic and financial data. From the vast amount of information available to decision makers, only the information most valuable and useful to them must be selected. In the case of economic–financial analyses, the literature describes a large number of grouping methods

used in various fields: risk analysis [14], financial risk analysis [15,16], selection of financial ratios [17], economic fraud activities [18], real estate portfolio analysis [19], financial performance analysis [20].

Cai et al. [21] provided great insight into clustering methods, underling their advantages and disadvantages for financial datasets. They demonstrated that density-based clustering does not suit financial datasets, whereas K-Mean gives the best number of clusters to help understand financial data classification.

Group disaggregation methods are also very suitable for grouping companies, as demonstrated by Serban et al. [22]. These authors applied clustering methods to 106 enterprises based on economic indicators (Economic Value Added, Net Income, Current Sales, Equity and Stock Price).

Kaur et al. [23] defined clustering “as a process of grouping data objects into disjoint clusters so that the data in each cluster are similar, yet different to the other clusters” (p. 42). There are several methods and techniques by which these clusters are created, such as K-Mean and K-Medoid. According to Kaur et al. [23], the K-Mean method is a simple and easy-to-use method. Each cluster is composed based on average values, and the values/elements attached to a cluster are the closest to this average. However, the same authors [23] also pointed out two major disadvantages of the method: sensitivity to extreme values and the lack of knowledge of the number of clusters. Medellu and Nugraha [24] described K-Means as a method that groups the values from a database into certain groups so that the level of similarity of the data within a group is as high as possible, whereas when compared to the other groups it is as small as possible. The level of similarity is determined based on the distance between the values of each element in the group and the group mean. The authors emphasized the simplicity of the method [24].

The K-Mean clustering method was analyzed by Ikotun et al. [25] in terms of advantages, improvements of the classic method, strengths and weaknesses of the existing implementation of hybrid K-means based on nature-inspired metaheuristic algorithms and identifying recent trends. The same authors presented different algorithms and optimizations which can be useful for the research community [25].

The K-Mean grouping method was used by Hernant et al. [26] to compare supermarkets based on various financial indicators. Kramaric et al. [27] also used the K-Mean clustering method to compare insurance companies based on the ROE financial indicator. The K-Mean cluster analysis and principal component analysis were applied to classify EU-25 countries and provide a comparative view of the interplay between digital entrepreneurship and sustainable development variables [28]. This method is considered the most popular clustering technique for partitioning a given dataset into a set of k-groups (clusters) [29].

Compared to K-Mean, the K-Medoid method works with representative values for each cluster. Instead of averages, the K-Medoid method selects representative values which are the central values in the cluster. The elements/values in the database are associated with those clusters whose medoids are most similar. Kaur et al. [23] defined the medoid as “the object of a cluster, whose average dissimilarity to all the objects in the cluster is minimal i.e., it is a most centrally located point in the given data set” (p. 43). The working method is the following: from the multitude of data, a certain number of medoids are chosen at random, around which the groups are formed with the elements that are most similar to the representative values. Once these clusters are formed, new medoids are chosen that better represent the formed group. It continues until no medoid changes its position [30]. This method overcomes perhaps the biggest disadvantage of the K-Mean method, i.e., the sensitivity to extreme values. It is therefore based on the representative/central values of the groups called medoids [24].

The two most popular clustering methods, K-Medoid and K-Mean, were analyzed by Arora et al. [31] and, as a result, they affirmed that the runtime is better for the K-Medoid method than for K-Mean, and also that K-Medoid is non-sensitive to outliers. Contrary to these results, Velmurugan [32] also compared K-Medoid and K-Mean clustering algorithms and the experimental results showed that the K-Mean algorithm yields the best results

compared with the K-Medoid algorithm. Moreover, Dsouza et al. [33] highlighted the positive and negative aspects of these two methods, concluding that K-Medoid is better in all aspects such as execution time, being non-sensitive to outliers and reduction of noise, but with the limitation that the complexity is greater as compared to K-Means.

Summarizing the opinions in the literature, the advantages and disadvantages of the K-Mean and K-Medoid methods are illustrated in Table 1.

Table 1. Advantages and disadvantages of K-Mean and K-Medoid methods.

	K-Mean Method		K-Medoid Method	
Advantages	✓	Well-known (popular) and commonly used method [29,33]	✓	Well-known and commonly used method [33]
	✓	Simple and easy to use method [23,24,33]	✓	Less sensitive to outliers [31,33]
			✓	Execution time per cluster is less in comparison with K-Mean [23,31,33]
Disadvantages	•	Execution time per cluster is more in comparison with K-Medoid [23,31,33]	•	Number of clusters is unknown
	•	It is sensitive to different distinct values [23,33]	•	Complexity is high as compared to K-Means [31,33].
	•	Number of clusters is unknown		

In both methods, the definition of the cluster number gets a lot of weight. There are several methods and procedures for this; the most frequently applied methods are the cluster elbow method and the silhouette method [29,34,35]. Another technique that can solve the problem of detecting the number of clusters is using VAT (Visual Assessment of Tendency) for clustering [36] and hierarchical cluster analysis. Kodinariya et al. [37] described the different cluster number determination methods applied to the K-Mean clustering algorithm in detail. Of the abovementioned methods, the Elbow method is considered to be the oldest and most commonly used. This is a visual method, the essence of which is to increase the number of groups by 1 to 1, starting with two. For each step we can see the variance, and when the graph shows a fracture (elbow) we can stop, as we have the optimal cluster number; the figure may show the break point (Elbow criterion: that is the elbow of this curve) where a “jump” occurs in the degree of heterogeneity [38].

3. Materials and Methods

3.1. Data and Sample

The database used for this study includes two complete databases, one for Hajdu-Bihar County in Hungary and one for Cluj County, Romania. Those databases include complete financial statements (balance sheet and income statement) for a period of five consecutive years for all active companies operating under the business activity code 4711—Retail sale in non-specialized stores with predominant sale of products food, beverages and tobacco. The financial statements were taken from the OPTEN database (a service providing complete information about economic entities) for Hungary, and for Romania they were received directly from the General Regional Directorate of Public Finances Cluj. Thus, the database includes 246 companies from Hungary and 1020 companies from Romania. In the study conducted on the financial performance of these companies, for the analysis we worked only with relative indicators because:

- the database includes all companies in this field of activity, from the smallest companies to large enterprises (in terms of turnover, value of assets, number of employees, etc.), so the absolute values would not have been representative and comparable.
- the financial statements were prepared in two different currencies (Hungarian forint and Romanian lei), so, by using the relative indicators, the results could be compared.

From the two databases, during the analysis and even in the first phase of data processing, the companies that recorded no revenue/sales (for the entire period analyzed, which means that they had no activity) were eliminated. There are many companies that no longer have activity, but have not officially closed their business and still submit financial

statements. For these companies, financial performance is not relevant (zero). Therefore, the companies with incomplete data and/or those with zero net sales were eliminated from the analysis. Thus, only 690 Romanian companies and 211 Hungarian companies were taken into account for the statistical analysis (see Table 2).

Table 2. Descriptive statistics of samples.

Variables	Hungarian Companies Sample (N = 211)				Romanian Companies Sample (N = 690)			
	Minimum	Maximum	Mean	Std. Deviation	Minimum	Maximum	Mean	Std. Deviation
ROS	−68.37	305.26	0.53	23.92	−34997.44	220.91	−65.48	1333.48
ROA	−657.22	640.85	−5.72	77.26	−4979.95	30459.91	−9.00	1208.53
ROE	−18442.93	529.92	−134.35	1297.74	−11850.59	14922.08	−11.51	917.09

Note: ROS—Return on Sales (%); ROA—Return on Assets (%); ROE—Return on Equity (%). Source: Based on own calculations.

3.2. Financial Performance Variables

Starting from the fact that we have analyzed the financial performance of companies which are very different in terms of size, sales and number of employees, we only worked with relative indicators for group breakdown. To evaluate the financial performance, we chose to analyze only relative financial ratios. So, as grouping criteria we chose the most common profitability ratios: ROS—Return on Sales (for evaluating cost management), ROA—Return on Assets (for assets efficiency) and ROE—Return on Equity (for efficiency of invested resources) [39,40]. These were used to evaluate the financial performance of the companies. These three ratios are calculated based on the formulas included in Table 3. All analyzed variables are numerical (percent).

Table 3. Financial performance indicators.

Indicators	Formula	Indicators (Mean Values) ¹ [41]			Recommended Values
		Italy	Sweden	Switzerland	
ROS—Return on Sales	$\frac{\text{Net Income}}{\text{Total Sales}} \times 100$	5.43%	5.92%	6.14%	Positive value; ROS > 5%
ROA—Return on Assets	$\frac{\text{Net Income}}{\text{Total Assets}} \times 100$	5.27%	3.80%	6.33%	Positive value; ROA > 5%
ROE—Return on Equity	$\frac{\text{Net Income}}{\text{Total Equity}} \times 100$	8.13%	7.29%	6.90%	Positive value; ROE > 10%

Note: ¹ The average values for ROS, ROA, ROE of companies in Italy, Sweden and Switzerland for three activity sectors (manufacturing, construction and other services, professional activities) based on [41].

Herciu et al. [40] considered that the ROA indicator, which reflects the profitability of using assets, should be at least 5% in every case. Regarding ROE, which is the most important indicator for investors and which reflects how effectively a company's management uses investors' money, most professional investors are looking for investments that have a return over 15% [40].

Hatem [41] conducted an international comparison of companies from three countries using ROS, ROA, ROE and other indicators. The average values determined in his study for ROS, ROA and ROE in three European countries (Italy, Sweden and Switzerland) for three activity sectors (manufacturing, construction and other services, and professional activities) are illustrated in Table 3. Nguyen et al. [42] analyzed the profitability of 58 listed companies in Vietnam, also based on ROS, ROA and ROE financial indicators, and the mean values for those companies were 90% for ROS, 1.50% for ROA and 3.94% for ROE.

For the top 10 retailers in the world, Deloitte [43] synthesized and analyzed performance indicators, of which, for ROA, the following values were recorded: Walmart Inc.—6.4%, Amazon.com Inc.—5.1%, Costco Wholesale Corporation—8.2%, The Kroger Co.—3.3%, Walgreens Boots Alliance, Inc.—5.9%, The Home Depot, Inc.—21.9%, Tesco PLC—1.9%.

For Romanian listed companies, Popa et al. [44] studied the ROA and ROE indicators and arrived at the following values: a mean of 2% for ROA, minimum 147%, maximum 29%; and a mean of 8% for ROE, minimum -1.201% , maximum 1.013%.

As shown above, the ROS, ROA and ROE values are very different, but in all cases the first requirement for these indicators is that they should be all positive, and it is recommended that they be over 5% for ROS and ROA, and at least 10% and increasing in value for ROE (Table 3).

Furthermore, there are many studies and economic–financial analyses which mainly use these three indicators to evaluate the financial performance of companies in various economic activities: in general, for listed firms [45–47], non-financial listed companies [34,48], transport and warehouses firms [39], agriculture [49,50], cosmetics industry [51], food and beverage [52], automotive industry [53] and other industries [54,55].

Popa et al. [44] also selected ROA and ROE indicators (in addition to six others) to build a composite financial index to determine the financial performance of listed companies. Pelloneova [56] used ROE (in determining EVA), ROA and ROS financial indicators to compare the financial performance of selected companies included in different clusters in the Czech and Slovak Republics. Due to the differences between the companies, Afrimayani and Devianto [57] (in terms of stock prices) and de Lima et al. [58] (in terms of financial performance) used the clustering method and ROA and ROE financial indicators to compare the financial performance of listed companies.

3.3. Statistical Methods

The starting point for the statistical analysis was the calculation of some basic statistical indicators, such as mean, standard deviation, minimum, maximum, skewness and kurtosis. These were chosen to examine the homogeneity of the database values. The results justified the need for group breakdown; the elements of the database were heterogeneous, showing large differences from each other. Following the formation of clusters, homogeneous groups were obtained. For homogeneous grouping, a cluster analysis with two widely used non-hierarchical grouping methods was used, namely the K-Mean and K-Medoid methods.

The clustering procedures can be hierarchical (tree-like structure) or non-hierarchical. The hierarchical method forms clusters gradually, and consequently this method has very high process cost because all objects are compared before every clustering step [59]. In the case of a big database, it can lead to a significantly high execution time. The algorithms of the K-Mean or K-Medoid methods (which are non-hierarchical methods) “generally change centers until all points are related to centers” [59] (p. 7). When compared with hierarchical classification, non-hierarchical classification is characterized by low cost in terms of calculation time [59]. Besides the execution time, in the case of larger samples the interpretation and use of the results of hierarchical cluster analysis are significantly more complicated. Therefore, it is advisable to use the K-Means method [60] or the K-Medoid method.

Taking into account the above aspects, and bearing the main purpose of this study in mind, we focused only on the two non-hierarchical methods to compare the results of the K-Mean and K-Medoid methods in evaluating financial performance in the case of two large databases containing data related to three main financial indicators (ROS, ROA, ROE) for five years of active retail food companies from one Hungarian and one Romanian county. The grouping of companies was based on the ROS (Return on Sales), ROA (Return on Assets) and ROE (Return on Equity) ratios, in both cases. We worked with predefined group numbers which were determined using the Elbow method (described in the previous section). Additionally, in order to find out the number of clusters, for both databases we applied hierarchical cluster analysis. Ward’s method and the squared Euclidian distance were employed. Finally, in the case of the Hungarian database, we chose five clusters, while in the case of the Romanian database we decided on fifteen clusters.

In the final step, we compared the specific data obtained with the K-Mean and K-Medoid clustering methods separately for each database, based on the five-year average values of the groups.

The statistical analysis of the databases and the editing of the graphs and diagrams were performed using Microsoft Excel spreadsheets, IBM SPSS Statistics 26.0 (IBM, Armonk, NY, USA) and R statistical software (version R, 3.5.0, R Core Team/R Foundation for Statistical Computing, Vienna, Austria).

4. Research Results

4.1. Descriptive Statistical Results

As shown in the previous section (Table 2), based on the descriptive analysis (mean, minimum—maximum values, standard deviation), the statistical results clearly show that the elements are heterogeneous, with a very large range of data for all financial indicators. Thus, in the case of the Hungarian database (N = 211), as well as in the case of the Romanian database (N = 690), very high differences between minimum and maximum values are noticed for all three financial indicators (ROS, ROA and ROE).

Therefore, before beginning the grouping process, outliers were identified. An “outlier” is defined by Hodge and Austin [61] as “one that appears to deviate markedly from other members of the sample in which it occurs” (p. 85). According to Hawkings [62], “outliers deviate significantly from the expectations”.

Extreme values were identified step by step using the BoxPlot chart, separately for each database and separately for each indicator (ROS, ROA, ROE). Many authors describe the BoxPlot diagram method as a possible (visually) method to detect outliers [61,63–65]. The BoxPlot chart helped us visually identify values that are outside the “normal” “value clouds”, values that can distort and tilt the group average.

Based on several boxplot charts, and eliminating the outliers step by step, we have reached the following minimum and maximum values for the three financial indicators (ROS, ROA, ROE) separately for the Romanian and the Hungarian companies (see Tables 4 and 5, Figures 1 and 2). For the Hungarian companies, the minimum and maximum values are as follows: for ROS, −36.36% and +32.00%; for ROA, −83.17% and +55.85%; and for ROE, −173.88% and +127.87%. The values of the financial indicators in the case of the Romanian companies are: −174.19% and +63.97% (for ROS); −370.43% and +148.75% (for ROA); and −487.92% and +216.63% (for ROE).

Table 4. Descriptive Statistics (N = 190)—Hungarian companies sample (without outliers).

	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis		Pearson Correlation (r)		
					Std. Error	Std. Error	ROS	ROA	ROE		
ROS	−36.36	32.00	−0.18	8.11	−0.93	0.17	6.14	0.35	1	0.66 *	0.39 *
ROA	−83.17	55.85	0.00	16.50	−1.47	0.17	5.85	0.35	0.66 *	1	0.61 *
ROE	−173.88	127.87	−4.94	45.08	−1.33	0.17	3.83	0.35	0.39 *	0.61 *	1

* Correlation is significant at the 0.01 level (2-tailed).

Table 5. Descriptive Statistics (N = 640)—Romanian companies sample (without outliers).

	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis		Pearson Correlation (r)		
					Std. Error	Std. Error	ROS	ROA	ROE		
ROS	−174.19	63.97	−7.59	24.43	−3.05	0.09	15.62	0.19	1	0.39 **	0.09 *
ROA	−370.43	148.75	−18.84	51.63	−3.07	0.09	13.13	0.19	0.39 **	1	0.09 *
ROE	−487.92	216.63	−25.19	70.68	−2.67	0.09	13.05	0.19	0.09 *	0.09 *	1

** Correlation is significant at the 0.01 level (2-tailed); * Correlation is significant at the 0.05 level (2-tailed).

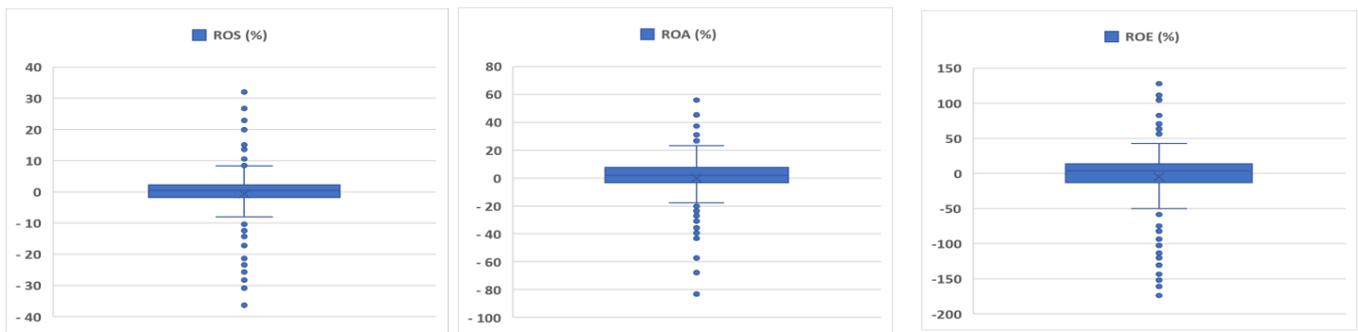


Figure 1. Distribution of the ROS, ROA and ROE averages (%) for Hungarian companies without outliers. Source: Authors' own editing, based on own calculations.

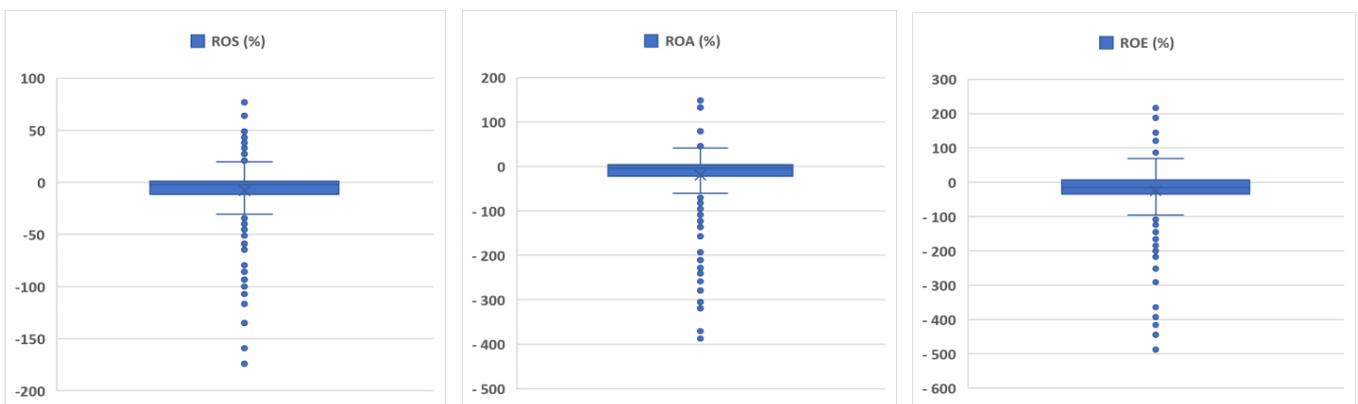


Figure 2. Distribution of the ROS, ROA and ROE averages (%) for Romanian companies without outliers. Source: Authors' own editing, based on own calculations.

The distribution of the mean values for ROS (Return on Sales), ROA (Return on Assets) and ROE (Return on Equity) without outliers is shown in Figures 1 and 2.

In the end, the numbers of the remaining companies in the two databases were as follows: the Romanian database contained 640, while the Hungarian database contained 190 companies.

The kurtosis and skewness indicators (Tables 4 and 5) also support what the previous indicators have shown: the companies are very different from the normal distribution. Examining the values of these statistical indicators, we can definitely conclude that the companies differ greatly from the normal distribution for both countries. These results confirmed our initial assumption that the differences between the individuals (companies) in the database are significant, so homogeneous grouping is required.

The results of the correlation analysis (Table 4) highlighted that, in the case of Hungarian companies, ROS was positively correlated with both the ROA ($r = 0.668$) and ROE ($r = 0.395$). The same positive correlation (but of a lower intensity) was seen between all three indicators in the case of Romanian companies (Table 5), confirming that the level of cost management, the efficiency of assets and the efficiency of invested resources are interlinked (Figure 3).

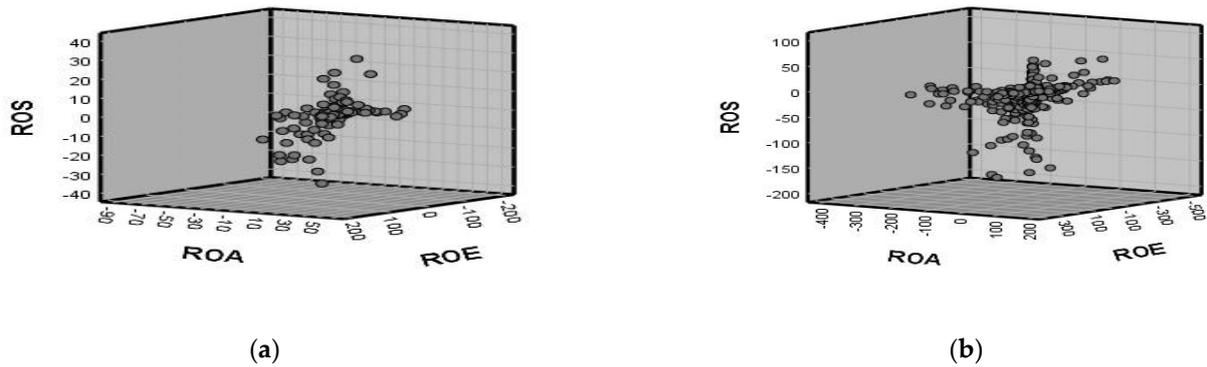


Figure 3. Link between financial performance indicators: (a) Hungarian companies. (b) Romanian companies. Source: Authors’ own editing, based on own calculations.

4.2. Cluster Analysis Results: An Overall Picture

The Figure 4a,b show the results of the Elbow method used to determine the cluster numbers (plotting of cluster numbers can show the break point—Elbow criterion, which represents the optimal cluster number). As a result, five clusters were used for Hungarian companies and fifteen for Romanian companies.

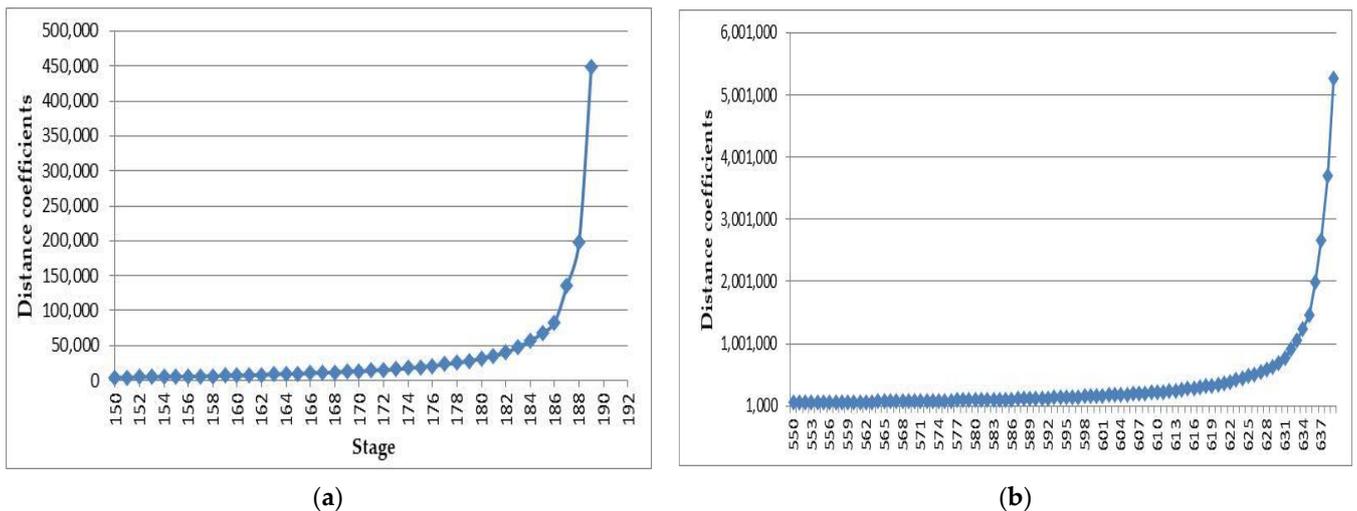


Figure 4. Number of clusters based on Elbow method: (a) Hungarian companies. (b) Romanian companies. Source: Authors’ own editing, based on own calculations.

Furthermore, the dendrograms (see Figure 5a,b), obtained using hierarchical cluster analysis (Ward’s method and squared Euclidian distance), have confirmed the number of clusters, five in the case of Hungarian companies, and fifteen in the case of Romanian companies.

In the next step, the structure of the clusters was based on both the K-Mean and the K-Medoid method. We have worked with the same number of clusters in the case of the K-Mean method as well as the K-Medoid method mainly due to the aim of this study, i.e., to perform a comparative analysis of the financial performance of companies based on the results of the two methods, focusing on the difference between them.

We can see quite a big difference in the number of companies included in the K-Mean and K-Medoid groups, as shown in Figure 6a,b. As can be seen, the K-Mean method has a much more varied number of companies within the group, and the K-Medoid has a more even distribution of companies, both for Hungarian (the database with fewer units) and Romanian companies (the bigger database).

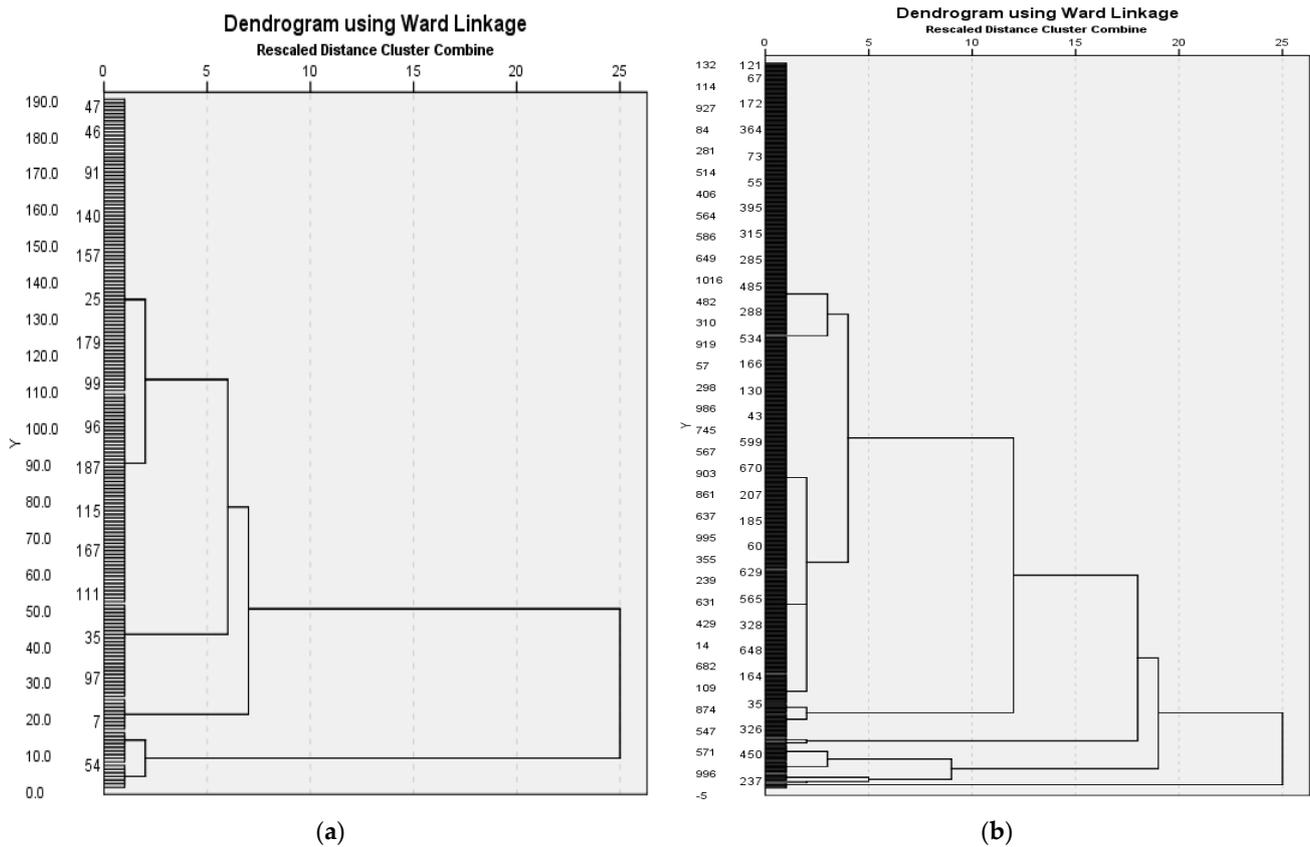
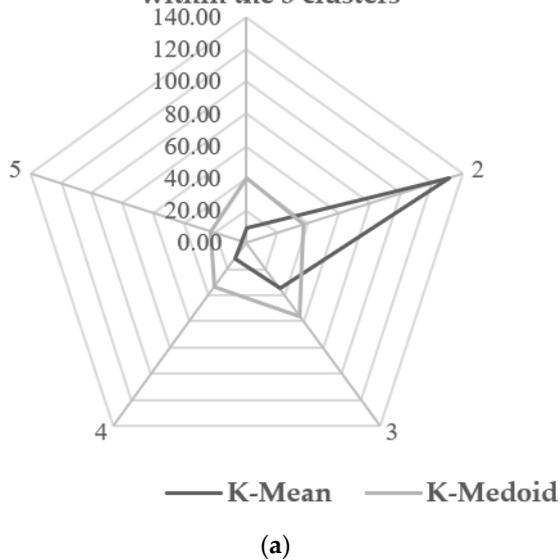


Figure 5. (a) Dendrogram using Ward Linkage method for Hungarian database. (b) Dendrogram using Ward Linkage method for Romanian database. Source: Authors’ own editing, based on own calculations.

Number of units for Hungarian database within the 5 clusters



Number of units for Romanian database within the 15 clusters

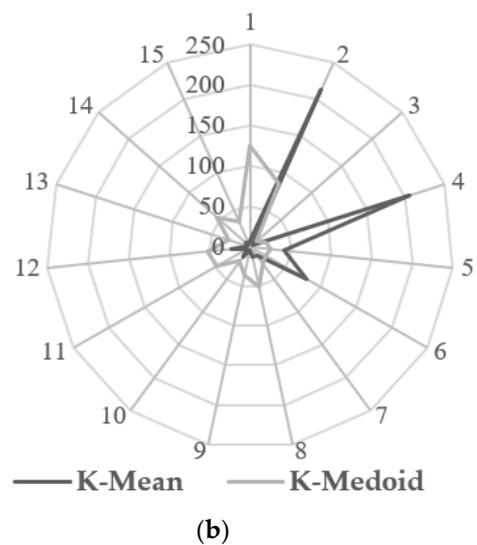


Figure 6. (a) Number of companies for Hungarian database within the clusters. (b) Number of companies for Romanian database within the clusters. Source: Authors’ own editing, based on own calculations.

The results obtained using the K-Mean method for the Hungarian companies highlighted that the number of companies included in cluster no. 5 (3 companies) is only 1.58% of the total companies (190 companies), while the number of companies included in cluster

no. 2 (131 companies), which contains more enterprises, is 68.95% of the total companies (190 companies). In contrast, the numbers of companies in the groups obtained by the group breakdown using the K-Medoid method is much more balanced: the group with the fewest companies included contained 12.11% (cluster no. 5 with 23 companies) of total companies (190 companies) and the group with the most companies included contained 29.47% (cluster no. 2 with 56 companies) of the total companies (190).

The results are similar for the Romanian companies, too: the K-Mean method has a much more varied number of companies within the group, compared to the K-Medoid method which has a more even distribution of companies. The number of Romanian companies included in the database was much higher (640 companies) and the differences between them were also larger, so we worked with 15 clusters. Grouping the companies with the K-Mean method, we obtained 5 clusters (clusters no. 3, 9, 11, 13, 15) with less than 1% of the companies included in these groups (which means less than 7 companies) from the total companies (640), and two clusters (clusters no. 2 and 4) with 32% (205 companies) and 33% (214 companies) of the companies included from the total.

4.3. Comparative Analysis of Financial Performance Results: K-Mean vs. K-Medoid

The average values of ROS (Return on Sales), ROA (Return on Assets) and ROE (Return on Equity) for the five groups of the Hungarian database (190 companies) are shown in Table 6.

Table 6. Average financial indicator values per cluster (%) for Hungarian companies.

Clusters	K-Mean			Clusters	K-Medoid				
	N ¹	ROS	ROA		ROE	N ¹	ROS	ROA	ROE
1	9	6.20	19.37	85.26	1	40	5.48	16.04	30.95
2	131	2.35	5.38	9.44	2	37	3.56	5.64	16.50
3	35	-9.61	-17.16	-30.22	3	56	0.56	1.39	2.89
4	12	-2.39	-5.79	-120.94	4	34	-2.71	-5.15	-50.43
5	3	-11.51	-69.47	-145.06	5	23	-14.18	-32.72	-53.73

Note: ¹ Number of companies in cluster. Source: Based on own calculations.

Also, the differences between the results obtained from the K-Mean and K-Medoid clustering methods are clearly illustrated in Figure 7. Thus, the results obtained based on the K-Mean method better reflect the “fluctuating” values: for each of the three ratios, the highest and lowest values are higher than the values obtained based on the K-Medoid method (ROE: 85.26% vs. 30.95% and -145.06% vs. -53.73%).

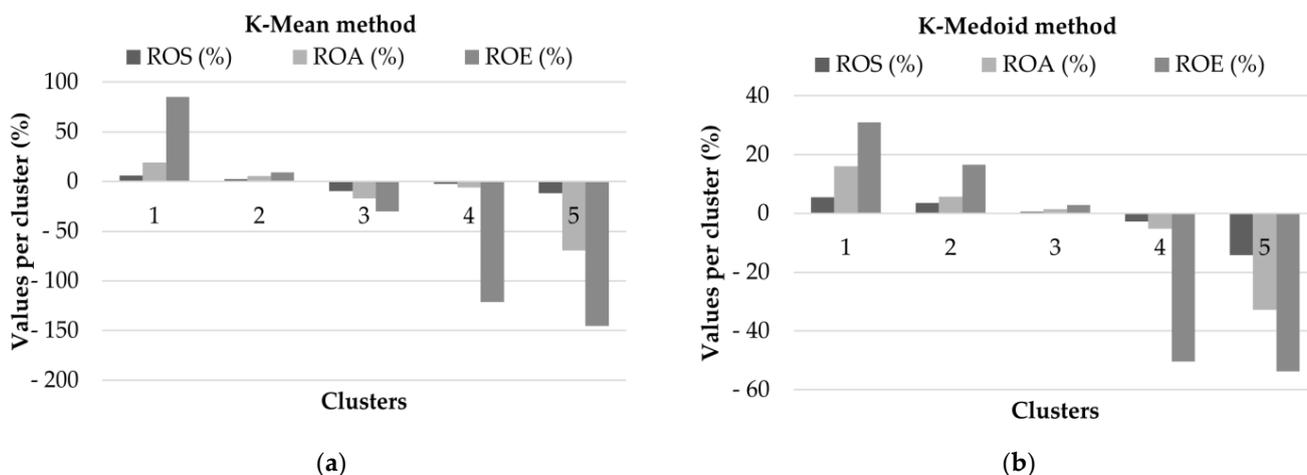


Figure 7. Average financial indicator values per cluster for Hungarian companies: (a) K-Mean method; (b) K-Medoid method. Source: Authors’ own editing based on own calculations.

Evaluating the financial results, it can be seen that there are two (K-Mean) and three (K-Medoid) groups with ROS, ROA and ROE financial ratios in the positive interval; this means that 73% of the companies grouped based on the K-Mean method and 70% of the companies grouped by the K-Medoid method have positive financial ratios. The fact that the indicators (about 70% of companies) have a positive value is the main requirement for these indicators and, obviously, the higher they are, the better the financial performance is. If compared with the existing values in the specialized literature, these positive values fall within the “normal/existing” values, but the problem appears in the case of clusters (companies) that register negative values.

The most visible differences are in case of the ROE indicator for all five clusters.

The number of Romanian companies included in the database and the differences between them were also larger, so we worked with fifteen clusters. The results are shown in Table 7 and Figure 8.

Table 7. Average financial indicator values per cluster (%) for Romanian companies.

K-Mean					K-Medoid				
Cluster	N ¹	ROS	ROA	ROE	Cluster	N ¹	ROS	ROA	ROE
1	7	2.96	2.92	167.4	1	126	8.59	12.14	34.61
2	214	5.51	8.01	18.99	2	85	1.94	1.95	9.48
3	3	20.82	120.44	-15.43	3	10	11.15	40.67	-18.22
4	205	-7.23	-8.41	-23.74	4	19	3.55	12.93	-54.02
5	42	-3.33	-3.14	-89.02	5	26	-1.20	-0.93	-8.68
6	80	-19.05	-61.59	-22.85	6	21	0.25	-1.44	-22.25
7	14	-98.36	-26.08	-33.01	7	26	-3.46	-12.04	-9.95
8	12	-20.39	-48.12	-113.77	8	51	-13.40	-14.18	-22.76
9	5	-137.43	-129.34	-27.34	9	36	-2.83	-7.45	-30.51
10	15	-23.77	-150.96	-21.00	10	22	-5.20	-6.44	-56.67
11	4	-20.43	-201.70	-85.19	11	48	-10.29	-49.15	-18.72
12	23	-1.83	-6.62	-204.89	12	52	-54.76	-30.63	-28.66
13	4	7.11	-2.55	-388.49	13	28	-7.58	-16.47	-99.20
14	8	-14.57	-272.52	-22.52	14	55	-31.27	-151.30	-32.74
15	4	0.35	7.32	-453.45	15	35	0.05	-4.95	-247.36

Note: ¹ Number of companies in cluster. Source: Based on own calculations.

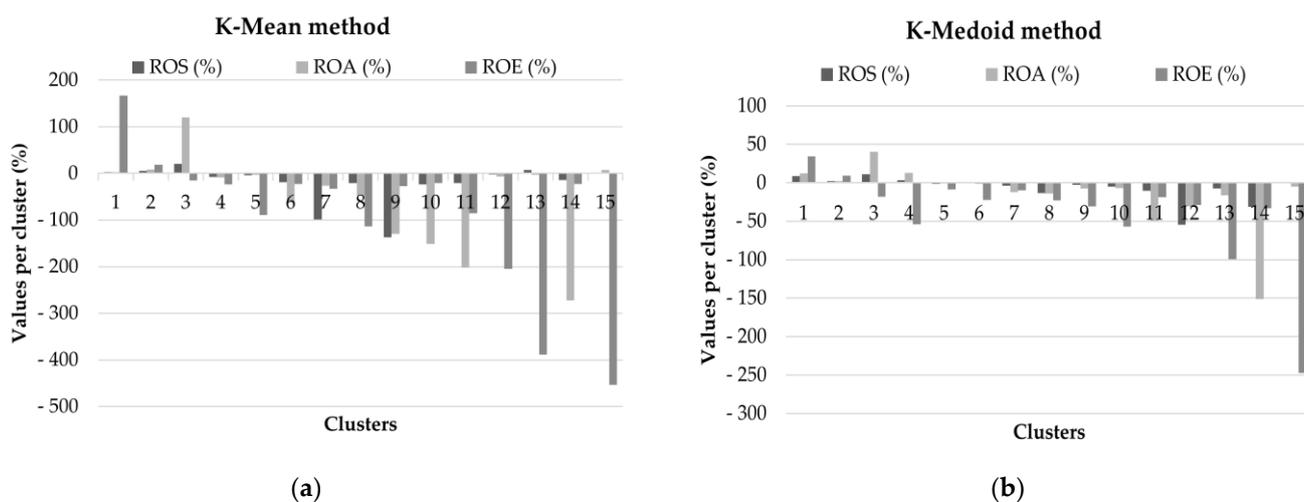


Figure 8. Average financial indicator values per cluster for Romanian companies: (a) K-Mean method; (b) K-Medoid method. Source: Authors’ own editing based on own calculations.

The characteristics of the Hungarian database are also present here: the results obtained with the K-Mean method better reflect values that differ significantly from the average, while the K-Medoid method compensates the fluctuations. The range of values

for the groups obtained with the K-Mean method is much larger than in the case of the groups obtained with the K-Medoid method.

The financial performance of Romanian companies is much weaker than that of Hungarian companies, with a lot of negative results (negative net profit), directly affecting ROS, ROA and ROE values. Thus, positive financial ratios were obtained by three groups containing about 35% of the analyzed companies in the case of the K-Mean grouping method, and by four groups containing 37.5% of the total companies when using the K-Medoid grouping method. Most companies recorded negative values for the three indicators, which means that the net result is negative, i.e., the activity is not profitable. This phenomenon is repeated every year, a fact which leads to a negative value for the equity of the enterprises. Although there is a legal provision in Romania regarding the increase of equity in case of consecutive losses [66], it seems that this legal provision is not applied by companies.

We grouped financial performance in three categories, namely *Good* (all three financial indicators—ROS, ROA and ROE—are above +5% or more), *Acceptable or Weak* (the financial ratios are all positive but close to zero) and *Negative* (all three financial indicators are below zero, i.e., they all have negative values). In this regard, the assessment of the financial performance of Hungarian and Romanian commercial enterprises is included in Table 8 (the proportion of enterprises in different categories).

Table 8. Financial performance of Hungarian and Romanian food retail companies.

ROS, ROA, ROE	Financial Performance of Hungarian Companies		Financial Performance of Romanian Companies	
	K-Mean	K-Medoid	K-Mean	K-Medoid
Good	4.74%	40.53%	33.44%	19.69%
Acceptable/Weak	68.95%	29.47%	1.56%	17.81%
Negative	26.31%	30.00%	65.00%	62.50%
Total	100.00%	100.00%	100.00%	100.00%

Source: Based on own calculations.

As shown in the previous table, the results for financial performance do differ: for Hungarian companies, when applying the K-Mean method, 68.95% of companies have an acceptable financial performance, while in the case of K-Medoid only 29.47%; for Romanian companies 1.56% of companies have an acceptable financial performance when applying the K-Mean method and 17.81% in the case of K-Medoid.

5. Discussion and Conclusions

The fundamental research hypothesis of this paper was as follows: in the case of large databases with variable quantitative data, cluster analysis is necessary in order to obtain accurate results and the clustering method chosen can yield different results.

Therefore, this study has pointed out that the results obtained using the two grouping methods (K-Mean and K-Medoid) are different. These findings are in line with other research papers [23,24,31,33] that have highlighted the advantages and disadvantages of the two methods. Moreover, our analysis identified significant differences between the results of these non-hierarchical clustering methods in terms of intra-group numbers: they are more uniform in case of the K-Medoid method than in case of the K-Mean method. Furthermore, this study emphasizes that the K-Medoid method is less sensitive to extremes and that it has better distribution within the group, with more balanced values. In the case of the K-Mean method, the grouping procedure is more sensitive and the range between the values of the units in the group is larger; at the same time, it gives a more detailed picture of the phenomenon under investigation.

Based on these results and the hypotheses assumed at the beginning of the research, we can state the following:

1. Processing the financial statements of the companies included in the database, we analyzed the most common statistical indicators (average, dispersion, deviation, standard deviation, kurtosis, skewness) for the financial performance indicators (ROS, ROA and ROE). The results have shown clearly that the companies are heterogeneous with a very large range of data for all indicators. An analysis of the standard deviation values was used to show the variability of the data and the differences between the individual data. These results confirmed our first hypothesis that the database contains highly variable elements.
2. The results of the descriptive statistical indicators confirmed our assumptions that we cannot examine the financial performance of the companies in the database based only on the average values of the financial indicators. Therefore, it is necessary to group the sample population into several homogeneous groups based on certain key indicators. The average values cannot be representative.
3. In order to achieve homogenous clusters, the key indicators were selected in accordance with the objectives of the research. Therefore, we selected the three main indicators having the most comprehensive characteristics of financial performance, which are the most accepted in the literature, too (ROS, ROA and ROE). The selection and determination of the indicators or characteristics based on which the groups will be formed are very important, as they can influence the obtained results. These indicators or characteristics must reflect the aim of the research. Of course, selecting other indicators can lead to a different group composition.
4. The results of the two applied grouping methods (K-Mean and K-Medoid) supported our main assumption that the chosen and applied method leads to different results. Thus, the results obtained with the K-Mean method better reflect values that differ significantly from the average (they better reflect the “fluctuating” values), while the K-Medoid method compensates the fluctuations. This characteristic manifested itself in both cases, both for companies in Hungary and for companies in Romania.
5. The focus of the research was not on the difference between the two “countries” but on the analysis of two “different databases”. As we have shown, the study identified significant differences both in the number of elements of the two databases and in the homogeneity of the databases. Therefore, there is a need for grouping the similar elements of a database. For smaller, more homogeneous, and larger, heterogeneous databases, the two grouping methods (K-Mean and K-Medoid) lead to different results. In terms of procedure, the two databases could be merged, but the authors aimed to demonstrate that even for different data (smaller, larger, more homogeneous, or less homogeneous), there is a significant difference between the results of the two procedures.

All the hypotheses presented at the beginning were successfully supported by empirical data.

It is indisputable that the two different grouping methods applied to the same database give different results. For financial ratios, these methods led to different performance evaluations of food retail companies. In the case of a large number of elements in databases from a statistical point of view, taking into account that the K-Medoid method is less sensitive to extreme and fluctuating values, this method has created balanced numbers of the units within the groups (each group has more than 10% of the total database). For K-Mean, it was not uncommon to have “only a few units groups”, which means 1%, 2% or 3% of the total database (this also supports the sensitivity of the K-Mean method for outliers).

Even though outliers were eliminated from the database, there are still several swing-ing values in the database which are not statistically representative values. The essence of cluster analysis is to find the common features of the sample population. From this perspective, the authors recommend the K-Medoid method for statistical analysis of large databases, while the K-Mean clustering method is recommended for the evaluation of more detailed analyses that also focus on fluctuating values.

Moreover, our findings regarding the financial performance of food retail businesses prove that their performance is unfavorable and even very weak. Especially in the case of Romanian companies, there are a lot of enterprises that achieve negative results, which produce losses not only in a given year, but year after year. In such circumstances, the question arises of how these businesses will manage to survive in the future and possibly develop even in the current difficult circumstances.

The expansion of hypermarkets and supermarkets, globalization and the online market have pushed “traditional” food retailers into a difficult situation characterized by low long-term investments in these businesses, accumulated losses from previous years and very low levels of profit margins applied. Taking all these into account, the level of financial performance of “traditional” food retailers is very low. It is worth mentioning that in this industry there is a significant number of companies (approx. 28% of total active companies in Romania) and a large number of employees (almost 1 million employees) [67].

Such challenges need to be managed by “traditional” retail businesses in order to survive and even to gain competitive advantages over their main competitors. In contrast to large shopping centers, the “traditional” shops are much more customer-centered, have a more familiar atmosphere and geographical location benefits, offering local products to consumers, and provide more direct relationships [52]. As regards pricing policy, they cannot compete with modern chain stores, but they can offer higher quality and local products. Furthermore, an important way to improve “traditional” retail businesses is to stimulate consumers to buy local and traditional products, to focus on the shorter supply chains, and to consume organic products as much as possible. This strategy is quite well adopted in Hungary due to the fact that local products and the consumption of food products originating in Hungary are constantly promoted in the media. Also, these products carry a certain label/emblem so that consumers can identify local and national products on store shelves more easily.

In order to improve the financial performance of companies grouped in clusters based on the K-Medoid method (which is the clustering method recommended based on our results), the authors propose the following specific measures for company management:

1. For companies included in the clusters with negative values for all three financial indicators (in the case of Hungary, clusters 4 and 5 which contain 57 companies, representing 30% of the total companies from the Hungarian database; in the case of Romania, clusters 5 to 15, containing 400 companies, which represent 62.5% of the total companies from the Romanian database), our recommendation is an injection of capital from the owners, primarily to regain the financial stability of the company. Also, these companies should adopt strategies for attracting customers which can lead to positive results in the future (e.g., modernizing locations thus attracting customers, loyalty programs, offering quality and local products, which can support a higher commercial margin). These measures should also be aligned with efforts made by the authorities to require companies, and their owners, to protect a company's equity [66], using different measures (e.g., capital injection) in the case of longer-term and significant negative equity. It is worth mentioning that the very high level of negative equity is primarily the result of the annual accumulation of negative after-tax profit. Furthermore, taxation in Romania (in the case of small businesses) should be improved taking into consideration that the current form of taxation is not effective due to the fact that the tax paid by companies is calculated based on the result (profit), but not on revenue–sales, and that companies strive to achieve a minimal or even negative result in order to avoid taxation [68]. It is possible that this will have long-term positive effects on equity.
2. For companies whose financial indicators are close to zero, or where ROE is negative (cluster 3 from Hungary, containing 56 companies, which represents 29.47% of the total companies from the database, and clusters 3 and 4 from Romania, including 29 companies, which represent 4.53% of the total companies from the Romanian

database), capital injection can be the solution that provides immediate “oxygen” and can restore the financial balance of a company.

3. In the case of companies with positive and also good financial results (clusters 1 and 2 from Hungary, which contain 77 companies, representing 40.52% of the total Hungarian companies from the database, and clusters 1 and 2, which include 211 Romanian companies, representing 32.96% of the total Romanian companies from the database), keeping and/or achieving even better financial performance in the future requires some specific measures, such as: long-term investments, very good inventory management, larger and varied product offerings, faster service, engagement in local social and cultural life (e.g., participation at local fairs), tastings, product presentations, promotions and continuous market research in order to retain customers [69].

The *originality of the paper* consists in the fact that the databases on which the study was based represent 100% complete databases for the chosen “sample”, i.e., they include all the active companies within the chosen field of activity from the two regions. A unique comparison was made in terms of the financial performance of retailers in non-specialist shops with predominant sales of food, beverage and tobacco from two different countries. The study has shown that the method chosen (in this case K-Mean or K-Medoid) for grouping companies in the database can lead to different results in terms of assessing financial performance.

Limitations and Future Research

It should be noted that this study, like any other study, has certain limitations and can be continued and improved upon in certain directions, and also presents original elements, undisputed by other specialists.

As limitations of the study, we can mention the number of financial indicators used in the research (only the most common financial indicators were chosen: ROS, ROA, ROE); the analysis period (the financial statements of the companies on which the analysis was performed comprised only five consecutive years); that only the trade sector was chosen as the field of activity of analyzed companies (the companies in the database are only companies whose main activity is retail trade in non-specialized stores with a predominant sale of food, beverages and tobacco); and that large supermarkets and hypermarkets (which have the same field of activity) have their registered office usually in the capitals of the countries where they operate and thus do not appear in these databases (the chosen regions of the two countries do not include the capitals of the countries).

Bearing these limitations of the study in mind, the research can be extended in the future in the following directions: extending the analysis period by at least another 5 years, thus reaching an analysis period of 10 years, which may be more representative; choosing several areas of activity besides trade (e.g., production, services); using several methods and techniques, in terms of descriptive data analysis, factorial analysis, the formation of homogeneous groups and analysis of financial performance (performance appraisal criteria); and constructing a composite index to characterize financial performance (including ROS, ROA and ROE alongside other indicators of liquidity, solvency, financial balance indicators, trade margin, turnover rate, etc.).

Author Contributions: Conceptualization, E.H., K.-E.Z. and V.F.; methodology, E.H., K.-E.Z. and V.F.; software, E.H., K.-E.Z. and V.F.; validation, E.H., K.-E.Z. and V.F.; formal analysis, E.H., K.-E.Z. and V.F.; investigation, E.H., K.-E.Z. and V.F.; resources, E.H., K.-E.Z. and V.F.; data curation, E.H., K.-E.Z. and V.F.; writing—original draft preparation, E.H., K.-E.Z. and V.F.; writing—review and editing, E.H., K.-E.Z. and V.F.; visualization, E.H., K.-E.Z. and V.F.; supervision, E.H., K.-E.Z. and V.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Lukác, J.; Teplická, K.; Culková, K.; Hrehová, D. Evaluation of the financial performance of the municipalities in Slovakia in the context of multidimensional statistics. *J. Risk Financ. Manag.* **2021**, *14*, 570. [CrossRef]
- Madhulatha, T.S. An overview on clustering methods. *IOSR-JEN* **2012**, *2*, 719–725. [CrossRef]
- Soni, K.G.; Patel, A. Comparative Analysis of K-means and K-medoids algorithm on IRIS Data. *Int. J. Comput. Intell. Syst.* **2017**, *13*, 899–906.
- Velmurugan, T.; Santhanam, T. Computational complexity between K-Means and K-Medoids clustering algorithms for normal and uniform distributions of data points. *J. Comput. Sci.* **2010**, *6*, 363–368. [CrossRef]
- Park, H.S.; Jun, C.H. A simple and fast algorithm for K-medoids clustering. *Expert Syst. Appl.* **2009**, *36*, 3336–3341. [CrossRef]
- Ben-Gal, I. Outlier Detection. In *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, 2nd ed.; Maimon, O., Rockach, L., Eds.; Springer: New York, NY, USA, 2010; pp. 117–130. Available online: https://link.springer.com/chapter/10.1007/978-0-387-09823-4_7 (accessed on 18 March 2022).
- Iqbal, M.Z.; Riaz, M.; Nasir, W. Multivariate Outlier Detection: A comparison among two clustering techniques. *Pak. J. Agric. Sci.* **2017**, *54*, 227–231. [CrossRef]
- Hennig, C. What are the true clusters? *Pattern Recognit. Lett.* **2015**, *64*, 53–62. [CrossRef]
- Szabo, Z.K.; Herman, E. Productive entrepreneurship in the EU and its barriers in transition economies: A cluster analysis. *Acta Polytech. Hung.* **2014**, *11*, 73–94.
- Esfahani, R.K.; Shahbazi, F.; Akbarzadeh, M. Three-phase classification of an uninterrupted traffic flow: A k-means clustering study. *Transp. B-Transp. Dyn.* **2019**, *7*, 546–558. [CrossRef]
- Windgassen, S.; Moss-Morris, R.; Goldsmith, K.; Chalder, T. The importance of cluster analysis for enhancing clinical practice: An example from irritable bowel syndrome. *J. Ment. Health* **2018**, *27*, 94–96. [CrossRef]
- Pham, D.T.; Afify, A.A. Clustering techniques and their applications in engineering. *Proc. Inst. Mech. Eng. Part C-J. Eng. Mech. Eng. Sci.* **2007**, *221*, 1445–1459. [CrossRef]
- Yadegari, R.; Rahmani, K.; Modarres Khiyabani, F. Providing a comprehensive model to measure the performance dimensions of industrial clusters using the hybrid approach of Q factor analysis and cluster analysis. *Int. J. Qual. Res.* **2018**, *13*, 235–248. [CrossRef]
- Lemieux, V.; Rahmdel, P.S.; Walker, R.; Wong, B.L.W.; Flood, M. Clustering techniques and their effect on portfolio formation and risk analysis. In Proceedings of the DSMM'14: International Workshop on Data Science for Macro-Modeling, Snowbird, UT, USA, 22–27 June 2014; pp. 1–6. [CrossRef]
- Kou, G.; Peng, Y.; Wang, G. Evaluation of clustering algorithms for financial risk analysis using MCDM methods. *Inf. Sci.* **2014**, *275*, 1–12. [CrossRef]
- Xiaojun, S.; Yalin, L. Research on financial early warning of mining listed companies based on BP neural network model. *Resour. Policy* **2021**, *73*, 102223. [CrossRef]
- Wang, Y.J.; Lee, H.S. A clustering method to identify representative financial ratios. *Inf. Sci.* **2008**, *178*, 1087–1097. [CrossRef]
- Sabau, A.S. Survey of clustering based financial fraud detection research. *Inform. Econ.* **2012**, *16*, 110–122.
- Hepsen, A.; Vatanserver, M. Using hierarchical clustering algorithms for Turkish residential market. *Int. J. Econ. Financ.* **2012**, *4*, 138–150. [CrossRef]
- Popa, D.N.; Bogdan, V.; Sabau Popa, C.D.; Belenesi, M. Performance mapping in two-step cluster analysis through ESEG disclosures and EPS. *Kybernetes* **2021**, *51*, 98–118. [CrossRef]
- Cai, F.; Le-Khac, N.A.; Kechadi, M.T. Clustering Approaches for Financial Data Analysis. In Proceedings of the 8th International Conference on Data Mining (DMIN 2012), Las Vegas, NV, USA, 15–18 December 2012. Available online: https://www.researchgate.net/publication/278409705_Clustering_Approaches_for_Financial_Data_Analysis/ (accessed on 27 August 2019).
- Serban, E.C.; Bogueanu, A.; Tudor, E. Clustering techniques in financial data analysis applications on the U.S. financial market. *Ann. “Constantin Brâncuși” Univ. Târgu Jiu Econ. Ser.* **2013**, *4*, 176–194.
- Kaur, N.M.; Kaur, U.; Singh, D. K-medoid clustering algorithm—A review. *Int. J. Comput. Appl. Technol.* **2014**, *1*, 42–45.
- Medellu, J.V.C.; Nugraha, E.S. K-means and k-medoid algorithm application in clustering stock data in Indonesia. In Proceedings of the Symposium on Data Science 2021, Petra, Jordan, 5–7 April 2021. Available online: <http://e-journal.president.ac.id/presunivojs/index.php/SDS/article/view/1726/965> (accessed on 24 July 2022).
- Ikotun, A.M.; Almutari, M.S.; Ezugwu, A.E. K-Means-Based Nature-Inspired Metaheuristic Algorithms for Automatic Data Clustering Problems: Recent Advances and Future Directions. *Appl. Sci.* **2021**, *11*, 11246. [CrossRef]
- Hernant, M.; Andersson, T.B.; Hilmola, O.P. Managing retail chain profitability based on local competitive conditions: Preliminary analysis. *Int. J. Retail Distrib. Manag.* **2007**, *35*, 912–935. [CrossRef]
- Kramaric, T.P.; Bach, M.P.; Dumcic, K.; Zmuk, B.; Zaja, M.M. Exploratory study of insurance companies in selected post-transition countries: Non-hierarchical cluster analysis. *Cent. Eur. J. Oper. Res.* **2018**, *26*, 783–807. [CrossRef]

28. Herman, E. The interplay between digital entrepreneurship and sustainable development in the context of the EU digital economy: A multivariate analysis. *Mathematics* **2022**, *10*, 1682. [CrossRef]
29. Pech, M.; Vrchota, J. Classification of Small- and Medium-Sized Enterprises Based on the Level of Industry 4.0 Implementation. *Appl. Sci.* **2020**, *10*, 5150. [CrossRef]
30. Vishwakarma, S.; Nair, P.S.; Rao, D.S. A comparative study of K-means and K-medoid clustering for social media text mining. *Int. J. Adv. Sci. Res. Eng. Trends* **2017**, *2*, 297–302. Available online: http://ijasret.com/VolumeArticles/FullTextPDF/95_IJASRET-A_Comparative_Study_of_K-means_and_K-medoid_Clustering_for_Social_Media_Text_Mining.pdf (accessed on 24 July 2022).
31. Arora, P.; Deepali, D.; Varshney, S. Analysis of K-Means and K-Medoids algorithm for big data. *Procedia Comput. Sci.* **2016**, *78*, 507–512. [CrossRef]
32. Velmurugan, T. Efficiency of K-Means and K-Medoids algorithms for clustering arbitrary data points. *Int. J. Comput. Appl. Technol.* **2012**, *3*, 1758–1764.
33. Dsouza, S.; Dsouza, J.D.; Vanitha, T. Analysis of data using k-means and k-medoids algorithms. *Int. J. Latest Trends Eng. Technol. Spec. Issue SACAIM* **2017**, 370–373. Available online: <https://www.ijltet.org/journal/151065795883.pdf> (accessed on 24 July 2022).
34. Szüle, B. Comparison of cluster number determination methods (Klaszterszám-meghatározási módszerek összehasonlítása). *Stat. Szle.* **2019**, *97*, 421–438.
35. Dzuba, S.; Krylov, D. Cluster analysis of financial strategies of companies. *Mathematics* **2021**, *9*, 3192. [CrossRef]
36. Pakhira, M.K. Finding Number of Clusters before Finding Clusters. *Proc. Technol.* **2012**, *4*, 27–37. [CrossRef]
37. Kodinariya, T.M.; Makwana, P.R. Review on determining number of cluster in K-Means clustering. *Int. J. Adv. Res. Comput. Sci. Manag. Stud.* **2013**, *1*, 90–95.
38. Simon, J. Applications of cluster analysis in marketing research (A klaszterelemzés alkalmazási lehetőségei a marketingkutatóban). *Stat. Szle.* **2006**, *84*, 627–650.
39. Do, D.T. A study on financial performance of transport & warehouses firms listed on the Hanoi stock exchange. *Econ. Financ. Lett.* **2021**, *8*, 44–52. [CrossRef]
40. Herciu, M.; Ogrea, C.; Belascu, L.A. Du Pont Analysis of the 20 Most Profitable Companies in the World. In Proceedings of the 2010 International Conference on Business and Economics Research, Kuala Lumpur, Malaysia, 26–28 November 2010; Volume 1, pp. 45–48. Available online: <http://www.ipedr.com/vol1/10-B00015.pdf> (accessed on 18 June 2022).
41. Hatem, B.S. Determinants of firm performance: A comparison of European countries. *Int. J. Econ. Financ. (IJEF)* **2014**, *6*, 243–249. [CrossRef]
42. Nguyen, V.C.; Nguyen, T.N.L.; Tran, T.T.; Nghiem, T.T. The impact of financial leverage on the profitability of real estate companies: A study from Vietnam stock exchange. *Manag. Sci. Lett.* **2019**, *9*, 2315–2326. [CrossRef]
43. Deloitte. Global Powers of Retailing. 2021. Available online: <https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Consumer-Business/gx-global-power-retailing-2021.pdf> (accessed on 18 June 2022).
44. Sabău Popa, C.D.; Popa, D.N.; Bogdan, V.; Simut, R. Composite financial performance index prediction—A neural networks approach. *J. Bus. Econ. Manag.* **2021**, *22*, 277–296. [CrossRef]
45. Zhussupova, Z.; Onyusheva, I.; El-Hodiri, M. Corporate governance and firm value of Kazakhstani companies in the conditions of economic instability. *Pol. J. Manag. Stud.* **2018**, *17*, 235–245. [CrossRef]
46. Bayaraa, B. Financial performance determinants of organizations: The case of Mongolian companies. *J. Compet.* **2017**, *9*, 22–33. [CrossRef]
47. Yen, M.-F.; Huang, Y.-P.; Yu, L.-C.; Chen, Y.-L. A two-dimensional sentiment analysis of online public opinion and future financial performance of publicly listed companies. *Comput. Econ.* **2022**, *59*, 1677–1698. [CrossRef]
48. Akhtar, M.; Yusheng, K.; Haris, M.; Ul Ain, Q.; Javaid, H.M. Impact of financial leverage on sustainable growth, market performance, and profitability. *Econ. Chang. Restruct.* **2022**, *55*, 737–774. [CrossRef]
49. Goral, J.; Soliwoda, M. On the profitability of Polish large agricultural holdings. *Acta. Oecon.* **2021**, *71*, 137–159. [CrossRef]
50. Hloušková, Z.; Lekešová, M. Farm outcomes based on cluster analysis of compound farm evaluation. *Agric. Econ.-Czech* **2020**, *66*, 435–443. [CrossRef]
51. De Blasio, V.; Pavone, P.; Migliaccio, G. Cosmetics companies: Income developments in time of crisis. *J. Small Bus. Enterp. Dev.* **2022**. ahead-of-print. [CrossRef]
52. Fenyves, V.; Zsido, K.E.; Bircea, I.; Tarnoczi, T. Financial performance of Hungarian and Romanian retail food small businesses. *Br. Food J.* **2020**, *122*, 3451–3471. [CrossRef]
53. Zainudin, R.; Mahdzan, N.S.; Mohamad, N.N. Internationalisation and financial performance: In the case of global automotive firms. *Rev. Int. Bus. Strategy* **2021**, *31*, 80–102. [CrossRef]
54. Khour, S.; Elexa, L.; Istok, M.; Rosova, A. A Study from Slovakia on the transfer of Slovak companies to tax havens and their impact on the sustainability of the status of a business entity. *Sustainability* **2019**, *11*, 2803. [CrossRef]
55. Pavelkova, D.; Zizka, M.; Homolka, L.; Knapkova, A.; Pelloneova, N. Do clustered firms outperform the non-clustered? Evidence of financial performance in traditional industries. *Ekon. Istraz.* **2021**, *34*, 3270–3292. [CrossRef]
56. Pelloneová, N. Are there differences in the financial performance of Czech and Slovak cluster organizations? *Ekon. Cas.* **2021**, *69*, 907–927. [CrossRef]

57. Afrimayani; Devianto, D. The time series clustering of stock price in LQ45 index and its financial performance analysis. *J. Phys. Conf. Ser.* **2021**, *1943*, 1–8. Available online: <https://iopscience.iop.org/article/10.1088/1742-6596/1943/1/012129/pdf> (accessed on 18 June 2022). [[CrossRef](#)]
58. Gülağız, F.K.; Şahin, S. Comparison of hierarchical and non-hierarchical clustering algorithms. *Int. J. Comput. Eng. Inf. Technol.* **2017**, *9*, 6–14.
59. Sajtos, L.; Mitev, A. *SPSS Research and Data Analysis Manual (SPSS Kutatási és Adatelemzési Kézikönyv)*; Alinea Publishing House: Budapest, Hungary, 2007.
60. de Lima, C.R.M.; Barbosa, S.B.; de Castro Sobrosa Neto, R.; Bazil, D.G.; de Andrade Guerra, J.B.S.O. Corporate financial performance: A study based on the Carbon Efficient Index (ICO2) of Brazil stock exchange. *Environ. Dev. Sustain.* **2022**, *24*, 4323–4354. [[CrossRef](#)]
61. Hodge, V.J.; Austin, J. A Survey of Outlier Detection Methodologies. *Artif. Intell. Rev.* **2004**, *22*, 85–126. [[CrossRef](#)]
62. Hawkins, D. *Identification of Outliers*; Springer: Dordrecht, The Netherlands, 1980. [[CrossRef](#)]
63. Wang, H.; Jaward, M.B.; Hammad, M. Progress in outlier detection techniques: A survey. *IEEE Access* **2019**, *7*, 107964–108000. Available online: <https://ieeexplore.ieee.org/abstract/document/8786096> (accessed on 18 June 2022). [[CrossRef](#)]
64. Portela, E.; Ribeiro, R.P.; Gama, J. The search of conditional outliers. *Intell. Data Anal.* **2019**, *23*, 23–39. [[CrossRef](#)]
65. Ur Rehman, A.; Belhaouari, S.B. Unsupervised outlier detection in multidimensional data. *J. Big Data* **2021**, *8*, 1–27. [[CrossRef](#)]
66. Romanian Company Law 31/1990. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/56732> (accessed on 4 June 2022).
67. National Institute of Statistics from Romania. Statistical Data. Available online: <https://insse.ro/cms/en> (accessed on 27 July 2022).
68. Romanian Tax Code, Law no. 16 from 15 July 2022. Available online: <https://legislatie.just.ro/Public/DetaliuDocument/257589> (accessed on 27 July 2022).
69. Zsidó, K.E. Comparative Analysis of the Financial Performance for Retail Businesses in Hajdú-Bihar and Cluj County (Hajdú-Bihar és Kolozs Megyei Élelmiszer Jellegű Kiskereskedelmi Vállalkozások Pénzügyi Teljesítményének Összehasonlító Elemzése). Ph.D. Thesis, University of Debrecen, Károly Ihrig Doctoral School of Management and Business, Debrecen, Hungary, 2018. Available online: <https://dea.lib.unideb.hu/dea/handle/2437/2265/browse?value=Zsid%C3%B3%2C+Kinga+Emese&type=author> (accessed on 27 July 2022).