Hierarchical classification of water service operators in Romania: an integrated analysis of financial and technical indicators

Octavia MOISE[1]*

To cite this article:

Moise, O., (2024). Hierarchical classification of water service operators in Romania: an integrated analysis of financial and technical indicators. Romanian Journal of Economics, 59(2), pp. 206 - 223

Abstract. *Objective: The main objective of this study is to apply hierarchical classification techniques to systematically classify water service operators based on several performance indicators, thereby identifying distinct categories of operational performance, information that can be used for strategic planning and intervention in the supply system with water. Method: The methodological approach of this study involves a hierarchical clustering analysis, a technique that groups water service operators into clusters based on the similarities between several predefined indicators. The analysis was facilitated by the use of statistical software, which allowed a robust assessment of similarities and differences between operators. Results: The hierarchical clustering analysis revealed significant differences between water service operators in Romania, resulting in the identification of two main clusters that reflect distinct operational and financial particularities. Cluster 1 includes most operators, characterized by moderate efficiency in water resource management. Operators in this group show a relatively average Non-Revenue Water (NRW) rate, indicating significant but manageable water losses. Cluster 2 consists of a smaller number of operators, which are distinguished by lower operational performance. The NRW rate is significantly higher, indicating substantial water losses and major inefficiencies in the distribution system. By understanding the profile and needs of each cluster, decision-makers can allocate resources efficiently and implement solutions adapted to local realities, thus contributing to the long-term sustainability of water services. Originality: The original approach of this study is to use hierarchical classification to structure a complex set of data into a comprehensible and actionable form, thereby providing a new insight into the operational efficiency of water services. By identifying clear patterns of performance among water operators, the research contributes directly to the literature and supports the development of evidence-based public policy.*

Keywords*: hierarchical classification, water service providers, non-revenue water (NRW)*

JEL classification*: C10, Q53, R10*

[1] Romanian Academy, School of Advanced Studies of the Romanian Academy, Doctoral School of Economic Sciences, National Institute of Economic Research "Costin C. Kirițescu", Institute of National Economy, Bucharest, Romania, *Corresponding author: moctavia2005@yahoo.com

1. Introduction

Regarding the dynamics of the water supply and sewage growth process network in Romania due to the major investments in this system in recent years, the efficient management of water and sewage services becomes essential for urban and rural sustainability against the background of increased pressure on natural resources.

Romania, with variable infrastructure across different regions, faces significant challenges in ensuring the equitable and efficient distribution of drinking water, as well as managing the treatment of worn-out water. This research attempts to investigate the geographical distribution and operational efficiency of water services in Romania, identifying areas of strength and vulnerability in order to propose customized solutions.

This study's primary goal is to evaluate the performance and efficiency of water services in the territory of Romania, using an approach based on cluster analysis to identify and characterize the different types of water management systems according to key variables such as water consumption, infrastructure, damage incidence and counting methods. The aim is to provide a comprehensive and up-to-date database to support the formulation of regional policies and strategies adapted to the specifics of each identified cluster.

To carry out this analysis, the following research hypotheses were approached:

1. Heterogeneity Hypothesis: It is assumed that there are significant variations in the performance of water services between the different regions of Romania, reflected in the efficiency of resource use and the frequency of infrastructural failures.
2. Efficiency Assumption: Water systems benefiting from recent investments and modern technologies are assumed to have higher operational efficiency and lower rates of unmetered water (NRW).
3. Correlation Hypothesis: It is anticipated that the total length of water networks is directly correlated with the number of failures and the age of the infrastructure, suggesting that areas with extensive and old networks are more susceptible to failures.

Hierarchical classification is a data analysis technique often used in automatic learning and statistical analysis. This method structures data into a hierarchy of groups or classes, which reflect subordination and superordination relationships between different objects or data sets.

The analysis was facilitated by the use of statistical software, SPSS, which allowed a robust assessment of similarities and differences between operators. The application of Hierarchical Clustering Techniques involved going through several stages: choosing the distance, opting for the Euclidean distance, frequently used in clustering studies due to its interpretability and effectiveness in measuring the similarities between cases; selecting the linkage algorithm, using Ward's method for linkage, because it minimizes the intra-cluster variance and is effective in forming compact and significantly different clusters; cluster validation through K-Means Cluster analysis.

The importance of this study is that the results obtained will provide valuable solutions for policy makers, utility companies and local communities, helping to optimize resources and plan investments in water infrastructure.

2. Literature review

Sustainable water management is a critical component of sustainable development and takes into account issues similar to sustainability.

According to Mays (2006), sustainable water management takes into account all available water needs without endangering future water supplies. Based on Loucks and Gladwell (1999), social, ecological, environmental, and hydrological factors should all be considered when implementing sustainable water management.

According to Loucks (1997), sustainable water resource systems are those that preserve ecological and hydrological integrity while meeting present and future socioeconomic demands. Based on Savic and Walters (1997), the fundamental requirements of a sustainable water supply system include minimizing energy consumption, water loss, and the frequency and effects of pipe failures, as well as making the best use of current resources and satisfying customer demands in terms of both quantity and quality.

Savic and Walters (1997) and Loucks (1997) both stressed the significance of the environment and the functionality of the infrastructure itself. According to a more recent definition (Wheater et al., 2010), sustainable water management refers to a practice that maintains the long-term capacity of the water resource to maximize its use for future users while preventing the irreversible degradation of the water resource and its related resources, such as land and ecosystem.

The sustainable management of urban water systems presents new problems for the water business. Many external variables, such as the effects of climate change, drought, and urban population development, make it more important than ever to manage the water sector in a more sustainable way (Sachidananda et al., 2016).

Water shortages might impact two-thirds of the world's population in a few decades. Building new infrastructure or expanding surface and groundwater storage and distribution are two ways that many nations still choose to alleviate water scarcity. This trend goes beyond the focus on reducing water demand, for example by stopping losses on distribution networks (Diep et al., 2017).

The water losses from the distribution systems have multiple implications, on the technological functioning regarding the need for equipment with increased capacities including purification, on the additional costs due to the increased consumption of energy and potable reagents, on conserving the environment by lowering the water supply, by energy pollution and through exfiltration *(Smits et al., 2011).*

Every water supply system faces the problem of **effectively managing Non-Revenue Water** (NRW) in the distribution network, water that, if managed correctly, can be instrumental in solving severe water shortages. Water losses in the distribution network are actually treated water that is generated and enters the water supply facility, but does not reach the final customer for billing. Cost recovery, monitoring non-revenue water (NRW) and understanding customer needs for revenue equity are some of the major challenges that water management must address *(Boyle et al., 2013).*

To accurately measure water loss in water systems, the Water Loss Working Group of the International Water Association (IWA) recommends using the term non-revenue water (NRW). Potable water that is produced and lost somewhere in the water distribution system without ever reaching its destination is known as non-revenue water. This implies that water is either paid for or not used, which has an impact on local economy and resources. NRW is calculated as the ratio of the amount of water not billed to the amount of water entering the system.

Reducing NRW levels can help achieve levels of economic development and reduce the proportion of the population without access to safe drinking water. As water loss is a local problem on a global scale, solutions must be tailored to local circumstances, as the causes of these losses and the mechanisms available to manage them vary (Ojo, 2011).

In the European Union, water management legislation is based on Directive 2000/60/EC, adopted by the European Parliament and the Council on October 23, 2000, which creates a framework for Community policy in this area. This directive, known as the Water Framework Directive, is central to water management and promotes river basin-based management. It establishes principles for the protection of waters, focusing on the prevention of their degradation, the conservation and improvement of aquatic ecosystems, the sustainable use of water resources in the long term, and the gradual reduction and prevention of groundwater pollution. Also, the European Union has a diverse set of policies that address different aspects of Sustainable Development Goal 6: Clean water and sanitation, the Water Framework Directive being supported by other more detailed directives.

According to article 4, paragraph 3 of Directive (EU) 2020/2184, adopted by the European Parliament and the Council on 16 December 2020, regarding the quality of drinking water (transposed into national legislation by Government Ordinance no. 7 of 18 January 2023), the member states are required to carry out water loss assessments from their infrastructure using the Infrastructure Loss Index (ILI) or another appropriate method. These assessments should measure current levels of water losses and identify possible improvements in efficiency in reducing them. Therefore, all Member States of the European Union must evaluate the levels of water losses and reduce them if they surpass certain thresholds in order to increase the efficiency of the water supply infrastructure and prevent overexploitation of the limited water resources meant for human consumption.

As a member of the European Union (EU) and the United Nations (UN), Romania has pledged to adhere to the 17 Sustainable Development Goals set forth in the 2030 Agenda, which were approved by Resolution A/RES/70/1 at the September 2015 UN Summit for Sustainable Development. Romania also took part in the most recent international meeting, COP28, which brought together 197 states to the UN Convention on Climate Change and took place from November 30 to December 12, 2023.

Romania consolidated its support for the 2030 Agenda and for the 17 development objectives by adopting the 2030 Sustainable Development Strategy, formalized by Government Decision no. 877/2018.

A point of national interest is the security of drinking water supply, essential for achieving Sustainable Development Goal 6, which aims to ensure universal access to water and sanitation. This aspect is particularly relevant for local public services, which play a crucial role in promoting sustainable development at local level and improving living conditions.

Access to EU funds for investment in the water sector is essential for the transformation of a large number of weak service providers into a few large and strong operators capable of providing sustainable services at affordable tariffs, thus ensuring full cost recovery and the continued development of water systems. Although the benefits of economies of scale and efficiencies gained through aggregation and regionalization will only be measurable after a few years of operation of these regional operators, it is anticipated that they will lead to improved quality of services, optimization of operational costs and investments, reduction of water losses and enhanced capacity for project implementation (Frone, 2008).

The regional approach adopted in Romania for the promotion of integrated water and sanitation systems aims at maximizing cost efficiency and optimizing investment and operating expenses. Essentially, the objective of regionalization is to ensure increased efficiency and quality in the provision of local public water services through sustainable financial investments and well-managed operations. At the same time, this process is designed to stimulate regional economic growth and sustainable social development (Frone & Frone, 2012).

In Romania, the management of water and sanitation services is decentralized and has been the main subject of reforms in the last decade. According to the National Regulatory Authority for Community Utilities, water and sewerage services are provided by 43 regional main publicly owned operators and two large mixed-owned operators, along with around 900 smaller operators. These operators cover over 82% of the population connected to water services and over 91% of the population served by the sewage system. There are also local operators of various sizes, either with private, mixed or public capital.

In Romania, the accepted limits for network water losses are defined in article 93 and article 116 of Order no. 88 issued on March 20, 2007 by the National Regulatory Authority for Public Utilities, which approves the framework regulation for water supply and sewerage services. According to these rules, water losses of up to 5% of the volume of water entering the system for adductions and up to 15% of the total water introduced into the distribution network are allowed. It is also mandatory to undertake rehabilitation or upgrading of the drinking water infrastructure if the overall water losses, from the catchment to the user, exceed 20%.

3. Methodology

This study is based on the analysis of specialized literature and uses the hierarchical classification method as a research method. The fundamental principles of this method are the following:

- Building the Hierarchy: hierarchical classification divides the data recursively into a number of groups or "clusters", each at a different level of granularity. The process can start from a single group containing all the data and progressively divide into smaller subgroups, or vice versa, it can start with each object in its group and gradually merge the groups until they are all consolidated into a single group (Florea & Aivaz, 2022).
  Types of Hierarchical Classification:
  - ✓ Agglomerative (bottom-up): begins with every component as a distinct cluster, and at every stage, the closest (or similar) clusters are combined.
  - ✓ Divisive (top-down): starts with a single cluster containing all the elements, which is then divided into smaller clusters (Johnson, 1967).
- Measuring Similarity: the choice of how the similarity or distance between clusters is measured is important.
  Popular methods include:
  - ✓ Euclidean Distance: useful for data with quantitative attributes.
  - ✓ Jaccard similarity: good for categorical data.
  - ✓ Complete Linkage: the greatest separation between any two elements in any cluster.
  - ✓ Simple link: the smallest separation between any two pieces in any cluster.
  - ✓ Average: the average of all element pairs' distances between two clusters
- Hierarchy view: the hierarchical structure can be visualized by means of a "Dendrogram" - a tree diagram that illustrates how each group is related to the others. This graphical representation often shows the levels of similarity at which groups are combined or split (Murtagh & Contreras, 2012).

This method has a number of advantages and disadvantages (Aivaz & Șerbănescu, 2024), which are detailed below:

- *Benefits*:
  - ✓ Provides a clear view of the grouping structure.
  - ✓ It does not require specifying the number of clusters at the beginning of the analysis.
  - ✓ Flexible in choosing distance measures.
- *Disadvantages*:
  - ✓ It can be computationally expensive for large datasets.
  - ✓ Decisions made at each step are final and may influence the final cluster structure.

4. Data analysis

a. Data source: the investigation was carried out using the information relating to the year 2022 existing in the database of the Romanian Water Association, information provided by 43 regional operators with public capital within a voluntary benchmarking exercise of the water and wastewater sector. The study's variables, which are listed in Table 1, are:

- NRW (non-revenue water) (%) = (Total raw and treated water paid -Total water billed)/Total raw and treated water paid x 100
- Gross water paid in the year at Romanian Water (m^3/year) = Total gross water paid during the evaluated year
- Imported raw water (m^3/year) = Imported raw water (m^3) in the evaluated year
- Imported drinking water (m^3/year) = Volume of imported drinking water in the evaluated year
- Total quantities of water supplied (m^3/year) = The total volume of metered and/or unmetered billed water that is bought by registered consumers for domestic, commercial, industrial or public purposes in the evaluated year. This includes exported water.
- Total length of water networks (km) = Total intakes and distribution networks (service connections not included) at the end of the year
- Pipeline failures (no.) = Number of pipeline failures in the evaluated year including valve and accessory failures
- Meters for household consumers (no.) = Total number of water meters for household consumers, at the end of the reported year.
- Meters at consumers, economic agents and institutions (no.) = Total number of meters at industrial water consumers at the end of the reported year

Table 1. Statistical description of the variables

	N	Mean	Std. Deviation	Std. Error Mean
NRW (non-revenue water)	43	46.72	12.05	1.83
Gross water paid in the year at Romanian Water	43	21272341.00	16681079.41	2543840.76
Imported raw water	43	89774.77	588692.51	89774.76
Imported drinking water	43	415893.60	1471475.82	224397.96
Total quantities of water supplied	43	11175720.72	8566305.97	1306349.41
Total length of water networks	43	1404.58	935.00	142.58
Pipeline failures	43	1757.19	2277.95	347.38
Meters for household consumers	43	53359.84	29838.24	4550.29
Meters at consumers, economic agents and institutions	43	4051.35	2331.55	355.55

Source: Own processing based on variables from the database of the Romanian Water Association

b. Average linkage (between groups)

To achieve hierarchical classification, several steps need to be followed:

- ✓ Data collection: ensuring that the necessary data is collected systematically and is of high quality.
- ✓ Data processing: cleaning and preparing data for analysis, including data normalization where necessary.
- ✓ Choosing the classification method: deciding whether to use an agglomerative or a divisive approach depending on the specifics of the data and the objectives of the analysis.
- ✓ Analysis and interpretation of results: evaluating the clusters formed and interpreting how they can be used to improve services.

Table 2 is an "Agglomeration Schedule" used in the context of a hierarchical classification analysis, used to study how the major regional operators of water and sewage services in Romania could be grouped based on specific coefficients. In this table, each "Stage" represents an iteration in the clustering process, where two clusters are combined into one cluster.

Table 2. Agglomeration schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
1	11	12	245251119.830	0	0	21
2	37	43	1138599952.028	0	0	3
3	35	37	1384070917.502	0	2	16
4	17	28	1452007161.246	0	0	18
5	27	41	7024051500.272	0	0	6
6	15	27	54675606251.481	0	5	10
7	19	26	111999906597.633	0	0	30
8	1	22	148507139179.200	0	0	13
9	39	40	177578401805.633	0	0	23
10	15	21	228741046097.695	6	0	12
11	10	30	285728849603.307	0	0	21
12	15	20	374225406099.026	10	0	16
13	1	3	404045372360.788	8	0	17
14	6	25	478757557096.805	0	0	24
15	5	32	690867237391.910	0	0	24
16	15	35	772763983417.384	12	3	25
17	1	7	800656770218.839	13	0	27
18	17	33	876270669735.820	4	0	31
19	2	8	921627665592.836	0	0	27
20	18	29	1312085181540.537	0	0	26
21	10	11	1573496442260.363	11	1	28
22	9	23	2833924020004.926	0	0	34
23	16	39	2835044222884.842	0	9	29
24	5	6	2991764394580.536	15	14	29
25	15	38	3717149861849.099	16	0	28
26	18	34	4733062178245.313	20	0	31
27	1	2	5992335836468.849	17	19	33
28	10	15	6774766490390.418	21	25	36
29	5	16	8374679893416.000	24	23	33
30	19	24	9315461192607.531	7	0	32
31	17	18	13278737112580.420	18	26	35
32	19	42	16038960412455.643	30	0	36
33	1	5	30679469445854.254	27	29	37
34	9	13	32740146728121.650	22	0	38
35	4	17	33131431526196.145	0	31	40
36	10	19	38761725829400.734	28	32	37
37	1	10	106339332295881.690	33	36	39
38	9	14	108065551119754.520	34	0	42
39	1	31	122537391517714.700	37	0	41

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
40	4	36	256107784380048.220	35	0	41
41	1	4	406789560908033.600	39	40	42
42	1	9	2388838449346657.500	41	38	0

Source: Own processing based on variables from the database of the Romanian Water Association

Next, I present a detailed interpretation of the table columns:

1. Stage: number of the stages in the agglomeration process.
2. Cluster Combined: The clusters that are combined in this stage; for example, in stage 1, operators 11 and 12 are combined.
3. Coefficients: the agglomeration coefficient for the combination of the respective clusters, which usually indicates the distance or similarity between the clusters; larger coefficients suggest greater difference between the combined clusters.
4. Stage Cluster First Appears: the stage at which the involved clusters appeared for the first time in the process; if '0', this indicates that the cluster was one of the original clusters.
5. Next Stage: the next stage in which the newly formed cluster will be involved in another combination (Olson, 1995).

The analysis can be used to understand the structure and relationships between different entities (here, regional water and sewage operators), allowing decision-makers to identify possible synergies or optimize the organizational structure (Aivaz et al., 2023). For example, in stage 1, operators 11 and 12 are combined with a coefficient of 245251119.83, and they were not previously combined in other stages. This combination will later be involved in stage 21. The process continues in this way until all the initial clusters are integrated into one large cluster, ending with stage 42 (Table 2).

c. Cluster analysis

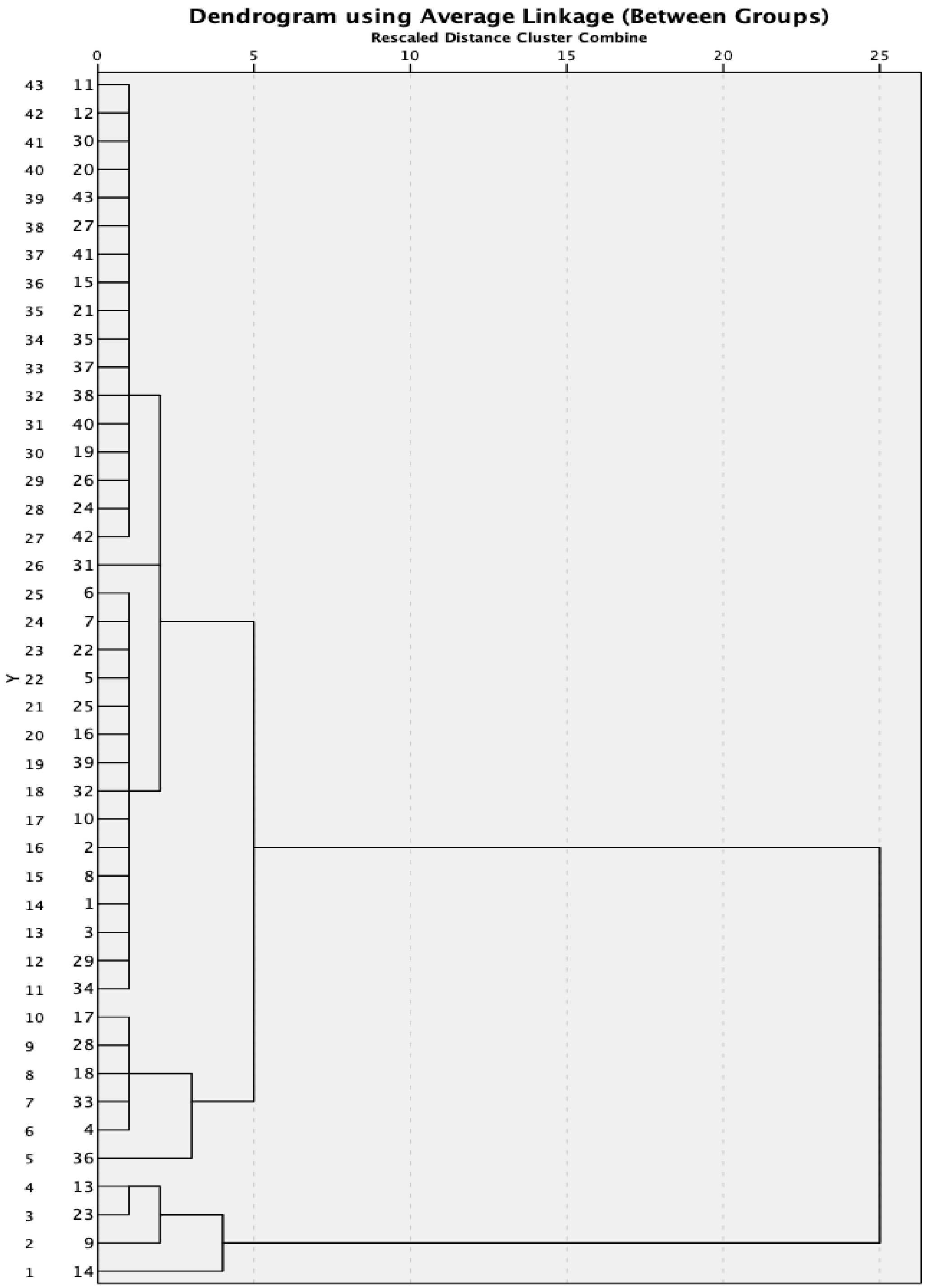
Figure 1 is a "Dendrogram" (tree diagram), a graphical representation used to visualize the results of a hierarchical clustering analysis. Hierarchical clustering is a data analysis method that organizes various objects (in this case, regional operators from Romania) into clusters based on the similarities between them.

Here is a detailed description and interpretation of this dendrogram:

A. Components of the dendrogram:
   - Axes: the vertical axis lists the regional operators, numbered according to a predefined scheme, and the horizontal axis shows the rescaled distance at which the clusters are combined. Distance indicates the level of similarity; higher values suggest greater similarity.
   - Branches: the horizontal and vertical lines connecting the regional operators represent how they are grouped together. The length of the horizontal line between two regional operators or groups of regional operators indicates how different those groups are.

B. Interpretation:
   - Cluster Formation: we observe that at small distances, many regional operators are grouped together individually, which indicates strong similarities between them. As we move to the right on the horizontal axis (increasing distance), smaller clusters merge into larger clusters, reflecting increased levels of overall similarity between regional operators.

Figure 1. Dendrogram



Source: Own processing based on variables from the database of the Romanian Water Association (https://h2obenchmark.org)

- Specific Clusters: regional operators OR 9, OR 13, OR 14 and OR 23 form a cluster at a relatively large distance, indicating that these regional operators, although different from the others, show some similarities between them. Another large group formed includes OR 18, OR 29, OR 34 and OR 4, and others, suggesting common characteristics among these regional operators.

- Strategic Interpretation: identifying these clusters can help local governments and regional planners understand how to distribute resources, implement policies, and coordinate efforts among counties with similar characteristics.
- Use in the Context of Water and Sewerage Services: if this dendrogram is related to water and sanitation services, it can be used to identify and implement effective practices in regions with similar needs and structures. For example, strategies to reduce water losses in one regional operator can be applied to other regional operators in the same cluster, given the structural or operational similarities between them. This dendrogram also provides a valuable visual basis for understanding the complex relationships between different regions, facilitating data-driven decision-making and policy coordination at the regional level (Herciu et al., 2023).

The last two clusters formed are the ones that join in the last step before forming a single global cluster. They represent the data clusters that are most different from each other and join only at the largest distance. The cluster on the left includes OR 9 and OR 14. The cluster on the right includes the remaining regional ORs, forming a much larger group.

Table 3 shows the evolution of the changes in the position of the centers of two clusters during 10 iterations within a clustering algorithm, more precisely K-means. The goal of this algorithm is to optimize the position of the cluster centers so as to minimize the distances between the points in each cluster and its center (Contreras & Murtagh, 2015). Each iteration adjusts these centers based on the data, trying to find a stable configuration where the adjustments become negligible (Florea & Aivaz, 2022).

Here is the step-by-step interpretation:

A. Details of iterations:
   - Iteration 1: The changes are very large, with approximately 13.9 million for Cluster 1 and 9.5 million for Cluster 2, indicating significant adjustments of centers in the first phase of the optimization process.
   - Iteration 2: The changes are dramatically reduced, to 355,384 for Cluster 1 and 1,577,715 for Cluster 2, suggesting that the centers are approaching an optimal position.
   - Iterations 3-5: The drops continue but get smaller, showing that the algorithm is starting to stabilize the center positions.
   - Iterations 6-10: The changes become extremely small, from 0.154 to values close to zero (6.923E-8), indicating that the centers have stabilized very close to their final position.

B. Final remarks:
   - Convergence: it claims that because the allotted maximum number of iterations was achieved, the iterations ceased and not because full convergence was reached. The last change mentioned is 0.939, which is the absolute maximum change distance of any center in the last iteration.
   - Distance between initial centers: A minimum distance between initial centers of 70,154,691,114 is indicated, suggesting that they were placed at large distances initially, possibly due to the high dispersion of the data in the feature space.

Preliminary conclusions: this historical iteration provides a detailed view of the clustering process, illustrating how the positions of the cluster centers gradually adjust to minimize the internal variability of the clusters. The fact that the algorithm did not fully converge in the maximum number of iterations may suggest the need to adjust the algorithm parameters or further investigate the data distribution to understand the reasons for this incomplete convergence.

Table 3. Iteration history

Iteration	Change in Cluster Centers	
	1	2
1	13859989.967	9466294.581
2	355384.358	1577715.763
3	9112.419	262952.627
4	233.652	43825.438
5	5.991	7304.240
6	0.154	1217.373
7	0.004	202.896
8	0.000	33.816
9	2.583E-6	5.636
10	6.923E-8	0.939

Source: Own processing based on variables from the database of the Romanian Water Association (https://h2obenchmark.org)

Table 4 illustrates the final values of the centers for two clusters, in the context of the analysis of water service providers. This table includes a number of indicators that reflect different aspects of water supply operations, including water losses, consumption, infrastructure and frequency of breakdowns.

This is a thorough examination of the information in the table:

1. NRW (Non-Revenue Water)
- Cluster 1: 46.21%
- Cluster 2: 50.68% NRW represents the percentage of water lost through leakage, measurement errors or theft, which does not generate revenue. The values shown show that both clusters experience significant loss problems, with Cluster 2 having a slightly worse situation.
2. Gross water paid per year at Apele Romane
- Cluster 1: 16,084,454
- Cluster 2: 60,700,283 These figures show the volume of raw (unprocessed) water purchased from Apele Romane. Cluster 2 acquires a much larger quantity, indicating significantly higher consumption, possibly due to a denser population or greater industrialization.
3. Imported raw water
- Cluster 1: 101,587
- Cluster 2: 0 Cluster 1 imports a small amount of raw water, which could indicate dependence on external sources to meet water needs.
4. Imported drinking water
- Cluster 1: 462,000
- Cluster 2: 65,487 Although both clusters import drinking water, Cluster 1 imports a much larger amount, suggesting limitations in purification capacity or higher quality requirements.
5. Total length of water mains
- Cluster 1: 1,171.80 km
- Cluster 2: 3,173.78 km Cluster 2 has a much more extensive water network, corresponding to a large urban area or a region with many connected localities.
6. Pipeline damage

- Cluster 1: 1,205
- Cluster 2: 5,957 The high number of failures in Cluster 2 indicates potential maintenance or wear and tear problems, possibly due to old infrastructure or urban density.

7. Meters for household consumers

- Cluster 1: 47,849
- Cluster 2: 95,246 The higher number of meters in Cluster 2 suggests a larger consumer base, consistent with a more urbanized and densely populated cluster.

8. Meters for consumers, economic agents and institutions

- Cluster 1: 3,512
- Cluster 2: 8,149 Similar to household meters, a higher number of meters for economic agents and institutions in Cluster 2 reflects more intensive economic and institutional activity.

Preliminary conclusions: the final values of the cluster centers indicate significant differences between the two groups analyzed, with Cluster 2 appearing to represent a large or heavily industrialized urban region with appropriate infrastructure and resource consumption. On the other hand, Cluster 1, although facing similar water loss problems, appears to be less dense and with a partial dependence on external water sources. This analysis can inform decisions about infrastructure investments, water resource management and strategies to improve operational efficiency.

Table 4. Final cluster centers

	Cluster	
	1	2
NRW (Non-Revenue Water)	46.21	50.68
Gross water paid per year at Apele Romane	16084454	60700283
Imported raw water	101587	0
Imported drinking water	462000	65487
Total length of water mains	1171.80	3173.78
Pipeline damage	1205	5957
Meters for household consumers	47849	95246
Meters for consumers, economic agents and institutions	3512	8149

Source: Own processing based on variables from the database of the Romanian Water Association

Table 5 shows the distances between the final centers of the two clusters identified in the clustering analysis. The values in the table represent the metric distance in a multidimensional space defined by the analyzed characteristics (such as water consumption, infrastructure, damage, etc.), between the center of Cluster 1 and the center of Cluster 2. The fact that the distance is identical regardless of the cluster perspective (1 vs. 2 or 2 vs. 1) is typical of symmetric distance matrices used in analyzes of this type.

Table 5. Distances between Final Cluster Centers

Cluster	1	2
1		44617732.389
2	44617732.389	

Source: Own processing based on variables from the database of the Romanian Water Association (https://h2obenchmark.org)

Interpretation of values:

- The distance of 44,617,732,389 is an indicator of the degree of differentiation between the two clusters. A very large value such as this suggests that the differences between cluster 1 and cluster 2 are significant, indicating that the two groups of data have very different characteristics.
- Implications:

✓ Clear differentiation: Such a large distance between the centers of the two clusters suggests that they represent entities with very different characteristics. For example, in the context of water services, one cluster might include regions with old infrastructure and important water losses, while the other might represent areas with modern infrastructure and efficient management.

✓ Customized strategies: The results show the need for customized approaches for each cluster, as needs and problems are very different. Resource planning, infrastructure investment and service improvement strategies should take these fundamental differences into account.

In conclusion, Table 5 underlines that there is a clear and significant separation between the two analyzed groups, providing a solid foundation for informed strategic decisions in the management or planning of the analyzed services.

Table 6. ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
NRW (Non-Revenue Water)	88.300	1	146.732	41	0.602	0.442
Gross water paid per year at Apele Romane	8795551611221140.000	1	70519551732278.530	41	124.725	<.001
Imported raw water	45599852812.806	1	353899345488.655	41	0.129	0.721
Imported drinking water	694702616063.663	1	2201107895529.674	41	0.316	0.577
Total length of water mains	17709435.932	1	463615.290	41	38.199	<.001
Pipeline damage	99777412.470	1	2882027.025	41	34.621	<.001
Meters for household consumers	9926497004.360	1	669926473.354	41	14.817	<.001
Meters for consumers, economic agents and institutions	95018816.225	1	3251189.940	41	29.226	<.001

Source: Own processing based on variables from the database of the Romanian Water Association (https://h2obenchmark.org)

Table 6 is a summary of an analysis of variance (ANOVA) for various variables associated with water services in order to assess differences between two or more clusters (Zhao, Karypis & Fayyad, 2005). Each row in the table represents a different variable (such as raw water consumption, length of water networks, etc.), and the columns provide statistical information about inter-cluster variation and error (intra-cluster variation), as well as F and p statistics -values (meaning) (Zhao & Karypis, 2002).

Interpretation of ANOVA results:

1. NRW (Non-Revenue Water)
    - F-statistic: 0.602
    - p-value: 0.442 This indicates that there are no statistically significant differences between the clusters regarding the percentage of non-revenue water.

2. Gross water paid per year at Apele Romane
    - F-statistic: 124.725
    - p-value: <.001 There are highly significant differences between clusters in the volume of raw water purchased, indicating significant variation in consumption or access to the resource.
3. Imported Raw Water and Imported Drinking Water
    - Both variables show high p-values (0.721 and 0.577, respectively), suggesting that there are no significant differences between the clusters regarding the amounts of raw and drinking water imported.
4. Total length of water mains
    - F-statistic: 38.199
    - p-value: <.001 Differences between clusters are highly significant, which may reflect large disparities in water infrastructure between different areas or systems.
5. Pipeline damage
    - F-statistic: 34.621
    - p-value: <.001 Significant differences in the frequency of pipe failures between clusters, possibly suggesting different levels of infrastructure age or quality.
6. Meters for household consumers and consumers of economic agents and institutions
    - Both categories of meters show significant statistical differences between clusters, indicating variations in the measurement and billing of water consumption.

Implications of the results:

- ✓ The significant differences in variables such as the volume of raw water paid, the length of the networks and the number of breakdowns suggest the need for strategies adapted to the specifics of each cluster in the management of infrastructure and water resources.
- ✓ Variables without significant differences (such as imported water) could indicate aspects that are managed consistently or that are influenced by similar external factors for both clusters.

The conclusion of the ANOVA analysis:

This ANOVA analysis is essential for understanding how different aspects of water services vary between clusters and provides a basis for informed decisions about optimization and interventions needed in different regions or systems.

Table 7 displays the distribution of cases between two clusters, as well as the total number of valid cases and the absence of missing cases.

Table 7. Number of cases in each cluster

Cluster	1	38.000
	2	5.000
Valid		43.000
Missing		0.000

Source: Own processing based on variables from the database of the Romanian Water Association (https://h2obenchmark.org)

The details of the table are as follows:

- Cluster 1 contains 38 cases, which indicates the number of analyzed units or subjects that were grouped in this cluster.
- Cluster 2 includes 5 cases, significantly fewer compared to Cluster 1, suggesting that this cluster may include cases with more specific or less frequent features.

- Valid: the total number of cases analyzed is 43, which confirms that all data collected were usable and included in the analysis.
- Missing: indicates that there are no missing cases in the data set, ensuring the integrity of the analysis.

Interpretation and implications:

- ✓ Improper Distribution between Clusters: the predominance of cases in Cluster 1 suggests that most subjects share a common set of characteristics on the basis of which they were grouped. The small number of cases in Cluster 2 could reflect a subset of subjects with distinctive characteristics that significantly differentiate them from the majority of cases.
- ✓ Analysis Planning: further analysis should investigate specific differences between clusters to understand why certain cases are so different as to form a separate cluster. It is important to examine whether the small size of Cluster 2 influences the representativeness or stability of the cluster, as models based on small numbers of cases may be less robust.
- ✓ No Missing Data: the absence of missing cases indicates that the data collection was complete and does not require further imputation or adjustment for further analysis, which increases the reliability of the results obtained.

Preliminary conclusions: the distribution of cases between clusters suggests that most of the collected data are homogeneous, with a small fraction showing distinctive features. This aspect is crucial for the formulation of intervention strategies or for the development of products/services adapted to the specific needs of identified subgroups (Bridges Jr, 1966). In the context of a water services analysis, for example, this distribution can influence how infrastructure investments are planned or how resources are allocated to improve services.

5. Conclusions and recommendations

The hierarchical clustering analysis revealed significant differences between water service operators in Romania, resulting in the identification of two main clusters that reflect distinct operational and financial particularities.

Cluster 1 includes most operators, characterized by moderate efficiency in water resource management. Operators in this group show a relatively average Non-Revenue Water (NRW) rate, indicating significant but manageable water losses. The infrastructure, although extensive, is periodically affected by breakdowns, reflecting a constant need for maintenance and investment in modernization. Despite these challenges, operators in this cluster manage to maintain an acceptable level of service due to the regular implementation of maintenance practices and strategic investments in technology and infrastructure. Cluster 2 consists of a smaller number of operators, which are distinguished by lower operational performance. The NRW rate is significantly higher, indicating substantial water losses and major inefficiencies in the distribution system. The infrastructure of these operators is often outdated and insufficiently maintained, with frequent breakdowns negatively impacting the quality of service. These characteristics are associated with low investment capacity and insufficient focus on modernization and efficiency (Stroe et al., 2023).

The contrast between the two clusters underlines the importance of tailored approaches in water management policies. While operators in the first cluster could benefit from programs to optimize efficiency and consolidate existing infrastructure, those in the second cluster require urgent interventions to rehabilitate water systems and implement innovative solutions to reduce NRW. The analysis showed that while there is progress in the water management sector, regional disparities and performance gaps remain significant. Investments in infrastructure and technology are crucial to improving overall efficiency, and tailoring to the specific needs of each cluster can accelerate this process. These results illustrate not only the current state of water resources management in Romania, but also provide a solid basis for the development of more efficient and better targeted strategies. By deeply understanding the profile and needs of each cluster, decision-makers can allocate resources more efficiently and implement solutions adapted to local realities, thus contributing to the long-term sustainability of water services.

The analysis carried out on the different datasets associated with water services in Romania highlighted some fundamental aspects regarding the operation, efficiency and distribution of resources in the water and sanitation sector. From the geographical distribution of infrastructure to operational efficiency and identifying the specific needs of different clusters, the analysis addressed multiple and complex aspects.

The main conclusions of this study are as follows:

- ✓ Significant differences between clusters: the analysis revealed the existence of two main clusters, which differed significantly in terms of water consumption, infrastructure, breakdown frequency and metering methods. Cluster 1, much more numerous, is characterized by a standard infrastructure and an average level of consumption and breakdowns. On the other hand, Cluster 2, with fewer cases, indicates a distinct set of characteristics, possibly due to higher urban or industrial density, or more modern or older infrastructure, requiring different approaches.
- ✓ The importance of efficient water management: indicators of unmetered water and significant losses highlight a major problem in the efficient management of water resources. This is a common challenge for many regions and requires innovative solutions and strategic investments to reduce losses and optimize consumption.
- ✓ The need for investments in water infrastructure: the differences in the length of the water networks and the high number of failures, especially in Cluster 2, underline the need to rehabilitate the existing infrastructure and expand it where necessary. Investments in modern monitoring and repair technologies can play a crucial role in preventing breakdowns and extending the life of networks.
- ✓ Implementation of personalized policies: discrepancies between clusters suggest that unified policies at the national level may not be sufficiently effective. Therefore, it is recommended to develop and implement personalized policies and strategies, adapted to the specific characteristics and needs of each cluster.
- ✓ Use of Data for planning and intervention: detailed data analysis provides a solid basis for strategic planning and intervention in water services. Smart use of this data can help anticipate problems, streamline responses, and allocate resources appropriately.
- ✓ Access to EU funds for investment in the water sector plays a major role in consolidating weaker service providers into a regionalized network of stronger and more efficient operators. This regionalization process not only optimizes costs and operational investments, but also contributes to improving the quality of services, reducing water losses and increasing the capacity to implement projects, while stimulating regional economic growth and sustainable social development.

Recommendations:

- ✓ Investment in technology: adopting new technologies for monitoring and managing water networks can reduce costs and improve service quality.
- ✓ Rehabilitation Program: initiate a national program to rehabilitate critical water infrastructure, with a focus on areas identified as having deficient or outdated infrastructure.

- ✓ Data-driven strategies: continue to collect and analyze data to refine intervention strategies and adapt policies to local and regional realities.

By implementing these conclusions and recommendations, the water services sector in Romania could significantly improve efficiency and sustainability, addressing both current needs and future challenges.

The original approach of this study is to use hierarchical classification to structure a complex set of data into a comprehensible and actionable form, thus providing a new perspective on the operational efficiency of water services. By identifying clear patterns of performance among water operators, the research contributes directly to the literature and supports the development of evidence-based public policy.

References

Aivaz, K.A., Munteanu, I., Rus, M.I., Chiriac (Matei), A., Leta (Mihai), F. (2023). Clarifying the Impact of Sanctions on Financial Indicators in Transports. An Empirical Comparative Analysis Using the Discriminant Model. *Transformations in Business & Economics*, Vol. 22, No 3A (60A), pp.933-953.

Aivaz, K. A., & Șerbănescu, M. (2024). Ecosystem Services Evaluation of the Danube Delta: An Analysis Using Hierarchical Multifactor Regression. *Studies in Business and Economics*, *19*(1), 5-21.

Boyle, T., Giurco, D., Mukheibir, P., Liu, A., Moy, C., White, S., & Stewart, R. (2013). Intelligent metering for urban water: A review. *Water (Switzerland)* (Vol. 5, Issue 3, pp. 1052–1081). MDPI AG. https://doi.org/10.3390/w5031052

Bridges Jr, C. C. (1966). Hierarchical cluster analysis. Psychological reports, 18(3), 851-854.

Contreras, P. & Murtagh, F. (2015). Hierarchical clustering. *Handbook of cluster analysis*, 103-123.

Diep, L., Hayward, T., Walnycki, A., Husseiki, M., & Karlsson, L. (2017). *Water, crises and conflict in MENA: how can water service providers improve their resilience?* International Institute for Environment and Development.

Florea, I-0, Aivaz, K.A. (2022). The dimension of the phenomenon of economic crime. A hierarchical classification of EU countries at the level of 2021. *Annals of "Dunărea de Jos" University of Galati fascicle I. Economics and applied informatics*, Years XXVIII – no3/2022, Page: 125-134, DOI: https://doi.org/10.35219/eai15840409296.

Frone, S., (2008). Factors and challenges of regionalization in the water and wastewater sector. Romanian Journal of Economics 27.2: 36: https://www.revecon.ro/articles/2008-2/2008-2-12.pdf

Frone, S., Frone, D.S., (2012). Factors and trends of economic efficiency in the water/wastewater sector. Procedia Economics and Finance 3: 1018-1023. https://core.ac.uk/download/pdf/82455499.pdf

Herciu, M., Ogrean, C., Mihaiu, D., Serban, R., Aivaz, K. A., & Tichindelean, M. (2023). Improving business models for the circular economy: Developing a framework for circular fashion. *Transformations in Business & Economics*, *22*(3).

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241-254.

Loucks, D. P. (1997). Quantification des tendances de la durabilité des systèmes. *Hydrological Sciences Journal*, *42*(4), 513–530. https://doi.org/10.1080/02626669709492051

Loucks, D. P., Gladwell, J. S. & International Hydrological Programme. (1999). *Sustainability criteria for water resource systems*. Cambridge University Press.

Mays, L.W. (2006). *Water Resources Sustainability*. United Kingdom: McGraw-Hill Education. ISBN 9780071462303, 0071462309.

Murrar, A., Batra, M., & Rodger, J. (2021). Service quality and customer satisfaction as antecedents of financial sustainability of the water service providers. *The TQM Journal*, *33*(8), 1867-1885.

Murtagh, F. & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *2*(1), 86-97.

Olson, C. F. (1995). Parallel algorithms for hierarchical clustering. *Parallel computing*, *21*(8), 1313-1325.

Ojo, V. O. (2011). *Customer satisfaction: A framework for assessing the service quality of urban water service providers in Abuja, Nigeria* (Doctoral dissertation, © Ojo, VO).

Sachidananda, M., Patrick Webb, D., & Rahimifard, S. (2016). A concept of water usage efficiency to support water reduction in manufacturing industry. *Sustainability (Switzerland)*, *8*(12). https://doi.org/10.3390/su8121222

Savic, D. A., & Walters, G. A. (1997). Elaboration de réseaux hydrauliques durables. *Hydrological Sciences Journal*, *42*(4), 549–564. https://doi.org/10.1080/02626669709492053

Smits, S., Verhoeven, J., Moriarty, P., Fonseca, C. & Lockwood, H. (2011). Arrangements and cost of providing support to rural water service providers. *IRC: The Hague, The Netherlands*.

Stroie, C., Duțescu, A., Munteanu, I. F., & Aivaz, K. A. (2023). The reorganisation decision test: A risk analysis model to increase competitiveness. *Journal of Competitiveness*, *15*(2).

Zhao, Y., Karypis, G., & Fayyad, U. (2005). Hierarchical clustering algorithms for document datasets. *Data mining and knowledge discovery*, *10*, 141-168.

Zhao, Y. & Karypis, G. (2002, November). Evaluation of hierarchical clustering algorithms for document datasets. In *Proceedings of the eleventh international conference on* Information and knowledge management (pp. 515-524).

XXX - European, P. (n.d.). *Directiva 2000/60/CE A Parlamentului European și a Consiliului din 23 octombrie 2000 de stabilire a unui cadru de politică comunitară în domeniul apei*.

Romanian Water Association database. https://h2obenchmark.org