

Technology for Managing and Grouping Datasets Regarding Protected Natural Areas and Species [†]

Delyana Dimova

Department of Mathematics and Informatics, Agricultural University, 12 Mendelev Blvd, 4000 Plovdiv, Bulgaria; ddimova@au-plovdiv.bg

[†] Presented at the 15th International Scientific Conference TechSys 2026—Engineering, Technologies and Systems, Plovdiv, Bulgaria, 14–16 May 2026.

Abstract

This article presents a technology for managing and grouping datasets regarding protected natural areas and species in Bulgaria for the period 2010–2024. The studied indicators include the area in hectares and number of the following objects: protected areas, natural landmarks, reserves and maintained reserves, national and natural parks. The other groups of considered elements are the number of protected plant and animal species, as well as protected venerable trees. This information is extracted from a created relational database and subsequently it is processed. The use of certain rules and the calculation of a given set of variables leads to the generation of the relevant solutions concerning the studied objects. Linear regression analysis is used to evaluate the examined data related to the protected natural areas in the considered period. This work also applies hierarchical cluster analysis to the mentioned data. The number of protected plant and animal species does not change in the whole examined time interval. During 2024 compared to the first year of the interval (2010), the number of protected venerable trees decreases by about 17.31%. The presented technology can also be applied when studying other economic indicators.

Keywords: database; datasets; grouping; protected natural areas and species; technology; time series

1. Introduction

In the modern world, the problems related to the protection and restoration of the environment, as well as the protection of plant and animal species, are of essential importance for the relevant society [1–3]. Various organizations and institutions collect and provide data on the mentioned objects. The use of information technologies [4,5] in this field allows users not only to analyze and interpret the results of data processing, but also to explore relationships between considered objects, to describe models and trends [6–9], and to make adequate decisions regarding the studied problems [10]. As a result, specialists in the field of ecology can use the obtained information to develop strategies, plans, and reports [11,12].

The current work studies time series data on protected natural areas and species in Bulgaria. They are provided by the Bulgarian National Statistical Institute (NSI) [13] and the Ministry of Environment and Water (MOEW) [14]. The information regarding the indicated objects is extracted from the websites of the mentioned organizations. Then it is presented and stored in a built relational database. The relationships between the designed tables are determined depending on the proposed hierarchy of the studied objects from



Academic Editors: Nikola G. Shakev,
Valyo N. Nikolov, Nikolay R.
Kakanakov and Sevil A.
Ahmed-Shieva

Published: 20 July 2026

Copyright: © 2026 by the author.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

the respective sets. As a result of this design process, the indicated database contains the following seven table schemes (Figure 1):

- Country (id_name, name);
- Total_Areas (id, year, total_areas, id_name);
- Area (id_area, area_name, id_name);
- Species (id_species, species, id_name);
- Number_object (id_3, year, number, id_area);
- Values_hectares (id_4, year, value_hectare, id_area);
- Number_species (id_2, year, number, id_species).

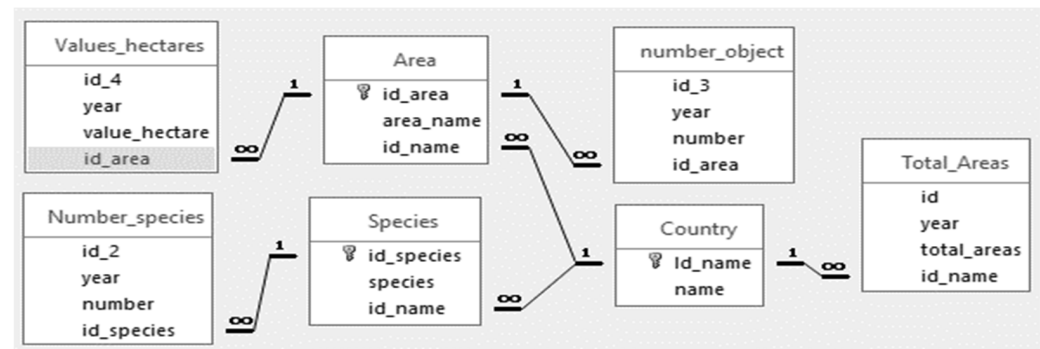


Figure 1. Scheme of the relational database.

The Country table is related to the following tables—Total_Areas, Species and Area. The attribute id_name is a primary key in the referenced table Country and a foreign key in each of the listed three referencing tables. The Area table is related to two others (Values_hectares and Number_object). Here the attribute id_area is a primary key for the Area table and a foreign key for the quoted two tables. It should also be noted that the Species table is related to the Number_species table. The attribute id_species is a primary key for the first of these mentioned tables and a foreign key in the second. Each of the respective indicated primary key and foreign key have the same data types. The requirement that each foreign key value has a corresponding value in the referenced key is met. Then in this case, the referential integrity [15] is guaranteed. In addition, the relationships between the mentioned tables of the database are of type one-to-many (Figure 1). The information concerning the names of protected natural areas and species is stored in two tables—Area and Species. The Country table includes the name of the considered country. Three other tables store the number and area (in hectares) of protected natural areas, as well as the number of species for each year of the studied time interval. These tables are Number_object, Values_hectares and Number_species. The Total_Areas table includes the total area (in hectares) of protected natural areas for the surveyed period. The database is updated annually. In this case, new data are entered into some of the listed tables.

The aim of this work is to present a technology for managing and grouping datasets regarding protected natural areas and species in Bulgaria for the period 2010–2024. The studied elements include the number and area (in hectares) of protected areas, natural landmarks, reserves and maintained reserves, national and natural parks, as well as the number of protected venerable trees, protected plant and animal species. This information is extracted from the presented database and subsequently it is processed. In addition, regression and hierarchical cluster analysis are also applied to these examined data.

2. Materials and Methods

Technology description: The current work mainly considers two datasets. The first of them contains six objects. They are the following:

- National parks;
- Natural parks;
- Protected areas;
- Natural landmarks;
- Reserves;
- Maintained reserves.

The second dataset includes protected species. In this case, these species are:

- Protected plant species;
- Protected animal species;
- Protected venerable trees.

A group of indicators for the above listed objects are studied and analyzed. This group contains:

- Area in hectares;
- Number of objects;
- Number of species;
- Indicated time interval.

The source for searching the listed data is the built database. The values of the indicators are stored in some of the designed tables (Figure 2). Quite naturally, the choice of a given indicator determines the choice of the corresponding related tables from which the data for the studied objects will be extracted.

Country

Id_name

name

1 Bulgaria

id_area

area_name

3 Protected areas

4 National parks

5 Natural parks

6 Maintained reserves

id_4

year

value_hectare

76

2010

4511

77

2011

4511

78

2012

4518

Country

Id_name

name

1 Bulgaria

id_species

species

1 Protected plant species

2 Protected animal species

3 Protected venerable trees

id_2

year

number

31

2010

1774

32

2011

1767

33

2012

1728

Figure 2. Structuring the information in the built database. Sources: Data from NSI [13] and MOEW [14].

The information for two of the indicators (areas in hectares and number of objects), related to the surveyed elements from the first set, is extracted from five database tables. These tables are the following:

- Country (id_name, name);
- Area (id_area, area_name, id_name);
- Number_object (id_3, year, number, id_area);
- Values_hectares (id_4, year, value_hectare, id_area);
- Total_Areas (id, year, total_areas, id_name);

The data for the next indicator (number of species) related to the elements of the second set are extracted from three tables of the database. The mentioned tables are:

- Country (id_name, name);

- Species (id_species, species, id_name);
- Number_species (id_2, year, number, id_objects);

The obtained groups of elements are formed based on the relevant defined criteria. These criteria include the studied object or species, one or more indicators, as well as the selected time interval. It should be noted that the present work considers a 15-year period from 2010 to 2024. The elements of the relevant mentioned groups must be processed, summarized and analyzed. For this purpose, it is necessary to find:

- Number of studied objects of the respective type ($c_{1k}, c_{2k}, \dots, c_{qk}; 1 \leq k \leq 6, 1 \leq q \leq 15$), whose values are the same over the whole considered period;
- The time subintervals in which the number of studied protected natural areas ($c_{qk}, c_{q+1k}, \dots, c_{q+nk}; 1 \leq k \leq 6, n \geq 1, q+n \leq 15$) and the respective number of studied species ($s_{i1}, s_{i2}, s_{q+n3}; n \geq 1, q+n \leq 15$) remains unchanged;
- The time subintervals where the areas (in hectares) of the protected natural areas ($b_{qk}, b_{q+1k}, \dots, b_{q+nk}; 1 \leq k \leq 6, n \geq 1, q+n \leq 15$) remain unchanged;
- The percentage change in the largest area (in hectares) of the studied object compared to the smallest area of the mentioned object in the considered period:

$$H_k = 100 - \frac{b_{\min_{qk}}}{b_{\max_{qk}}} \cdot 100, \quad (1)$$

where $b_{\min_{qk}}$ and $b_{\max_{qk}}$ are the smallest and largest area of the respective object;

- The percentage change in the largest number of given objects compared to the smallest number of the indicated objects in the studied time interval:

$$L_k = 100 - \frac{c_{\min_{qk}}}{c_{\max_{qk}}} \cdot 100, \quad (2)$$

$$N_k = 100 - \frac{s_{\min_{qw}}}{s_{\max_{qw}}} \cdot 100, \quad (3)$$

where $c_{\min_{qk}}$ and $c_{\max_{qk}}$ are the smallest and largest number of the studied protected natural areas; $s_{\min_{qw}}$ and $s_{\max_{qw}}$ are the smallest and largest number of the respective protected species, $1 \leq w \leq 3$;

- The share (v_{qk}) of the area of the respective protected natural area to the total area of the mentioned protected natural areas for the considered year of the time period:

$$v_{qk} = \frac{b_{qk} \cdot 100}{\sum_{a=1}^6 b_{ia}}, \quad (4)$$

where $1 \leq k \leq 6, 1 \leq q \leq 15$;

- The share (z_{qk}) of the number of given protected natural area to the total number of protected natural areas for a given year of the time period:

$$z_{qk} = \frac{c_{qk} \cdot 100}{\sum_{a=1}^6 c_{ia}}, \quad (5)$$

where $1 \leq k \leq 6, 1 \leq q \leq 15$.

Applying the listed rules and calculating the above-mentioned set of variables leads to the generation of the relevant solutions. In addition, this work uses regression analy-

sis [16–18] and hierarchical cluster analysis [19,20] for the statistical processing [21,22] of the surveyed data.

3. Results and Discussion

The studied datasets regarding protected natural areas are extracted from the following three sequences of related tables—Country, Area and Number_object; Country, Area and Values_hectares; and Country and Total_Areas. The information concerning protected species is extracted from the tables Country, Species and Number_species.

The obtained results show that the values of one of the studied elements of the first set remain the same for whole considered interval. In this case, this is the number of reserves. A similar situation is observed for three other elements belonging to the mentioned set. These are national and natural parks, as well as maintained reserves. It should be noted here that the number of national parks is significantly smaller than the number of the other three mentioned objects. The calculations show that the share (z_{qk}) of the number of national parks with respect to the total number of protected natural areas is 0.31% for each year of the time interval. Higher values of the variable z_{qk} are obtained for national parks, maintained reserves and reserves. They are 1.15%, 3.67% and 5.77%, respectively, as can be seen from Table 1. The situation is quite different for the remaining two elements (natural landmarks and protected areas) from the mentioned set. The number of the first indicated objects (natural landmarks) does not change in three subintervals of the considered period. The first of them includes two consecutive years (2010–2011). During 2012 compared to 2011, a slight decrease (about 0.58%) in the values of this studied indicator is calculated. The years from 2012 to 2015 form the second subinterval, in which the indicator values are the same. An insignificant increase (about 0.29%) in the number of these mentioned objects is observed in 2016. The two consecutive years 2016 and 2017 are included in the third subinterval, where the indicator values are again the same. In addition, the share (z_{qk}) of the number of natural landmarks to the total number of protected natural areas during the considered period varies in range from 36.10% to 37.46% (Table 1). Significantly higher values of variable z_{qk} are calculated for protected areas in the examined period. They vary from 52.78% to 62.85% (Table 1). Here it should be pointed out that the number of the mentioned objects remains the same in a subinterval including only two consecutive years (2014–2015).

Table 1. Presentation of the obtained values of the variable z_{qk} concerning the studied objects.

Years	Reserves	Natural Landmarks	Protected Areas	National Parks	Natural Parks	Maintained Reserves
2010	5.77%	36.31%	52.78%	0.31%	1.15%	3.67%
2011	5.77%	36.31%	52.89%	0.31%	1.15%	3.67%
2012	5.77%	36.10%	55.09%	0.31%	1.15%	3.67%
2013	5.77%	36.10%	58.87%	0.31%	1.15%	3.67%
2014	5.77%	36.10%	59.18%	0.31%	1.15%	3.67%
2015	5.77%	36.10%	59.18%	0.31%	1.15%	3.67%
2016	5.77%	36.20%	59.08%	0.31%	1.15%	3.67%
2017	5.77%	36.20%	59.29%	0.31%	1.15%	3.67%
2018	5.77%	36.10%	59.60%	0.31%	1.15%	3.67%
2019	5.77%	35.99%	59.81%	0.31%	1.15%	3.67%
2020	5.77%	36.31%	60.13%	0.31%	1.15%	3.67%
2021	5.77%	36.41%	60.23%	0.31%	1.15%	3.67%
2022	5.77%	36.94%	61.18%	0.31%	1.15%	3.67%
2023	5.77%	37.15%	61.80%	0.31%	1.15%	3.67%
2024	5.77%	37.46%	62.85%	0.31%	1.15%	3.67%

Sources: Own calculations on the basis of data from NSI [13] and MOEW [14].

Moreover, the present work calculates the percentage change (L_k) of the largest number of the respective studied objects to the smallest number of these indicated objects in the investigated time period. The obtained values of the variable L_k for natural landmarks and protected areas are 3.91% and 16.03%, respectively. For the remaining four objects belonging to the above-mentioned set, this studied variable has a value of zero, since their number does not change during the considered period.

Summarizing the data regarding the studied objects, it is found that in the time subinterval from 2010 to 2019, the areas in hectares occupied by national parks are the same. A similar situation is observed for the indicated object in the next subinterval including the years 2020–2023. Here, it should be noted that there is a slight increase in the examined indicator (area in hectares). It is about 0.2901%. The indicator values for another examined object (in this case natural parks) remain unchanged in two separate subintervals of the period. The first of them contains the years 2010–2012, while the second includes 2013–2024. However, an insignificant decrease in the indicator (area in hectares) is observed in the second subinterval. It is about 0.0058%. The calculations show that the percentage change (H_k) of the largest area of this studied object compared to the smallest area of the same indicated object during the considered period is 0.006%. For the studied data related to national parks, the value of the variable H_k is 2.02%. The case is quite different for the examined data concerning natural landmarks. Here, the values of the indicator (area in hectares) remain the same only in two consecutive years (2016–2017). In addition, the calculated variable H_k is 9.12%. For the next two considered objects (reserves and maintained reserves), different subintervals are formed where the studied indicator remains unchanged. For the first of the listed objects, the time segments contain 2018–2021 and 2023–2024, respectively. During 2023 compared to 2021, the indicator (area in hectares) increases by 0.15%. For the studied data related to the second object (maintained reserves), three subintervals are formed. They include the following years: 2010–2011, 2014–2016, and 2018–2021. The values of the investigated indicator (area in hectares) for maintained reserves in the third time segment are slightly higher than those in the first and second segment. The variable H_k is also calculated for the data regarding the other three objects (reserves, maintained reserves, and protected areas). Its values are 0.38%, 0.73%, and 4.84%, respectively. The values of the studied indicator (area in hectares) change in each year of the considered period only for protected areas.

Table 2 visualizes the calculated share (v_{qk}) of the area of the respective protected natural area to the total area of the protected natural areas for the indicated year of the time interval. The values of the variable v_{qk} for natural parks are significantly higher compared to those for the other five studied objects. Relatively high values of v_{qk} are also calculated for the studied data concerning national parks. They vary in the range from 25.815% to 25.890% in the period 2010–2024. Close values of the indicated variable are obtained for two of the considered objects. In this case, these objects are reserves and protected areas. As can be seen from Table 2, the values of the variable are lowest for the other two examined objects (natural landmarks and maintained reserves).

Regression analysis is used to evaluate the studied data related to the protected natural areas in the considered time period. The constructed linear regression equation is the following:

$$y = 23.024x + 560,911 \quad (6)$$

where x is the total number of protected natural areas, y is the total area (in hectares).

The result shows that the presented linear model is adequate, because it has statistically significant coefficients at the 5% level of significance. The coefficient of correlation between these two indicators is 0.72, indicating a strong positive relationship. This means that as the total number of protected natural areas increases, their total area (in hectares) also

tends to grow. The calculated coefficient of determination is 0.5810. Therefore, 53.10% of the variation in total area (in hectares) of protected natural areas can be explained by the variation in their total number.

Table 2. Results for the variable v_{qk} during the studied time period.

Years	Reserves	Natural Landmarks	Protected Areas	National Parks	Natural Parks	Maintained Reserves
2010	13.223%	2.897%	13.259%	25.815%	44.030%	0.774%
2011	13.226%	2.897%	13.200%	25.815%	44.030%	0.774%
2012	13.227%	2.904%	13.491%	25.815%	44.030%	0.777%
2013	13.231%	2.892%	13.622%	25.815%	44.027%	0.776%
2014	13.230%	2.890%	13.612%	25.815%	44.027%	0.776%
2015	13.232%	2.890%	13.615%	25.815%	44.027%	0.776%
2016	13.234%	2.891%	13.607%	25.815%	44.027%	0.776%
2017	13.236%	2.891%	13.615%	25.815%	44.027%	0.776%
2018	13.255%	2.861%	13.675%	25.815%	44.027%	0.779%
2019	13.255%	2.825%	13.721%	25.815%	44.027%	0.779%
2020	13.255%	3.093%	13.148%	25.890%	44.027%	0.779%
2021	13.255%	3.094%	13.154%	25.890%	44.027%	0.779%
2022	13.258%	3.109%	13.344%	25.890%	44.027%	0.780%
2023	13.274%	3.059%	13.491%	25.890%	44.027%	0.780%
2024	13.274%	3.064%	13.818%	25.366%	44.027%	0.779%

Sources: Own calculations on the basis of data from NSI [13] and MOEW [14].

The current work also applies hierarchical cluster analysis to the studied data. The indicated protected natural areas are grouped according to the values of their area (in hectares). As a result, the following clusters are formed (Figure 3):

- One of the surveyed objects (in this case natural parks) is presented in a separate cluster. Here, the values of the indicator (area in hectares) are the highest for each year of the time segment.
- Two of the considered objects are included in one cluster. In this case, these are reserves and protected areas. Subsequently national parks join to them.
- Two other objects form a separate cluster. These are natural landmarks and main-tained reserves.

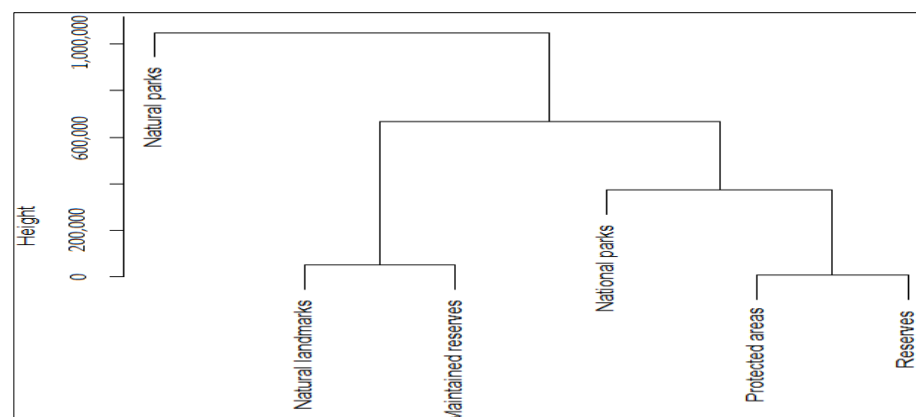


Figure 3. Clustering the studies objects according to the values of their area (in hectares). Sources: Own calculations on the basis of data from NSI [13] and MOEW [14].

The processed data related to the protected species from the second considered set of objects are summarized. The results show that the number of two of the studied objects does not change for the whole considered period. In this case, these are protected plant

species and protected animal species. Only the number of protected venerable trees changes in the time segment from 2010 to 2024. As a whole, the percentage change (N_k) of the largest number of these listed objects (protected venerable trees) to the smallest number of the indicated objects in the mentioned period is 24.69%. It should be pointed out that during 2024 compared to the first year of the time interval (2010), the number of protected venerable trees decreases by about 17.31%.

In addition, the presented technology can also be applied when studying other economic indicators.

4. Conclusions

The current work presents a technology for managing and grouping datasets regarding protected natural areas and species in Bulgaria in the time interval from 2010 to 2024. The surveyed indicators include the area (in hectares) and number of the following six objects: protected areas, natural landmarks, reserves and maintained reserves, national and natural parks. The other sets of considered elements are the number of protected plant and animal species, as well as protected venerable trees. This information is extracted from a built relational database and subsequently it is processed.

The obtained results show that the number of four of the studied objects (reserves and maintained reserves, national and natural parks) from the first set does not change during the considered period. The number of national parks is significantly smaller than the number of the other two objects (in this case protected areas and natural landmarks). The calculated share (z_{qk}) of the number of protected areas to the total number of protected natural areas is the largest. It varies from 52.78% to 62.85%. The values of the variable z_{qk} are lower for natural landmarks in the studied period. They vary in range from 36.10% to 37.46%. The percentage change (L_k) in the largest number of protected areas to the smallest number of the indicated objects in the considered time period is 16.03%. The value of the variable L_k for natural landmarks is 3.91%. The values of the indicator (area in hectares) change in each year of the examine period only for protected areas. The percentage change (H_k) in the largest area (in hectares) of the studied object compared to the smallest area of the mentioned object in the considered period is also found. The value of the variable H_k for national parks is 2.02% and 0.006% for natural parks. For the remaining four objects the value of H_k is 9.12% for natural landmarks, 0.38% for reserves, 0.73% for maintained reserves and 4.84% for protected areas. The share (v_{qk}) of the area of the respective protected natural area to the total area of the mentioned protected natural areas for the given year of the time period is also calculated. The obtained values of the variable (v_{qk}) for natural parks are significantly higher than those for the other five studied objects. Relatively high values of v_{qk} are also calculated for the data concerning national parks. They vary from 25.815% to 25.890% in the period 2010–2024. Linear regression analysis is used to evaluate these data related to the protected natural areas in the considered time period. In addition, grouping the listed protected natural areas according to the values of their area (in hectares) leads to the formation of three clusters.

The results regarding the data related to the protected species from the other studied set show that the number of two of these mentioned objects does not change in the whole considered period. Only the number of protected venerable trees changes in the studied 15-year interval. The percentage change (N_k) of the largest number of the examined objects (in this case protected venerable trees) to the smallest number of the mentioned objects in the indicated period is 24.69%.

In addition, this technology can also be applied when studying other economic indicators.

Funding: This research is supported by Agricultural University, Bulgaria, Plovdiv, under project 17-12 Support of the publication activity of the university lecturers in Agricultural University.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available on the website of the National Statistical Institute, Bulgaria, <https://www.nsi.bg>, accessed on 5 December 2025. Other data sources included in the investigation are referenced in the text.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Olmos-Martínez, E.; Romero-Schmidt, H.L.; Blázquez, M.d.C.; Arias-González, C.; Ortega-Rubio, A. Human Communities in Protected Natural Areas and Biodiversity Conservation. *Diversity* **2022**, *14*, 441. [CrossRef]
2. Aneva, I.; Zhelev, P.; Lukanov, S.; Peneva, M.; Vassilev, K.; Zheljazkov, V.D. Influence of the Land Use Type on the Wild Plant Diversity. *Plants* **2020**, *9*, 602. [CrossRef] [PubMed]
3. Mandić, A. Nature-based solutions for sustainable tourism development in protected natural areas: A review. *Environ. Syst. Decis.* **2019**, *39*, 249–268. [CrossRef]
4. Seyidova, M. Mathematical and Statistical Methods in the Application Fields of Data Mining Technology. *Luminis Appl. Sci. Eng.* **2025**, *2*, 81–87. [CrossRef]
5. Olszak, J.; Skublewska-Paszkowska, M. The impact of relational and non-relational databases on application performance. *J. Comput. Sci. Inst.* **2025**, *35*, 184–190. [CrossRef]
6. Wu, Y. Linear regression in machine learning. In Proceedings of the International Conference on Statistics, Applied Mathematics, and Computing Science (CSAMCS 2021), Nanjing, China, 26–28 November 2021; Volume 12163, pp. 1253–1264. [CrossRef]
7. Hall, T.; Rasheed, K. A Survey of Machine Learning Methods for Time Series Prediction. *Appl. Sci.* **2025**, *15*, 5957. [CrossRef]
8. Kadam, V.S.; Kanhere, S.; Mahindrakar, S. Regression Techniques in Machine Learning & Applications: A Review. *Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET)* **2020**, *8*, 826–830. [CrossRef]
9. Rahman, A. Statistics-Based Data Preprocessing Methods and Machine Learning Algorithms for Big Data Analysis. *Int. J. Artif. Intell.* **2019**, *17*, 44–65.
10. Lian, J. Research on Data Quality Analysis Based on Data Mining. *Comput. Artif. Intell.* **2025**, *2*, 1–7. [CrossRef]
11. Drogkoula, M.; Kokkinos, K.; Samaras, N. A Comprehensive Survey of Machine Learning Methodologies with Emphasis in Water Resources Management. *Appl. Sci.* **2023**, *13*, 12147. [CrossRef]
12. Ionescu, M.S.; Negoitǎ, O. Data Analysis, Fundamental Factor in the Elaboration of the Top Organizational Managerial Decision. In Proceedings of the 14th International Management Conference “Management Sustainable Organizations”, Bucharest, Romania, 5–6 November 2020; pp. 677–687. [CrossRef]
13. National Statistical Institute, Bulgaria. Available online: <https://www.nsi.bg> (accessed on 5 December 2025).
14. Ministry of Environment and Water, Bulgaria. Available online: <https://www.moew.government.bg> (accessed on 4 November 2025).
15. Roman, S. *Access Database Design & Programming*, 3rd ed.; O'Reilly Media: Sebastopol, CA, USA, 2002.
16. Akomodì, J.O. The Efficacy of Statistics in All Major Fields of Research: A Focus on Regression Analysis. *Open J. Stat.* **2025**, *15*, 53–72. [CrossRef]
17. Desboulets, L.D.D. A Review on Variable Selection in Regression Analysis. *Econometrics* **2018**, *6*, 45. [CrossRef]
18. Klees, S.J. Inferences from regression analysis: Are they valid? *Real-World Econ. Rev.* **2016**, *74*, 85–97.
19. Yim, O.; Ramdeen, K.T. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *Quant. Methods Psychol.* **2015**, *11*, 8–21. [CrossRef]
20. Fox, J. The R Commander: A Basic-Statistics Graphical User Interface to R. *J. Stat. Softw.* **2005**, *14*, 1–42. [CrossRef]
21. Ratner, B. The correlation coefficient: Its values range between +1/−1, or do they? *J. Target. Meas. Anal. Mark.* **2009**, *17*, 139–142. [CrossRef]
22. Zhang, D. A Coefficient of Determination for Generalized Linear Models. *Am. Stat.* **2017**, *71*, 310–316. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.