Energy cluster analysis based on consumption data in different weather condition

Dimitrios Th. Kazolis[1*], *Panagiotis* G. Kogias[1], and *Nikolaos* I. Roumeliotis[2]

[1]International Hellenic University, Department of Physics, Agios Loukas, 65403, Kavala, Greece
[2]HENDO Public's Electricity Company, Kavala, Greece

Abstract. The main aim of this effort is the discovery of knowledge from data, concerning consumption of electric energy, during the year 2022, based on unattended learning methods. These data were collected from the Public Electricity Company of Kavala and the methods used are, at first the Factor analysis and second the K-means clustering algorithm. The overhead methodologies are realized by the use of Statistica Data Miner software.

1 Introduction

The purpose of all Electricity Companies is to provide stable and sufficient energy to all of their customers at any time of the year. A decisive factor in achieving this is not only the production of energy, but the prediction of future consumption and the identification of the factors that affect it. This has as result to constructive energy modeling and modifications that are essential, since nowadays; there is an increase in the energy needs of man. For this reason, there has been a growing interest in use of various techniques for extraction of knowledge from databases related to energy consumption. [1, 2, 3].

The pursuit of this knowledge is, to reveal and create conclusions, which are useful for the energy consumption process [4], [5]. Thus, there are articles aiming to the extraction of knowledge, for example in the energy consumption of buildings [6], for the characterization of electricity customers [7] or for the Analysis to Power Grid System [8]. The production of the electricity from renewable energy sources as wind farms and photovoltaic systems is also appropriate to be study [9, 10].

Generally, many data mining techniques have been implemented in the development of predicting and solving the various problems of energy consumption. [11, 12].

As a continuation of previous research, the current effort aims to contribute to the development of the specific scientific field. So, the rationale behind this work is, to exploit information that may arise from large databases.

This procedure applies in practice all key stages of the extracting knowledge process from large databases [13, 14].

Thus, a main file was created, which had 17521 lines and 20 columns. Initially, these data were changed to a numeric form. Then, they fed the process of factor analysis. This resulted in three factors and after that, the total sample led the clustering algorithm which divided it

* Corresponding author: author@email.org

into 60 groups. Those groups had a logical relevance in their categorization and proved the validity of the procedures used.

As a conclusion this article proves that, the used methods, can cooperate together perfectly in constructing understandable relationships, during the process of knowledge discovery.

2 Data and methods

2.1 Data selection

Initially, data from eight energy distribution points in the city of Kavala were recorded, with a half an hour frequency, for the entire year 2022. So, an initial table was created with 17521 lines. Then, this table combined with more variables such as, the month, the time, the days of the week, the minimum and the maximum temperature of the day, the average temperature of the month, the average humidity, the wind, the rain height, the average pressure, the sunny and rainy days and the sun intensity. [15]

As a result of the above, a final table with a total of 17521 lines and 32 columns was constructed.

Although many formats could be imported directly into Statistica, it was considered appropriate to change the format of data as many of them had a text formatting. So they were converted into Byte and Numeric type, for the better operation of the clustering algorithm.

2.2 Factor analysis

As mentioned, the result of the above actions was to create a table of 17521 lines and 32 columns. This makes a total of 543151 entries, a fact which creates great difficulty in the identification of relationships and in addition, has a negative effect on the clustering algorithm to be used. So, a multivariate exploratory technique as factor analysis was used in order to reduce the number of variables available [14]. The default minimum eigenvalues was left to be 1, and by the use of the Scree-plot and checking the Kaiser criterion, the number of the chosen factors was three. This is shown in Figure 1.

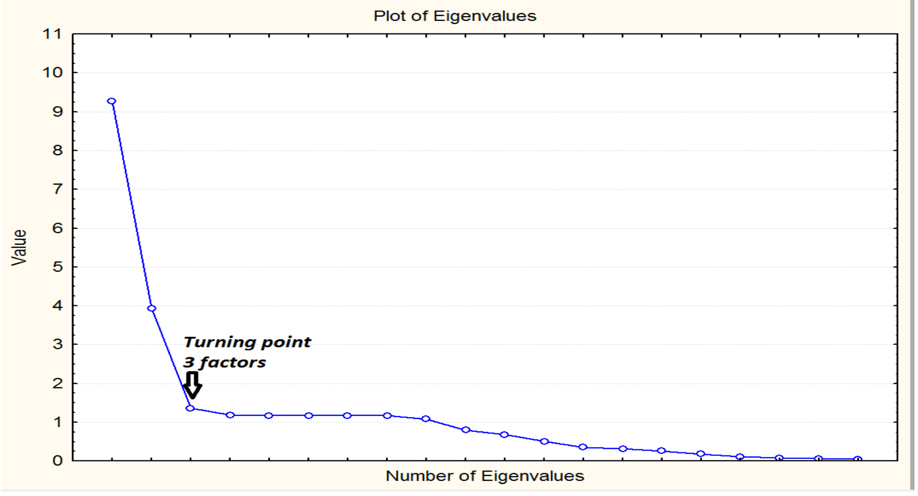

Fig. 1. Scree plot.

Furthermore a part of the correlation Matrix is depicted in the following Figure 2.

 https://doi.org/10.1051/e3sconf/202340401005

	1 TIME	2 HUM%AVG	3 hPa AVG PRESSURE	4 SUN ENERGY	TEMP. daily MAX	TEMP. daily Min	7 SATURDAY	8 SUNDAY	9 MONDAY
TIME	1,00000	-0,00000	-0,00000	-0,00000	0,00018	0,00028	-0,00000	-0,00000	-0,00000
HUM%AV	-0,00000	1,00000	0,72556	-0,83503	-0,83724	-0,80835	0,01090	-0,00390	-0,00126
hPa AVG	-0,00000	0,72556	1,00000	-0,58646	-0,71874	-0,72808	0,00802	0,00795	0,00083
SUN ENE	-0,00000	-0,83503	-0,58646	1,00000	0,69379	0,66322	-0,02057	-0,00015	-0,00184
TEMP. da	0,00018	-0,83724	-0,71874	0,69379	1,00000	0,93559	0,00403	0,01149	-0,01978
TEMP. da	0,00028	-0,80835	-0,72808	0,66322	0,93559	1,00000	0,01474	0,04619	0,00620
SATURD/	-0,00000	0,01090	0,00802	-0,02057	0,00403	0,01474	1,00000	-0,16799	-0,16799
SUNDAY	-0,00000	-0,00390	0,00795	-0,00015	0,01149	0,04619	-0,16799	1,00000	-0,16613
MONDAY	-0,00000	-0,00126	0,00083	-0,00184	-0,01978	0,00620	-0,16799	-0,16613	1,00000
TUESDAY	0,00000	0,00059	-0,00051	0,00939	-0,02433	-0,01011	-0,16799	-0,16613	-0,16613
WEDNES	0,00000	-0,00429	-0,00783	0,00939	-0,01785	-0,01484	-0,16799	-0,16613	-0,16613
THURSD/	0,00000	0,00283	0,00526	0,00367	0,01736	-0,02221	-0,16799	-0,16613	-0,16613
FRIDAY	0,00000	-0,00495	-0,01379	0,00028	0,02905	-0,02010	-0,16799	-0,16613	-0,16613
Καβάλα Σ	0,39908	0,06106	0,09774	-0,10936	-0,26231	-0,25642	-0,04758	-0,13750	0,01552
Καβάλα Σ	0,40249	0,34619	0,34258	-0,39371	-0,56199	-0,55348	-0,00846	-0,07156	0,00083
Καβάλα Σ	0,30604	-0,17564	-0,16623	0,13345	0,22656	0,21345	-0,00655	-0,15247	-0,01374
Καβάλα Σ	0,35881	-0,23160	-0,16167	0,10313	-0,02841	-0,03234	0,00903	-0,03576	-0,00661
Καβάλα Σ	0,41060	0,43312	0,39782	-0,47664	-0,62385	-0,59080	0,00833	0,01244	-0,00165
Καβάλα Σ	0,28839	0,00933	0,06760	-0,20646	-0,28596	-0,28592	0,00748	-0,00707	-0,00366
Καβάλα Σ	0,41278	0,02801	0,03771	-0,28354	-0,07996	-0,02306	-0,02143	-0,06156	0,01946
Καβάλα Σ	0,34397	0,27067	0,26237	-0,35354	-0,47160	-0,45546	-0,00551	-0,14299	0,00705
Καβάλα Σ	0,42451	0,11825	0,13028	-0,33581	-0,25192	-0,24855	0,01521	-0,06735	-0,00958
Καβάλα Σ	0,41774	0,41482	0,39029	-0,45179	-0,59244	-0,56657	0,00977	-0,01176	0,00068
Καβάλα Σ	0,41883	0,00059	0,04117	-0,11780	-0,21467	-0,20165	-0,03933	-0,15952	0,00841
Καβάλα Σ	0,40282	0,19696	0,20714	-0,36511	-0,39577	-0,35731	-0,01745	-0,07070	0,00991
Means	0,48958	69,22603	1017,43205	570,73973	20,60097	10,77123	0,14521	0,14247	0,14247
Std.Dev.	0,28862	5,93911	4,07379	184,91585	8,01529	7,44957	0,35232	0,34954	0,34954
No.Cases	17520,00000								
Matrix	1,00000								

Fig. 2. The correlation matrix.

For a factor rotation method, the Varimax raw method was used, and in the final stage of the process the loads of the factors were created. Those characterize the point that every factor interprets every variable. If a load is greater than 0.3 then it can be considered significant, but the larger it is, the more the factor interpret the corresponding variable.

So, the interpretation of the first factor correlates the time with the consumptions. As it can be observed here, there is an abnormality at the energy distribution point P230.

This is because initially, the distribution point P230 was used to distribute electric energy to the industrial and not the residential area of the city of Kavala.

The second factor interprets the change of temperature and moreover, the effect of the sun on the electricity consumptions.

Finally, the third factor interprets the variables connected with the days of the week. Here, a decline can be observed during Sundays and on the other hand, during Thursdays and Fridays the consumptions are increasing.

After the factor analysis, the 32 columns of the original file were replaced by 8 contributing to the reduction of data.

The above file is now ready to be led to the process of data mining by the use of generalized cluster analysis.

2.3 Cluster process

The main purpose of this step is to find clusters with common features, which can be interpreted by a human.

For this purpose, the Generalized k-Means Cluster Analysis is used, as, it can easily handle categorical variables. In these, the maximum frequency category grows into the center of the corresponding cluster, and all distances can be zero and one.

In this process, before the algorithm is applied, the number of clusters and the number of iterations is defined by the user. Here, the number of clusters is set to be 60 and the number of repetitions, defining group centers, is chosen to be 600. The selection of this big number was made just in order to avoid interrupting the process before it is properly completed.

Moreover, for the initial cluster centers the k observations were selected randomly and the Manhattan method is defined. This method determines the distances of non-uniformity and is the average of the modification among the elements. This is done in order to prevent the analysis from being influenced from variables with different gradient.

After all the above manipulations, a final record was created. Here, every specific hour of every month was grouped into a cluster and also was presented its distance from the cluster's center. The histogram in Figure 3 is showing the number of members of each cluster.

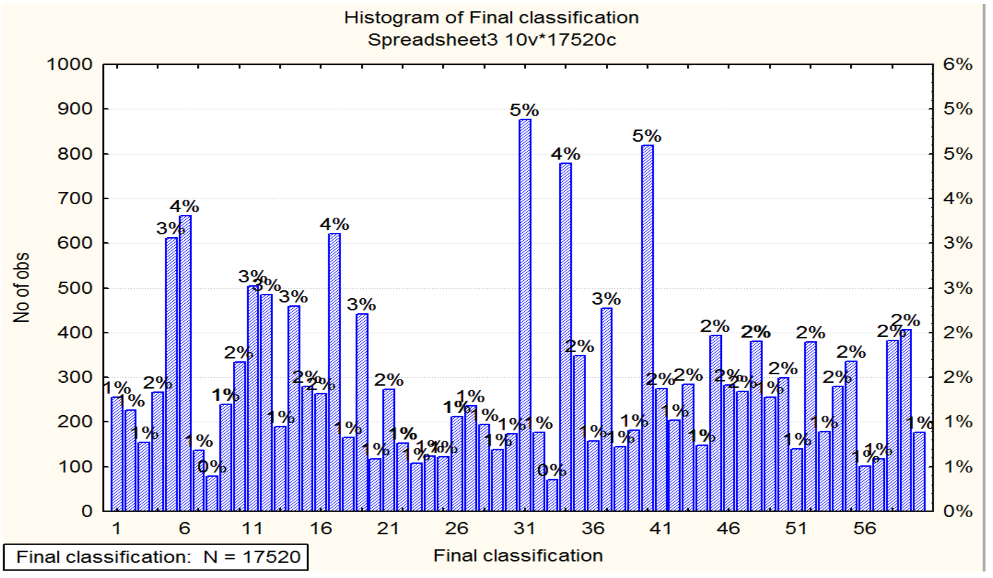

Fig. 3. Histogram of final classification.

At the present the Cluster Analysis process has been finalized. As a continuation, the examination of the similarity of each clusters' member characteristics must validate the knowledge discovery effort.

However, it should be emphasized that one of the characteristics of unsupervised learning and the K-Means algorithm is that it gives similar but not exactly the same results whenever the process occurs from the opening. And this is because the selection of the initial K centers is random.

3 Validity control

As mentioned above, the number of clusters was set to be 60, a number which can help with the interpretation of results. So, in the specific execution, some of the clusters will be examined in order to justify the reliability of this knowledge discovery process.

To begin with, cluster 2, is selected. As it can be seen in the following Figure 4, this cluster includes days of month May except Sundays. This group consists of high energy demand hours, from 8-9 o'clock in the morning until late in the evening.

	1 WIND	RAIN HEIGHT	3 SUNNY days	4 RAIN DAYS	5 tempAVG	6 FACTOR1	7 FACTOR2	8 FACTOR3	Final	Distance to
28/5/2022 9:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,51700	-0,75651	0,00997	2,000	0,106
28/5/2022 10:00:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,47415	-0,74594	0,04301	2,000	0,091
28/5/2022 10:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,45551	-0,75260	0,04853	2,000	0,088
28/5/2022 11:00:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,45191	-0,72873	0,07773	2,000	0,085
28/5/2022 11:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,44569	-0,72378	0,05629	2,000	0,089
28/5/2022 12:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,35702	-0,73870	0,10572	2,000	0,063
28/5/2022 12:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,28901	-0,76721	0,16382	2,000	0,048
28/5/2022 1:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,22323	-0,74233	0,16454	2,000	0,054
28/5/2022 1:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,11941	-0,77295	0,16967	2,000	0,073
28/5/2022 2:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,21023	-0,73587	0,29114	2,000	0,032
28/5/2022 2:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,21145	-0,76247	0,24468	2,000	0,042
28/5/2022 3:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,35900	-0,72228	0,16738	2,000	0,055
28/5/2022 3:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,44293	-0,73988	0,09619	2,000	0,077
28/5/2022 4:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,46536	-0,74250	0,01156	2,000	0,096
28/5/2022 4:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,45517	-0,75226	0,02618	2,000	0,093
28/5/2022 5:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,40830	-0,80342	-0,00296	2,000	0,104
28/5/2022 5:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,29411	-0,85900	-0,06076	2,000	0,115
28/5/2022 6:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,21531	-0,89229	-0,08883	2,000	0,140
28/5/2022 6:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,13850	-0,92382	-0,05324	2,000	0,152
28/5/2022 7:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,08366	-0,98031	-0,04777	2,000	0,173
28/5/2022 7:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,06356	-0,95647	-0,05568	2,000	0,172
28/5/2022 8:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,03078	-0,99162	-0,08723	2,000	0,191
28/5/2022 8:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,00084	-1,00064	-0,07864	2,000	0,196
28/5/2022 9:00:00 μμ	4,6	30,2	9,7	8,9	18,9	0,19081	-1,04073	-0,13786	2,000	0,246
28/5/2022 9:30:00 μμ	4,6	30,2	9,7	8,9	18,9	0,19698	-1,04467	-0,10027	2,000	0,240
28/5/2022 10:00:00 μμ	4,6	30,2	9,7	8,9	18,9	0,05796	-1,00821	-0,10198	2,000	0,211
28/5/2022 10:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,04139	-0,98508	-0,12063	2,000	0,195
28/5/2022 11:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,15889	-0,96055	-0,14955	2,000	0,177
28/5/2022 11:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,20214	-0,94340	-0,11363	2,000	0,160
30/5/2022 9:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,34992	-0,70323	0,03702	2,000	0,084
30/5/2022 10:00:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,34430	-0,69055	0,03332	2,000	0,087
30/5/2022 10:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,32138	-0,68265	0,05895	2,000	0,081
30/5/2022 11:00:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,34457	-0,66997	0,12390	2,000	0,074
30/5/2022 11:30:00 πμ	4,6	30,2	9,7	8,9	18,9	-0,31754	-0,66027	0,18038	2,000	0,062
30/5/2022 12:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,29646	-0,66453	0,13917	2,000	0,067
30/5/2022 12:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,20194	-0,68019	0,23769	2,000	0,058
30/5/2022 1:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,13141	-0,67340	0,21432	2,000	0,074
30/5/2022 1:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,12187	-0,67882	0,19889	2,000	0,077
30/5/2022 2:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,20500	-0,63681	0,14924	2,000	0,085
30/5/2022 2:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,28260	-0,65761	0,15217	2,000	0,068
30/5/2022 3:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,40428	-0,63107	0,07012	2,000	0,103
30/5/2022 3:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,48992	-0,58856	0,02673	2,000	0,135
30/5/2022 5:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,43845	-0,65379	-0,05033	2,000	0,127
30/5/2022 6:00:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,29318	-0,68454	0,00418	2,000	0,089
30/5/2022 6:30:00 μμ	4,6	30,2	9,7	8,9	18,9	-0,25935	-0,72535	-0,00309	2,000	0,086

Fig. 4. The number 2 cluster.

Next the cluster 38, in Figure 5, was selected. Here are days of February but the group consists of low energy demand hours, from midnight up to early in the morning.

	1 WIND	RAIN HEIGHT	3 SUNNY days	4 RAIN DAYS	5 tempAVG	6 FACTOR1	7 FACTOR2	8 FACTOR3	Final	Distance to
7/2/2022 5:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,50202	0,78273	-1,38959	38,000	0,046
7/2/2022 5:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,45615	0,78970	-1,38789	38,000	0,037
7/2/2022 6:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,29398	0,75504	-1,36189	38,000	0,056
7/2/2022 6:30:00 πμ	4,8	49,3	7,4	8,9	6,3	0,00285	0,69872	-1,26842	38,000	0,132
8/2/2022	4,8	49,3	7,4	8,9	6,3	0,08878	0,85660	-1,41379	38,000	0,080
8/2/2022 12:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,21795	0,88014	-1,43654	38,000	0,036
8/2/2022 1:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,36519	0,88385	-1,43514	38,000	0,016
8/2/2022 1:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,50057	0,89806	-1,41538	38,000	0,035
8/2/2022 2:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,57850	0,88792	-1,42301	38,000	0,043
8/2/2022 2:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,30303	0,90677	-1,41396	38,000	0,035
8/2/2022 3:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,33926	0,89659	-1,42709	38,000	0,025
8/2/2022 3:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,33669	0,89317	-1,41843	38,000	0,026
8/2/2022 4:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,41167	0,88417	-1,41132	38,000	0,020
8/2/2022 4:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,38267	0,86240	-1,40465	38,000	0,014
8/2/2022 5:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,29794	0,84562	-1,35927	38,000	0,034
8/2/2022 5:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,24144	0,84040	-1,32529	38,000	0,050
8/2/2022 6:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,01262	0,81737	-1,30186	38,000	0,094
8/2/2022 6:30:00 πμ	4,8	49,3	7,4	8,9	6,3	0,26855	0,77742	-1,32067	38,000	0,141
9/2/2022	4,8	49,3	7,4	8,9	6,3	0,35322	0,89394	-1,42632	38,000	0,126
9/2/2022 12:30:00 πμ	4,8	49,3	7,4	8,9	6,3	0,03285	0,93120	-1,44436	38,000	0,084
9/2/2022 1:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,17476	0,92926	-1,43104	38,000	0,056
9/2/2022 1:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,29860	0,91417	-1,47350	38,000	0,033
9/2/2022 2:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,39460	0,92097	-1,46381	38,000	0,020
9/2/2022 2:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,13630	0,93477	-1,47932	38,000	0,064
9/2/2022 3:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,21928	0,96057	-1,49487	38,000	0,061
9/2/2022 3:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,23949	0,94965	-1,46937	38,000	0,050
9/2/2022 4:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,31225	0,94465	-1,47065	38,000	0,038
9/2/2022 4:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,28022	0,92421	-1,42709	38,000	0,040
9/2/2022 5:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,22494	0,91205	-1,41753	38,000	0,047
9/2/2022 5:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,13960	0,91067	-1,41938	38,000	0,059
9/2/2022 6:00:00 πμ	4,8	49,3	7,4	8,9	6,3	0,03575	0,87401	-1,34430	38,000	0,090
9/2/2022 6:30:00 πμ	4,8	49,3	7,4	8,9	6,3	0,34303	0,82484	-1,40742	38,000	0,123
10/2/2022	4,8	49,3	7,4	8,9	6,3	0,28052	0,84315	-1,32387	38,000	0,126
10/2/2022 12:30:00 πμ	4,8	49,3	7,4	8,9	6,3	0,02106	0,88632	-1,24794	38,000	0,110
10/2/2022 1:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,22002	0,86902	-1,29202	38,000	0,062
10/2/2022 1:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,36378	0,87421	-1,27359	38,000	0,046
10/2/2022 2:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,46872	0,88654	-1,23868	38,000	0,063
10/2/2022 2:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,23409	0,93813	-1,25887	38,000	0,084
10/2/2022 3:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,24033	0,91424	-1,23842	38,000	0,081
10/2/2022 3:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,31720	0,91797	-1,25598	38,000	0,067
10/2/2022 4:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,38702	0,90320	-1,28162	38,000	0,048
10/2/2022 4:30:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,37206	0,89413	-1,22762	38,000	0,059
10/2/2022 5:00:00 πμ	4,8	49,3	7,4	8,9	6,3	-0,30625	0,89395	-1,17389	38,000	0,079

Fig. 5. The number 38 cluster.

The following, cluster 43 in Figure 6, has the same characteristics as the previous cluster, with low energy demand hours, from midnight up to early in the morning. The clustering algorithm differentiated them because of the month August.

13/3/2022 10:00:00 μμ	5,3	36,6	6	9,2	9	1,60736	0,80396	-0,36542	42,000	0,227
1/8/2022 1:30:00 πμ	3,9	17	18,5	3,4	25,5	0,14827	-1,48801	-0,42243	43,000	0,096
1/8/2022 2:00:00 πμ	3,9	17	18,5	3,4	25,5	0,01589	-1,46348	-0,45243	43,000	0,077
1/8/2022 2:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,08995	-1,43382	-0,47413	43,000	0,058
1/8/2022 3:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,12504	-1,42488	-0,47198	43,000	0,051
1/8/2022 3:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,19126	-1,41874	-0,50573	43,000	0,046
1/8/2022 4:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,25603	-1,42435	-0,48766	43,000	0,048
1/8/2022 4:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,28078	-1,39287	-0,41271	43,000	0,035
1/8/2022 5:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,29200	-1,39811	-0,41127	43,000	0,035
1/8/2022 5:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,27217	-1,41472	-0,40602	43,000	0,032
1/8/2022 6:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,33136	-1,41377	-0,37580	43,000	0,034
1/8/2022 6:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,30957	-1,42595	-0,37986	43,000	0,035
1/8/2022 7:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,17730	-1,46624	-0,26757	43,000	0,027
2/8/2022 1:30:00 πμ	3,9	17	18,5	3,4	25,5	0,02477	-1,43817	-0,45338	43,000	0,072
2/8/2022 2:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,09174	-1,40870	-0,44797	43,000	0,047
2/8/2022 2:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,18076	-1,40740	-0,47950	43,000	0,040
2/8/2022 3:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,26052	-1,38433	-0,46860	43,000	0,046
2/8/2022 3:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,31987	-1,36483	-0,45677	43,000	0,057
2/8/2022 4:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,33872	-1,38200	-0,47121	43,000	0,058
2/8/2022 4:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,36495	-1,36908	-0,42975	43,000	0,057
2/8/2022 5:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,41354	-1,35795	-0,42823	43,000	0,067
2/8/2022 5:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,35504	-1,37985	-0,36782	43,000	0,041
2/8/2022 6:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,24881	-1,41583	-0,35199	43,000	0,018
2/8/2022 6:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,34558	-1,40451	-0,32906	43,000	0,025
2/8/2022 7:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,13236	-1,46954	-0,22017	43,000	0,043
3/8/2022 1:00:00 πμ	3,9	17	18,5	3,4	25,5	0,06883	-1,47227	-0,34846	43,000	0,066
3/8/2022 1:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,08999	-1,42835	-0,37644	43,000	0,038
3/8/2022 2:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,20577	-1,40859	-0,36634	43,000	0,014
3/8/2022 2:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,27630	-1,40004	-0,38823	43,000	0,028
3/8/2022 3:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,32982	-1,37373	-0,33510	43,000	0,032
3/8/2022 3:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,35472	-1,38547	-0,35189	43,000	0,036
3/8/2022 4:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,40696	-1,37999	-0,35607	43,000	0,046
3/8/2022 4:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,41783	-1,37213	-0,33724	43,000	0,046
3/8/2022 5:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,41852	-1,36815	-0,28811	43,000	0,043
3/8/2022 5:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,40264	-1,38319	-0,29728	43,000	0,035
3/8/2022 6:00:00 πμ	3,9	17	18,5	3,4	25,5	-0,30367	-1,40421	-0,24094	43,000	0,026
3/8/2022 6:30:00 πμ	3,9	17	18,5	3,4	25,5	-0,38734	-1,39836	-0,23513	43,000	0,041
3/8/2022 7:00:00 πμ	3,9	17	18,5	3,4	25,5	[illegible]	[illegible]	[illegible]	43,000	[illegible]

Fig. 6. The number 43 cluster.

A similarity can be observed in the following cluster 57, in Figure 7. As with the previous groups this consists of the low energy demand hours, from midnight up to early in the morning of November days. Furthermore, the low consumption can be observed here until midday during Sundays.

	1 WIND	2 RAIN HEIGHT MM	3 SUNNY days	4 RAIN DAYS	5 tempAVG	6 FACTOR1	7 FACTOR2	8 FACTOR3	9 Final classification	10 Distance to centroid
7/11/2022	4,1	62,3	8,2	9,7	10,9	-1,22639	0,82991	0,46176	57,000	0,044
7/11/2022 12:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,33937	0,83264	0,42237	57,000	0,049
7/11/2022 1:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,45597	0,84856	0,40099	57,000	0,066
7/11/2022 1:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,51310	0,83741	0,34244	57,000	0,089
7/11/2022 2:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,55815	0,83789	0,38158	57,000	0,088
7/11/2022 2:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,54741	0,84133	0,38209	57,000	0,085
7/11/2022 3:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,57331	0,85078	0,37696	57,000	0,088
7/11/2022 3:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,55868	0,84578	0,39097	57,000	0,084
7/11/2022 4:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,55508	0,84967	0,36691	57,000	0,087
7/11/2022 4:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,54913	0,84169	0,43524	57,000	0,075
7/11/2022 5:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,52379	0,84403	0,47012	57,000	0,064
7/11/2022 5:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,45443	0,83003	0,48516	57,000	0,054
7/11/2022 6:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,32788	0,81367	0,58569	57,000	0,036
7/11/2022 6:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,17976	0,79521	0,64572	57,000	0,063
7/11/2022 7:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,11218	0,81178	0,77825	57,000	0,095
8/11/2022	4,1	62,3	8,2	9,7	10,9	-1,22275	0,88650	0,52476	57,000	0,018
8/11/2022 12:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,34452	0,87572	0,47956	57,000	0,028
8/11/2022 1:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,42020	0,88332	0,45987	57,000	0,041
8/11/2022 1:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,46820	0,88060	0,45072	57,000	0,050
8/11/2022 2:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,51068	0,88374	0,42166	57,000	0,061
8/11/2022 2:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,52205	0,88707	0,40012	57,000	0,067
8/11/2022 3:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,52831	0,90330	0,42841	57,000	0,058
8/11/2022 3:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,52435	0,89300	0,41318	57,000	0,063
8/11/2022 4:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,53431	0,89186	0,43630	57,000	0,060
8/11/2022 4:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,49529	0,87547	0,46665	57,000	0,052
8/11/2022 5:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,47768	0,87083	0,51628	57,000	0,041
8/11/2022 5:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,42221	0,85596	0,52977	57,000	0,034
8/11/2022 6:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,27863	0,84918	0,60396	57,000	0,027
8/11/2022 6:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,14316	0,82535	0,67240	57,000	0,067
8/11/2022 7:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,07290	0,83916	0,80905	57,000	0,101
8/11/2022 7:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-0,87664	0,85613	0,93952	57,000	0,151
8/11/2022 7:00:00 μμ	4,1	62,3	8,2	9,7	10,9	-2,89183	1,08267	0,13615	57,000	0,359
9/11/2022	4,1	62,3	8,2	9,7	10,9	-1,18079	0,96271	0,47864	57,000	0,044
9/11/2022 12:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,31996	0,97602	0,43953	57,000	0,043
9/11/2022 1:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,41989	0,98044	0,40643	57,000	0,065
9/11/2022 1:30:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,48542	0,99311	0,39423	57,000	0,080
9/11/2022 2:00:00 πμ	4,1	62,3	8,2	9,7	10,9	-1,51320	0,98600	0,38698	57,000	0,084

Fig. 7. The number 43 cluster.

Similarly the cluster 9 in Figure 8, consists of days of April. There is a low energy demand period from midnight up to 7-7:30 o'clock in the morning. Moreover during Sundays, this low energy demand period, reaches the early afternoon hours.

	1 WIND	2 RAIN HEIGHT	3 SUNNY days	4 RAIN DAYS	5 tempAVG	6 FACTOR1	7 FACTOR2	8 FACTOR3	9 Final	10 Distance to
3/4/2022 5:00:00 μμ	4,9	33,2	5,7	9,3	13,7	-0,50114	0,00401	-0,98947	9,000	0,102
4/4/2022	4,9	33,2	5,7	9,3	13,7	-0,63727	0,12882	-1,28384	9,000	0,072
4/4/2022 12:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,77948	0,14903	-1,33095	9,000	0,066
4/4/2022 1:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,95569	0,16559	-1,36304	9,000	0,079
4/4/2022 1:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,02418	0,16822	-1,37038	9,000	0,091
4/4/2022 2:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,07922	0,17285	-1,37164	9,000	0,100
4/4/2022 2:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,97296	0,19274	-1,35370	9,000	0,086
4/4/2022 3:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,02794	0,18652	-1,36624	9,000	0,095
4/4/2022 3:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,05388	0,19486	-1,37127	9,000	0,102
4/4/2022 4:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,08607	0,17305	-1,35530	9,000	0,098
4/4/2022 4:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,07216	0,17063	-1,29616	9,000	0,084
4/4/2022 5:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-1,05420	0,16515	-1,32357	9,000	0,085
4/4/2022 5:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,97703	0,14436	-1,28187	9,000	0,060
4/4/2022 6:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,84201	0,12156	-1,27264	9,000	0,038
4/4/2022 6:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,65513	0,09451	-1,19971	9,000	0,044
4/4/2022 7:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,51962	0,11025	-1,10077	9,000	0,069
4/4/2022 7:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,15297	0,09012	-0,91684	9,000	0,154
5/4/2022	4,9	33,2	5,7	9,3	13,7	-0,41346	0,03607	-1,10492	9,000	0,084
5/4/2022 12:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,60521	0,06543	-1,16221	9,000	0,041
5/4/2022 1:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,80860	0,07540	-1,18643	9,000	0,015
5/4/2022 1:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,90196	0,07898	-1,24627	9,000	0,026
5/4/2022 2:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,99157	0,09802	-1,23762	9,000	0,042
5/4/2022 2:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,88272	0,11344	-1,26182	9,000	0,035
5/4/2022 3:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,91497	0,10991	-1,22337	9,000	0,031
5/4/2022 3:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,94252	0,09756	-1,28692	9,000	0,045
5/4/2022 4:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,97527	0,08472	-1,26487	9,000	0,042
5/4/2022 4:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,93682	0,07560	-1,23505	9,000	0,028
5/4/2022 5:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,90680	0,06959	-1,21914	9,000	0,021
5/4/2022 5:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,83983	0,05496	-1,19213	9,000	0,015
5/4/2022 6:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,69695	0,01628	-1,17179	9,000	0,042
5/4/2022 6:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,51171	0,01297	-1,09142	9,000	0,078
5/4/2022 7:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,30964	-0,01607	-1,01737	9,000	0,130
6/4/2022	4,9	33,2	5,7	9,3	13,7	-0,28134	0,06522	-1,07627	9,000	0,102
6/4/2022 12:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,53926	0,09068	-1,17556	9,000	0,056
6/4/2022 1:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,73428	0,10162	-1,21045	9,000	0,037
6/4/2022 1:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,85858	0,11149	-1,23683	9,000	0,026
6/4/2022 2:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,90884	0,10434	-1,23858	9,000	0,032
6/4/2022 2:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,81020	0,12610	-1,22181	9,000	0,034
6/4/2022 3:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,83790	0,11248	-1,26405	9,000	0,035
6/4/2022 3:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,87990	0,13275	-1,24579	9,000	0,036
6/4/2022 4:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,91262	0,12716	-1,27163	9,000	0,045
6/4/2022 4:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,89588	0,10507	-1,23079	9,000	0,029
6/4/2022 5:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,85117	0,08549	-1,20708	9,000	0,015
6/4/2022 5:30:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,80452	0,08989	-1,18670	9,000	0,019
6/4/2022 6:00:00 πμ	4,9	33,2	5,7	9,3	13,7	-0,68918	0,05233	-1,12697	9,000	0,035

tor Analysis Results: δε... Generalized Cluster A...

Fig. 8. The number 9 cluster.

Similar examination can be performed in all of the created clusters. In general it can be noted that members of each cluster, have many common features, while on the other hand, the clusters have several differences among them, so they can be separated by the algorithm.

In general, it can be observed that, all of the clusters contain time periods with similarities in the energy consumption.

4 Conclusions

The conclusions from all of the above procedure can be derived from each step of it.

Initially it can be concluded that, a very effective statistical method for multivariate exploration, is factor analysis. With it, the large dimensional data, which must be processed, can be transformed into a relative smaller group of variables by replacing many of them, with factors. As a result the difficulties which could be raised from the very large dimensional data can be erased and moreover, the created factors include the essential meaning of the replaced variables

It can be inferred that, in such cases, Factor Analysis is considered to be necessary, as it is a reliable and effective method.

Moreover, the above procedure has also included the K-Means clustering algorithm. This is a very active method of classification. It can easily handle categorical variables especially in cases without learning examples.

To conclude with, this paper is one more proof that these two methods can cooperate together perfectly. They are following their theoretical background in a predictable way and they can be applied for further knowledge discovery research.

So, for further research, one could investigate the same data model with the method of V-fold cross validation, or to discover the effectiveness of different data mining algorithms in this common data model.

References

1. L.G.B. Ruiz, M.C. Pegalajar, R. Arcucci, M. Molina Solana, Expert Systems with Applications, **160**, 113731 (2020)

2. Zhun. Yu, F. Haghighat, B.C.M. Fung, Sustainable Cities and Society, 25, pp. 33-38 (2016)
3. Y. Zhang, J. Liu, Sensors, **22** (10), 3704 (2022)
4. V. Kamaev, A. Finogeev, A. Finogeev and S. Shevchenko, *Knowledge Discovery in the SCADA Databases Used for the Municipal Power Supply System*. In: Kravets, A., Shcherbakov, M., Kultsova, M., Iijima, T. (eds) Knowledge-Based Software Engineering. JCKBSE 2014. Communications in Computer and Information Science, vol 466. Springer, Cham. (2011)
5. J.Kogan, C. Nicholas, M.Teboulle, Grouping Multidimensional Data. Recent Advances in Clustering (Springer-Verlag, Berlin Heidelberg, 2006)
6. A. Abdelaziz, V. Santos, M. Dias, Energies, **14** (22), 7810 (2021)
7. S. Ramos, J. Soares, S. Cempranel, I. Tavares, Z. Vale, Procedia Computer Science, **186**, pp. 475-488 (2021)
8. G. Grigoraş, M. Raboaca, C. Dumitrescu, D. Manea, T. Mihaltan, V. Niculescu, B. Neagu, Sensors, **23**(4), 1895 (2023)
9. C. Genç, A. Sakalli, I. Stoyanov, T. Iliev, G. Mihaylov, I. Beloev, E3S Web of Conferences, **207**, 02013 (2020)
10. C. Genç, A. Sakalli, I. Stoyanov, T. Iliev, G. Mihaylov, I. Beloev, E3S Web of Conferences, **207**, 02014 (2020)
11. S. Singh, A. Yassine, Energies, **11**(2), 452 (2018)
12. U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, AI Magazine, **17** (3), pp. 37-54 (1996)
13. W. Liu, J. Zhao, D. Wang, WIREs Data mining and Knowledge Discovery, **11**(4), e1406 (2021)
14. N. Shrestha, American Journal of Applied Mathematics and Statistics, **9** (1), pp. 4-11 (2021)
15. http://www.emy.gr/emy/el/climatology/climatology_city?perifereia=East%20Macedonia%20and%20Thrace&poli=Kavala_Chryssoupoli (accessed on 10 March, 2023)