Knowledge discovery from energy consumption data

Dimitrios Th. Kazolis[1], *Jacob G.* Fantidis[1*], and *Nikolaos* Roumeliotis[2]

[1]Democritus University of Thrace, Department of Physics, 64003 Kavala, Greece
[2]Hellenic Electricity Distribution Network Operator S.A., Kavala area, Greece

Abstract. The acquisition of information and thus, the knowledge extraction from large databases, is a constantly developing modern scientific field, and a particularly important aspect of Information Technology. Different techniques and methodologies have been applied in combination with different types of data for obtaining the optimal result. This paper is a continuation of the effort to discover knowledge, in the form of correlations, from data concerning electricity consumption. The innovative part of this attempt is, the way that data was associated with time, and moreover, the combination of the used methods. Specifically, analytical consumption data was used, which were taken at a frequency of half an hour, throughout the year 2023. This consumption, which covers an entire city, concerning the indications of the distribution transformers found in different areas of the city of Kavala, in Greece. The data, was further combined with the time subdivisions of the whole year with the aim, to draw conclusions about the variation and association of consumption in relation to the hours, days and seasons of the year. In order to carry out the process, both statistical methods, such as factor analysis, normalization, and data mining techniques, such as cluster analysis were implemented. The final conclusion of the above process is that the methods used cooperate perfectly with each other. Furthermore, the analysis reveals that consumption is greatly influenced by certain periods of time during the year and this result seems strongly reasonable.

1 Introduction

The modern era in which we find ourselves is characterized by the use of electricity [1]. It is supplied indiscriminately by a complex distribution network and is based on instantaneous generation, since central storage is not efficient [2, 3]. On the other hand, it is easy to understand that the demand for electricity is not constant either during the day or even during the months of the year. It is influenced by factors that will not be examined in detail in this paper. For example, it will be mentioned that the Greek electricity company, knowing that demand falls during the night hours, has shown the possibility of reduced charges for consumption during the late evening hours [4].

For the above reasons, it is important to be able to know the variation in consumption in a specific region and also at central level, throughout the year.

* Corresponding author: ifantidi@physics.duth.gr

This can help both in predicting and in taking decisions concerning the management and also in modifications for the electricity generation and distribution network.

To this end, much work has been done in order to extract knowledge from data concerning electricity consumption.

Various extraction techniques have been applied, and the data bases used, have been enriched with many additional data, e.g. temperature, humidity, etc. [5-8].

This is therefore, the purpose of the present study. An analysis of both total and partly consumption data in a specific city was carried out. The data for this procedure was provided by the public's electricity company of Greece and concerned a specific area. That, after being enriched with temporal data, was processed in order to extract intelligible relationships between them. All of the procedures required for a correct and accurate extraction of knowledge were applied. The whole process is described in detail in the following section.

2 Materials and methods

2.1 Data processing

As mentioned above, the data for this procedure were obtained from the public's electricity company and concerned the area of the city of Kavala, in Greece. This included continuous recordings at half-hourly frequency, during the year 2023, of both the central and the individual transformers supplying the different points of the area. The original archive contained a total of 12 columns, from equal distribution points, and a total of 17520 lines corresponding to the frequency of the half-hourly recordings for the whole year. This was then checked and cleared of missing or incorrect data. That process was done manually with the utmost care so as not to distort the correctness of the overall results. For example, the missing entries that existed, for reasons unknown, were filled in, with the values of the average of the above and below existing prices. Then the file was enriched with temporal data, and in particular columns were placed which included the time during the day, the days of the week and the months. All of the above data were assigned by numerical values. For example, the column corresponding to Monday had a value of 1, when it was actually Monday, while all the other days had a value of 0.

The whole process eventually created an archive with a total of 21 columns. This archive was then subjected to a normalization process to ensure that, all the data are on the same scale. In general, all the procedures required to prepare the basis for a proper analysis were followed [9]. After all of the above, the file was ready for the beginning of the knowledge extraction process.

2.2 Factor Analysis

As described previously, a file was created which contained 21 columns and 17520 lines. This corresponds to a total of 367920 records. In order to assist the clustering algorithm to be used in the following, it was considered constructive for the overall procedure to use a multivariate exploratory statistical technique, in order to reduce the number of data. The archive was processed using the software Statistica. Thus, the application of Factor Analysis was carried out.

The factors will replace all the consumption data in the file, i.e. the original 12 columns, in order to reduce the volume of data without having a negative impact on the accuracy of the process. The number of factors was determined by Kaiser criterion and by using the

Scree-plot, as shown in the following Figure 1. From this it is easy to be seen that the inflection point which determines the number of factors to be extracted, is at 2 factors.

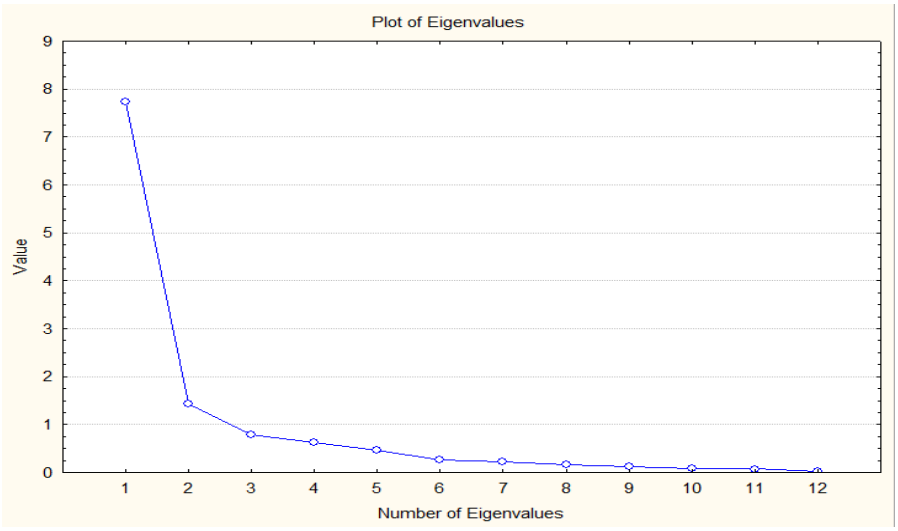

Fig.1. Screen-plot.

The following picture also shows the result of the process with the Factor loadings, where, it's also easy to be seen the impact that every Factor has, at the specific distribution points. Moreover, Varimax raw was used as Factor rotation and Principal components for the extraction [10]. The above are depicted in Figure 2.

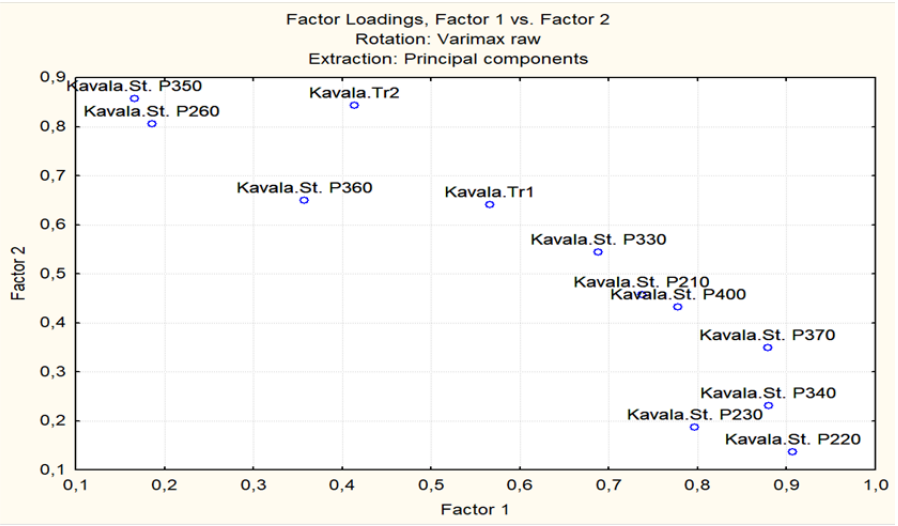

Fig.2. Factor loadings.

After the end of this process the final resulting file has only 9 columns and a total of 157680 records. So, there has been a significant reduction of these records, and in particular, they are 210240 less .The process of data mining can now follow.

2.3 Data mining

After the completion of the above stages, the archive is now ready for the final stage of processing, with the aim of acquiring knowledge. The method of data mining to be followed is Cluster Analysis, because it can effectively group the results according to the similarities between them [11].

In the selection of variables, therefore, only the time data and the two variables were selected. Next in the options of the method, the k-mean algorithm and the random selection of k observations, is chosen for the determination of the initial centers of the clusters. Moreover, the City-block (Manhattan) method was used for distance measure. In addition, the V-fold cross-validation method, with a smallest percentage decrease 5%, was used to determine the number of clusters and this resulted in the lower value of the set limit, which was 12. And finally, the number of iterations was set at a high number (500), so that, the whole process not to be interrupted before its completion, by this setting. After completing the adjustments, the software constructed the 12 clusters as shown in Figure 3:

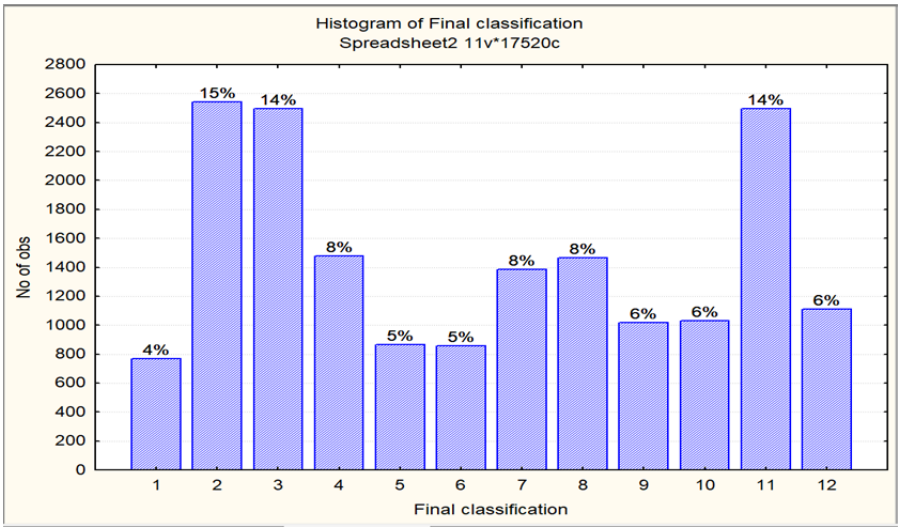

Fig. 3. Final classification.

It is worth pointing out here that, each time the procedure is repeated, the clusters will have a different form; however, they will still provide similar information.

Additional, the following Figure 4 and Figure 5 depict the distributions for the factors 1 and 2 respectively.

From Figure 4 it can be observed that the distributions for the Factor1 are not as centralized as those of Factor2, but on the other hand, the probability density is above 0.4 for most of the clusters, except for the numbers 2 and 10.

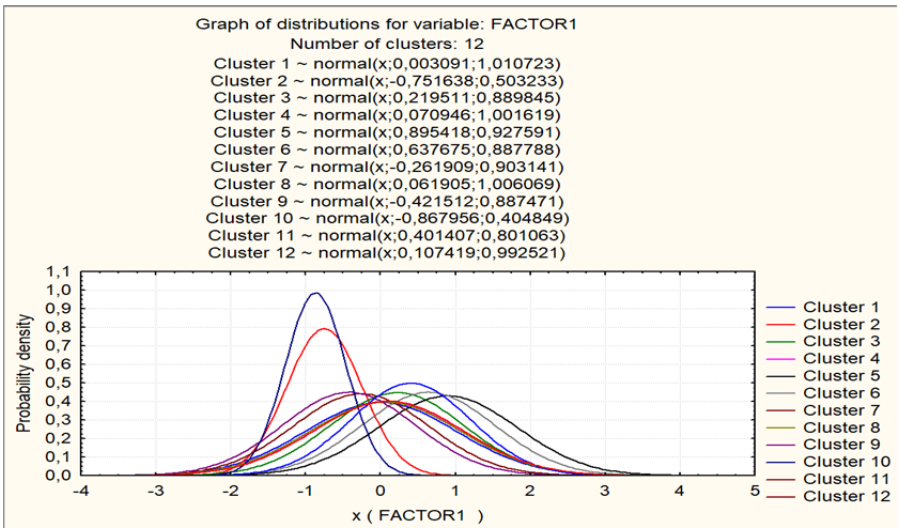

Fig.4. Distributions for the factor 1.

Similarly, in the Figure 5, it can be noticed that the distributions for Factor 2 are more centralized, but on the other hand, the probability density is under 0.4 for most of the clusters. On the contrary, here it can be seen that the higher probability densities are for clusters 2.10 and 5.

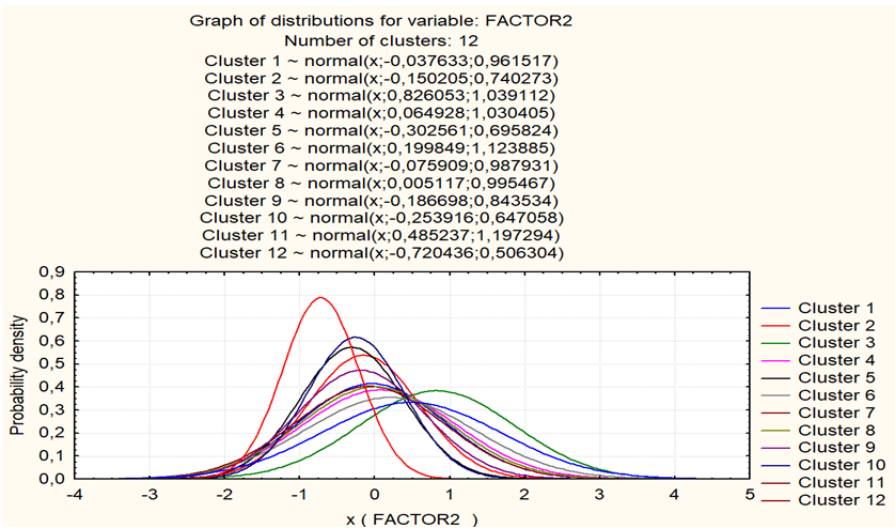

Fig.5. Distributions for the factor 2.

After all the above, the whole process is complete. The next stage that will follow is to verify the possibility of knowledge extraction from the analysis of the results.

3 Results

After the completion of all the above stages, this section will examine some of the clusters that have been created, in order to seek relevance in the elements they contain. t is decided to present in this chapter, only tree representative clusters, based on the number of its

members, with the help of the diagram in Figure 3. Thus, first of all, the cluster 1 will be examined. Part of it is shown in Figure 6.

Cluster members (ΚΑΒΑΛΑνεο ΔΕΗ meFactors (B2:YWW17521))
Number of clusters: 12
Total number of training cases: 17520

Case No.	Final classification	Time	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	month	FACTOR1	FACTOR2	sum.con.	Distance to centroid
82	1	0,687500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,56525	-0,29893	738,000	0,822724
83	1	0,708333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,58767	-0,19327	772,000	0,769757
84	1	0,729167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,70302	0,33295	875,000	0,632342
85	1	0,750000	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,89625	0,50195	932,000	0,625711
86	1	0,770833	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,96097	0,51387	930,000	0,612693
87	1	0,791667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,02273	0,45393	938,000	0,626215
88	1	0,812500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,06804	0,45265	942,000	0,657935
89	1	0,833333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,99786	0,53977	948,000	0,658895
90	1	0,854167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,03678	0,34214	920,000	0,696669
91	1	0,875000	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,02647	-0,03167	876,000	0,769685
92	1	0,895833	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,82648	0,19913	868,000	0,729830
93	1	0,916667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,64322	0,10268	828,000	0,757581
94	1	0,937500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,40304	-0,03038	769,000	0,826578
95	1	0,958333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,16497	-0,15435	724,000	0,933668
96	1	0,979167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,04677	-0,34521	684,000	1,026228
411	1	0,541667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,71295	0,13671	988,000	1,063740
412	1	0,562500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,93856	0,06437	1005,000	1,101719
413	1	0,583333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,89089	0,02462	1020,000	1,086982
414	1	0,604167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,62236	0,02346	981,000	0,995545
415	1	0,625000	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,48340	-0,08467	941,000	0,940462
416	1	0,645833	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,89763	0,16324	994,000	0,989193
417	1	0,666667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,77302	0,03778	995,000	0,964325
418	1	0,687500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,75286	0,06422	1002,000	0,940285
419	1	0,708333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,80421	0,12598	1016,000	0,928024
420	1	0,729167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,62107	0,52085	1068,000	0,853217
421	1	0,750000	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,83482	0,48083	1085,000	0,884718
422	1	0,770833	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,79330	0,51422	1095,000	0,857905
423	1	0,791667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,79782	0,49812	1091,000	0,848137
424	1	0,812500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,71316	0,53602	1076,000	0,840367
425	1	0,833333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,81036	0,44149	1077,000	0,891959
426	1	0,854167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,72372	0,48012	1070,000	0,888699
427	1	0,875000	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,48372	0,46388	1026,000	0,843697
428	1	0,895833	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,19851	0,33084	966,000	0,797057
429	1	0,916667	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,91870	0,12752	897,000	0,758853
430	1	0,937500	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,66730	0,05011	848,000	0,777840
431	1	0,958333	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,40286	-0,10169	788,000	0,845918
432	1	0,979167	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,14394	-0,31602	725,000	0,980136

Fig.6. Cluster number 1.

As can be seen from the final classification of Figure 3, this cluster does not contain many members, only the 4% of total, which means that it does not have a long duration in time. But, it is characterized by moments of high electricity consumption during Mondays mainly between 12:30 and 23:00. A typical point, near the center of this cluster (the distance from the center is 0.206), is on line 8487 in the archive, where at 18:30 the amperes were 802.The cluster number 2 is following and depicted in Figure 7.

Cluster members (ΚΑΒΑΛΑνεο ΔΕΗ meFactors (B2:YWW17521))
Number of clusters: 12
Total number of training cases: 17520

Case No.	Final classification	Time	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	month	FACTOR1	FACTOR2	sum.con.	Distance to centroid
38	2	0,770833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,60836	0,26811	845,000	1,123014
39	2	0,791667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,66721	0,25191	851,000	1,155902
40	2	0,812500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,67948	0,31200	862,000	1,194344
41	2	0,833333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,67044	0,39848	871,000	1,231507
42	2	0,854167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,70490	0,38195	871,000	1,256442
43	2	0,875000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,76878	0,21367	848,000	1,251372
44	2	0,895833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,68842	0,17088	836,000	1,245525
45	2	0,916667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,50523	0,07577	799,000	1,199341
46	2	0,937500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,27441	-0,00294	761,000	1,146597
47	2	0,958333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,06322	-0,04638	731,000	1,107023
48	2	0,979167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,01794	-0,25455	687,000	1,107582
337	2	0,000000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,00216	-0,41932	670,000	1,123141
338	2	0,020833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,19326	-0,58679	623,000	1,064321
339	2	0,041667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,43086	-0,65209	580,000	1,095869
340	2	0,062500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,57501	-0,71015	557,000	1,121588
341	2	0,083333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,71004	-0,78351	525,000	1,153410
342	2	0,104167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,47375	-0,76443	562,000	1,065978
343	2	0,125000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,55646	-0,79022	543,000	1,074225
344	2	0,145833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,62777	-0,83001	530,000	1,078721
345	2	0,166667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,69702	-0,84591	519,000	1,078344
346	2	0,187500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,74925	-0,82427	513,000	1,066792
347	2	0,208333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,76230	-0,83775	514,000	1,048592
348	2	0,229167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,78647	-0,74341	523,000	1,013352
349	2	0,250000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,71428	-0,71301	537,000	0,966821
350	2	0,270833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,68269	-0,61127	558,000	0,913150
351	2	0,291667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,52974	-0,44502	602,000	0,815515
352	2	0,312500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,34747	-0,60907	598,000	0,787887
353	2	0,333333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	-0,13928	-0,54257	625,000	0,747462
354	2	0,354167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,05560	-0,64968	615,000	0,780282
355	2	0,375000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,31155	-0,82963	593,000	0,841346
356	2	0,395833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,43301	-0,89541	558,000	0,871648
357	2	0,416667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,62795	-1,04651	532,000	0,920734
358	2	0,437500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,74569	-1,13679	519,000	0,940146
359	2	0,458333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,91497	-1,29896	501,000	0,981388
360	2	0,479167	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,96314	-1,21694	502,000	0,956517
361	2	0,500000	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,95743	-1,20309	486,000	0,963534
362	2	0,520833	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,01835	-1,32411	484,000	1,013140
363	2	0,541667	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,08705	-1,26248	509,000	1,022302
364	2	0,562500	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	1,04712	-1,22340	514,000	1,028253
365	2	0,583333	1,000000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,00000	0,88200	-1,21828	504,000	1,026614

Fig.7. Cluster number 2.

Here it can be easily noticed that this cluster has many members (15% of summary) and includes all of the Sundays of the year.

Another characteristic of this is its relatively low consumption throughout the day. As a distinctive feature of the cluster 2 will be mentioned the position 6408 in the archive, with a distance from the centre of the cluster 0.17, and 628 amperes.

Finally, the cluster 7 will also be presented (Figure 8).

Cluster members (ΚΑΒΑΛΛΑΡΙΟ ΔΕΗ meFactors (B2:YWW17521))
Number of clusters: 12
Total number of training cases: 17520

Case No.	Final classification	Time	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	month	FACTOR1	FACTOR2	sum.con.	Distance to centroid
324	7	0,729167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,96821	0,34113	921,000	0,725205
325	7	0,750000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,16546	0,56210	981,000	0,847873
326	7	0,770833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,26285	0,49400	977,000	0,874076
327	7	0,791667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,24482	0,57492	987,000	0,909549
328	7	0,812500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,34344	0,43300	979,000	0,923352
329	7	0,833333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,27011	0,51117	975,000	0,940274
330	7	0,854167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,18728	0,36935	949,000	0,911744
331	7	0,875000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,97840	0,25679	901,000	0,851940
332	7	0,895833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,91819	0,13318	877,000	0,860017
333	7	0,916667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,69222	0,12482	835,000	0,817586
334	7	0,937500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,50183	-0,00073	795,000	0,798602
335	7	0,958333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,26330	-0,10659	754,000	0,851017
336	7	0,979167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,19114	-0,24431	718,000	0,925890
642	7	0,354167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,51141	-0,32860	781,000	0,961245
643	7	0,375000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	0,98401	-0,27427	841,000	1,050923
644	7	0,395833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,29631	-0,34879	869,000	1,110918
645	7	0,416667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,36426	-0,58810	835,000	1,113143
646	7	0,437500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,37477	-0,34024	853,000	1,070784
647	7	0,458333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,40078	-0,33167	859,000	1,056519
648	7	0,479167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,50951	-0,29525	885,000	1,065031
649	7	0,500000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,46507	-0,06956	922,000	1,028024
650	7	0,520833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,70645	-0,13529	949,000	1,073995
651	7	0,541667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,86766	-0,01422	982,000	1,084246
652	7	0,562500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	2,13438	-0,35875	965,000	1,145622
653	7	0,583333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	2,13579	-0,50582	946,000	1,133022
654	7	0,604167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,97571	-0,57572	899,000	1,064723
655	7	0,625000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,66657	-0,45104	886,000	0,965115
656	7	0,645833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,99452	-0,22204	965,000	1,017692
657	7	0,666667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,96571	-0,25499	957,000	0,991033
658	7	0,687500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,74275	-0,03995	961,000	0,904362
659	7	0,708333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,81411	-0,10596	971,000	0,913621
660	7	0,729167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,61115	0,23188	1014,000	0,880218
661	7	0,750000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,79173	0,37154	1062,000	0,979550
662	7	0,770833	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,82314	0,32308	1069,000	1,003907
663	7	0,791667	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,83237	0,33804	1072,000	1,030682
664	7	0,812500	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,72642	0,34549	1057,000	1,025477
665	7	0,833333	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,72686	0,32350	1050,000	1,039479
666	7	0,854167	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,61647	0,17483	1016,000	1,017862
667	7	0,875000	0,000000	0,000000	0,000000	0,000000	0,000000	0,000000	1,000000	1,00000	1,47429	0,07477	976,000	1,003668

Fig.8. Cluster number 7.

This group includes 8% of the total and only the Saturdays, between 12:00 and 23:30. During these periods, consumption is at relatively middle levels. A typical line from the archive, will also be mentioned here the 9058, with a distance from the center of 0.22 and has 709 A.

The remaining clusters created, were also examined in detail. But, for reasons of convenience they will not be shown here. There will only be a general report on them. Thus, the cluster No. 3 includes the Thursdays and the cluster No. 11 the Tuesdays. Moreover the cluster No. 12 involves also the Saturdays, as cluster No.7, but between 0:00 and 11:30. It can be noticed here, the similarities and the differences with cluster No.7, examined above, which includes different hours of the same day, namely from 12:00 until 23:30.

Furthermore, the Monday is interpreted by 3 clusters, the cluster No. 1 that was examined above and the clusters No.5 and 6 .Of them, the cluster No.6 consistently represents the night hours, while the cluster No.5 represents the morning hours of some days. In the same way it was found that clusters No. 4 and 9, correspond to Wednesday but at different time periods of the same day, namely from 00:00 to 08:30 for cluster No. 9 and from 09:00 to 23:30 for cluster No.4. Similarly, it was observed that clusters No. 8 and 10, represent Friday but at separate time intervals, specifically from 00:00 to 08:30 for cluster No. 10 and from 09:00 to 23:30 for cluster No.8. It is therefore generally observed that consumption is affected both by the day and by the hours.

4 Conclusion

After the completion of all the above procedures applied, and the analysis of the results, the possibility of logical correlations from the data and the extraction of knowledge is evident. The methods used proved to be effective and reliable, and moreover, the cooperation between them was constructive.

In particular, the Factor Analysis accurately represented a large volume of data with two single factors. This method proved to be accurate and constructive. Accurate because, as the results showed, the two Factors represented with high reliability the twelve variables they replaced, and constructive because by using them, the size of the database drastically reduced. This procedure therefore, can be justifiably applied to the initial processing of the database, in order to considerable assist the subsequent data mining process. As a follow-

up, Cluster Analysis worked computationally easier on a smaller data volume and, divided the samples into groups that contained strong similarities in the data they contained. On the other hand, each group had significant differences among them, which justified the separation.

As a general conclusion from the whole effort, it is evidenced that electricity consumption is highly influenced not only by the days of the week but also by the specific hours of each day. Generally, consumption is reduced during the midnight hours as well as during the morning hours on Sundays. On the contrary, in the morning hours on the other days of the week, especially on Mondays, after 7 a.m., there is a significant increase of the demand.

The information that can be extracted using this methodology can be a valuable tool, both for taking relevant decisions, and as a basis for further analysis.

In particular, as a continuation of this paper, a separate analysis could be carried out for each one of the days, in order to establish the variation of consumption, during the 24-hour period, within a given day. In addition, the same procedure could be used to select more clusters, in order to create more detailed groups of data, so that it would be possible to examine how the records are affected by specific time periods, e.g. Christmas time, months, holiday periods, periods of extreme climatic conditions, etc.

References

1. L. Jin, K.Duan, X. Tang, Empirical Analysis Based on Provincial Panel Data from China. Sustainability, 10(2), **145** (2018)
2. P. Pálenský, D. Dietrich, IEEE Transactions on Industrial Informatics, **7**(3), 381–388. (2011)
3. O. Schmidt, S. Melchior, A. Hawkes, I. Staffell, Joule, **3**(1), 81–100. (2019)
4. Electricity for households, Available at: https://www.dei.gr/en/home/electricity/g1-g1n/ (accessed on 15 April 2024)
5. D. Kazolis, P. Kogias, N. Roumeliotis, Energy cluster analysis based on consumption data in different weather condition. E3S Web of Conferences, **404**, 01005, (2023)
6. S. Singh, A.Yassine, Big data mining of energy time series for behavioral analytics and energy consumption forecasting. Energies, 11(2), **452** (2018)
7. H. Beloev, I. Stoyanov, T. Iliev, *Good Practices in Implementing Energy Efficiency Measures in „Angel Kanchev" University of Ruse*, in the Proceeding of the 2022 8th International Conference on Energy Efficiency and Agricultural Engineering (EE&AE), Ruse, Bulgaria, pp. 1-4, (2022), doi: 10.1109/EEAE53789.2022.9831424
8. D. Kazolis, I. Gerontidis. Knowledge Mining from Student Data. Journal of Engineering Science and Technology, Special Issue, pp. 299–302, (2020)
9. I. Vlahavas, I. Kefalas, P. Bassiliades, N. Kokkoras, F.Sakellariou, *Artificial Intelligence*, 4th ed., University of Macedonia Press. Thessaloniki, Greece, (2020)
10. J. Horn, A rationale and test for the number of factors in factor analysis, Psychometrika, **30**(2), 179–185. (1965)
11. A. Ikotun, A. Ezugwu, L. Abualigah, B. Abuhaija, H. Jia. Information Sciences, **622**, 178–210 (2023)