ATTEST

WP5

Development of a smart environment for asset management in power grids

Tool for the definition of condition indicators based on heterogeneous information sources

D5.3

DOCUMENT CONTROL PAGE

DOCUMENT	D5.3 – Tool for the definition of condition indicators based on heterogeneous information sources
TYPE	Tool Manual
DISTRIBUTION LEVEL	Restricted
DUE DELIVERY DATE	31 / 12 / 2021
DATE OF DELIVERY	20/ 01 / 2022
VERSION	V2.0
DELIVERABLE RESPONSIBLE	Comillas
AUTHOR(S)	Gopal Lal Rajora, Miguel A. Sanz-Bobi, Rafael Palacios, Carlos Mateo, Rafael Cossent
OFFICIAL REVIEWER(S)	

DOCUMENT HISTORY

VERSION	AUTHORS	DATE	CHANGES
0.1	Gopal Lal Rajora COMILLAS	10 /12 / 2021	First complete version
0.2	Miguel A. Sanz-Bobi COMILLAS	22 / 12 / 2021	Second revision
0.3	Dino Žanić HOPS	24 / 12 / 2021	Received comments from first reviewer
1.3	Gopal Lal Rajora COMILLAS	05/01/2022	Third consolidated revision
1.4	Hrvoje Keko Koncar Digital	19/01/2022	Received comments from second reviewer
2.0	Gopal Lal Rajora COMILLAS	20/01/2022	Final version

Table of Contents

List of Figures

List of Tables

Abbreviations and Acronyms

DSO	Distribution System Operator
SOM	Self-Organized map
TSO	Transmission System Operator

1. INTRODUCTION

The primary goal of work package 5 is to develop instruments which can help electric utility companies to make decisions regarding management of assets typical for TSO/DSO power systems. This goal will be reached by design and evaluation of a certain indicators which characterize the asset during its exploitation period. The strategy for defining these indicators and their interpretation of them will be implemented in three different tools that will support the whole process of asset management in the context of the ATTEST project.

This deliverable describes the User Manual of the second tool of Work Package 5. The goal of this tool is the development of methods for estimating condition indicators for the assets based on the results obtained in the first tool described in the deliverable 5.2. These methods will be based on the heterogeneous information/data sources feeding the first level of the asset management. The idea behind the tool described in this deliverable is to obtain comparable condition indicators for different assets. This introduces an important advance in the current state-of-the-art. This tool is able to supply a perspective of the current condition of the assets studied building an ordered list of them according to four dimensions of analysis previously defined in the deliverable D5.2.

2. Tool for the definition of condition indicators based on heterogeneous information sources.

2.1. SCOPE

The goal of this tool is to define a set of the condition indicators of the assets studied based on heterogeneous information sources. This tool defines an innovative approach to translate the results obtained from the previous application (tool reported in deliverable D5.2) into a set of harmonized, easily measurable, and comparable indicators for different types of assets, which will allow identification of elements which require special attention from an asset management point of view.

2.2. METHODOLOGY

This section describes the methodologies that run behind the tool to explain how the results are obtained and explained. The main objective of this tool is to use simultaneously all four dimensions obtained in the previous level to integrate this information in comparable condition indicators for the assets. A new clustering algorithm is to be applied for combining the information collected from the four dimensions in a weighted formula. The user selects the weights according to the desired strategy of the electrical company. Each dimension contributes to the assessment of each asset according to predefined weights as shown in Figure 1.

2.2.1.1. Clustering process. The number of clusters

The development of methods for estimating condition indicators for the assets is carried out through a clustering process in which assets with similar features are grouped. Clustering refers to the process of automatically grouping together data points with similar characteristics and assigning them to "clusters." The total number of clusters (patterns) is determined through a well-known clustering algorithm: k-means; and two morphological assessment methods: elbow method and gap statistic method.

In this tool, the same process followed as tool 5.1 for find the optimum number of clusters. We are using Elbow method and the Gap Statistic method to find the optimum number of clusters. Elbow method is based on the distortion (within-cluster dispersion value), based on the Euclidean distance of each sample to the center of the cluster, to which each observation belongs. Gap Statistic method is based on comparing the dispersion value obtained for the input dataset and an expected one for an equally distributed idealistic dataset. (Tibshirani, 2001)

2.2.1.2. Clustering process. Normalization and data preprocessing.

The basic indicators obtained from tool 5.1 are normalized before computing each of the clusters that define the SOM. This normalization is carried out using their maximum and minimum value. The min-max normalization is defined by the following equation:

$$v_{y,norm} = \frac{v_y - min(V_y)}{max(V_y) - min(V_y)} \qquad 1$$

where $v_{y,norm}$, is the normalized value; y, is the variable to be normalized; v_y is an observation of the variable y and V_y is the set of observations of the variable y.

2.2.1.3. Weighting Process.

A new clustering algorithm is applied to combine the information collected from the four dimensions in a weighted formula. The user selects the weights according to desired strategy of the electrical company. Each dimension contributes to the assessment of each asset according to predefined weights as shown in Figure. 1.
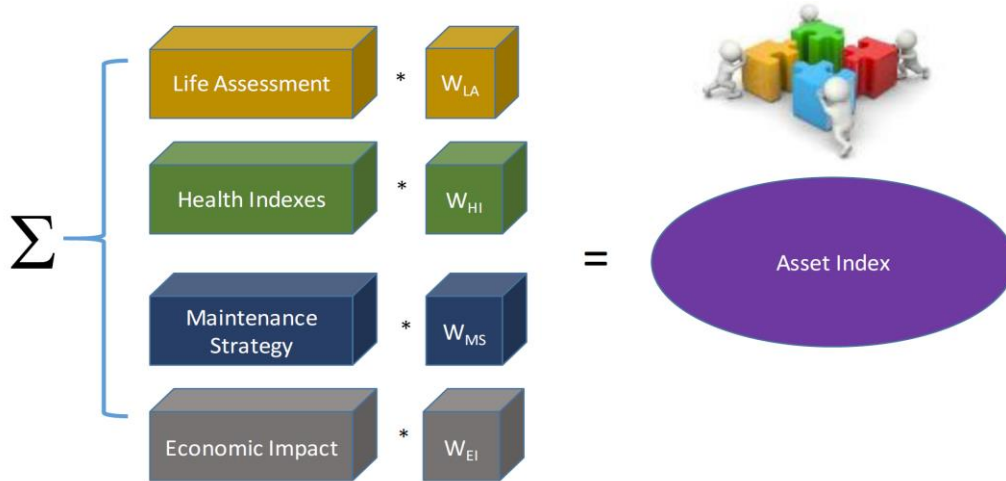

FIGURE 1 COMBINATION OF ALL DIMENSIONS

Weights will be assigned according to each feature's importance. The weight of a particular feature will depend on the importance factor, higher importance corresponds to higher weight factor.. The sum of all the weights will be 1. For example, we have three features such as x1, x2 and x3 where the importance of the feature goes in ascending order respectively. We assign weight to each feature such as w1 for x1, w2 for x2, and w3 for x3.

$$w = 1 = w1 * X1 + w2 * X2 + w3 * X3 \qquad 2$$

$$Indicator = w1 * X1 + w2 * X2 + w3 * X3 \qquad 3$$

So, the same process is done for all dimensions and we will generate indicators for each of them.

2.2.1.4. Clustering process. SOM training process.

The optimal number of clusters is not an exact indicator. This means that there might be some variations between the initial optimal number and the final number of clusters. These variations are caused by the initial optimum number of clusters being be too large, producing some empty clusters. An empty cluster means that the layout is not optimum, therefore the number of clusters will be reduced by one until all the clusters are full. This iterative process is depicted in Figure 2.
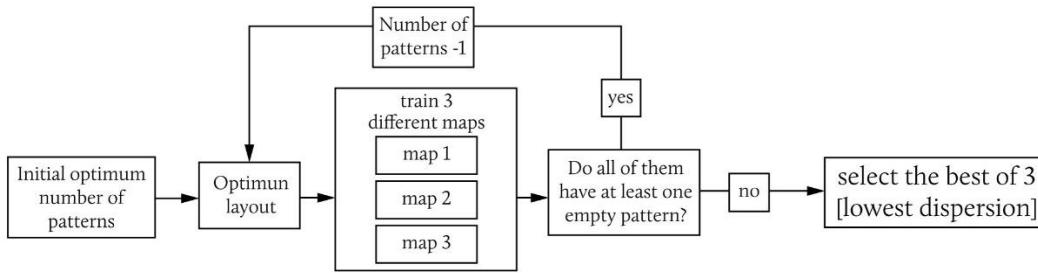
FIGURE 2 FLOW CHART OF THE TRAINING PROCESS OF THE SOM

2.2.1.5. Clustering process. Denormalization and visualization.

The information contained in each pattern is denormalized to its original scale. The equation applied is the inverse of the previous one:

$$v_y = v_{y,norm} \cdot \left(\max \left(V_y \right) - \min \left(V_y \right) \right) + \min \left(V_y \right)$$

The information of each pattern is represented through box-plot diagrams like the one shown in the following figure 3:
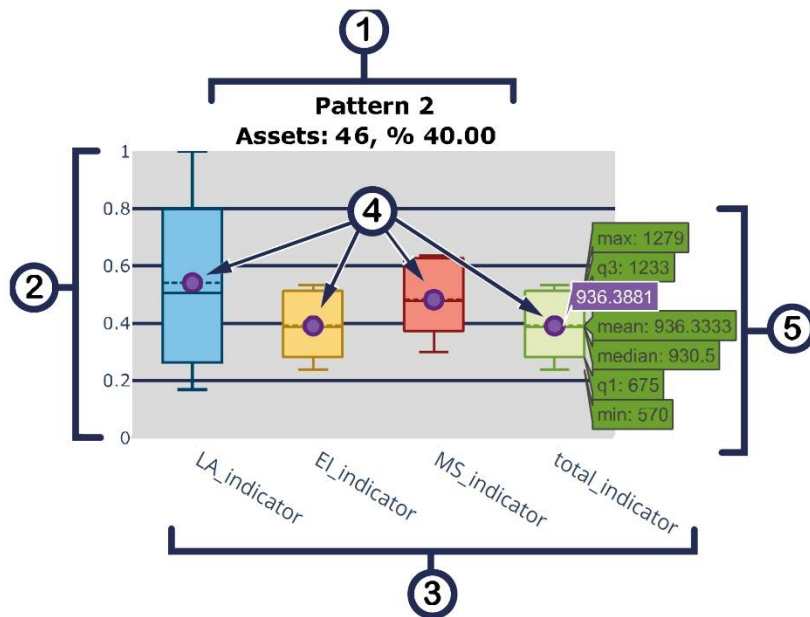
FIGURE 3 ELEMENTS OF THE BOX PLOTS

The main elements of each plot are:

1) The header of the pattern. It includes the number of the pattern, the number of assets within the pattern, and the percentage of assets included within the pattern regarding the whole set of assets.

2) The main scale. This scale allows the user to compare variables with different maximum and minimum values.
3) The name of, indicators, or measurements included in the study.
4) The centroid of the patterns. The centroids are the most representative values of each pattern.
5) Box-plot. They show how the assets are distributed within the same pattern. The values of the box plots are not normalized. Each box contains information about the minimum value, the first quartile, the mean value, the median, the third quartile, and the maximum value of each group of assets.

3. INPUTS AND OUTPUTS

This section presents the formats of the input files loaded and output files generated by the tool.

3.1. INPUTS

Input files are:

- **Assessment file (Life assessment, Maintenance strategy, Economic Impact):** It includes the features that characterize the state of each asset. This input file might contain numerical and non-numerical variables but only the numerical values are included within the state patterns.

    Comments:

    o Selected variables from Dimension files must be numerical.
    o Dimension files must not contain empty or NaN values. It such cases, it is recommendable to fill empty values with Ø, in case of numbers; or "not available", in case of text strings.

    File format:
    o Compatible formats: .CSV or Excel

- **Coordinate file:** This input file must contain Identifier, Latitude and Longitude.

    Comments:
    o Auxiliary files must share a common identifier column with the Dimension file.

    File format:
    o Compatible formats: .CSV or Excel

CSV or Excel format comments:

CSV files can present different formats. This can hinder the parsing process of the file, producing errors in the reading process of the file. Therefore, it is important to take into account that this tool only supports the following CSV coding:

- Two types of **delimiter characters** are accepted: **semicolon** (;) or **comma** (,)
- One type of **decimal separator** is accepted: **dot** (.)

3.2. OUTPUTS

3.2.1. OUTPUT FILE FORMATS

The outputs of the tool are:

- Pattern dashboard. The patterns obtained for each type of asset are shown in separate dashboards for each dimension included in the assessment. Such dashboards are visualized as an interactive webpage in HTML format.

- Assets location visualization. The geographical location of assets can be seen in a separate HTML output file.

3.2.2.OUTPUT FILE LOCATIONS

The outputs of the tool are located within the folder: *static\results_asset_management*

3.2.3.OUTPUT VISUALIZATION

Output files can be visualized through the tool; by contrast, it can be visualized outside the tool opening the generated HTML file.

Note: To see the geographical locations of the assets, this tool requires a file (CSV or Excel) that contains asset names and coordinates (Latitude and Longitude) of the assets.

4. USER INTERFACE

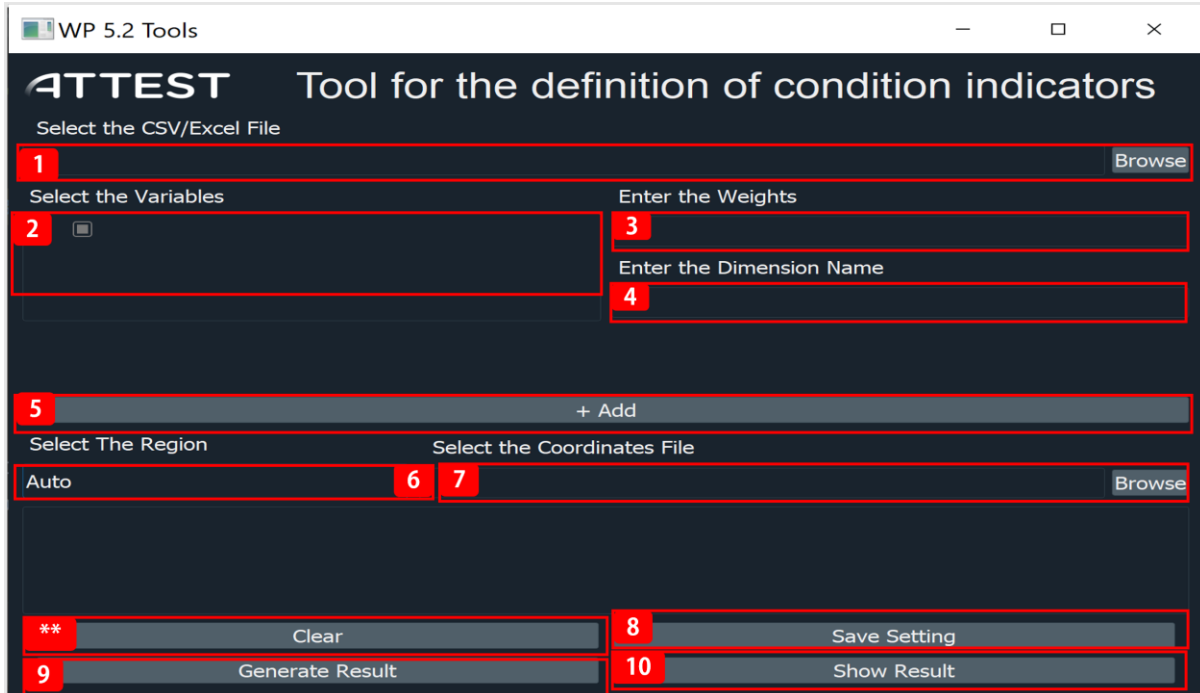This section describes the main elements of the user interface, including the steps to load the inputs for tool.

FIGURE 4 USER INTERFACE OF TOOL

The steps to load the input files of the Asset Management Tool are the following:

1) Enter the path to the dimension input file.

----------------- Configuration of a dimension -----------------
Repeat these steps for each dimension.

2) Select the variables for the weighting process.
3) Enter the weights for selected variables and separate by commas.
4) Enter the Dimension name.
5) Add another dimension file and repeat the step 1 to 3.
6) Select the Region.
7) Enter the path of the coordinates file.
8) Save dimension configuration. Once each dimension is configured it has to be saved by clicking on this button.

----------- End of the configuration of a dimension -----------

9) Generate results. Starts the process that generates the results with the configuration selected.

10) Show results. Starts the local server and shows the corresponding dashboards.

The results generated from this tool are divided in 3 parts.

- **List of the total indicators:** this output contains the tabular data format for all the indicators, sorted in descending order.

name	LA_indicator	MS_indicator	EI_indicator	total_indicator
48	0.351	0.8	0.551	0.567
19	0.504	0.541	0.652	0.566
12	0.663	0.509	0.506	0.559
39	0.096	0.842	0.682	0.54
44	0.181	0.808	0.611	0.533
72	0.626	0.509	0.432	0.522
28	0.375	0.336	0.693	0.468
5	0.352	0.477	0.512	0.447
11	0.55	0.4	0.342	0.43

FIGURE 5 LIST OF THE TOTAL INDICATORS

- **Dashboards.** One example of dashboard is shown in Figure 6. Each dashboard is made up of:

1) Drop down menu which allows the user to see full information of the assets and its centroid.
2) Pattern graph. Each of the pattern contains information about how the assets contained in such pattern are distributed. Clicking on the title of the pattern it shows a table with the information available of all the assets that belong to that pattern. The title also includes the number and the total percentage of assets included in the pattern.
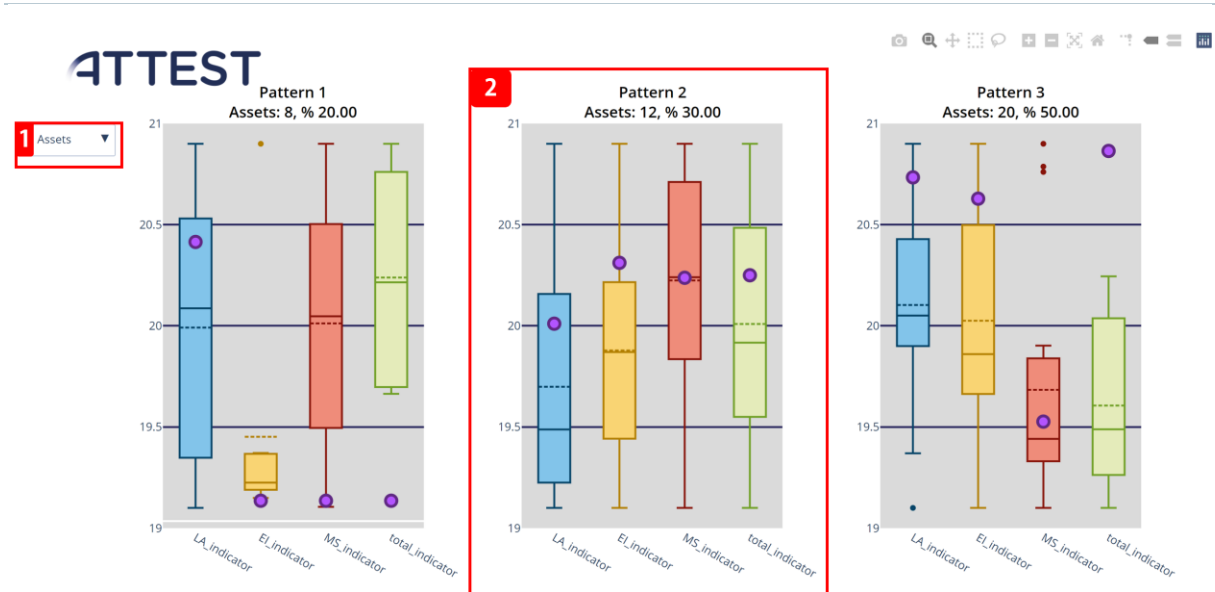
FIGURE 6 EXAMPLE OF DASHBOARD

- Geolocation of the Assets.

Assets location and condition can be viewed on a map. There are 3 color types for location points, red, orange and green. Red represents the indicator with highest risk where value of total indicator is greater then 75%. Orange color shows the indicator locations with medium risk where total indicator is between 50% to 75% and the green color shows the location of indicators with minimal risk with indicator value less then 50%. Each asset is represented by a marker which also shows the assets name/number, indicators and total indicator.
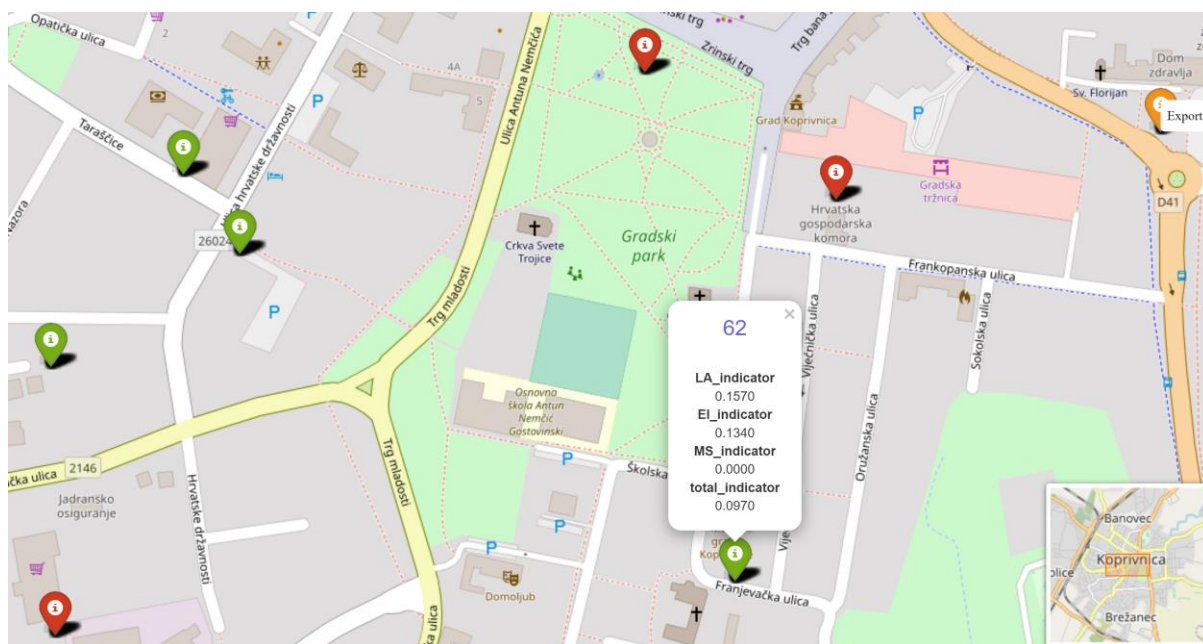

FIGURE 7 GEOLOCATION OF THE ASSETS

5. INSTALLATION

The software is developed as a portable tool. It only has to be unzipped within the desired folder. The files contained in the zip file are:
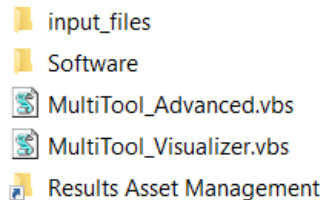

FIGURE 8 MAIN FILES OF THE TOOL

- Input_files: folder with the case studies included in this manual.
- Software: folder with all the Scripts and Virtual Environment files of the tool.
- MultiTool_Advanced.vbs: Script to call the advanced tool to generate and visualize the results contained in the corresponding results folder.
- MultiTool_Visualizer.vbs: Script to call the tool to visualize the results contained in the corresponding results folder.
- Results Asset Management: Opens in a new window the folder that contains the results generated by the Asset Management Tool.
- Results Condition Monitoring: Opens in a new window the folder that contains the results generated by the Condition Monitoring Tool.

The PDFs of the clustering process for each dimension (Optimal number of clusters and training process) are not opened automatically but they are accessible through the path:

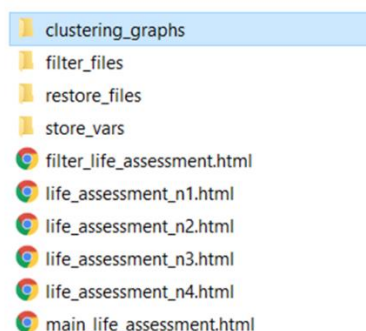- o Results Asset Management >> clustering_graphs

FIGURE 9 FOLDER WITH THE RESULTS OF THE CLUSTERING PROCESS

5.1. Software – Hardware requirements

- This software has been developed for a system with these minimum requirements:
  - OS: Windows 7 SP1/8.1/10 - 64bit.
  - Screen resolution. 1920 x 1080.
  - Python 3.8
  - RAM: 4GB
  - Hard Drive Space Required: 3 GB

6. CASE-STUDIES

In order to illustrate the type of the results obtained, several real cases were analyzed through this tool.

6.1. HEP – ASSETS

6.1.1. Transformers

- o   Number of assets: 40 transformers
- o   Results:
  - ▪ 3 patterns identified.
  - ▪ Identifier 48 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
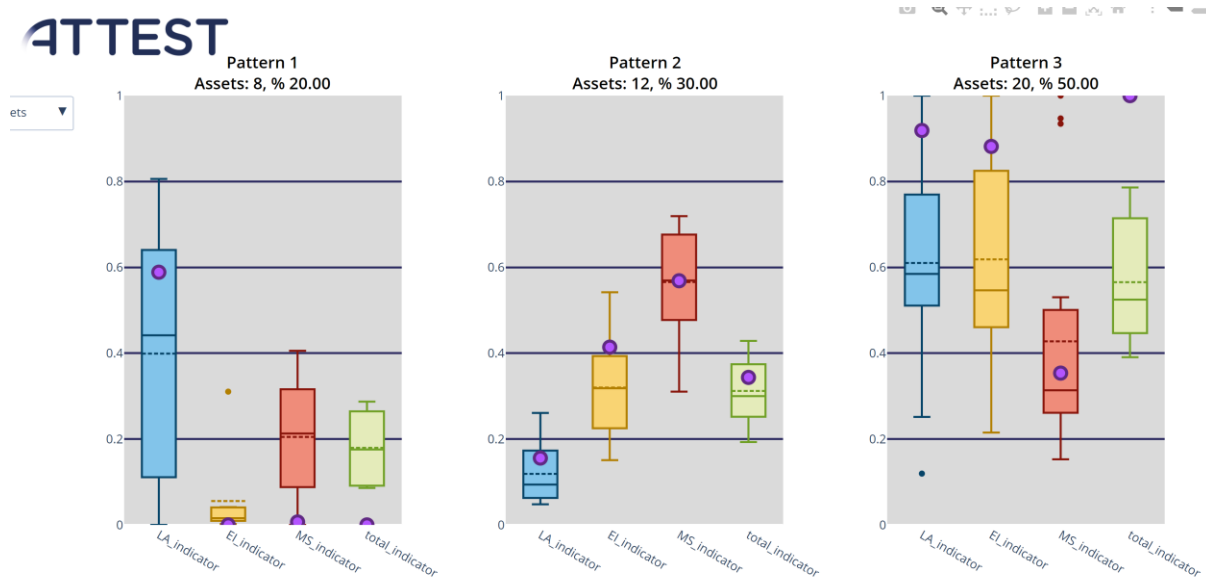

FIGURE 10 RESULTS: HEP TRANSFORMERS

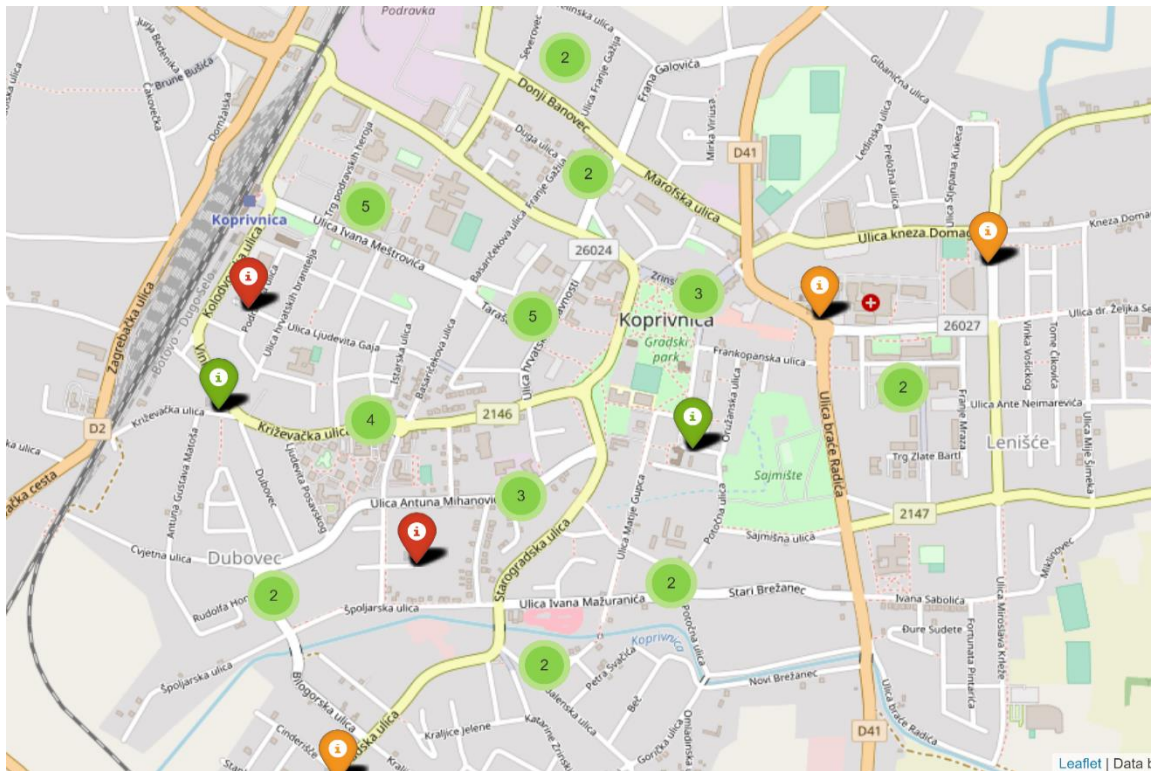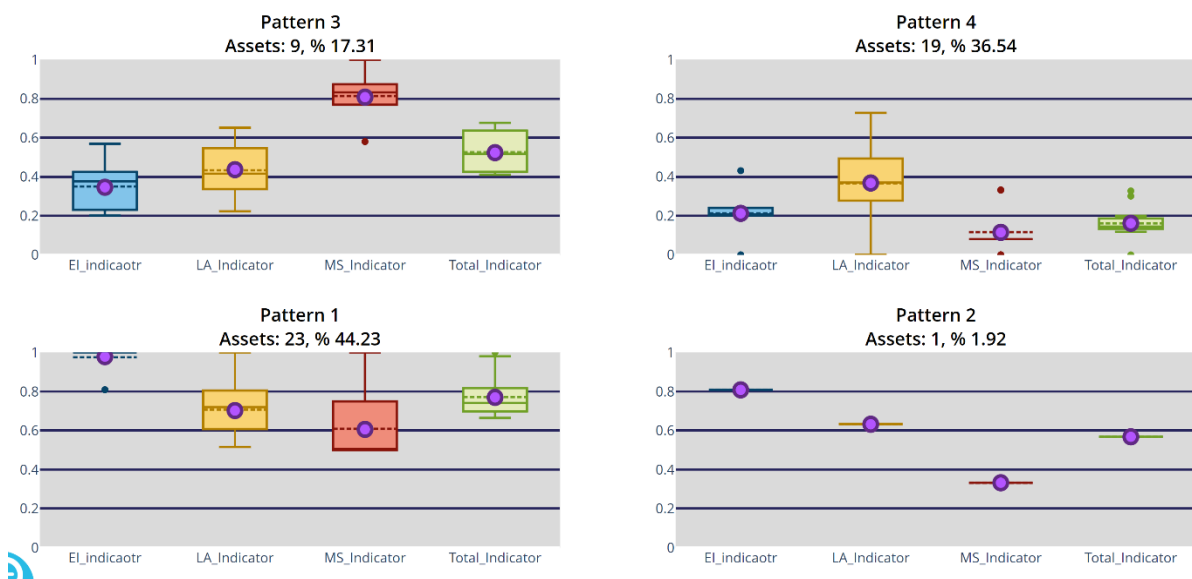| name	LA_indicator	MS_indicator	El_indicator	total_indicator
48	0.351	0.8	0.551	0.567
19	0.504	0.541	0.652	0.566
12	0.663	0.509	0.506	0.559
39	0.096	0.842	0.682	0.54
44	0.181	0.808	0.611	0.533
72	0.626	0.509	0.432	0.522
28	0.375	0.336	0.693	0.468
5	0.352	0.477	0.512	0.447
11	0.55	0.4	0.342	0.43

FIGURE 11 RESULTS: GEOLOCATION OF HEP TRANSFORMERS

6.1.2. Transmission Lines

- o Number of assets: 52 lines
- o Results:
  - ▪ 4 patterns identified.
  - ▪ Identifier 1KV3/10 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.

Name	El_indicator	LA_Indicator	MS_Indicator	Total_Indicator
1KV3/10	1.0	0.816	1.0	0.704
1KV3/4	1.0	0.77	1.0	0.692
1KV3/2	1.0	0.757	1.0	0.689
1KV17/10	0.832	0.942	0.832	0.651
1KV17/12	0.832	0.812	0.832	0.619
1KV17/9	0.832	0.727	0.832	0.598
1KV3/1	1.0	0.828	0.5	0.582
1KV3/17	1.0	0.787	0.5	0.572

6.2. Spanish – ASSETS

6.2.1.Transformer

o   Number of assets: 92 transformers.
o   Results:
  ▪   4 patterns identified.
  ▪   Identifier 28CDR2 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.

FIGURE 12 RESULTS: SPANISH GRID TRANSFORMERS

name	LA_indicator	MS_indicator	EI_indicator	total_indicator
28CDR2	0.867	0.752	0.91	0.843
28CHHS	0.51	0.781	0.477	0.589
28CQPQ	0.491	0.458	0.4	0.45
28CDR5	0.254	0.749	0.155	0.386
28CYL3	0.297	0.513	0.263	0.358
28CHQD	0.362	0.032	0.563	0.319
28CIA7	0.399	0.306	0.246	0.317
28CNTW	0.253	0.3	0.383	0.312

FIGURE 13 RESULTS: GEOLOCATION OF SPANISH GRID TRANSFORMERS

6.2.2. Support

- o Number of assets: 380 towers/supports.
- o Results:
  - ▪ 4 patterns identified.
  - ▪ Identifier 74-32 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
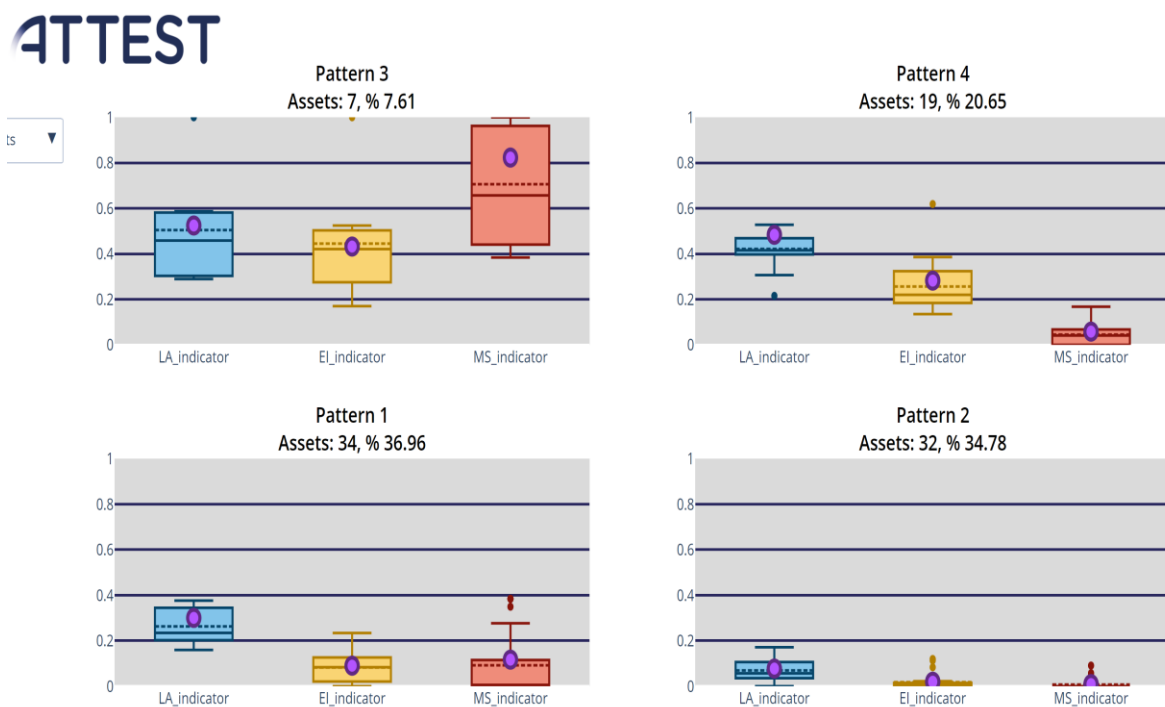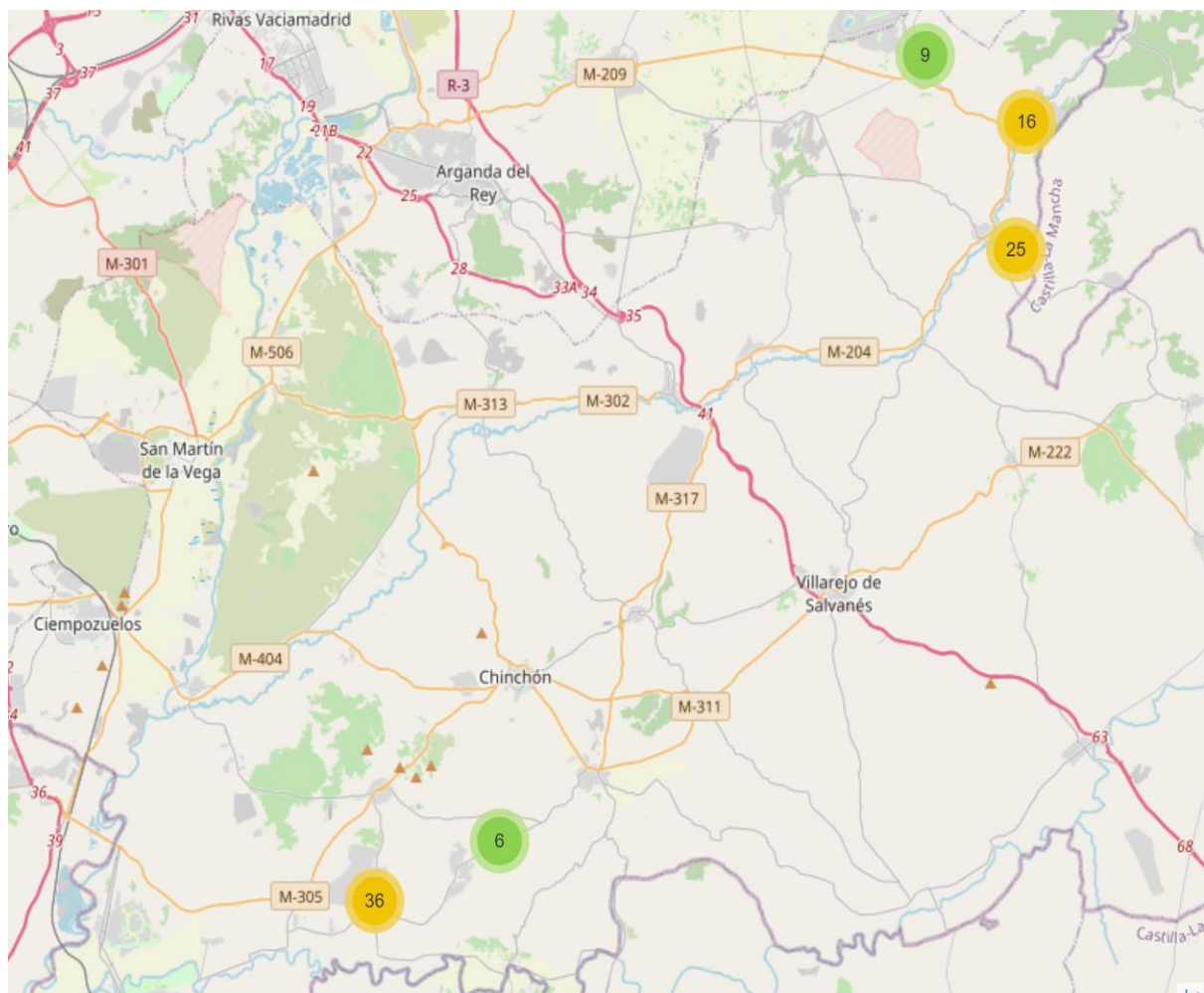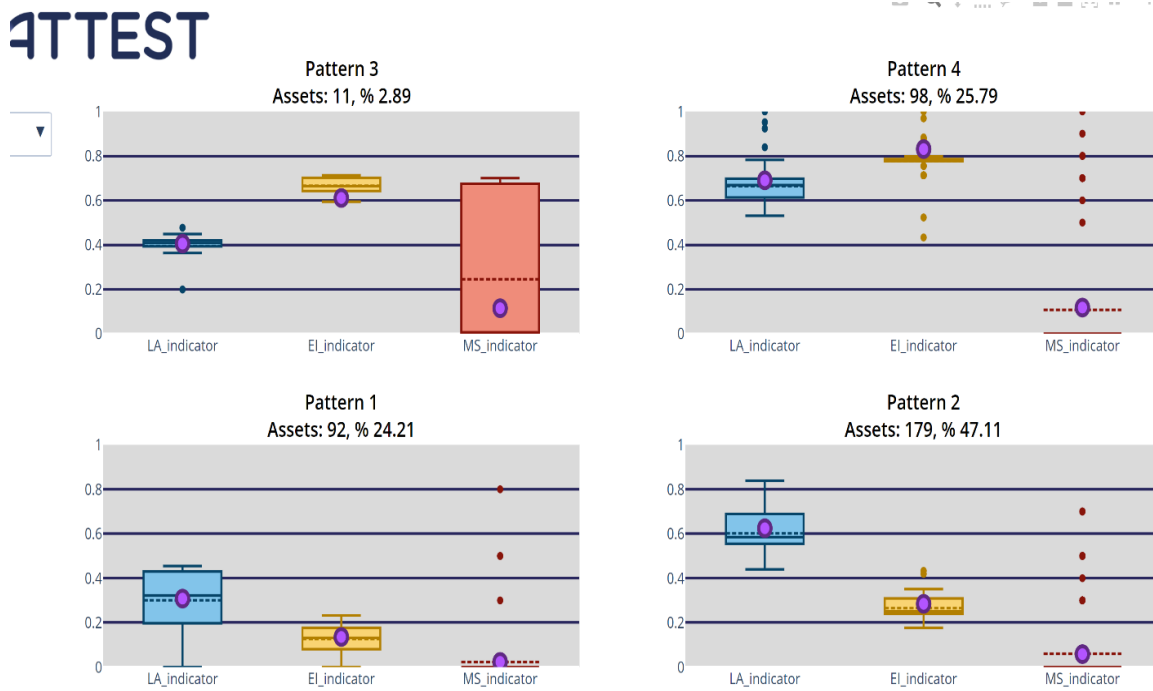

FIGURE 14 RESULTS: SPANISH GRID SUPPORTS

Support_ID	LA_indicator	MS_indicator	El_indicator	total_indicator
RQ9W9OJR//74-32	0.885	0.8	0.904	0.863
RQ8UCGAG//74-33-4	0.729	1.0	0.819	0.849
RQAUFPTH//74-33-2	0.729	0.9	0.819	0.816
RQBQDFTA//74-33	0.751	0.8	0.824	0.792
SAM60AF0//D59-70	0.912	0.5	0.933	0.782
RQA9HCG7//74-33-3	0.729	0.7	0.819	0.749
RQ3K0TFG//82	0.958	0.8	0.488	0.749
RQ37A8RE//74-28	0.724	0.7	0.804	0.743

FIGURE 15 RESULTS: GEOLOCATION OF SPANISH GRID SUPPORTS

6.2.3. Transmission Lines

- o Number of assets: 219 transmission lines.
- o Results:
  - 3 patterns identified.
  - Identifier CD07031339 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
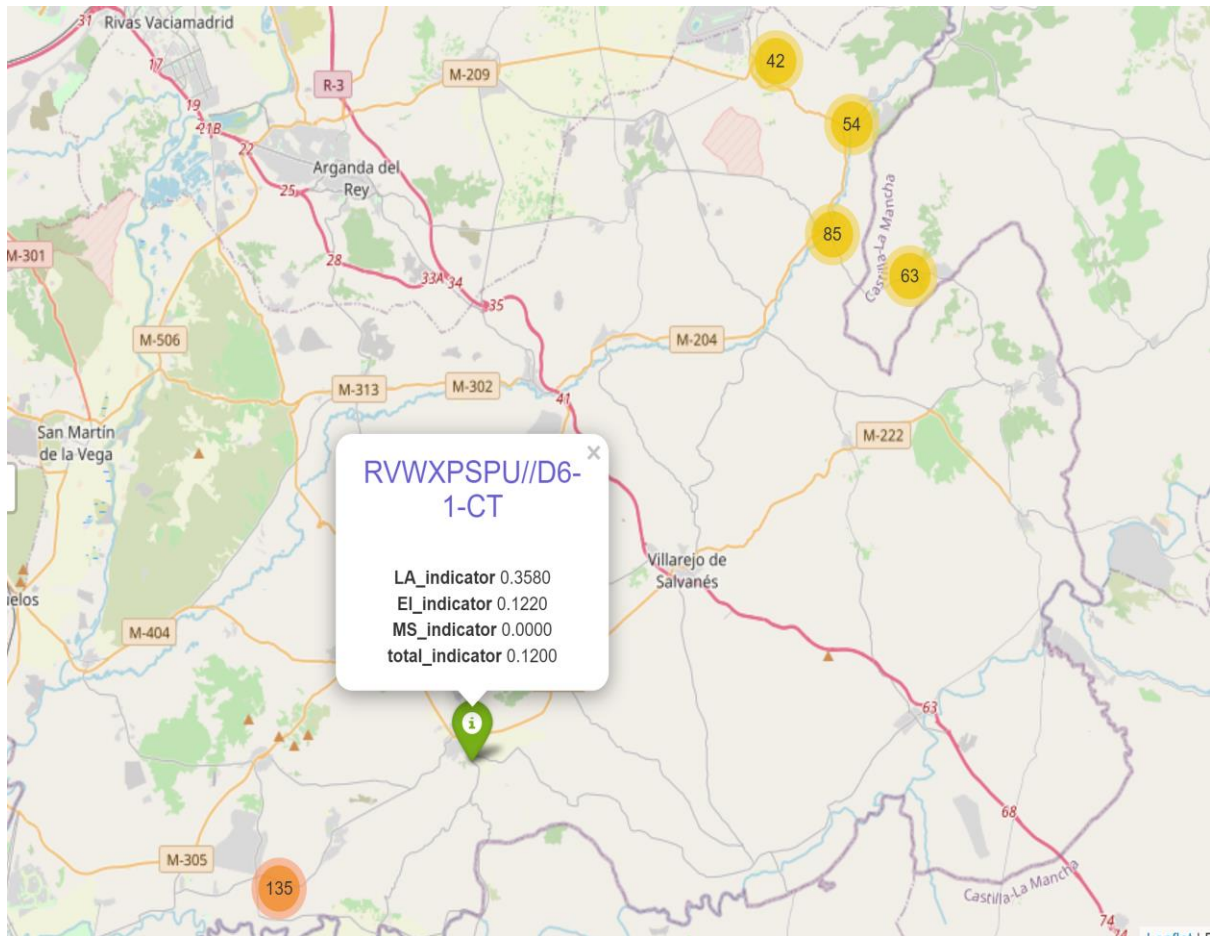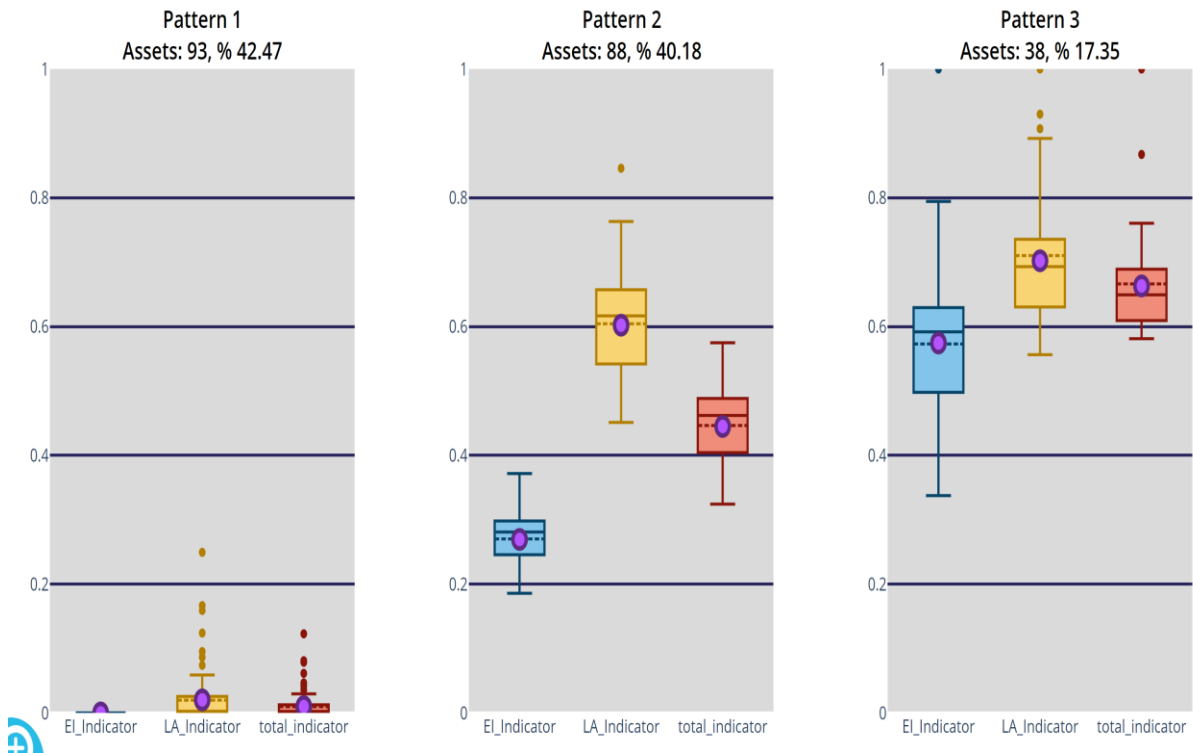
FIGURE 16 RESULTS: SPANISH GRID TRANSMISSION LINES

Segment_ID	El_Indicator	LA_Indicator	Total_Indicator
CDO7031339	0.904	0.724	0.814
CDO7031336	0.718	0.695	0.706
NBZ7080868	0.651	0.587	0.619
CDO7031361	0.587	0.621	0.604
CDO7031327	0.463	0.689	0.576
CDO7031335	0.595	0.555	0.575
NBZ7080889	0.569	0.561	0.565
NBZ7080869	0.569	0.559	0.564

6.3. HOP – ASSETS

6.3.1.Transmission Line

o   Number of assets: 10 transmission lines.
o   Results:
  ▪   4 patterns identified.
  ▪   Identifier SUBST42 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
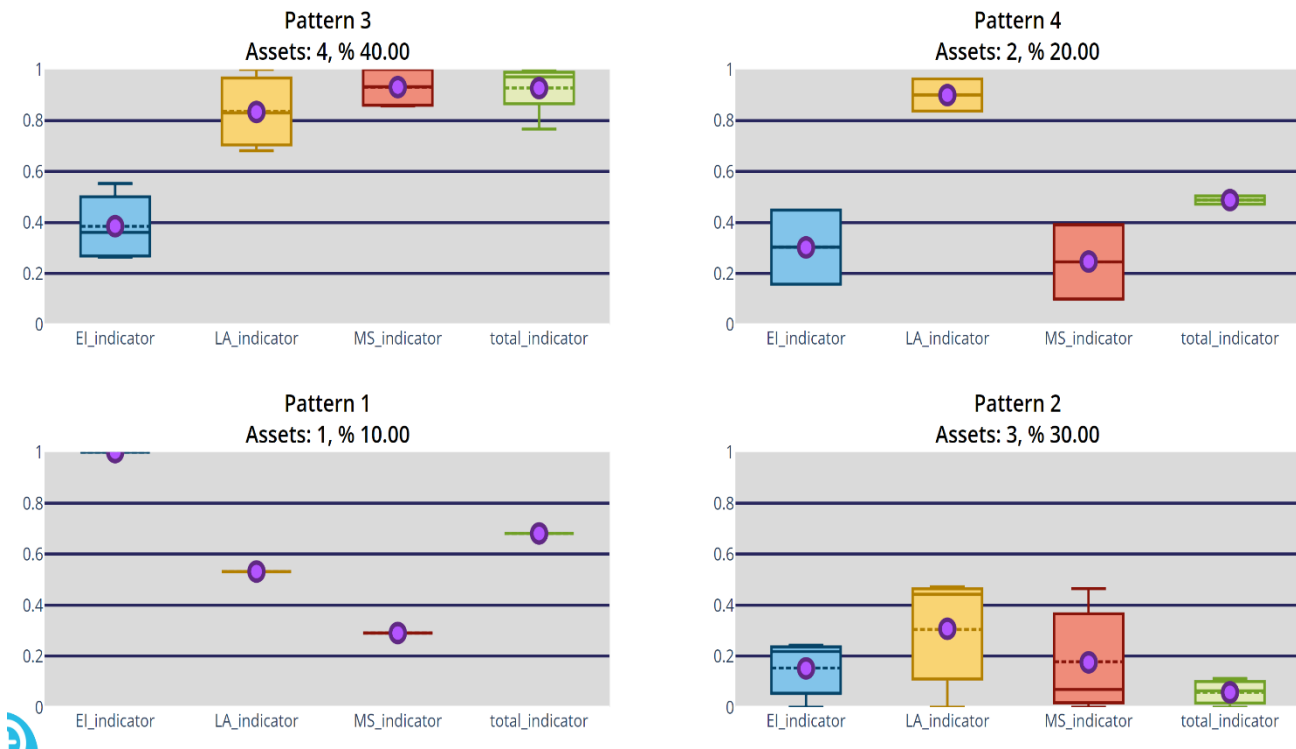

FIGURE 17 RESULTS: HOPS TRANSMISSION LINES

name	LA_indicator	MS_indicator	EI_indicator	total_indicator
SUBST42-SUBST05	0.963	1.0	0.222	0.728
SUBST26-SUBST55	0.698	1.0	0.448	0.716
SUBST05-SUBST01	0.907	0.856	0.364	0.709
SUBST01-SUBST55	0.736	0.863	0.214	0.604
SUBST36-SUBST01	0.574	0.291	0.811	0.559
SUBST34-SUBST29	0.932	0.1	0.364	0.465
SUBST29-SUBST01	0.827	0.391	0.128	0.448
SUBST19-SUBST01	0.132	0.465	0.177	0.258

6.3.2. Transformers

- o Number of assets: 15 transformers
- o Results:
  - 4 patterns identified.
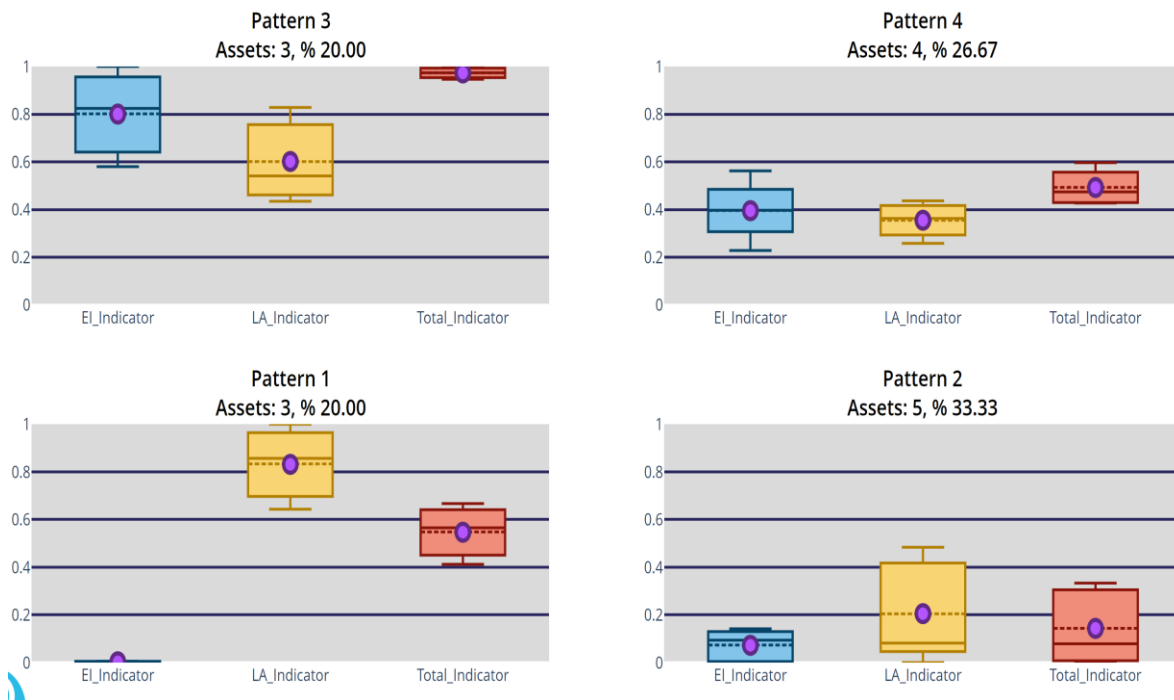 - All the assets are in good condition and don't require immediate attention.

FIGURE 18 RESULTS: HOPS TRANSFORMERS

Name	EI_Indicator	LA_Indicator	Total_Indicator
SUBST01	0.5	0.224	0.241
SUBST29	0.29	0.416	0.235
SUBST42	0.412	0.276	0.229
SUBST29	0.0	0.5	0.167
SUBST35	0.281	0.172	0.151
SUBST35	0.002	0.43	0.144
SUBST05	0.193	0.206	0.133
SUBST26	0.204	0.138	0.114
SUBST19	0.114	0.225	0.113

6.3.3.CIRCUIT BREAKERS

- o Number of assets: 39 circuit breakers
- o Results:
  - 3 patterns identified.
  - Identifier SUBST34 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
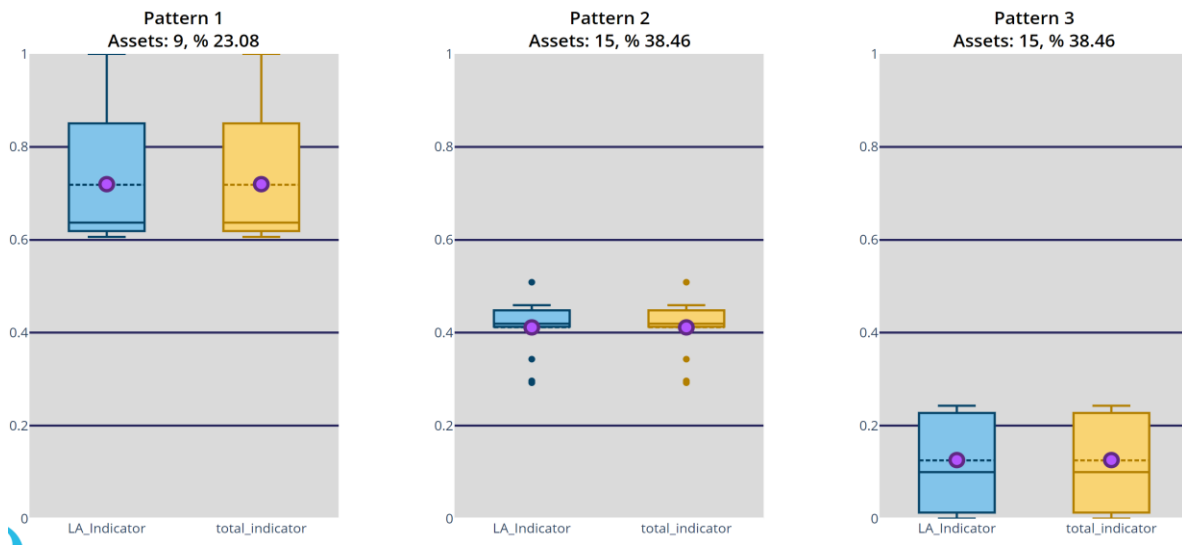
FIGURE 19 RESULTS: HOPS CIRCUIT BREAKERS

Substation_ID	LA_Indicator	Total_Indicator
SUBST34	0.829	0.829
SUBST36	0.717	0.717
SUBST36	0.717	0.717
SUBST19	0.568	0.568
SUBST05	0.557	0.557
SUBST34	0.557	0.557
SUBST05	0.545	0.545
SUBST35	0.539	0.539

6.4. SYNTHETIC GRID - ASSETS

6.4.1.TRANSMISSION LINES

- o Number of assets: 8414 lines.
- o Results:
  - ▪ 4 patterns identified.
  - ▪ Identifier TBT406 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.


FIGURE 20 RESULTS: SYNTHETIC GRID TRANSMISSION LINES

Name	LA_Indicator	Total_Indicator
TBT406	0.842	0.842
TBT70	0.836	0.836
TBT241	0.804	0.804
TBT548	0.804	0.804
TBT558	0.797	0.797
TBT414	0.792	0.792
TBT631	0.78	0.78
TBT4095	0.758	0.758

6.4.2. TRANSFORMERS

- o Number of assets: 391 transformers.
- o Results:
    - 4 patterns identified.
    - Identifier Trafo_61 has highest value of total_indicator which makes it fall into critical condition and requires immediate attention.
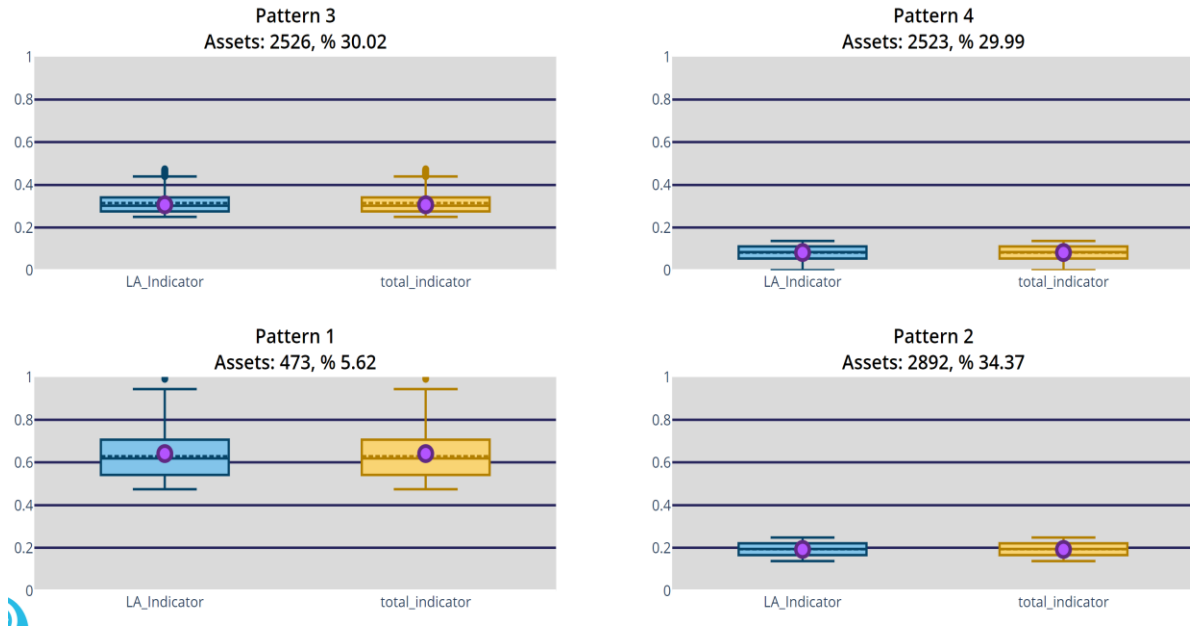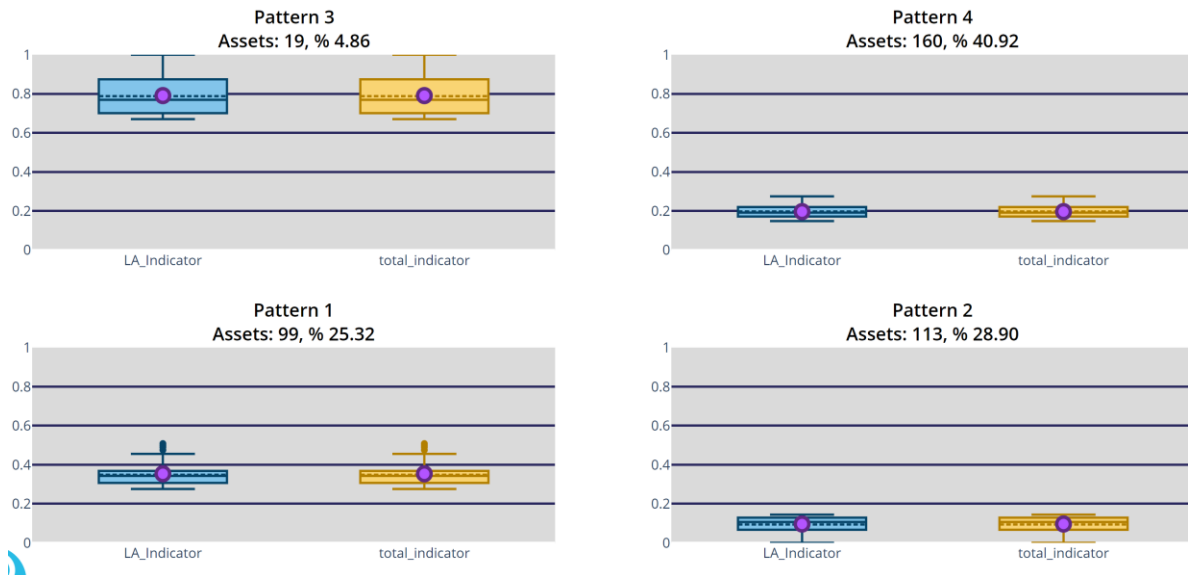

FIGURE 21 RESULTS: SYNTHETIC GRID TRANSFORMERS

Name	LA_Indicator	Total_Indicator
Trafo_61	0.86	0.86
Trafo_262	0.806	0.806
Trafo_245	0.802	0.802
Trafo_164	0.782	0.782
Trafo_188	0.757	0.757
Trafo_247	0.739	0.739
Trafo_24	0.698	0.698
Trafo_237	0.682	0.682
Trafo_311	0.663	0.663
Trafo_71	0.663	0.663

7. Conclusions.

This report has documented the main features of the tool T5.2 developed in the ATTEST project. This tool defines an innovative approach to translate the results obtained from the previous application (tool reported in deliverable D5.2) into a set of harmonized, easily measurable, and comparable indicators for different types of assets, which will allow identification of elements that require special attention from an asset management point of view.

The report includes examples of applications in real cases supplied by the partners contributing in the development of this tool and in WP5. This report is part of the effort developed in this task, and the rest is the code developed for the implementation of the tool.