Aden Abreham
Heidar Zirak

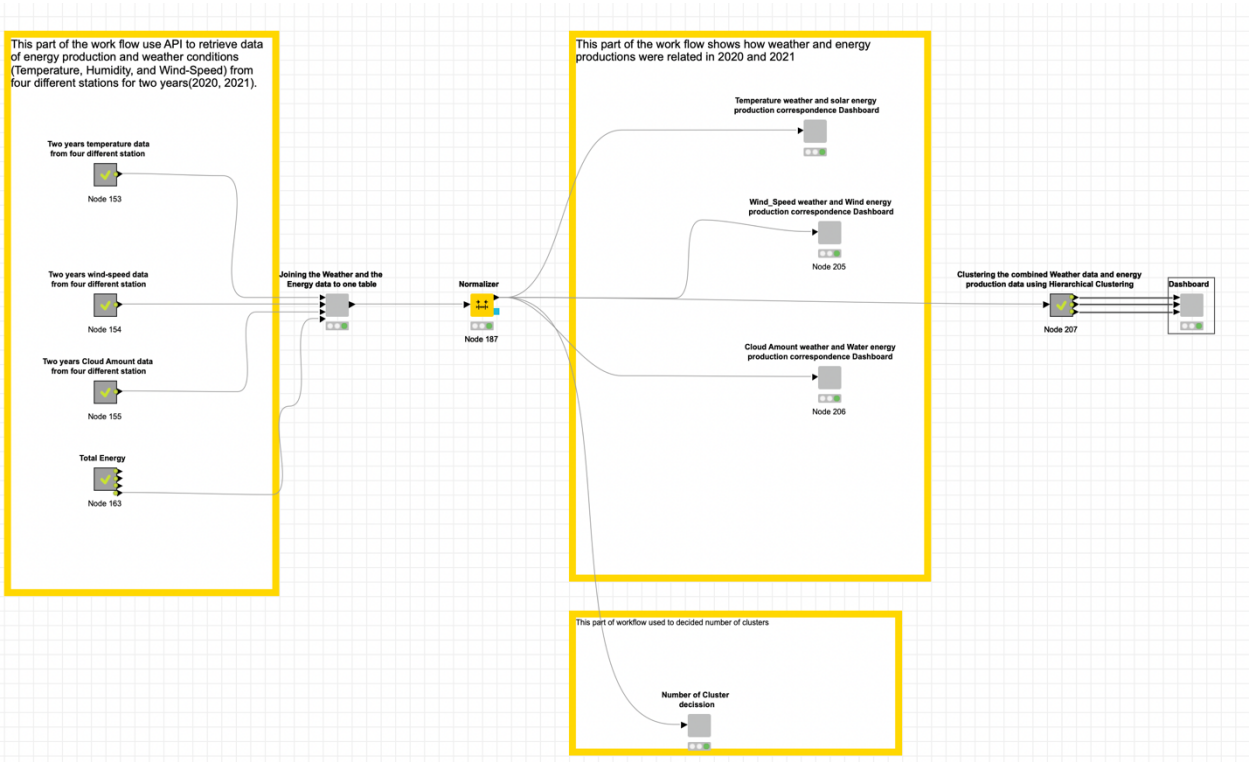
Introduction

The goal of this project is the depicting the relationship between weather and energy production in Sweden. In this case, the KNIME Analytics Platform has been used which is open-source software for data analytics, reporting, and integration platform. KNIME integrates various components for machine learning and data mining.

Meteorological Observations data was fetched from four different stations around Sweden using SMHI API. Targeted data were collected monthly in both years, 2020 and 2021. Meanwhile, Cloud Amount, Temperature, and wind speed are three different variables gathered from the mentioned stations in that period. Similarly, for energy production, the necessary data was retrieved from the SCB API monthly for two years. Three different energy data used for the analysis: Solar, wind, and water power.

To clarify the relationship between weather and energy production there are several data joining and aggregation, and all the statistics are shown by different plots and charts. To identify groups of similar objects in datasets we used the hierarchical clustering algorithm.

Methodology



We used the Metanodes to clean up messy workflow, and components to encapsulate flow variables and get an interactive outputs view.
In energy production Metanode, we collected data from SCB API using POST request node & performed a high level of data manipulation like filtering unwanted data, translation, date formatting, and joining them.

First, the data from four different stations around Sweden is provided in an API by SMHI. Since we needed a two-year period for this measurement the only available data type was CSV, therefore we used CSV Write and reader to have this information. The weather data was processed as in the previous section, and some statistics were added to find the proper station, as most were out of data at some point.

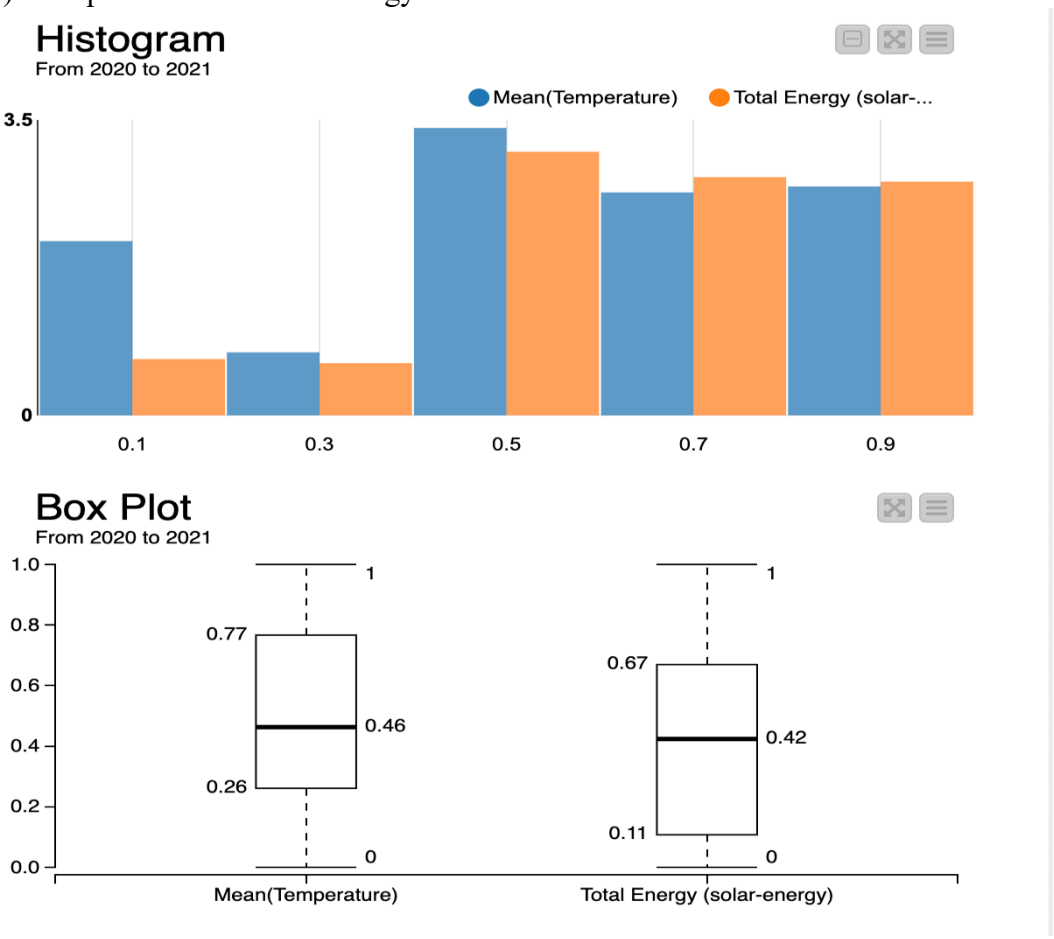
Then, normalizing the output data from data preparation, we send them into the Classification component which inside it, we have produced different plots to compare in an interactive dashboard. We depict the below items in relation:

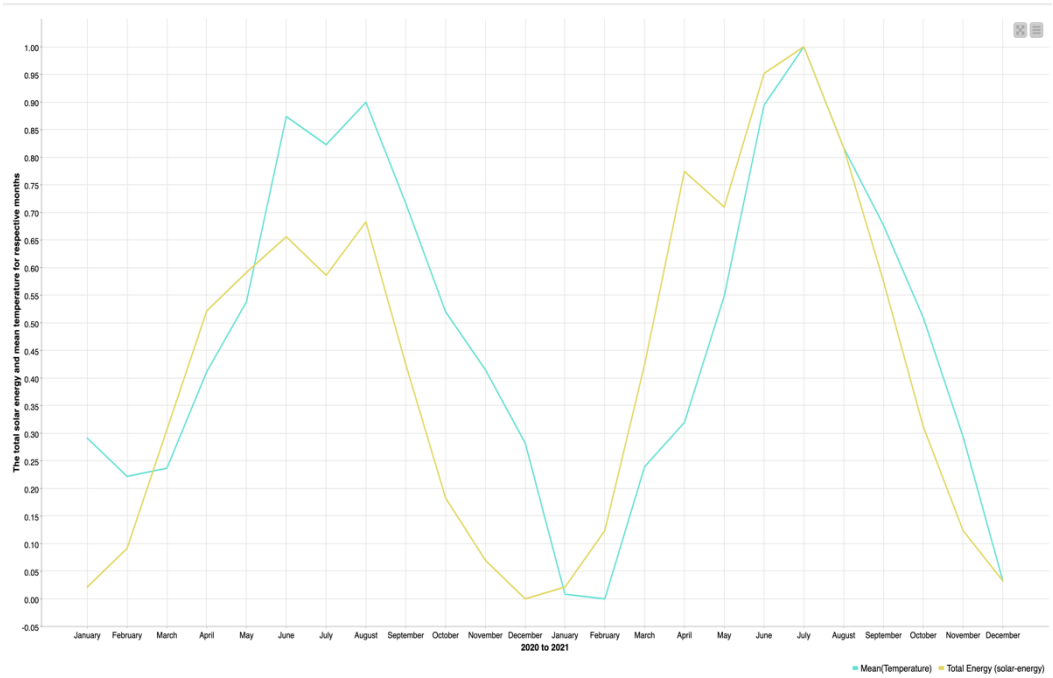
- Temperature weather and solar energy production.
- Wind-speed weather and wind energy production.
- Could amount weather and water energy production.

Finally, we have used the hierarchical clustering method for grouping objects. In the clustering configuration, the number of output clusters is set to 3. We attained this number from Dendrogram by analyzing the distance between clusters, the Euclidean method was selected, and finally the linkage type "average" has been chosen.

Result

A) Temperature and Solar energy





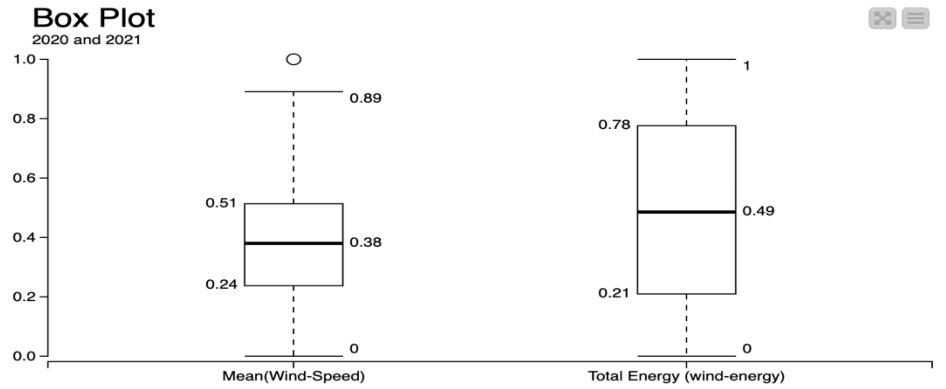
The Line graph presents the correspondence between solar energy and air temperature for each month in 2020 and 2021.

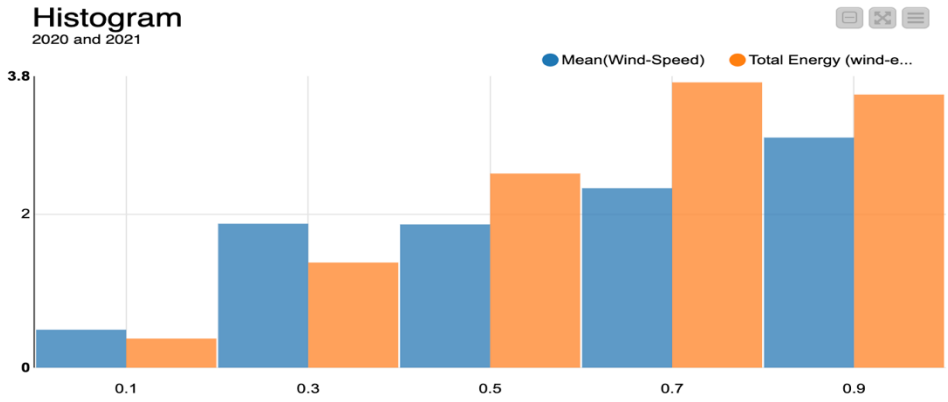
As the graph depicts, the winter season shows a direct relationship between air temperature and solar energy production, which was much lower compared to other seasons, although January to February showed an increase in solar energy production when the air temperature decreased in both years. Which could indicate the maintenance time for weather stations or indicates the time when the sensors were broken.

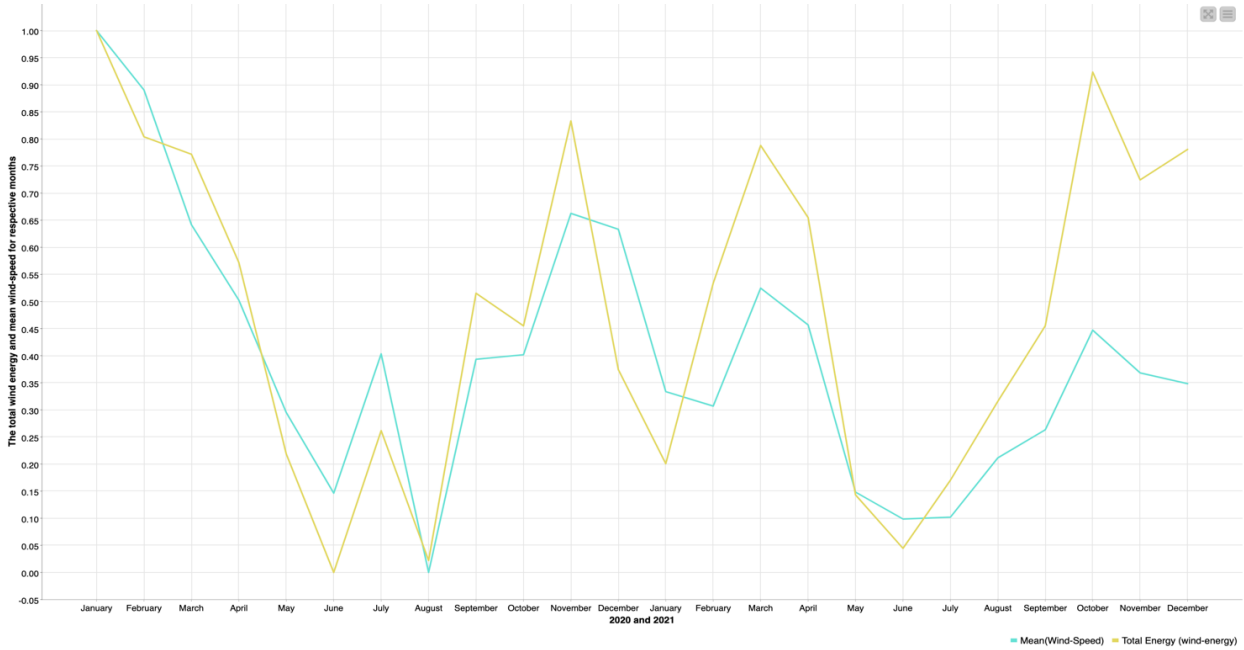
In the spring and summer periods, solar energy production increased as air temperature increased, even though in July 2020 air temperature and solar energy production decreased slightly. However, the interrelation between solar energy production and air temperature was strictly direct. In addition, September, and October show that the air temperature is gradually decreasing, and solar energy is strictly reduced.

Therefore, solar energy production and air temperature have direct connections except for January and February.

B) Wind-Speed and Wind Energy





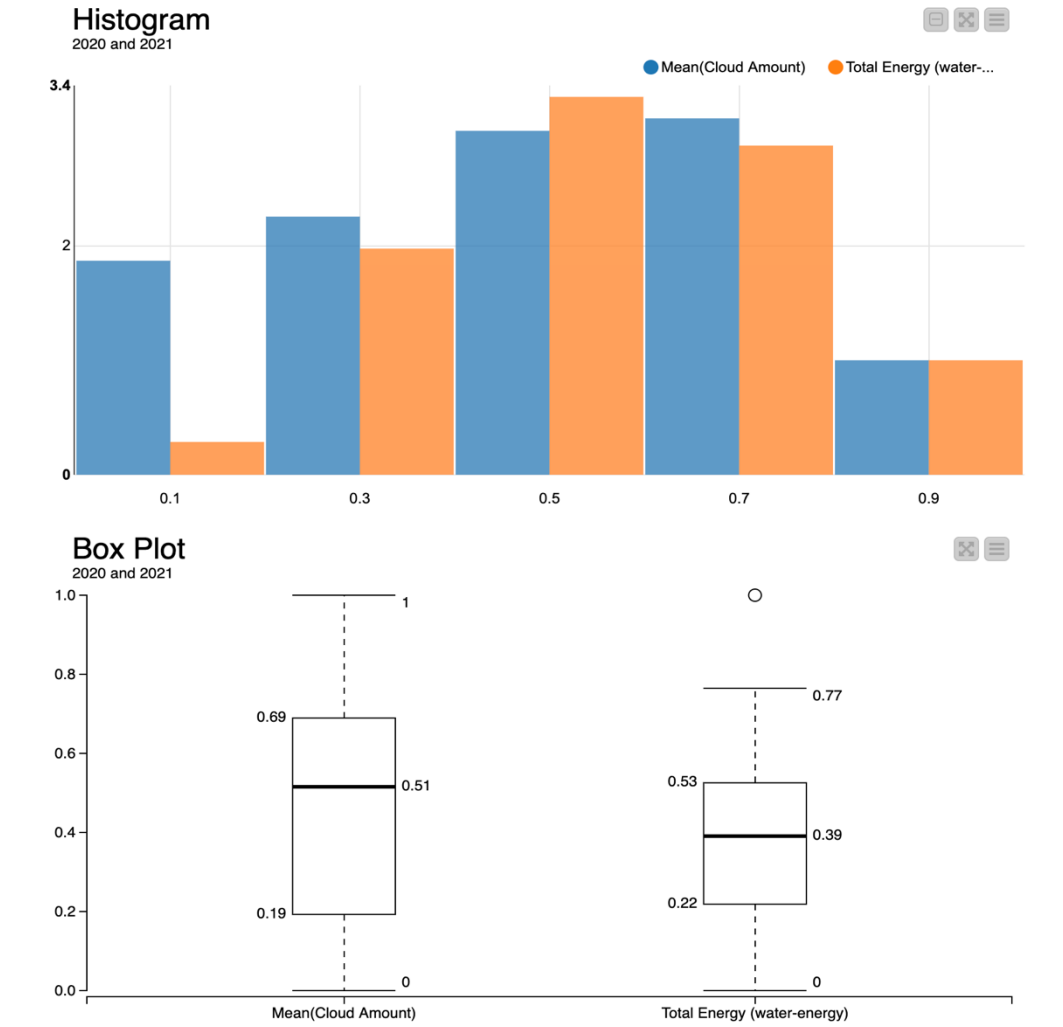


The line plot shows the relationship between wind-speed and wind-energy in 2020 and 2021.

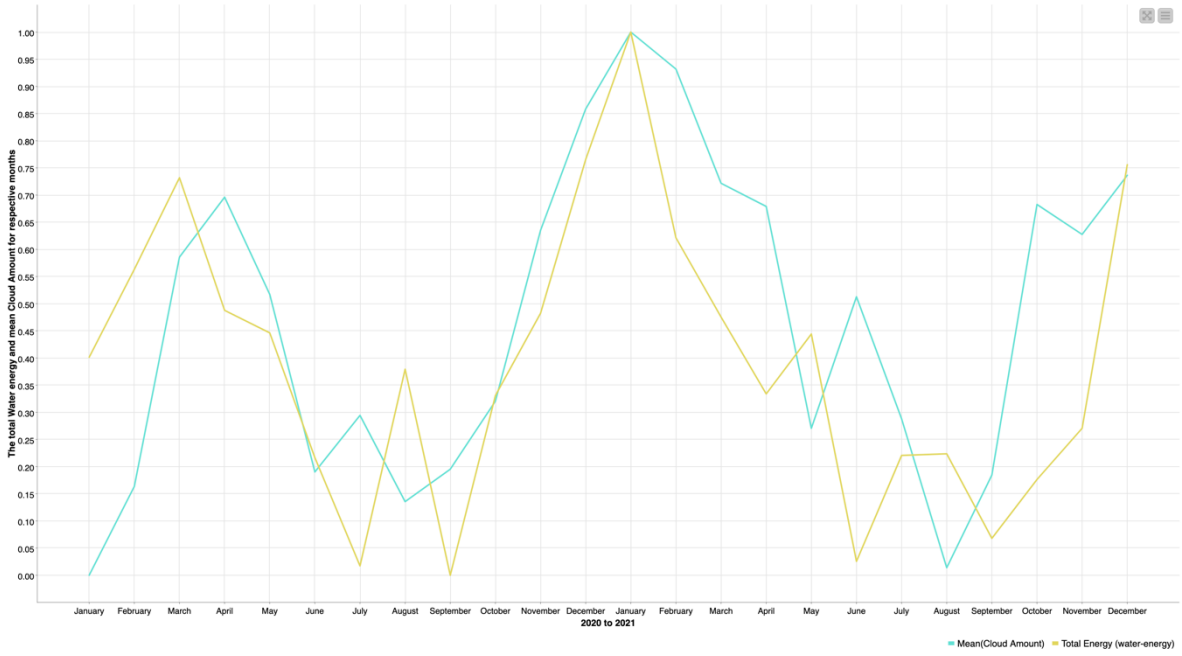
The graph shows that in both years, the production power was higher in winter harvest. The relationship between energy and weather was direct. However, from September 2020 to October 2020, wind power production decreased, but the wind speed remained the same, indicating a sensor error at the weather stations. Until January, energy production was fairly straight. In contrast, from January 2021 to February 2021, energy production was increasing sharply but weather was strongly decreasing. This shows that there was a sensor error in measuring the wind speed or the maintenance time of the stations during this period.

In the spring, both weather and wind energy production were strongly reduced for 2020 and 2021. For the summer of 2020, the relationship was strictly direct between energy production and weather. Whereas in 2021, energy production was increasing significantly, but the June and July weather measurements were similar, which could be due to errors in the wind sensors or maintenance time for the stations.

In summary, the relationship between wind speed and wind power was strictly linear except for the months of September 2020 to October 2020, January 2021 to February 2021, and June 2021 to July 2021.



C) Cloud Amount and water energy



The relationship between cloud and hydropower showed in the line plot is for 2020 and 2021.

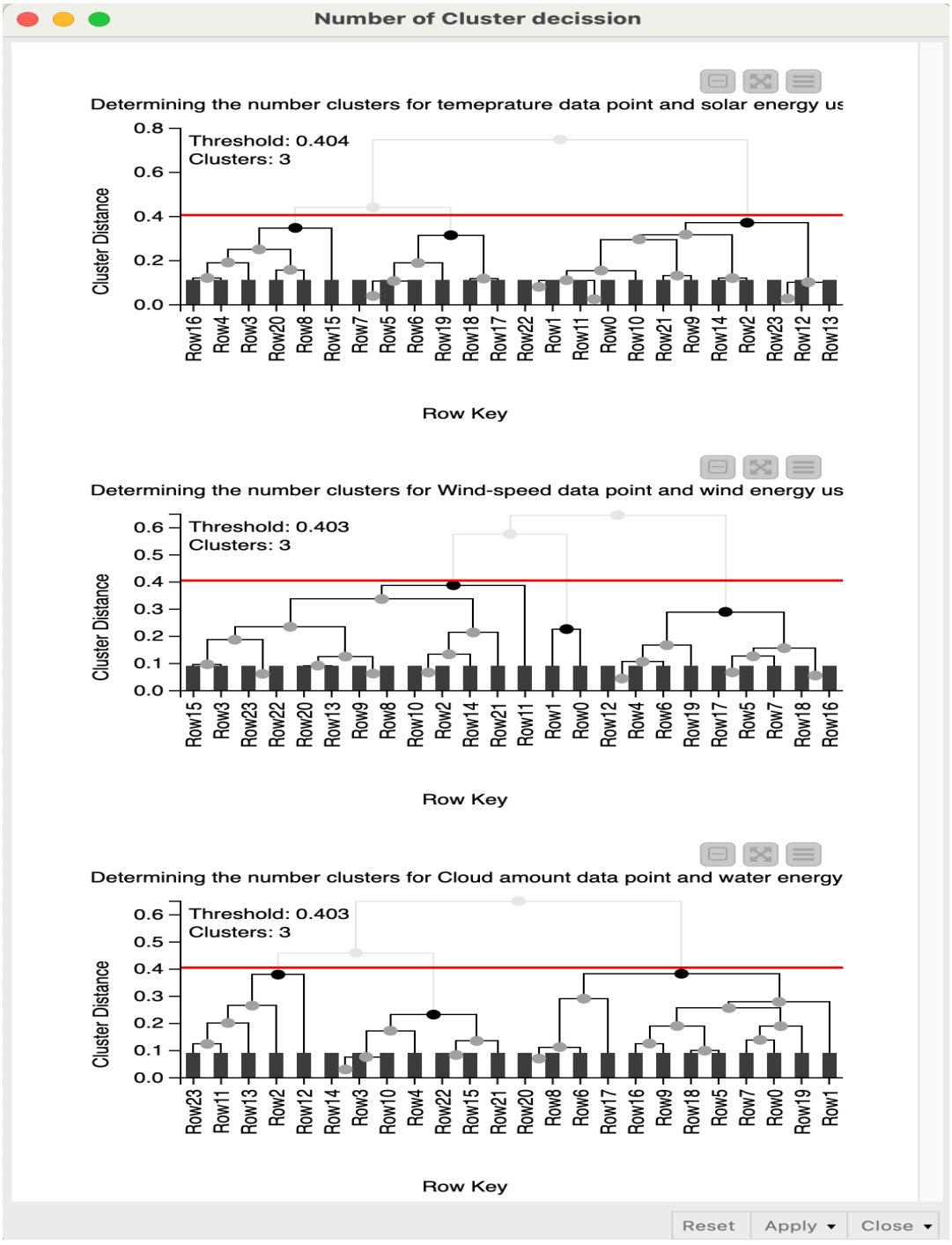
As the graph presents, the relationship between water-energy and cloud-amount was strictly direct in the winter season. In addition, their correspondence was strictly direct in the autumn period.

Besides, the relationship between energy and weather was an inverse relation for the summer period for both years. And, the inverse relation between cloud-amount and water-energy was the result of a faulty sensor or maintenance period.

Also, for spring 2020, energy and weather, except for March through April, energy was decreasing linearly, but weather was increasing significantly. This could be the result of a broken sensor. At the same time next year, their relationship is the same as last year, but both are decreasing. However, in the period from April to May, the power was high and the weather decreased significantly. The reason for this may be during maintenance or due to broken sensors

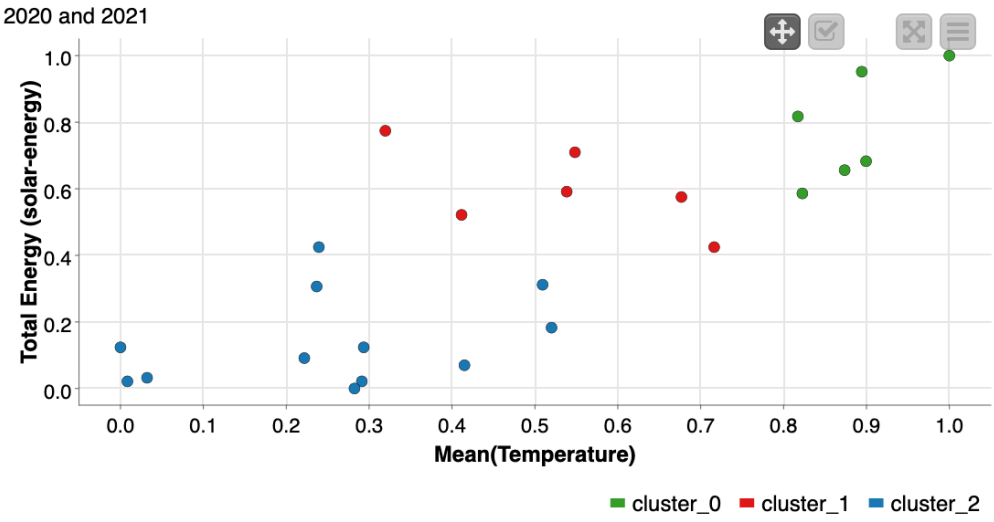
In conclusion, the correspondence between weather and energy production was positive in years except for spring and summer.

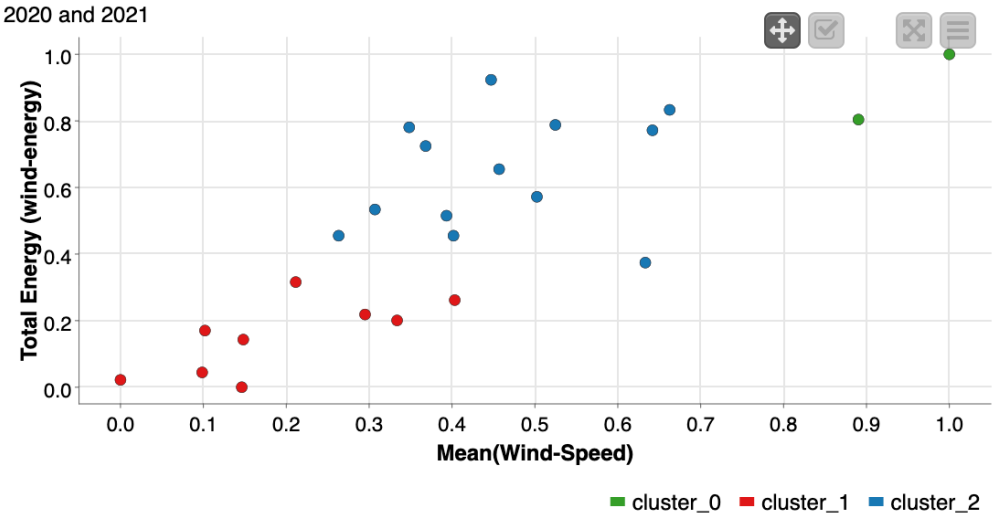
D) Clusters

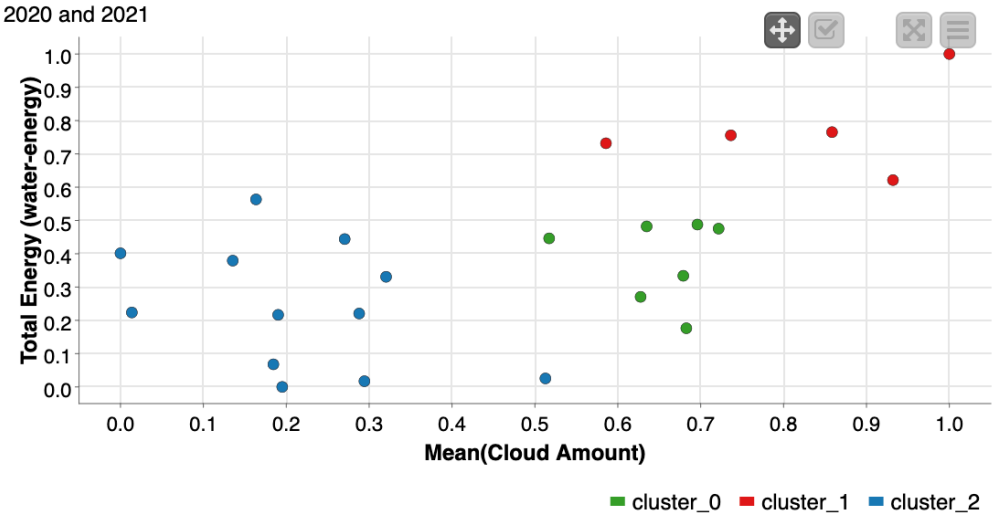


The data points are clustered into three clusters and the distance between the clusters was considered for deciding the number of clusters.

Dashboard







Reset | Apply | Close