MASTER'S DEGREE IN BIG DATA TECHNOLOGIES AND ADVANCED ANALYTICS

FINAL MASTER´S PROJECT

ANALYSIS OF THE SPATIAL AND TEMPORAL CORRELATIONS BETWEEN SOLAR, WIND AND HYDRO POWER GENERATION IN SPAIN

Autor: Jeonghyeon Lee

Director: Eugenio F. Sánchez Úbeda

Madrid

Junio 2023

MASTER'S DEGREE IN BIG DATA TECHNOLOGIES AND ADVANCED ANALYTICS

FINAL MASTER´S PROJECT

ANALYSIS OF THE SPATIAL AND TEMPORAL CORRELATIONS BETWEEN SOLAR, WIND AND HYDRO POWER GENERATION IN SPAIN

Autor: Jeonghyeon Lee

Director: Eugenio F. Sánchez Úbeda

Madrid

Junio 2023

Declaro, bajo mi responsabilidad, que el Proyecto presentado con el título

ANALYSIS OF THE SPATIAL AND TEMPORAL CORRELATIONS BETWEEN SOLAR, WIND AND HYDRO POWER GENERATION IN SPAIN

en la ETS de Ingeniería - ICAI de la Universidad Pontificia Comillas en el

curso académico 2022/23 es de mi autoría, original e inédito y

no ha sido presentado con anterioridad a otros efectos.

El Proyecto no es plagio de otro, ni total ni parcialmente y la información que ha sido

tomada de otros documentos está debidamente referenciada.

Fdo.: Jeonghyeon Lee Fecha: 28/ 06/2023

Autorizada la entrega del proyecto

EL DIRECTOR DEL PROYECTO

Fdo.: Eugenio F. Sánchez Úbeda Fecha: 28/ 06/2023

Acknowledgements

I would like to take this opportunity to express my heartfelt appreciation to the individuals who have played a crucial role in the completion of this article.

First and foremost, I am deeply grateful to my supervisor, Eugenio F. Sánchez Ubeda. Your expertise, guidance, and unwavering support throughout the duration of my research have been invaluable. Your insightful feedback, constructive criticism, and dedication to my academic growth have significantly influenced the quality of this work. I am truly fortunate to have had the opportunity to work under your supervision, and I extend my sincere thanks for your patience, wisdom, and commitment.

I would also like to express my heartfelt thanks to Carlos Morrás, my master's director. I am genuinely grateful for your unwavering faith in me and for the opportunities you have provided me to grow and develop as a scholar.

Lastly, I want to express my deepest appreciation to my family. Your love, encouragement, and understanding have been a constant source of strength throughout my academic journey. Your unwavering belief in my abilities and the sacrifices you have made to support my aspirations have been instrumental in my accomplishments. I am forever grateful for your unwavering support and the countless ways in which you have shaped my life.

To my supervisor, Eugenio F. Sánchez Ubeda, master's director Carlos Morrás, and my family, I extend my heartfelt thanks for your invaluable contributions. Each of you has played a significant role in my academic and personal development, and I am truly fortunate to have had you by my side throughout this endeavor.

Thank you.

ANALYSIS OF THE SPATIAL AND TEMPORAL CORRELATIONS BETWEEN SOLAR, WIND AND HYDRO POWER GENERATION IN SPAIN

Author: Jeonghyeon, Lee.
Supervisor: Eugenio F. Sánchez Úbeda.
Collaborating Entity: ICAI – Universidad Pontificia Comillas

ABSTRACT

The purpose of this project is to analyze the public information available in ESIOS related to renewable power generation. We will study the series of wind, solar and hydroelectric production using machine learning techniques. The characteristics of each variable will be analyzed individually and jointly, taking in consideration of the temporal and spatial dimensions. To this end, we automate the collection of all the information needed for the study in an internal warehouse specifically designed to facilitate analysis.

Keywords: Renewable energy, Solar energy, Wind energy, Hydro energy

1. Introduction

This research project is motivated by the global imperative to transition towards renewable energy and the limited progress of renewable energy projects in Korea. Inspired by Spain's success in renewable energy, the study aims to understand and learn from their advancements. By analyzing the dynamics of solar and wind energy in Spain, the research seeks to gain valuable insights applicable to renewable energy development in Korea.

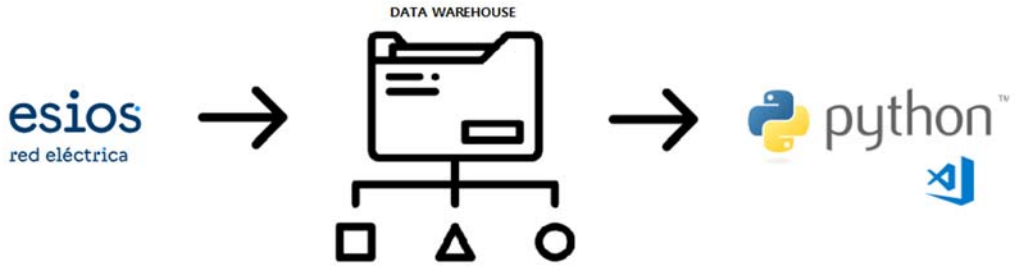

Illustration 1 – Main components of the system

2. Data Ingestion

ESIOS serves as a valuable resource for accessing reliable and up-to-date information on solar, wind, and hydro power generation in Spain for this study (Illustration 1). Through web scraping, we efficiently collect large volumes of data from ESIOS, enabling us to explore spatial and temporal correlations among renewable energy sources. This automated approach saves time and effort compared to manual data collection, allowing for comprehensive analysis of historical data and the identification of significant trends in renewable energy generation across Spain.

3. Exploratory Data Analysis and Data Normalization

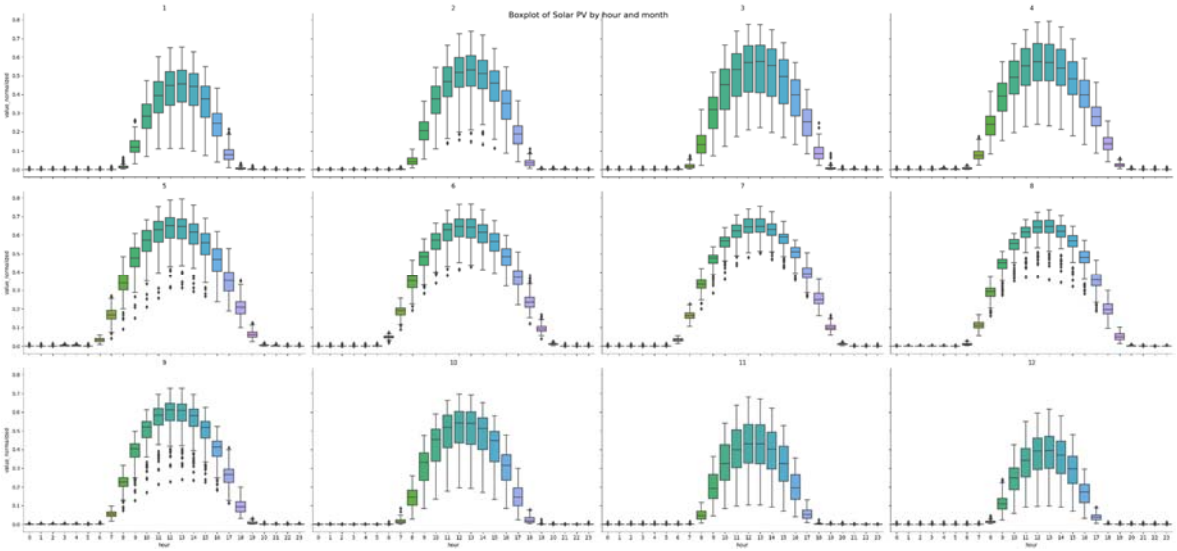

Illustration 2 – Box-plot by month and hour of Solar PV generation

Solar PV energy box-plots (Illustration 2) provide an ideal and visually intuitive representation for interpreting energy production, showing higher production during the summer season and peaking during central daylight hours.

Wind energy mainly shows a consistent hourly profile, with some temporal fluctuations, especially in summer.

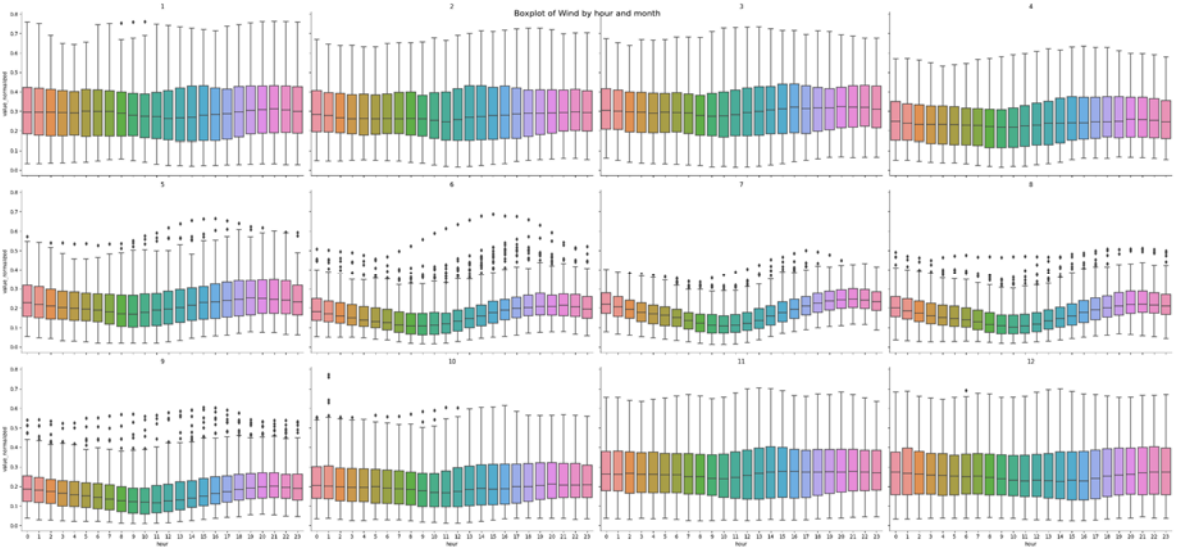

Illustration 3 – Box-plot by month and hour of Wind generation

4. Photovoltaic and Solar Thermal generation

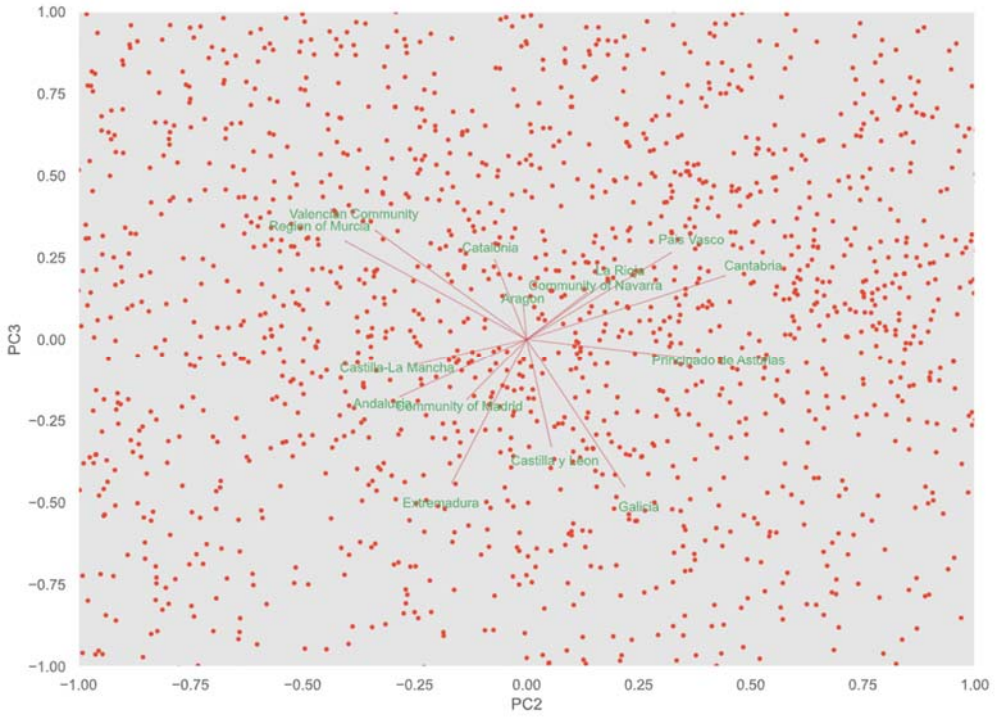

Illustration 4 – Solar PV Bi-plot with PC2 and PC3

The Principal Component Analysis (PCA) exposes useful information about the dynamics. For example, the second and third principal components reveals north-south and east-west differences in solar radiation across provinces in Spain. The bi-plot of these components provides a spatial representation of solar power generation, showcasing a rotated Spanish map that highlights regional variations and patterns (Illustration 4). This visualization aids in understanding the geographical distribution of solar energy generation and identifying clusters of provinces with similar solar profiles.

5. Wind energy generation

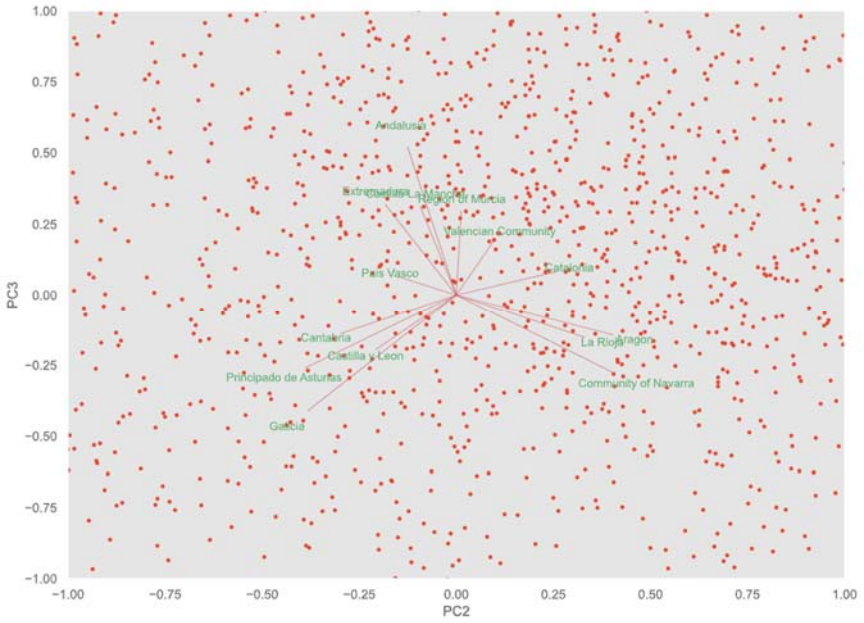

Illustration 5 – Wind Bi-plot with PC2 and PC3

PCA of wind energy generation reveals valuable insights into regional dynamics and variations. The first principal component (PC1) highlights the significant contributions of Valencia and Castilla-La Mancha to the overall trend of wind power generation, while Andalusia and Extremadura have relatively lesser influence. The second principal component (PC2) indicates regional variations within the northern provinces, with an emphasis on the eastern and western divisions. PC3 reflects the north-south disparities in wind energy generation, with Andalusia playing a prominent role in the south. PC4 underscores the significance of coastal regions in wind energy production, while PC5 highlights the relevance of Extremadura and Cantabria. The bi-plot of PC2 and PC3 (Illustration 5) reveals a symmetrical pattern across the Spanish map, suggesting shared wind energy characteristics among provinces such as Galicia, Aragon, and Andalusia.

6. Photovoltaic with Wind energy generation

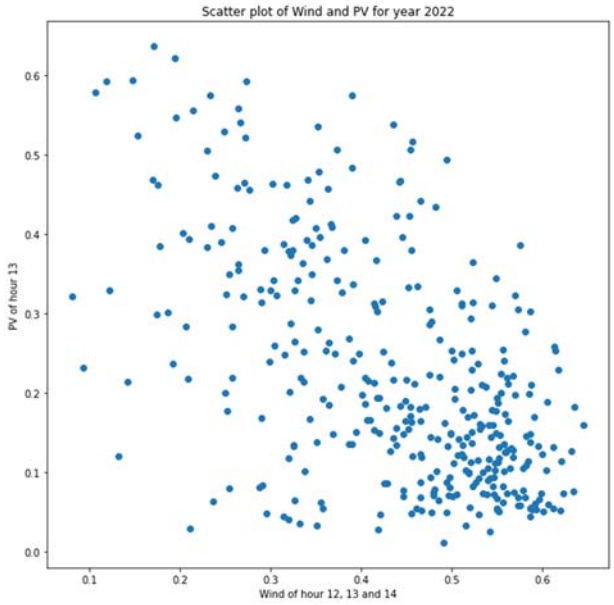

Illustration 6 – Scatter plot of Wind and PV for year 2022

Illustration 6 presents a scatter plot showing the negative correlation between solar radiation at 13:00 and the mean of wind energy at 12, 13, and 14. The correlation coefficient of -0.44 indicates a moderate negative relationship between the two variables, suggesting that as solar radiation increases, wind energy tends to decrease, and vice versa. This finding suggests the presence of cloud cover as a potential factor influencing this relationship. Additionally, the dendrogram analysis confirms the independent properties of solar and wind energy generation, supporting earlier observations based on correlation coefficients.

7. Conclusions and Future Work

In this study, the patterns and dynamics of solar and wind power generation were comprehensively analysed, finally excluding the hydro energy analysis due to the strong management to which this generation technology is subjected thanks to the existence of a high number of dams in Spain. The study identified clear patterns of solar intensity, time variations, and regional differences in solar power generation. Similarly, significant regional and temporal changes were observed in wind power generation, providing insight into the temporal progression and regional characteristics of wind power generation. This study highlights the need for future research exploring the complexity and synergies of hydroelectric energy with solar and wind energy to enhance understanding of renewable energy landscapes and promote the development of sustainable energy strategies.

8. References

[1] ESIOS (System Operator's Information System). (n.d.).Glossary. ESIOS (System Operator's Information System). https://www.esios.ree.es/en/glossary

[2] G. James, D. Witten, T. Hastie & R. Tibshirani (2013). An Introduction to Statistical Learning with Applications in R. Springer (see http://www-bcf.usc.edu/~gareth/ISL/). 2nd ed. (2021)

[3] T. Hastie, R. Tibshirani & J. Friedman (2009). The Elements of Statistical Learning. Data Mining, Inference and Prediction. 2nd Ed. Springer.

Contents

List of Figures

List of Tables

Chapter 1. INTRODUCTION

1.1 MOTIVATION OF THE PROJECT

The motivation behind this project stems from the increasing global interest in renewable energy sources as a sustainable solution to mitigate climate change and meet the growing energy demands of modern society. Amidst this global movement, the limited progress of renewable energy projects in Korea has sparked a personal interest in exploring successful renewable energy models from other countries. In particular, Spain has emerged as a frontrunner in the field of renewable energy with significant investments and remarkable performance.

From a personal standpoint, the observation of Spain's advancements in renewable energy has fueled a strong curiosity to delve deeper into the country's renewable energy landscape. By studying and analyzing the dynamics of solar and wind energy in Spain, this research aims to gain valuable insights that can be applied in the context of renewable energy development in Korea. By identifying successful practices, challenges, and potential solutions, this study aspires to contribute to the wider growth and progress of the renewable energy industry.

In summary, this research project is driven by the global imperative to transition towards renewable energy and the personal aspiration to learn from successful renewable energy models like Spain. By exploring the dynamics of renewable energy in Spain and their potential applicability in Korea, this study aims to contribute to the wider renewable energy discourse and support the development of sustainable energy solutions.

1.2 OBJECTIVE OF THE PROJECT

The objective of this study is to comprehensively examine the patterns and interrelationships among solar, wind, and hydro energy generation in Spain. Through a meticulous analysis of diverse datasets encompassing solar photovoltaic (PV), solar thermal, wind, and hydro

power, our primary aim is to discern the intricate temporal and spatial dynamics of renewable energy generation. Our research specifically focuses on elucidating the fluctuations in energy production across various timeframes, including daily and seasonal patterns, while also delineating the geographical distribution of renewable energy sources across distinct regions within Spain.

Also, this study necessitates the automation of data collection and integration processes, enabling the seamless ingestion of all relevant study information into a dedicated internal data warehouse. This purpose-built repository has been specifically designed to streamline and facilitate the analysis of the collected data, ensuring efficient access, organization, and retrieval of information for subsequent analysis. By employing automated data management techniques, we aim to enhance the effectiveness and efficiency of our research, enabling comprehensive and insightful analysis of the solar, wind, and hydro energy generation data in Spain.

1.3 RESOURCES OF THE PROJECT

The project was conducted using Python programming language in Jupyter Notebook, utilizing the Visual Studio Code (VS Code) development environment. Python, a widely-used language in data analysis and scientific computing, provided a robust foundation for the implementation of various analytical techniques and statistical models. The Jupyter Notebook environment facilitated an interactive and collaborative workflow, allowing for seamless integration of code, visualizations, and explanatory text. Additionally, the Visual Studio Code platform offered a comprehensive set of features, including code editing, debugging, and version control, enhancing the efficiency and effectiveness of the project. The combination of Python, Jupyter Notebook, and Visual Studio Code provided a powerful and user-friendly toolkit for data exploration, analysis, and visualization in this research endeavor.

Chapter 2. DATA INGESTION

2.1 DATA SOURCE

ESIOS (System Operator's Information System), developed by Red Eléctrica (www.ree.es), serves as a comprehensive platform designed to manage and provide information related to the electricity market in Spain. It plays a vital role in handling tasks such as information management and process coordination specifically related to the electricity market operations. ESIOS gathers, manages, and provides real-time data on energy production, consumption, and market prices, making it a crucial tool for ensuring the efficient operation and optimization of the energy market. This platform enables stakeholders to monitor and analyze key indicators and trends, facilitating informed decision-making in the energy sector.

For this study, ESIOS serves as an invaluable resource for accessing reliable and up-to-date information on solar, wind, and hydro power generation in Spain. By utilizing the data provided by ESIOS, we can investigate the spatial and temporal correlations between these renewable energy sources, gaining a better understanding of their interconnectedness and potential for combined utilization.

The following data sets were utilized:

Table 1. Data sets of ESIOS

	SOLAR PV	*SOLAR THERMAL*	*WIND*	*HYDRO*
Installed capacity generation	O	O	O	O
Measured generation	O	O	O	O
Operating Hourly Program P48	O	O	O	O

According to ESIOS, data sets are defined as the following ways:

- Installed capacity is defined as the maximum active power that an installation can reach.
- Generation measures depending on the production type used.
- Energy corresponding to the operative program that each SO establishes in each period till the end of the daily scheduling horizon (ESIOS, n.d.)

2.2 DATA STORAGE

To gather the necessary data for our research, we employed a technique called web scraping. Web scraping involves automating the extraction of data from websites by parsing the HTML code underlying web pages. This methodology enables us to collect large volumes of data efficiently and systematically, saving considerable time and effort compared to manual data collection.

Web scraping simplifies the data collection process by providing efficient access to historical data from ESIOS and facilitates the collection of data on a larger scale, enabling us to explore comprehensive spatial and temporal trends in renewable energy generation across Spain. By automating the data collection process, we can gather a significant amount of historical data, which enhances the robustness and depth of our analysis. Also this approach allows us to focus on analyzing the data rather than spending excessive time on manual data gathering.

```
solar_pv_measured_hour_geo = data_funtions.data_hour_geo(1161,'01-01-2014T00%3A00','28-
02-2023T23%3A55','hour','peninsular',list_geo.geo_ids_sm)

solar_pv_measured_hour_geo.to_csv('D:/master/TrabajoFindeMaster/TFM/save_data/solar/sola
r_measured_hour_allgeo.csv', index=False)
```

The code snippet provided demonstrates how we used web scraping to download the data. We called a function, which we specifically designed for web scraping, to retrieve information on measured solar power generation at an hourly resolution across different geographical locations. The function accepts parameters such as the specific dataset identifier, the start and end dates for the data collection period, the time resolution, geographical scope, and a list of geographic identifiers.

A tailored approach was employed by utilizing different Python functions based on the type of data, enabling the effective handling and analysis of the data at various temporal resolutions, including minutes, hours, days, and months. By employing these distinct Python functions, specifically designed for each data type, accuracy, and efficiency in the research were ensured, allowing for the effective handling and analysis of the collected data. This

approach enhanced the ability to explore comprehensive spatial and temporal trends in renewable energy generation across Spain, providing valuable insights for the analysis.

Once the data was extracted using web scraping techniques, we saved it as a CSV file for further analysis. The collected data were then transformed into a structured format suitable for subsequent analysis.

During the data downloading process, the acquired dataset was categorized into specialized groups within a designated folder. Within the 'save_data' directory, distinct subfolders were established to systematically organize the data into 'solar', 'hydro', and 'wind' categories. By implementing this systematic classification approach, the data organization was enhanced, and the subsequent analysis process was streamlined. This meticulous organization not only ensures efficient data management but also provides a foundation for comprehensive investigation and insightful observations in our research.

Chapter 3. EXPLORATORY DATA ANALYSIS AND DATA NORMALIZATION

The dataset used in this study consists of comprehensive solar, wind, and hydroelectric data from various provinces. Data serves as a valuable resource for investigating and understanding the dynamics of energy production across different regions. It aims to leverage this dataset to contribute to existing knowledge bases in the field of renewable energy and to provide insight into the temporal and regional patterns of energy generation.

3.1 DATA OVERVIEW

In the context of energy analysis, the choice between measured data and P48 data depends on the specific research objectives and the level of accuracy required.

Measured data refers to actual observations of energy generation obtained from on-site sensors or monitoring systems. It provides precise information about the real-time performance of energy installations. Measured data is generally preferred when the analysis requires accurate and reliable data for studying the actual energy generation patterns, assessing system performance, or validating models. On the other hand, P48 data provides insights into the expected energy generation based on forecasted weather conditions. It is commonly used for short-term or long-term energy forecasting, energy management, and system planning.

We focus is on understanding the actual performance of energy installations or evaluating system efficiency, measured data are preferred. To evaluate the appropriateness of each dataset, we conducted a comparative analysis by plotting the solar data per hour, comparing the P48 data against the measured data.

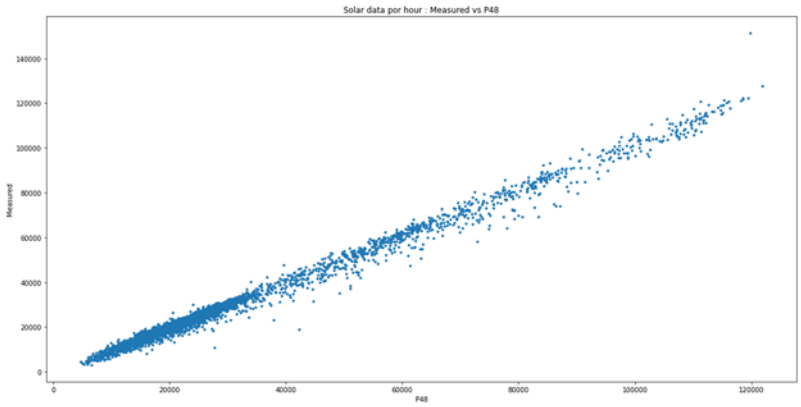

Figure 1. Scatter plot with Measured and P48 for Solar gemeration

As depicted in Figure 1, a scatter plot was generated, with the P48 data represented on the x-axis and the measured data represented on the y-axis. Notably, a clear strong linear relationship is observed between the two datasets. This finding suggests that the measured data aligns closely with the P48 data, indicating its reliability and accuracy.

Based on this analysis, we have determined that utilizing the measured data is more appropriate for our study, not requiring data cleansing due to the high correlation with P48. The measured data provides a more accurate representation of the actual solar energy generation, ensuring the reliability of our findings and enhancing the credibility of our research outcomes. Therefore, the data sets used in the analysis are as follows.

- INSTALLED CAPACITY GENERATION SOLAR PV (MW): Data is available since 2015-01-01, and the units available for download are year and month, which are divided by community.
- MEASURED GENERATION SOLAR PV (MWh): Data is available since 2014-01-01, and the units available for download are year, month, day and hour, which are divided by city.
- MEASURED GENERATION SOLAR THERMAL (MW): Data is available since 2015-01-01, and the units available for download are year and month, which are divided by community.

- MEASURED GENERATION SOLAR THERMAL (MWh): Data is available since 2014-01-01, and the units available for download are year, month, day and hour, which are divided by province.
- INSTALLED CAPACITY GENERATION WIND (MW): Data is available since 2015-01-01, and the units available for download are year and month, which are divided by community.
- MEASURED GENERATION WIND (MWh): Data is available since 2014-01-01, and the units available for download are year, month, day and hour, which are divided by province.
- INSTALLED CAPACITY CONVENTIONAL GENERATION HYDRO UGH (MW): Data has been available since 2015-01-01, and the units available for download are year, month, day and hour, which are divided by province.
- MEASURED HYDRO UGH GENERATION (MWh): Data has been available since 2014-01-01, and the units available for download are year, month, day and hour, which are divided by province.
- INSTALLED CAPACITY GENERATION HYDRO (MW): Data has been available since 2015-01-01, and the units available for download are year, month, day and hour, which are divided by province.
- MEASURED HYDRO NON UGH GENERATION (MWh): Data has been available since 2014-01-01, and the units available for download are year, month, day and hour, which are divided by province.

Prior to the detailed analysis, a quick review of the seasonal behavior of each measured energy can provide a useful view of the variables.

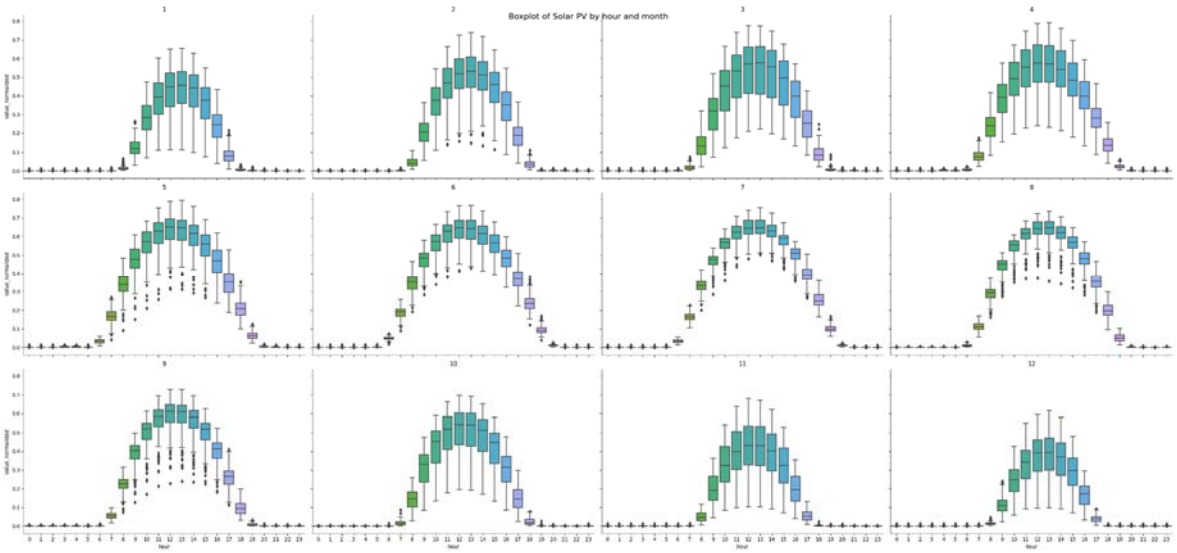

Figure 2. Box-plot of Solar PV

Figure 2 shows that box-plots are ideal for interpreting the seasonal daily patterns of Solar PV energy. During the year, energy production is high in summer, and among them, energy production is the highest during the central hours of the day.

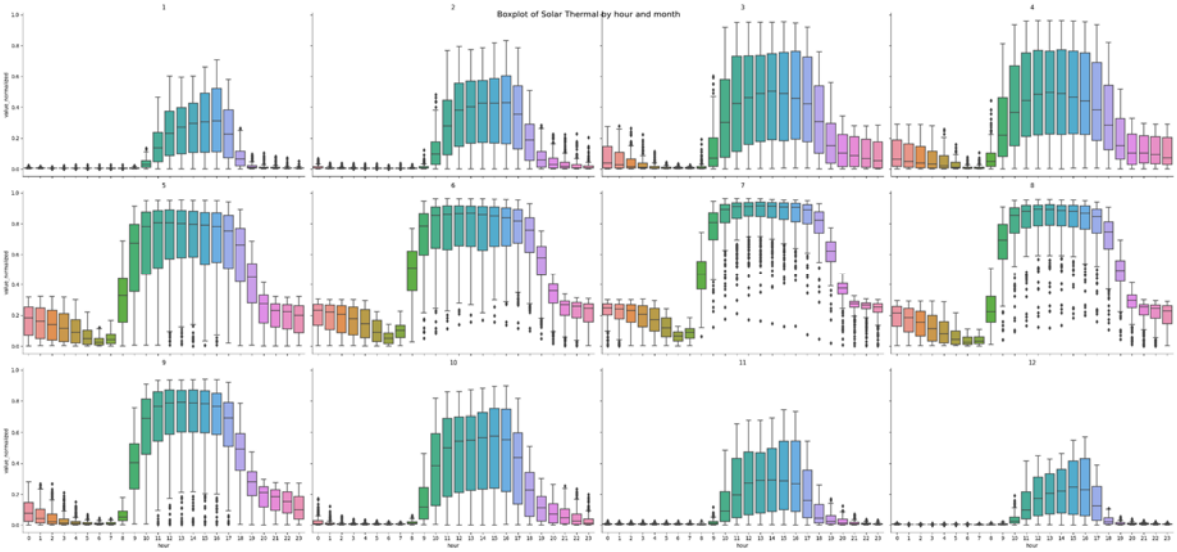

Figure 3. Box-plot of Solar Thermal

Solar Thermal energy exhibits a similar pattern of energy production (see Figure 3). However, it is noteworthy that there are a higher number of outliers compared to solar PV energy sources. Despite this variability, the data clearly demonstrate the characteristic of

decreased energy production during the winter season. This decline during winter can be attributed to factors such as reduced sunlight exposure and lower ambient temperatures, which adversely affect the efficiency of solar thermal systems.

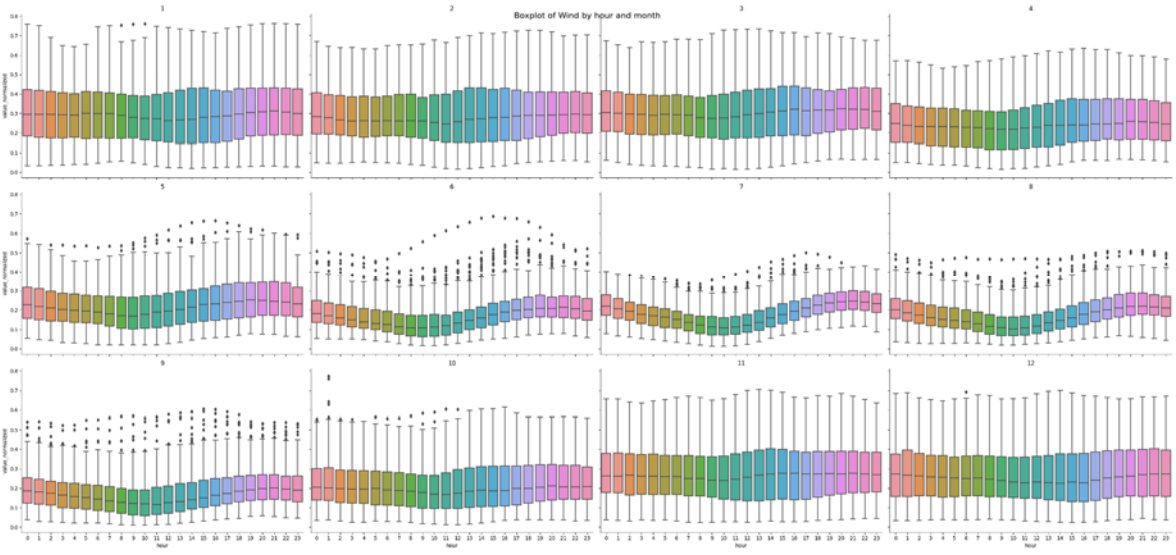

Figure 4. Box-plot of Wind

Concerning wind energy, Figure 4 shows that wind energy generation exhibits a generally constant profile, but with slight hourly temporal variations. During nighttime hours, temperature differentials between land and sea can create favorable conditions for increased wind flow. The cooling of land surfaces leads to the formation of land breezes, which can enhance the overall wind speed. This phenomenon contributes to the slight time-varying nature of wind energy production, with higher output observed during specific times, especially in the summer months.

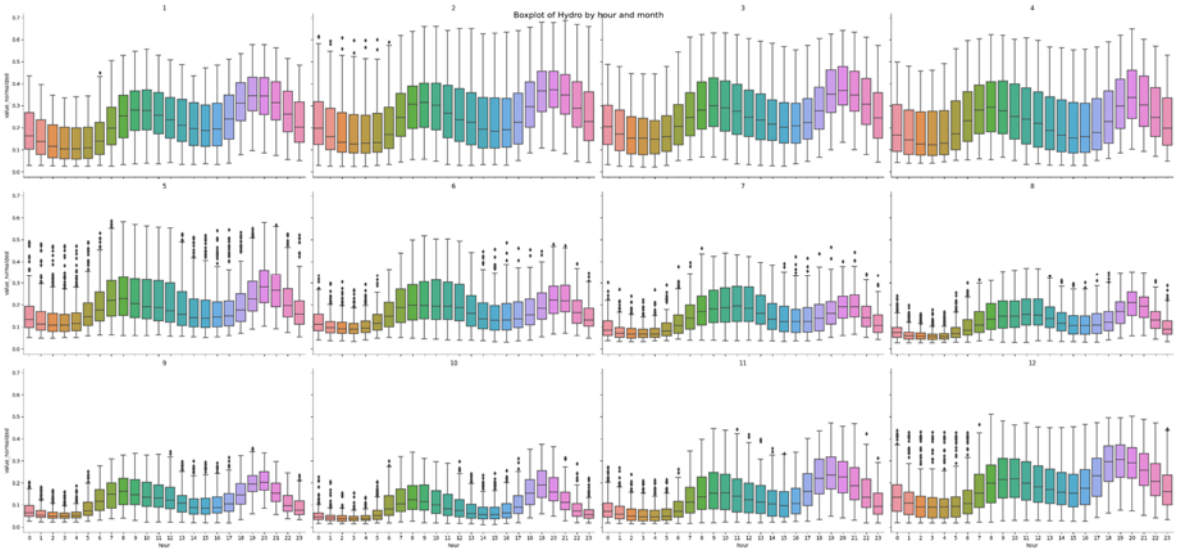

Figure 5. Box-plot of Hydro UGH

The distinctive profile observed in hydroelectric generation, characterized by two prominent peaks in the morning and afternoon, is not solely determined by the availability of water resources. Instead, it reflects a deliberate decision to ramp up electricity production during these specific time periods when the energy market offers more favorable pricing conditions. By adjusting the release of water and the operation of turbines, hydropower operators can align their generation capacity with the demand for electricity, maximizing their revenue potential. Therefore, it is difficult to understand the dynamics of hydroelectric energy from this data alone. Based on our analysis, we have chosen to prioritize the investigation of solar and wind energy sources.

3.2 DATA NORMALIZATION

The data will be standardized before the analysis is performed. Standardizing measurements by dividing them by installed capacity is a common practice in renewable energy data analysis. Standardizing measurements allows you to compare values across multiple installations and periods. The installed capacity may be increased or decreased, so if the

measured data is divided by the installed capacity, it becomes a common scale and meaningful comparison is possible.

In addition, when dealing with spatial analysis data, the geographical units of the installed capacity data and the measured value data are different. In our case, the installed capacity data is reported at the regional level (Community), while the measured values are provided at a more granular level (Province). To address this discrepancy and establish a consistent framework for analysis, it was necessary to associate the Provinces with their respective Communities.

The below code is updating the `'geoname'` column in the `solar_measured_d DataFrame`. If the value in the `'geoname'` column matches any of the cities in the `andalucia_cities` list, it is replaced with the string 'Andalusia'. This is a way to group the solar measurement data from different cities in Andalusia under a single label.

```
andalucia_cities = ['Almería', 'Cádiz', 'Cordova', 'Granada', 'Huelva', 'Jaén',
'Málaga', 'Sevilla']
solar_measured_day.loc[solar_measured_day['geoname'].isin(andalucia_cities), 'geoname']
= 'Andalusia'
```

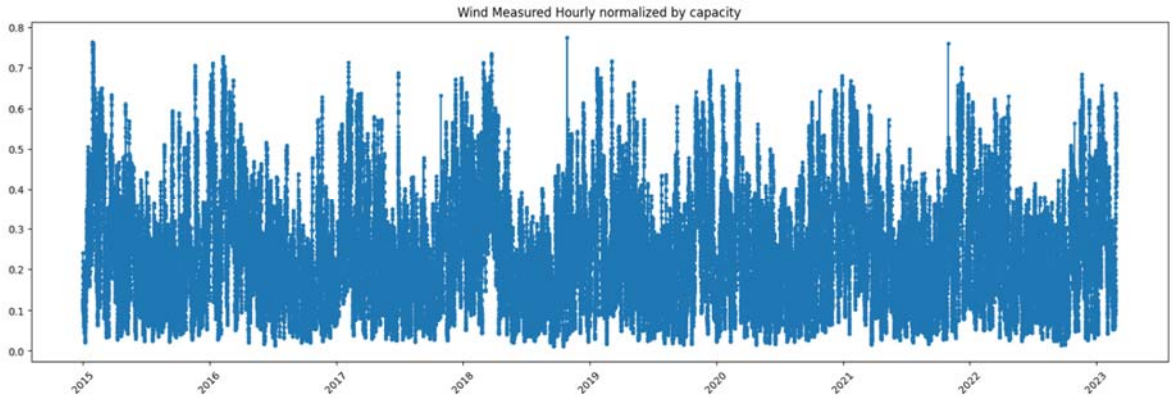

Figure 6. Wind Measured Hourly normalized by capacity

Figure 6 displays the standardized total wind energy data. This graph demonstrates the consistency and reliability of the standardized total wind energy data, highlighting its effectiveness in capturing the underlying trends and fluctuations in wind energy generation.

The standardized values enable a fair comparison across different time periods, as they remove the influence of varying scales and capacities. The importance of data standardization can be further demonstrated by looking at the two primary graphs in Figure 7.

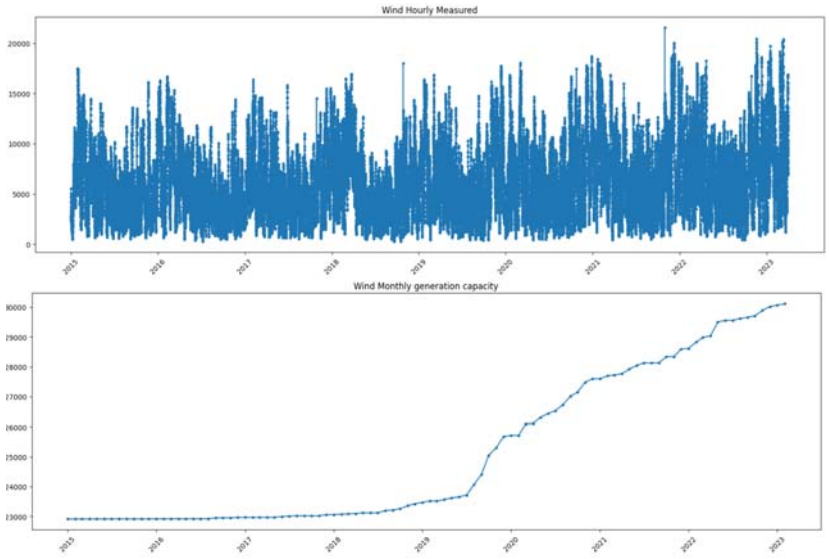

Figure 7. Wind - Measured / Capacity

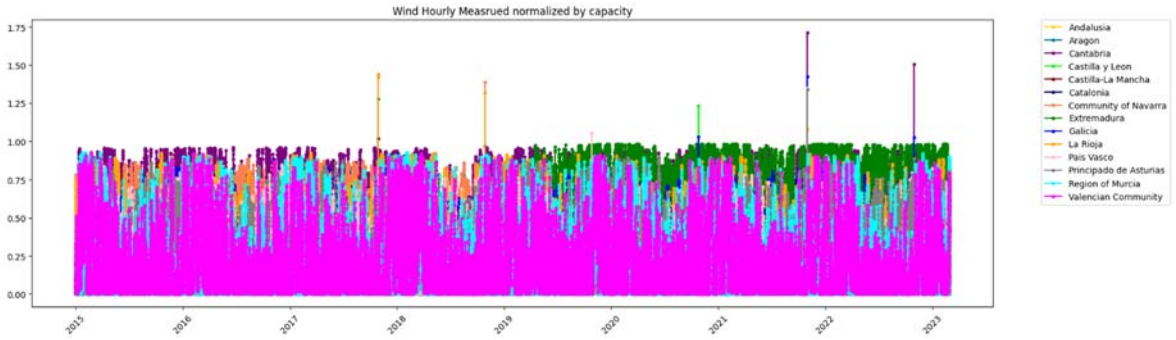

Figure 8. Wind Measured Hourly spatial normalized by capacity

Figure 8 demonstrates the standardization of spatial data, showcasing the application of standardized values to analyze and compare wind energy generation across different geographical regions (Figure 9).

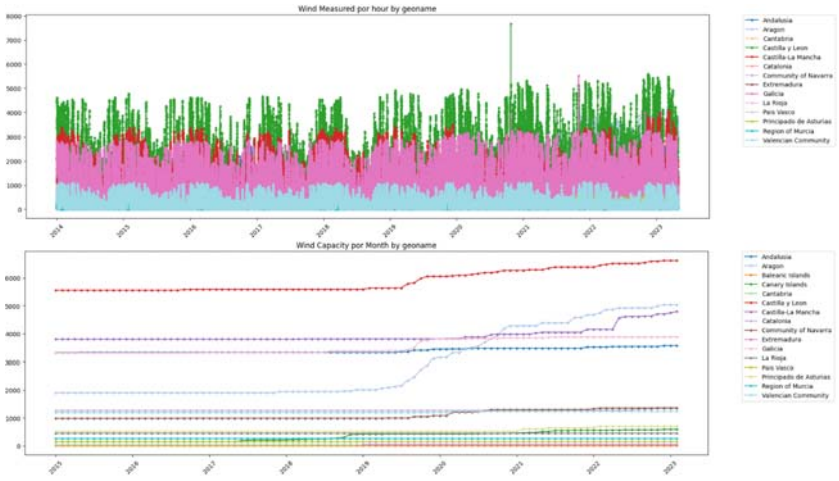

Figure 9. Wind spatial- Measured / Capacity

The analysis will be conducted using the standardized data, allowing for a comprehensive examination of energy generation patterns and trends across various regions.

Chapter 4. PHOTOVOLTAIC AND SOLAR THERMAL GENERATION

4.1 SOLAR PHOTOVOLTAIC ENERGY GENERATION

Prior to conducting an analysis of solar energy, it is crucial to assess the relevance and significance of two prominent technologies within this domain: photovoltaic (PV) and thermal systems. By evaluating their importance and determining the value derived from analyzing these methods, we can establish a strong justification for their comprehensive examination.

To explore the relationship between the two main methods of solar energy, photovoltaic (PV) and thermal systems, we can employ a dendrogram analysis. This approach allows us to study shedding light on their interdependencies and shared characteristics.

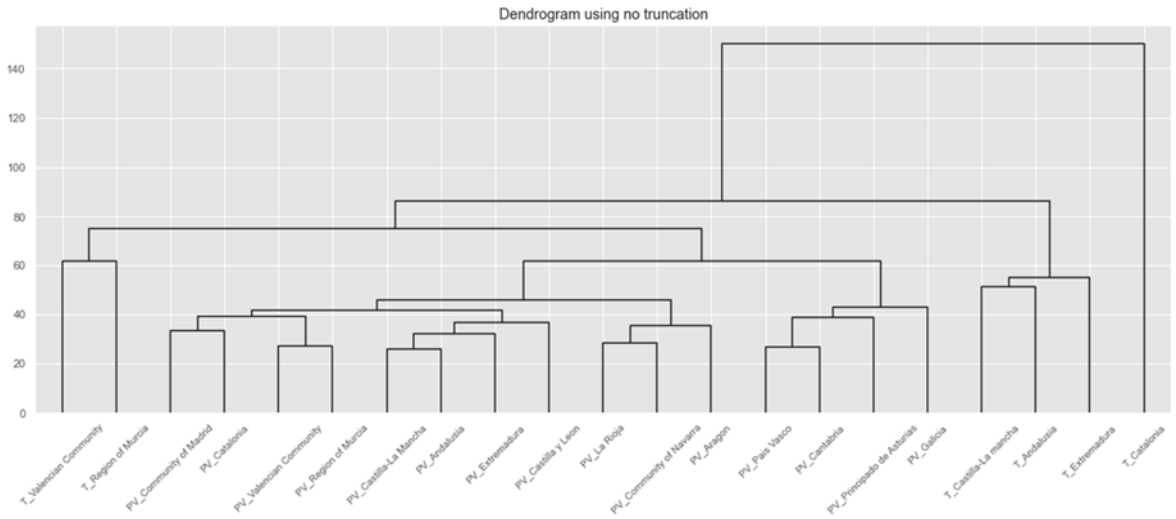

Figure 10. Regional Correlation of solar PV and Thermal

The application of hierarchical clustering (Figure 10) with the Euclidean metric reveals a distinct separation between PV and Thermal (T) factors, suggesting that these two basic solar technologies are different. This observation emphasizes the importance of considering PV and thermal systems as separate entities when studying solar energy.

Furthermore, given the stagnant installation capacity of thermal systems over the years and the limited number of variables associated with it, it is prudent to prioritize our analysis on solar photovoltaic (PV) systems. Accordingly, our research endeavors will be centered on the analysis of solar photovoltaic (PV) systems.

4.1.1 DAILY DATA

In this chapter, the daily data of solar power generation energy across each province will be analyzed. The objective of this analysis is to gain comprehensive insights into the daily patterns and changes in solar power at the regional level. The dataset employed for this analysis includes a 'DATE' row and a 'PROVINCES' column, facilitating the exploration and examination of solar power generation trends at a granular level.

Significant emphasis is placed on understanding the temporal dynamics and regional aspects of solar energy production. By scrutinizing the daily patterns, the factors influencing solar power production on a daily basis, such as seasonal variations, weather conditions, and regional peculiarities, can be discerned.

4.1.1.1 Energy generation by provinces

The correlation between provinces is examined to gain insights into the relationships and dependencies among them in terms of solar power generation. By analyzing the correlation matrix, we can identify the degree of association between different provinces and assess the similarities or differences in their solar power generation patterns. This information is valuable for understanding the interconnectedness and potential influences between provinces in terms of solar energy production.

geoname	Andalusia	Aragon	Cantabria	Castilla y Leon	Castilla-La Mancha	Catalonia	Community of Madrid	Community of Navarra	Extremadura	Galicia	La Rioja	Pais Vasco	Principado de Asturias	Region of Murcia	Valencian Community
Andalusia	1.000000	0.754453	0.513403	0.772433	0.897510	0.710705	0.838214	0.697564	0.859444	0.629493	0.680899	0.610140	0.529367	0.783675	0.763389
Aragon	0.754453	1.000000	0.678282	0.837072	0.845758	0.878302	0.849319	0.894870	0.772166	0.690950	0.863913	0.759017	0.640508	0.665393	0.780182
Cantabria	0.513403	0.678282	1.000000	0.674942	0.570763	0.612832	0.602968	0.723171	0.519520	0.654804	0.759254	0.855042	0.789471	0.423126	0.500144
Castilla y Leon	0.772433	0.837072	0.674942	1.000000	0.839048	0.755774	0.892774	0.835814	0.875030	0.828692	0.825691	0.736567	0.676079	0.587105	0.665929
Castilla-La Mancha	0.897510	0.845758	0.570763	0.839048	1.000000	0.790568	0.924155	0.779479	0.870240	0.660233	0.759675	0.661466	0.567317	0.794059	0.862659
Catalonia	0.710705	0.878302	0.612832	0.755774	0.790568	1.000000	0.794066	0.819492	0.703149	0.620253	0.794784	0.706927	0.593241	0.677666	0.811763
Community of Madrid	0.838214	0.849319	0.602968	0.892774	0.924155	0.794066	1.000000	0.824018	0.877404	0.719388	0.803497	0.695761	0.604634	0.691595	0.786926
Community of Navarra	0.697564	0.894870	0.723171	0.835814	0.779479	0.819492	0.824018	1.000000	0.725977	0.709808	0.964321	0.841353	0.650279	0.601734	0.701533
Extremadura	0.859444	0.772166	0.519520	0.875030	0.870240	0.703149	0.877404	0.725977	1.000000	0.707045	0.699851	0.588507	0.569008	0.623977	0.681229
Galicia	0.629493	0.690950	0.654804	0.828692	0.660233	0.620253	0.719388	0.709808	0.707045	1.000000	0.691773	0.663739	0.690385	0.455079	0.497236
La Rioja	0.680899	0.863913	0.759254	0.825691	0.759675	0.794784	0.803497	0.964321	0.699851	0.691773	1.000000	0.879014	0.666319	0.582707	0.683575
Pais Vasco	0.610140	0.759017	0.855042	0.736567	0.661466	0.706927	0.695761	0.841353	0.588507	0.663739	0.879014	1.000000	0.695697	0.518722	0.591207
Principado de Asturias	0.529367	0.640508	0.789471	0.676079	0.567317	0.593241	0.604634	0.650279	0.569008	0.690385	0.666319	0.695697	1.000000	0.425341	0.482768
Region of Murcia	0.783675	0.665393	0.423126	0.587105	0.794059	0.677666	0.691595	0.601734	0.623977	0.455079	0.582707	0.518722	0.425341	1.000000	0.871723
Valencian Community	0.763389	0.780182	0.500144	0.665929	0.862659	0.811763	0.786926	0.701533	0.681229	0.497236	0.683575	0.591207	0.482768	0.871723	1.000000

Figure 11. Regional Correlation of Solar PV

According to the correlation matrix shown in Figure 11, all pairs of provinces analyzed exhibit positive correlations. Navarra and La Rioja demonstrate a high level of similarity with a correlation coefficient of 0.964321. Madrid and Castilla-La Mancha exhibit a significant correlation of 0.924155, while Andalusia and Castilla-La Mancha display a substantial correlation of 0.897510. Aragon and Navarra show a strong positive correlation of 0.894870, and Castilla y Leon and Madrid demonstrate a notable positive correlation of 0.892774. Furthermore, the province of Castilla y Leon exhibits a high correlation (above 0.8) with six other provinces, indicating a strong interdependence in terms of solar power generation.

Prior to conducting PCA, performing a dendrogram analysis serves a crucial purpose in enhancing our understanding of the underlying data structure and facilitating informed decisions regarding variable selection, data transformation, and interpretation of PCA outcomes.

By assessing the hierarchical structure and identifying potential clustering patterns within the dataset, a dendrogram visually depicts the similarities and dissimilarities between data points or variables. This representation enables the identification of distinct groups or clusters that may exist within the data.

Furthermore, the dendrogram analysis assists in determining the optimal number of clusters or groups based on the similarity of variables or observations. This knowledge aids in guiding subsequent PCA by providing valuable insights into the dataset's inherent structure.

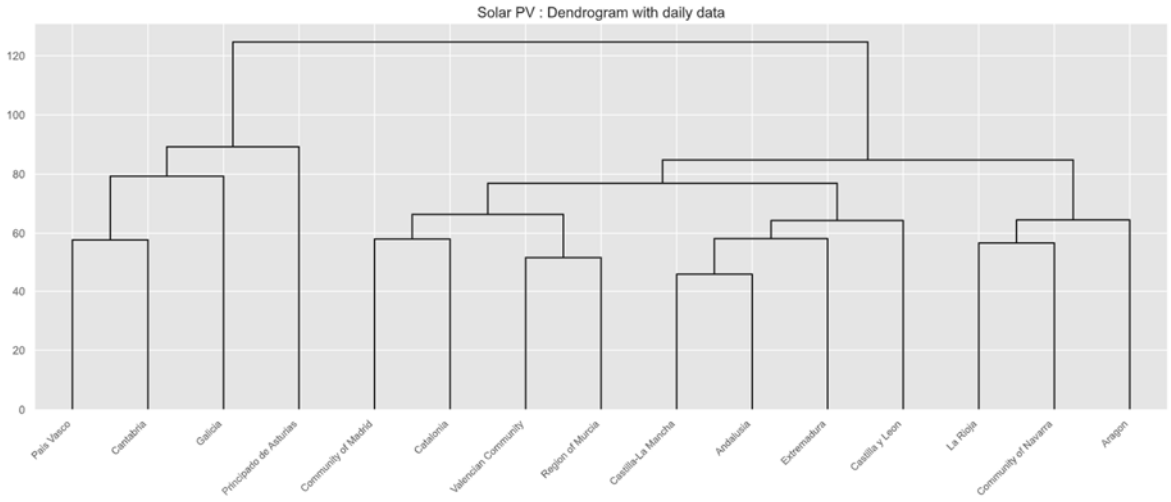

Figure 12. Dendrogram by province with Solar PV daily data

The dendrogram shown in Figure 12 reveals the presence of four distinct groups within the dataset. Interestingly, these groups primarily consist of provinces that are geographically close to each other, indicating a regional clustering pattern. However, Madrid stands out as an exception, as it appears to have a closer relationship with provinces such as Catalunia, Valencia, and Murcia rather than Castilla León or Castilla-La Mancha. This grouping pattern suggests a unique association between Madrid and these specific provinces, highlighting the potential influence of geographical proximity and other factors on solar power generation dynamics.

This pattern in the dendrogram aligns, obviously, with the findings from the correlation matrix analysis. This further supports the notion that geographical proximity and other factors contribute to the similarities in solar power generation patterns among neighboring provinces. Additionally, the outlier status of Madrid, which demonstrates stronger correlations with provinces outside its immediate geographical vicinity, suggests unique factors influencing solar energy generation in the region.

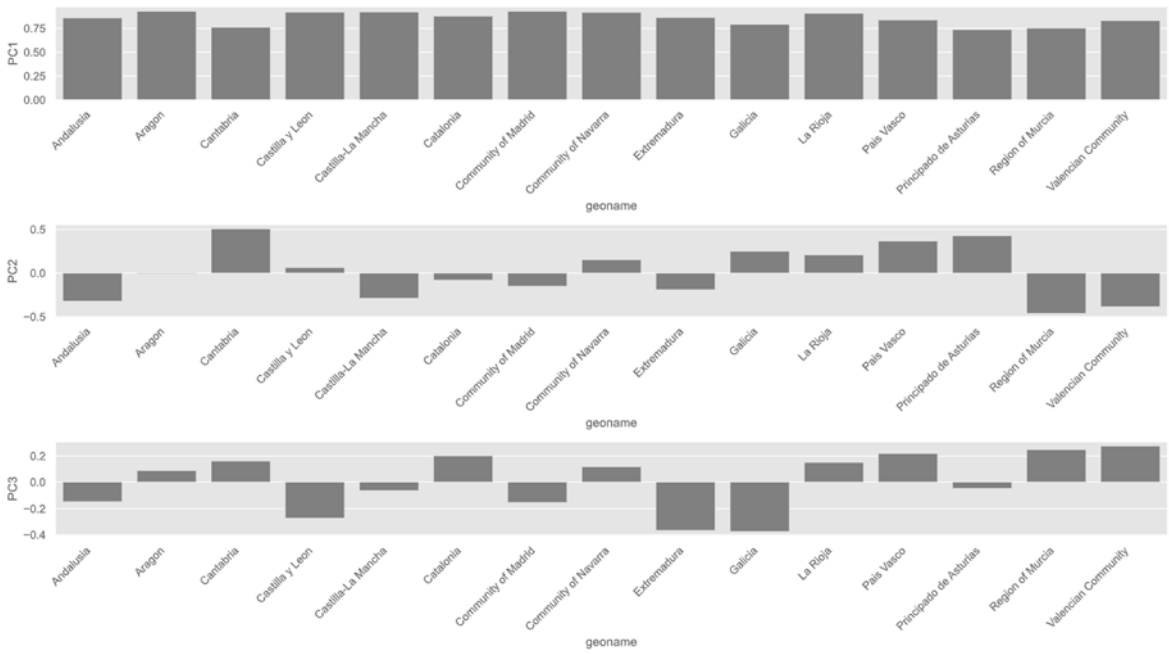

Figure 13. PCA Loadings for Solar PV daily generation

The analysis of the principal components reveals valuable insights into the regional variations in solar power generation. According to Figure 13, the first principal component (PC1) represents a weighted average, with Madrid and Castilla y Leon having the highest weight and Asturias having the lowest weight. This indicates that Madrid and Castilla y Leon play significant roles in determining the overall trend of solar power generation across provinces.

The second principal component (PC2) captures the north-south differences, assigning positive weight to the north and negative weight to the south. This component highlights the regional differences in solar radiation between the northern and southern provinces.

Similarly, the third principal component (PC3) captures additional east-west changes, assigning positive weight to the east and negative weight to the west. This component reflects the regional differences in solar radiation between the eastern and western provinces.

Together, these principal components provide a comprehensive understanding of the regional dynamics of solar power generation, incorporating factors such as geographic location, solar radiation intensity, and other underlying variables.

	Exp_variance	cum_Exp_variance
PC1	0.737960	0.737960
PC2	0.088315	0.826275
PC3	0.046639	0.872914
PC4	0.033177	0.906091
PC5	0.020621	0.926713
PC6	0.017142	0.943854
PC7	0.011599	0.955453
PC8	0.010500	0.965953
PC9	0.007520	0.973473
PC10	0.006812	0.980285
PC11	0.005500	0.985785
PC12	0.005206	0.990991
PC13	0.004066	0.995058
PC14	0.003034	0.998091
PC15	0.001909	1.000000

Figure 14. Cumulative explained variance of PCA for Solar PV daily generation

Figure 14 shows the relative and cumulative importance of each principal component. The cumulative explained variance of the three main principal components (PC1, PC2, and PC3) surpasses 87%, demonstrating their ability to capture a substantial portion of the underlying variability in the solar photovoltaic energy data. This high cumulative explained variance signifies that these principal components successfully capture the fundamental patterns and relationships within the dataset, providing valuable insights into the dynamics of solar power generation.

In a biplot, the variables are represented by vectors, and the observations are depicted as points in the plot. The direction and length of the vectors indicate the contribution and importance of each variable to the overall data structure. The closer the vectors are to each other, the higher the correlation between the variables. By visualizing the variables and observations together, the biplot provides a compact and intuitive representation of the data, facilitating the exploration and interpretation of complex multivariate datasets. It is a

valuable tool for dimensionality reduction, pattern recognition, and identifying key variables that drive the structure of the data.

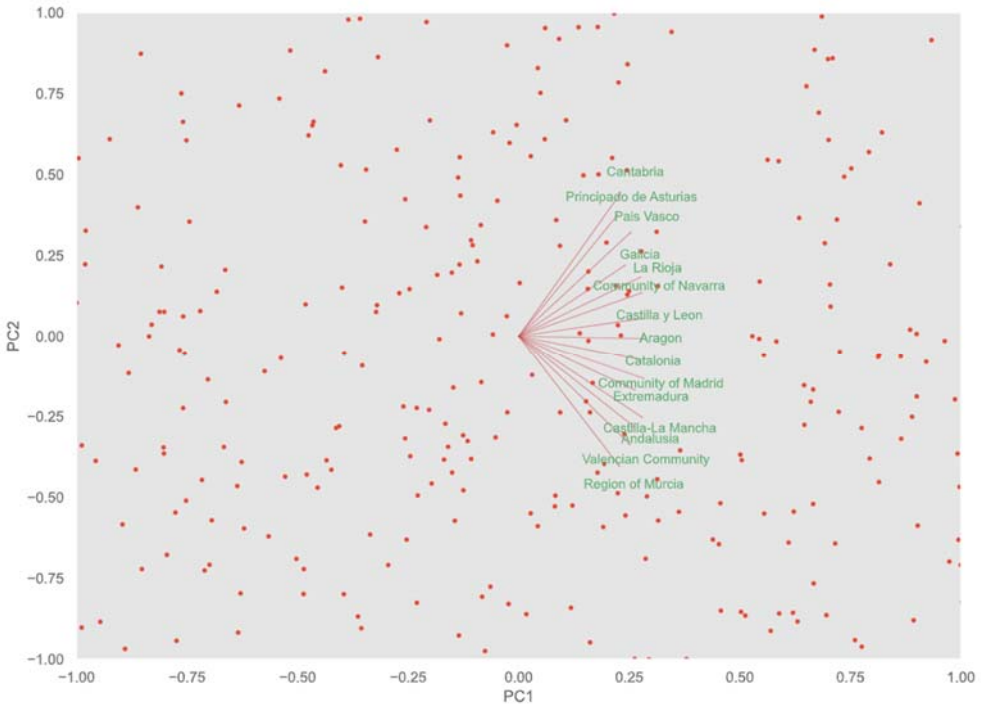

Figure 15. Biplot with PC1 and PC2 for Solar PV daily generation

As observed from the biplot of Figure 15, PC1 exhibits a positive side that corresponds to periods with typically higher solar radiation. On the other hand, PC2 displays a clear distinction between the northern and southern regions, with its positive side representing the northern areas and the negative side representing the southern areas. These findings provide valuable insights into the relationship between solar radiation intensity and geographical location, enhancing our understanding of the spatial patterns of solar energy generation.

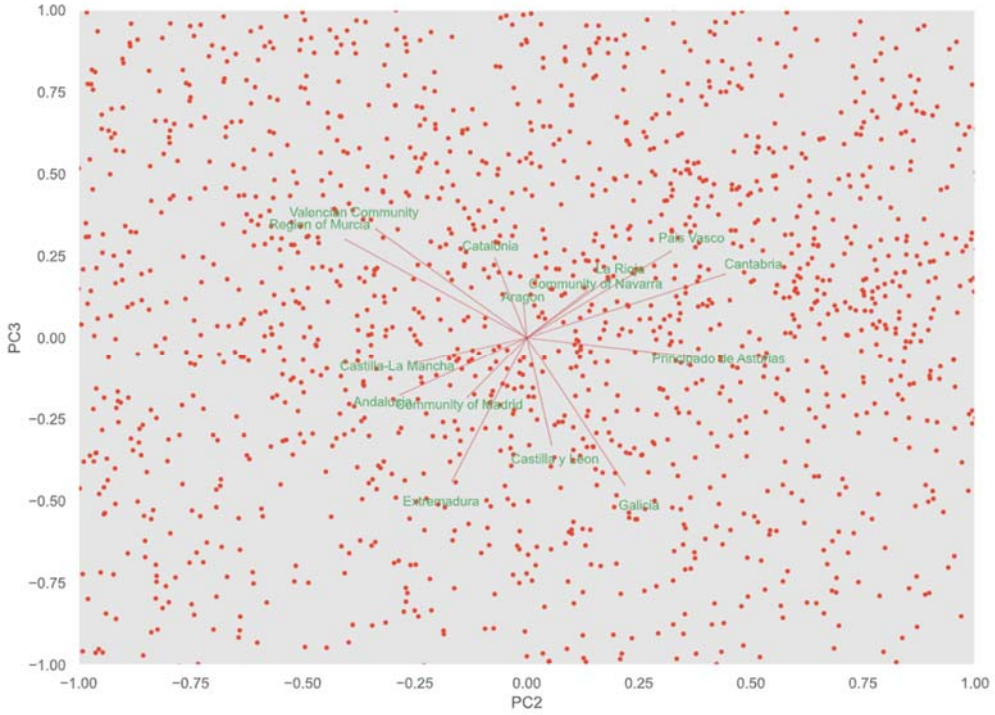

Figure 16. Biplot with PC2 and PC3 for Solar PV daily generation

Figure 16 shows the biplot for PC2 and PC3, presenting a spatial representation of solar power generation across the provinces of Spain. Notably, the biplot exhibits a rotated Spanish map, where the provinces are horizontally symmetrical. This visualization allows for a clear understanding of the geographical distribution and patterns of solar energy generation within the country. The rotation of the map provides insights into the regional variations in solar radiation intensity and highlights the distinct solar power generation characteristics in different parts of Spain. This representation is instrumental in identifying clusters of provinces with similar solar energy profiles and facilitating further analysis of the underlying factors influencing solar power generation.

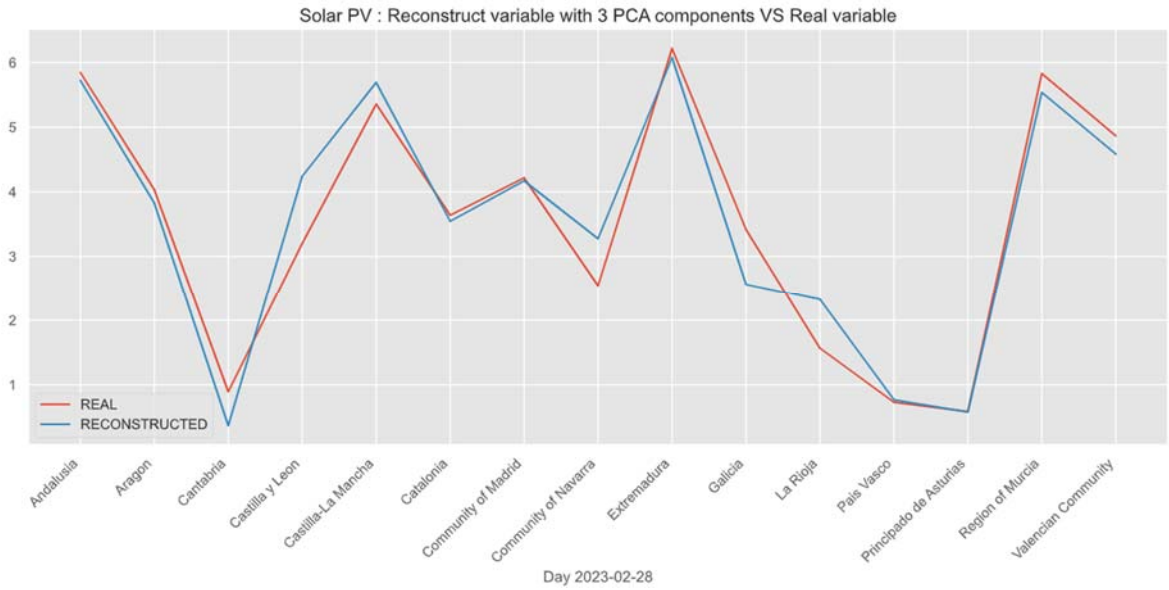

Figure 17. Reconstruct variables with the three main principal components for Solar PV daily generation

Figure 17 represents a specific day's solar power generation data. Upon reconstructing the data using the three main principal components PC1, PC2, and PC3 (which represents a cumulative explained variance larger than 87%), we find that the reconstructed data closely aligns with the actual data. The high degree of similarity observed between the reconstructed and original datasets highlights the effectiveness of the principal component analysis (PCA) in capturing the underlying patterns and variability present in the solar photovoltaic energy data. The ability to accurately reconstruct the data using the principal components further reinforces the validity and reliability of our findings and supports the notion that PC1, PC2, and PC3 effectively capture the primary drivers and distinctive patterns in solar energy generation.

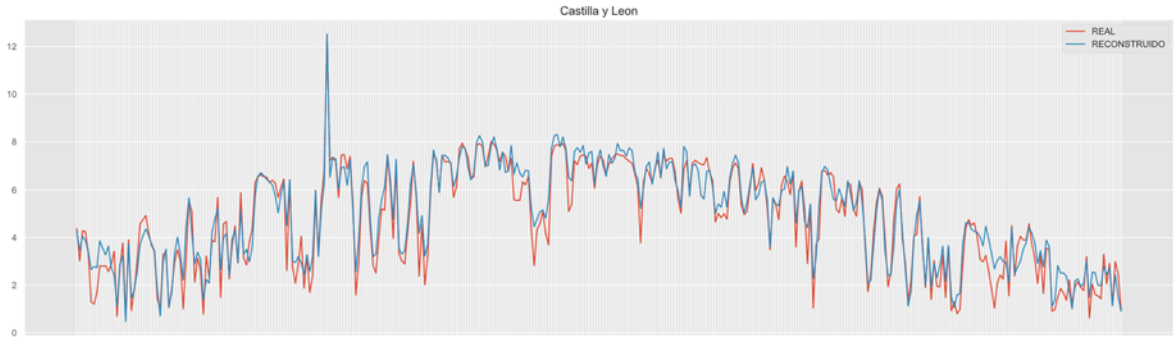

Figure 18. Reconstruct variables of Castilla y Leon with the first three principal components for Solar PV daily generation

Also, Figure 18 displays the solar photovoltaic data of Castilla y Leon for the entire year of 2015. The close correspondence between the reconstructed data and the actual data demonstrates the effectiveness of the principal component analysis in capturing and reproducing the underlying patterns and variability in solar power generation.

This comprehensive analysis of daily photovoltaic generation data offers valuable insights into local patterns, dependencies, and dynamics, providing essential background knowledge for further detailed and in-depth analyses.

4.1.2 HOURLY DATA

4.1.2.1 Energy generation

In this chapter, the hourly data of solar photovoltaic energy is analyzed. PCA was employed to identify and understand the underlying patterns and relationships in the solar photovoltaic energy data. The data frame used consists of the 'DATE' row and the 'HOUR' column.

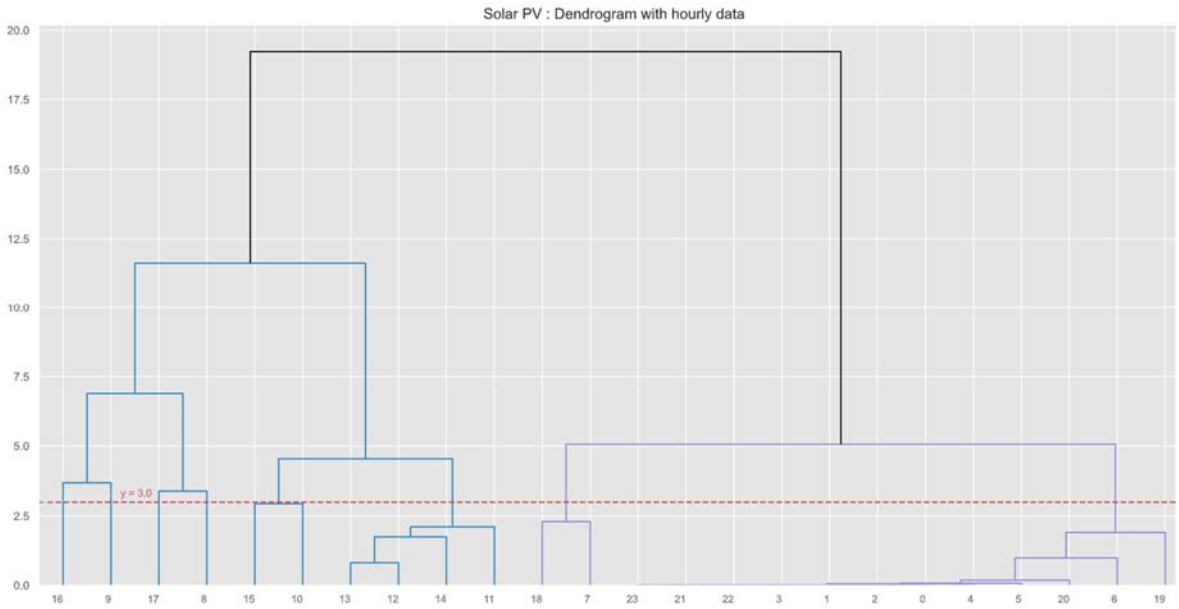

Figure 19. Dendrogram by hour for Solar PV generation

In hierarchical clustering, the choice of distance metric is an important consideration as it determines how the distances between samples or clusters are calculated. The Euclidean distance used here measures the straight-line or Euclidean distance between two points in a multidimensional space. In the case of your data, it calculates the Euclidean distance between pairs of samples based on their feature values. Samples that are closer together in the Euclidean space will be considered more similar, while samples that are farther apart will be considered more dissimilar.

The dendrogram analysis uncovers clear divisions between time periods characterized by the presence or absence of solar radiation. Furthermore, within the time periods marked by solar radiation, revealing the existence of subgroups delineated by central and other time zones.

The observed divisions and clustering patterns provide valuable insights into the temporal dynamics of solar radiation and its impact on energy generation. By identifying these distinct groups, we can discern variations in solar radiation patterns and better understand the underlying factors influencing solar energy generation during different time intervals.

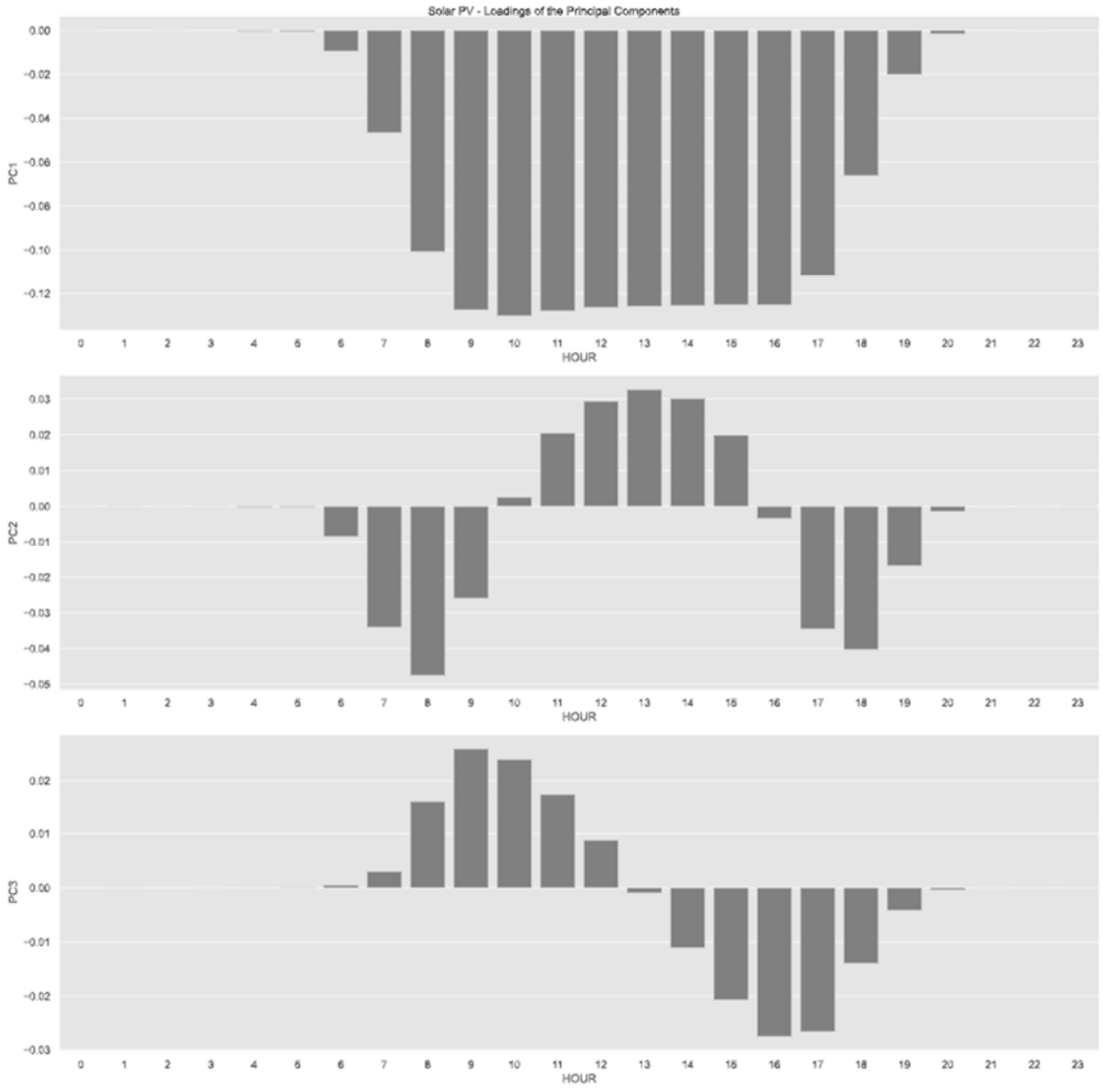

Figure 20. PCA Loadings for hourly Solar PV generation

According to Figure 20, the first principal component (PC1) corresponds to the predominant direction in the data space that captures the most significant variation in solar photovoltaic energy data. It can be interpreted as the principal component that effectively captures changes in solar radiation levels. The magnitude of PC1 serves as an indicator of the solar impact on a given day. A more negative value of PC1 means a higher intensity of solar radiation, implying favorable conditions for solar energy generation. Conversely, a less negative value of PC1 suggests diminished solar irradiation, indicating a reduced potential for solar energy production. The analysis of PC1 provides valuable insights into the

dependency of solar photovoltaic energy generation on the availability and strength of sunlight throughout the day, contributing to a comprehensive understanding of the influential factors affecting solar power output.

The second principal component (PC2) represents an orthogonal direction to PC1, capturing the remaining main variability in the data that is not accounted for by PC1. It reveals additional patterns and relationships in the solar photovoltaic energy data that are distinct from those captured by PC1. Analyzing the loadings and contributions of variables to PC2 enables us to uncover secondary factors that influence energy generation and identify any complementary or contrasting trends compared to PC1.

PC2 offers valuable insights into the temporal variations in sunlight intensity throughout the daylight hours. Positive values of PC2 indicate stronger solar radiation occurring during the central hours of the day, signifying instances of heightened sunlight exposure and potential peak energy generation. Conversely, negative values indicate stronger solar radiation at other times, suggesting fluctuations in solar intensity throughout the day. Analyzing PC2 enables us to unravel the temporal dynamics of solar radiation and its implications for energy generation, contributing to a comprehensive understanding of solar power production.

The third principal component (PC3) serves to enhance our understanding of the solar photovoltaic energy data by capturing additional distinctive patterns and variations that were not accounted for by PC1 and PC2. Through a careful analysis of the variable loadings and contributions to PC3, we can unveil further nuances in the data and gain valuable insights into additional factors influencing solar energy generation.

Specifically, PC3 provides indications regarding the temporal distribution of solar radiation intensity relative to the time of day. Positive values of PC3 suggest stronger solar radiation occurring before 13:00, highlighting periods of increased solar intensity during the morning. Conversely, negative values of PC3 indicate stronger solar radiation after 13:00, pointing towards heightened solar radiation during the afternoon with respect to the morning, for example when there are clouds during the morning that disappear during the afternoon. By examining PC3, we can discern the temporal dynamics of solar radiation patterns and their

implications for solar energy generation, contributing to a comprehensive understanding of the factors shaping renewable energy production.

	Exp_variance	cum_Exp_variance	PC13	9.164438e-05	0.999943
PC1	9.010941e-01	0.901094	PC14	4.187445e-05	0.999984
PC2	6.228340e-02	0.963378	PC15	5.734636e-06	0.999990
PC3	2.354082e-02	0.986918	PC16	4.346261e-06	0.999995
PC4	6.981026e-03	0.993899	PC17	3.465635e-06	0.999998
PC5	2.439144e-03	0.996338	PC18	1.150696e-06	0.999999
PC6	1.396529e-03	0.997735	PC19	3.407455e-07	0.999999
PC7	8.972781e-04	0.998632	PC20	2.007168e-07	1.000000
PC8	4.420269e-04	0.999074	PC21	1.087711e-07	1.000000
PC9	3.525086e-04	0.999427	PC22	9.132550e-08	1.000000
PC10	1.715476e-04	0.999598	PC23	7.436247e-08	1.000000
PC11	1.559938e-04	0.999754	PC24	6.431499e-08	1.000000
PC12	9.652824e-05	0.999851			

Figure 21. Cumulative explained variance of PCA for hourly Solar PV generation

The cumulative explained variance of the three main principal components (PC1, PC2, and PC3) exceeds 98% (see Figure 21). This indicates that these three components collectively capture a significant portion of the overall variability present in the solar photovoltaic energy data. The high cumulative explained variance suggests that these principal components effectively represent the essential patterns and relationships within the dataset.

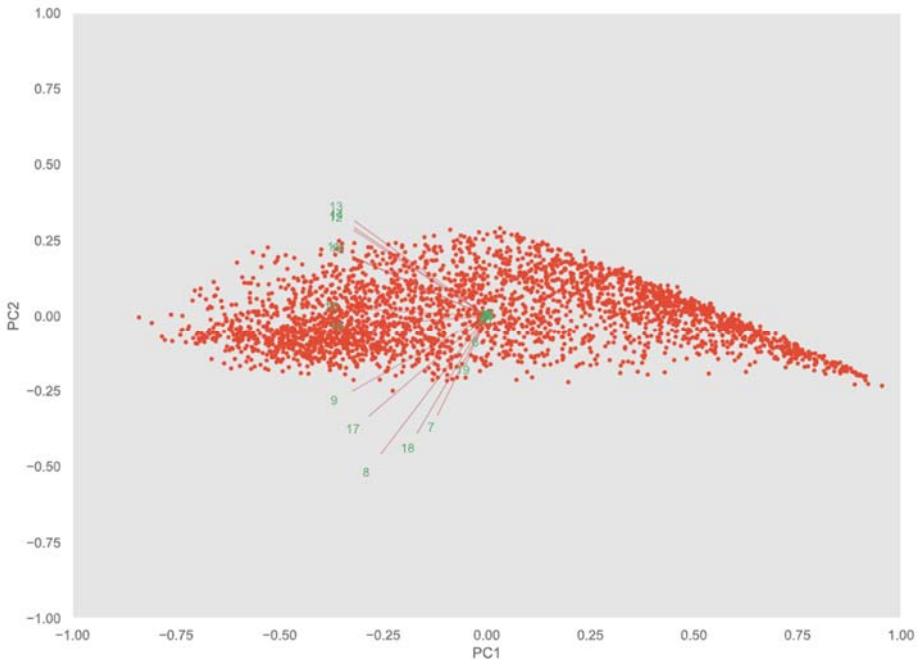

Figure 22. Biplot with PC1 and PC2 for hourly Solar PV generation

As observed from the biplot shown in Figure 22, PC1 exhibits a negative side that corresponds to periods with typically higher solar radiation. In contrast, PC2 shows a positive side representing daylight hours and a negative side indicating other times of the day.

These findings suggest that PC1 captures variations in solar radiation intensity, with negative values indicating periods of stronger solar radiation. On the other hand, PC2 highlights the temporal distribution of solar radiation, with positive values aligning with daylight hours and negative values corresponding to non-daylight periods.

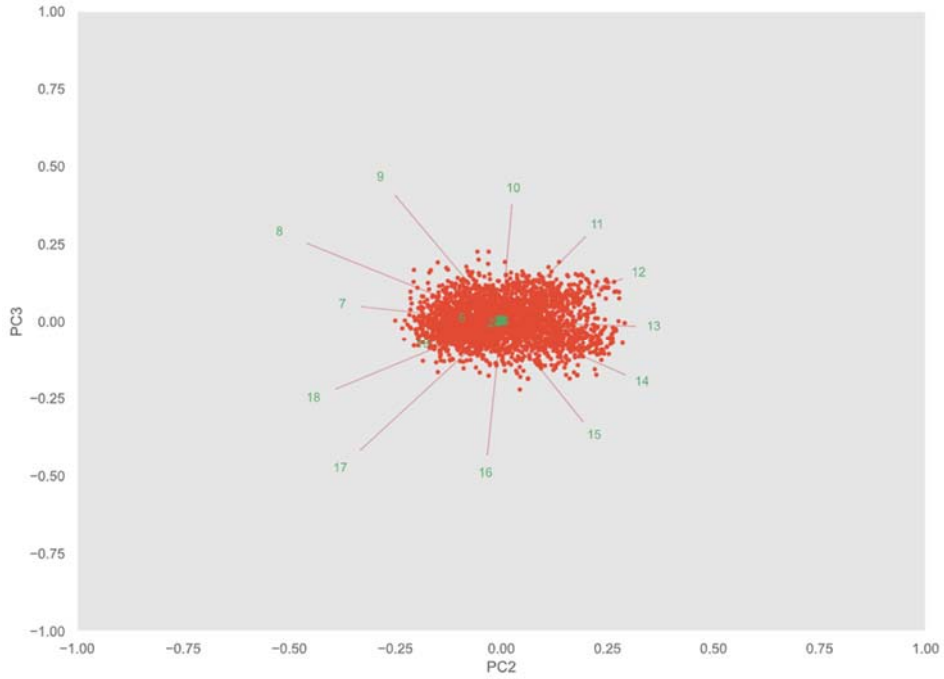

Figure 23. Biplot with PC2 and PC3 for Solar PV generation

A notable and intriguing observation emerges from the biplot of PC2 and PC3 (see Figure 23), revealing a distinct clock-shaped pattern that aligns with specific time zones associated with solar radiation. Particularly, when examining the PC3 axis, the resemblance to a conventional clock is strikingly apparent.

This intriguing finding implies a strong correlation between the solar radiation patterns and the temporal progression throughout the day. The clock-shaped pattern observed in the bi-plots suggests a consistent and predictable cyclic variation in solar radiation intensity over time. This regularity reinforces the notion of solar energy generation being influenced by the diurnal cycle, with distinct time zones exhibiting varying levels of solar radiation.

The clock-shaped pattern observed in the biplot serves as a valuable visual representation of the temporal dynamics of solar radiation, offering insights into the temporal patterns and potential opportunities for optimizing solar energy utilization.

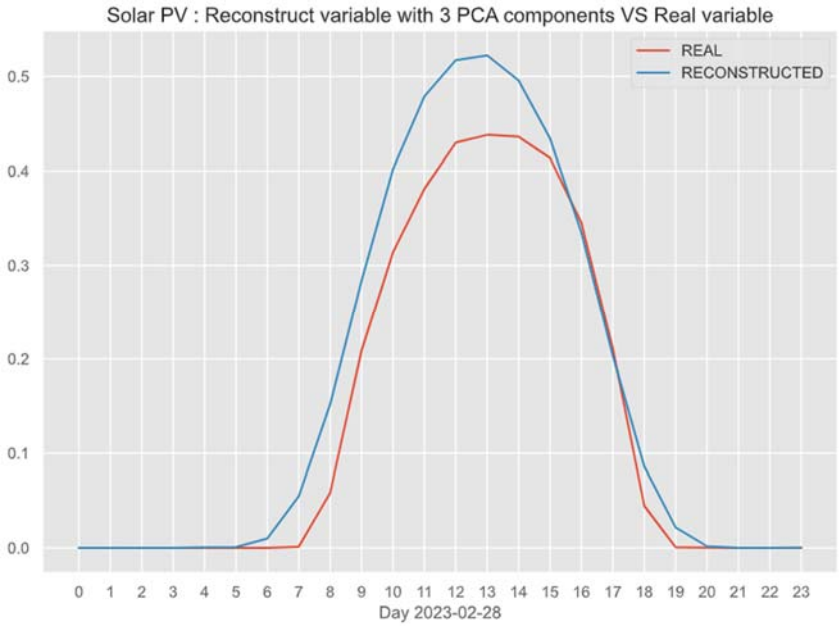

Figure 24. Reconstruct variables with the three main principal components for Solar PV generation

Reconstruction of the data using principal components PC1, PC2, and PC3 closely aligns with the actual data, validating the effectiveness of PCA in capturing underlying patterns and variability in solar photovoltaic energy data (see Figure 24). The accurate reconstruction reinforces the validity of our findings, highlighting that PC1, PC2, and PC3 effectively capture primary drivers and distinctive patterns in solar energy generation.

Let us look at the graph of how the characteristics of each PCA appear in real solar energy data.

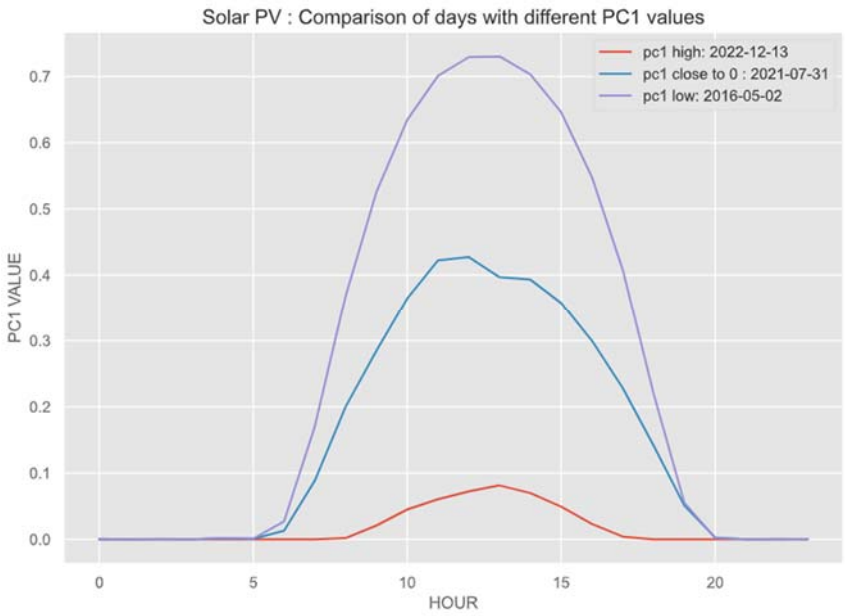

Figure 25. PC1 values with dates for Solar PV generation

Figure 25 illustrates three distinct curves with varying heights. It is evident that the highest purple curve corresponds to lower values of PC1, indicating a greater amount of solar radiation. Conversely, the lowest curve is observed when PC1 has higher values. An intermediate curve represents average conditions when PC1 is close to zero.

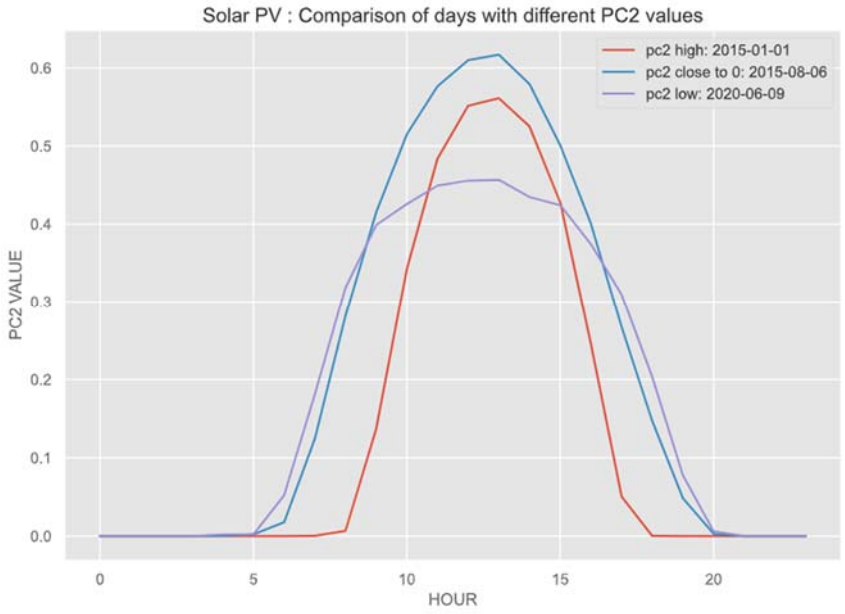

Figure 26. PC2 values with dates for Solar PV generation

Figure 26 reveals three distinct curves with varying widths. The narrowest red curve corresponds to higher values of PC2, indicating an increased amount of solar radiation during the day. This pattern suggests that solar energy generation is concentrated during daylight hours, with energy generation starting later and ending relatively quickly.

Conversely, the widest purple curve is observed when PC2 values are low. This indicates that energy generation begins earlier and continues until later hours, extending the duration of solar energy production. The width of the curve reflects the extended period of sunlight exposure, allowing for a more sustained generation of solar power.

The intermediate blue curve represents average conditions when PC2 approaches zero. It signifies a balanced distribution of solar energy generation throughout the day, reflecting the mean conditions for energy production.

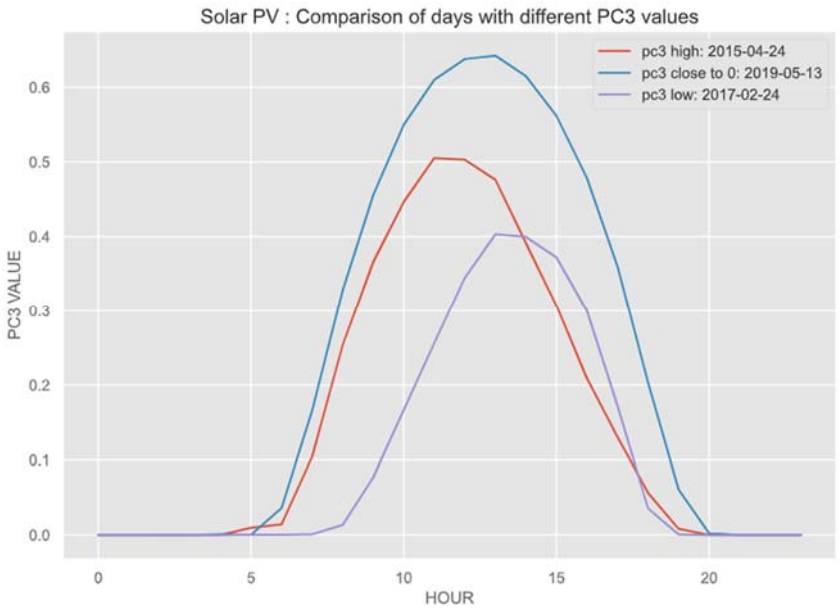

Figure 27. PC3 values with dates for Solar PV generation

Figure 27 exhibits a normal curve along with two asymmetric curves. The red curve, displaying positive skewness, corresponds to higher values of PC3, signifying increased solar radiation prior to 13:00. Consequently, energy generation starts relatively quickly and concludes within a shorter duration.

Conversely, the purple curve, displaying negative skewness, is observed when PC3 values are low, indicating higher solar radiation levels after 13:00. Energy generation initiates later and extends for a longer period.

The mid-blue curve represents average conditions when PC3 approaches zero, resulting in a normal curve shape. This signifies a balanced distribution of solar radiation throughout the day, leading to a relatively consistent pattern of energy generation.

By integrating the information obtained from the dendrogram analysis and PCA, we have covered two essential aspects of data exploration and dimensionality reduction. The dendrogram analysis helps us understand the hierarchical structure and potential clustering patterns within the dataset, while the PCA analysis allows us to capture the most relevant information and identify key components driving the variability in the data. The combination of these techniques provides a comprehensive perspective on the dataset and aids in the initial estimation of the optimal number of clusters. The dendrogram analysis, combined with the PCA analysis, provides valuable insights into the initial estimation of the optimal number of clusters.

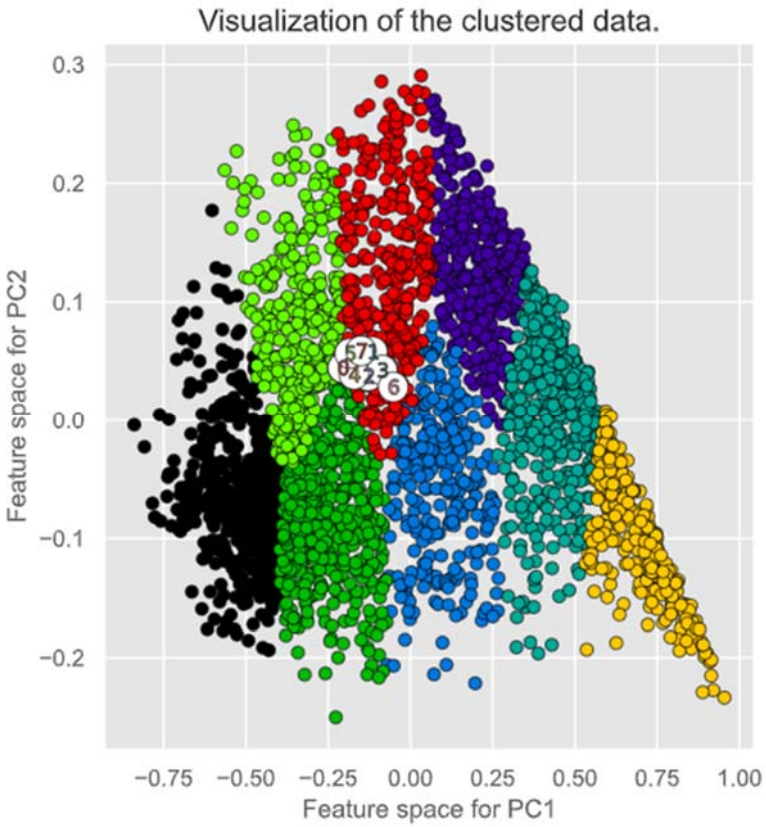

Figure 28. Clustering on data with 8 clusters for Solar PV generation

Figure 28 reveals the identification of eight distinct groups based on the values of PC1 and PC2, representing an appropriate clustering solution. The visual representation of these clusters illustrates a clear division that corresponds to the intensity of solar radiation and the occurrence of solar radiation during the day.

This clustering analysis enables the identification of groups exhibiting similar patterns of solar energy generation. By visually examining the clusters, we can discern variations in solar radiation levels and the temporal distribution of sunlight, providing valuable insights into the relationship between solar radiation intensity and its temporal dynamics in relation to energy generation.

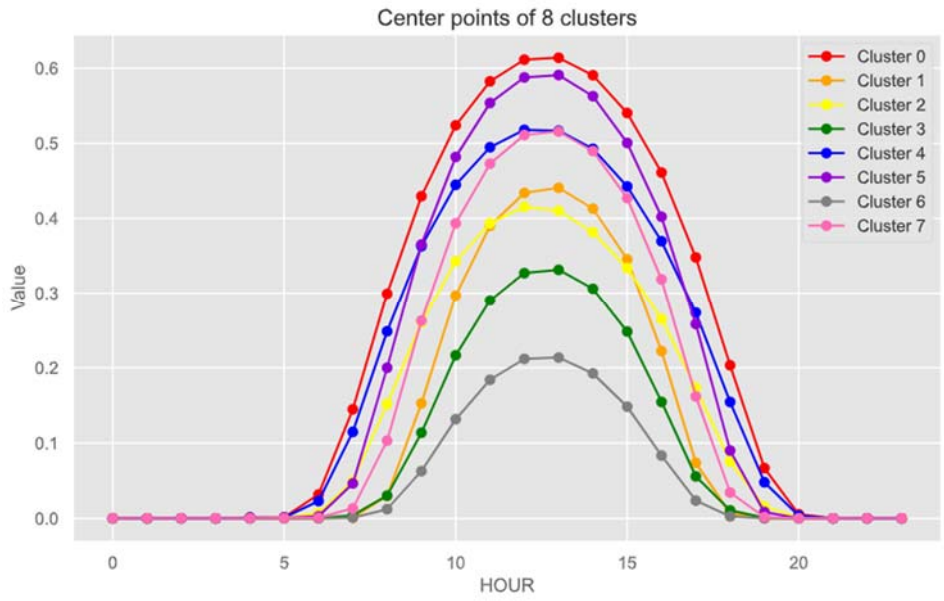

Figure 29. Centers of the 8 clusters for Solar PV generation

Figure 29 displays the average values for each cluster, i.e. the centers or prototypes of the clusters. The blue and pink curves (Cluster 7) exhibit a similar total area, indicating comparable levels of solar radiation. However, both curves display distinct characteristics in terms of positive and negative skewness. This aligns with the earlier description of the PC3 component, highlighting the role of clouds in modulating solar radiation levels. The yellow (Cluster 2) and orange curves (Cluster 1) also demonstrate a similar pattern, further supporting the impact of cloud cover on solar radiation variability.

The red curve (Cluster 0) represents days with early sunrise and late sunset, resulting in the highest solar radiation levels (i.e., summer days). Moreover, these days exhibit minimal cloud interference, leading to a consistent measurement of solar radiation throughout the day. On the other hand, the green curve (Cluster 3) and gray curve (Cluster 6) correspond to days with relatively late sunrise and early sunset, indicating a shorter duration of sunlight (i.e. winter days). These days are more susceptible to cloud effects, influencing the observed solar radiation patterns.

This analysis provides valuable insights into the relationship between cloud cover, solar radiation, and the temporal dynamics of energy generation. Understanding these patterns contributes to a comprehensive understanding of solar power production and aids in the development of strategies for maximizing energy generation efficiency under varying atmospheric conditions.

4.1.2.2 Energy generation by provinces

In this chapter, an in-depth analysis is conducted on the hourly data of solar power generation energy across different provinces. This analysis aims to provide comprehensive insights into the temporal patterns and variations in solar energy generation at a regional level.

By examining this data, we can explore the dynamics of solar power generation within each province and identify any unique characteristics or trends that may emerge. This analysis will enable a deeper understanding of the factors influencing solar energy generation, such as geographical location and climate conditions.

Through this investigation, valuable insights can be gained regarding the temporal distribution of solar energy generation across different provinces. These insights will contribute to the development of more effective strategies for optimizing solar power generation, enhancing energy planning and management, and promoting sustainable energy practices at the regional level.

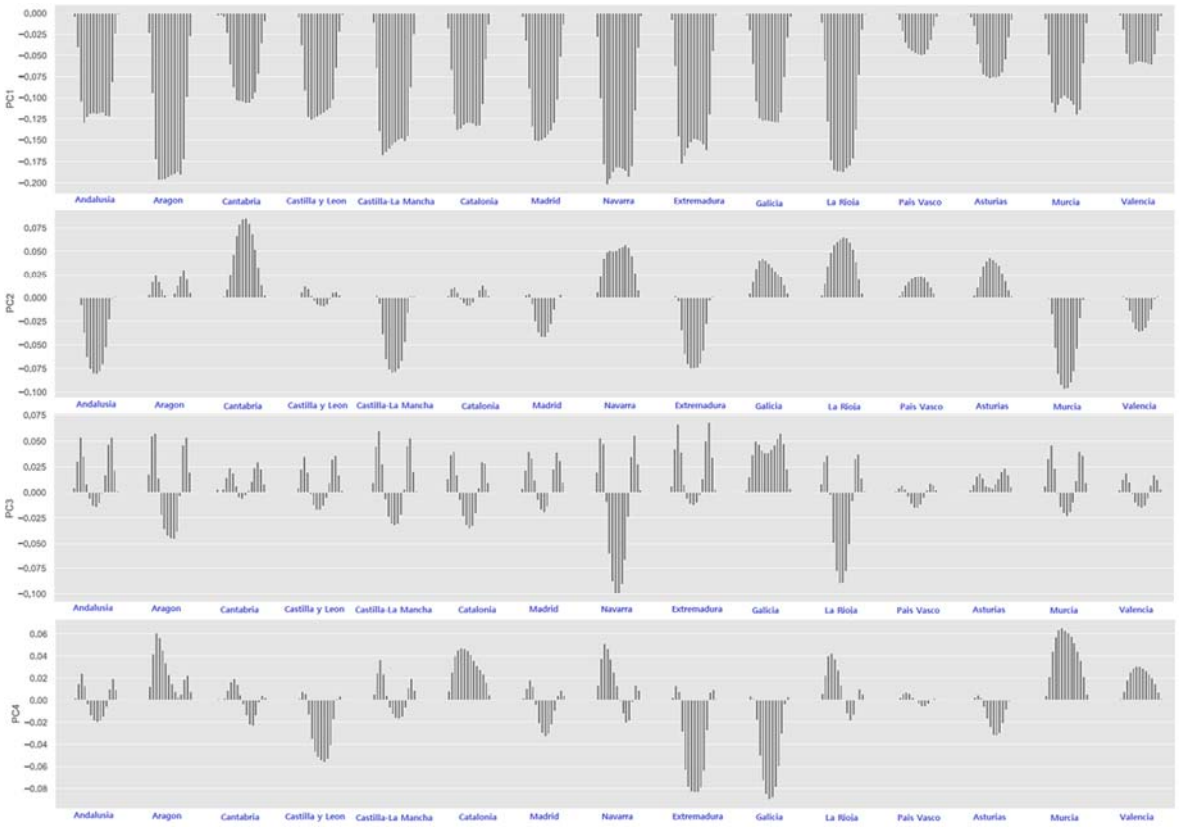

Figure 30. PCA Loadings for hourly Solar PV generation by provinces

Figure 30 shows the loadings of the first four principal components of the hourly solar PV generation of all provinces. PC1 reveals a distinct curved pattern representing the concentration of solar power generation during daytime hours across different provinces. This implies that solar energy production is predominantly influenced by the availability of sunlight during daylight hours, where some provinces contribute more than others to PC1. The magnitude of PC1 serves as an indicator of the intensity of solar radiation.

Furthermore, the weights assigned to each province in PC1 analysis demonstrate regional variations. Notably, Navarra and Aragon exhibit the highest weights, indicating their significant contribution to solar power generation.

A more negative value of PC1 indicates a higher intensity of solar radiation, representing favorable conditions for solar energy generation. Analyzing PC1 offers valuable insights

into the dependency of solar power generation on sunlight availability and intensity throughout the day.

PC2 provides a clearer representation of the concentration of solar power generation during the daytime, building upon the observations made in PC1 analysis. PC2 captures additional patterns and relationships in the solar energy data that are distinct from those captured by PC1, enhancing our understanding of solar power generation dynamics.

Notably, PC2 allows us to differentiate between the southern and northern provinces based on their solar power generation characteristics. A positive value of PC2 indicates stronger solar radiation in the northern region, while a negative value of PC2 indicates stronger solar radiation in the southern region.

This distinction highlights regional variations in solar energy generation and emphasizes the role of geographical factors in influencing solar radiation intensity. The analysis of PC2 offers valuable insights into the spatial distribution of solar power generation within the study area, contributing to a comprehensive understanding of the factors shaping solar energy production patterns across different provinces.

PC3 has a clear focus on energy generation during daytimes and specific areas, especially Navarra and La Rioja. As PC3 values become more negative, the concentration of solar power generation becomes increasingly centered around these two areas.

The PC4 component shows various characteristics. In some areas, solar radiation weighs the difference between day and afternoon. In addition, the focus was on specific areas: Galicia, Extremadura, Castilla y Leon and Murcia. These areas show distinct changes in solar power generation patterns, indicating a clear difference between the east and west of the analyzed areas. The more negative the PC4 is, the greater the emphasis on solar power in Galicia, Extremadura, and Castilla y Leon in the west.

	Exp_variance	cum_Exp_variance
PC1	0.623124	0.623124
PC2	0.071784	0.694908
PC3	0.051463	0.746371
PC4	0.047656	0.794027
PC5	0.027311	0.821337
PC6	0.018996	0.840334
PC7	0.018144	0.858477
PC8	0.015859	0.874336
PC9	0.009885	0.884221
PC10	0.009749	0.893970
PC11	0.008799	0.902770

Figure 31. Cumulative explained variance of PCA for Solar PV generation by provinces

According to Figure 31, the cumulative explained variance of the four main principal components (PC1, PC2, PC3, and PC4) exceeds 80%, indicating that these components capture a significant portion of the variability in the solar photovoltaic hourly energy data of all the provinces.

By considering these principal components, we account for a substantial amount of the underlying information and patterns in the data. This enables us to gain a comprehensive understanding of the factors influencing solar energy generation, such as solar radiation intensity, temporal distribution, regional variations, and the impact of clouds.

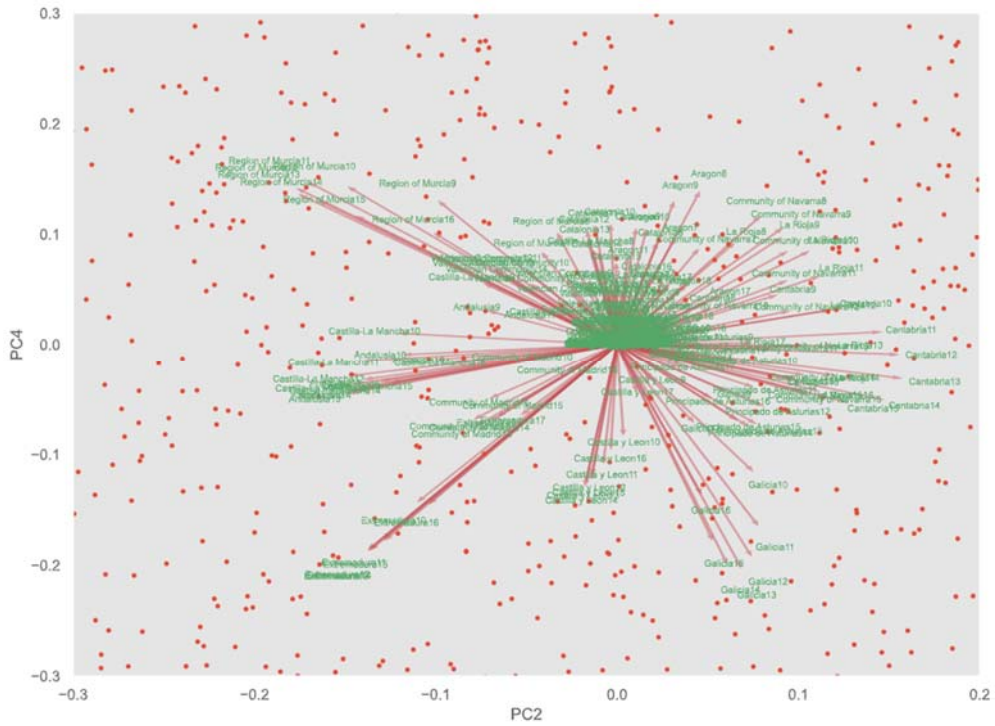

Figure 32. Biplot with PC2 and PC4 for Solar PV generation by provinces

Based on the analysis carried out, the biplot of PC2 and PC4 (Figure 32) in this context reveal additional insights into the spatial distribution and patterns of solar power generation within the study area.

In the case of PC2, the biplot helps visualize the differentiation between the southern and northern provinces based on their solar power generation characteristics. It provides a clearer representation of the concentration of solar power generation during the daytime and highlight the regional variations in solar energy generation. The biplot would show how different provinces contribute to PC2 and how they are positioned relative to each other based on solar radiation intensity.

Regarding PC4, the biplot lights on the focus regions of Galicia, Extremadura, Castilla y León, and Murcia. It illustrates the distinct solar power generation patterns in these regions and further emphasize the differentiation between the eastern and western parts of the analyzed area. The biplot would show the contributions of these regions to PC4 and their

spatial relationships, providing insights into the unique characteristics of solar energy production in different parts of the study area.

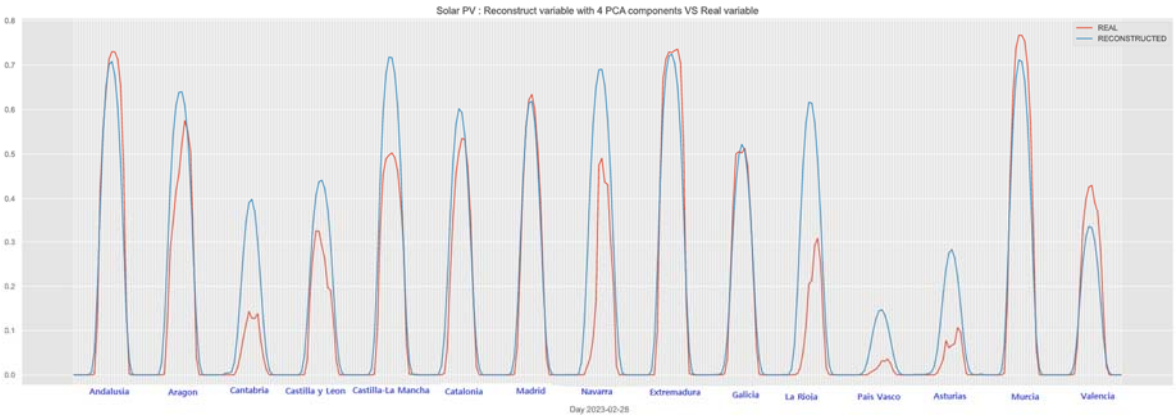

Figure 33. Reconstruct variables with the four main PCs for hourly Solar PV generation by provinces

Upon reconstructing the data using the principal components PC1, PC2, PC3, and PC4, we have observed a remarkable consistency and close alignment between the reconstructed data and the original dataset. Figure 33 shows the reconstruction of a particular day for all the provinces in on shot. This high degree of similarity highlights the robustness and efficacy of the principal component analysis (PCA) in capturing the underlying patterns and variability inherent in the solar photovoltaic energy data.

However, Cantabria, La Rioja, Pais Vasco, and Asturias do not align as closely with the reconstructed data compared to the other provinces. While the majority of provinces demonstrate a strong match between the reconstructed and actual data, these specific regions exhibit distinctive characteristics that deviate from the general trend.

Notably, the observed discrepancy in the reconstructed data for Cantabria, La Rioja, Pais Vasco, and Asturias appears to be closely linked to the installed capacity of solar PV in these provinces. It is intriguing to observe that these four provinces have relatively lower installed capacity for solar PV generation and display no significant increase over time. This suggests that the smaller installed capacities in these regions may render them more susceptible or less resilient to the influence of external variables.

This association between installed capacity and the observed deviations in the reconstructed data implies that the magnitude of installed capacity plays a critical role in shaping the dynamics and responses of solar energy generation to various factors. Provinces with larger installed capacities may possess a more robust infrastructure and resources, allowing them to better withstand and adapt to the impact of different variables, thereby exhibiting a stronger alignment between the reconstructed and actual data. This finding raises important considerations regarding the significance of installed capacity as a determining factor in the performance and reliability of solar energy generation systems.

Let us look at the graph of how the characteristics of each PCA appear in real solar energy data.

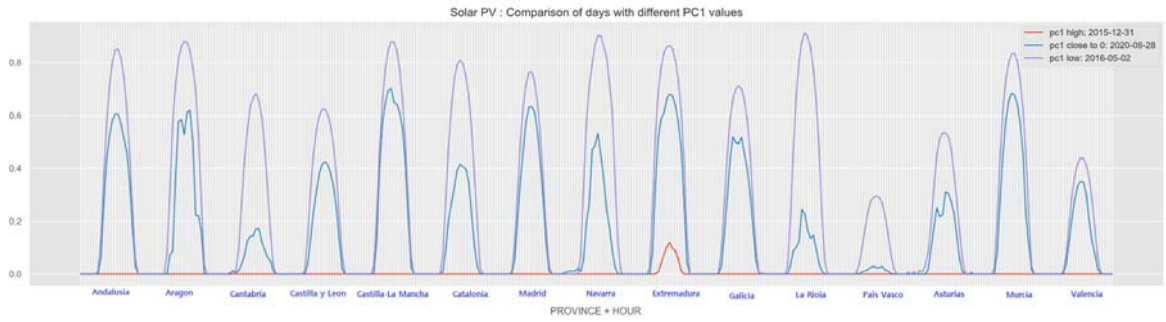

Figure 34. Three representative days according to PC1 for Solar PV generation by provinces

Figure 34 shows three distinct curves corresponding to three different days, selected via PC1, showing regional differences. The highest purple curve in all provinces corresponds to the lower value of PC1, indicating a higher amount of solar radiation. In particular, Pais Vasco has very low values in all respects. Conversely, the lowest red curve is observed when the value of PC1 is high, indicating that there is no amount of solar radiation in Spain. The mid-blue curve represents the average condition when PC1 is close to zero.

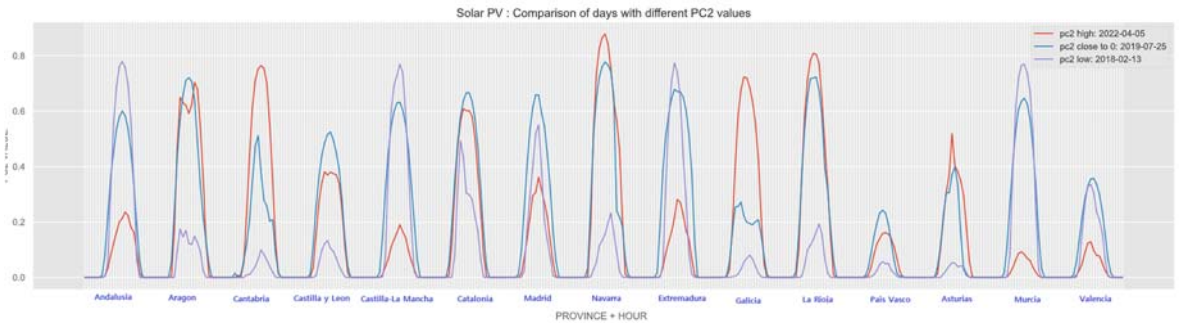

Figure 35. Three representative days according to PC2 values for Solar PV generation by provinces

Figure 35 shows the impact of the value of PC2 in the solar PV generation spatial patterns. In particular, varying PC2 three distinct and complex curves has been selected, providing insight into solar radiation patterns throughout the day, particularly with respect to geographical characteristics in the southern and northern regions. The high purple curve corresponds to the day with the lower value of PC2, indicating that the amount of solar radiation in the South is higher. It is consistent with the above analysis that a negative PC2 indicates energy production in the southern region, and a positive PC2 indicates energy production in the northern region.

In addition, in the northern region of Pais Vasco, it is observed that the amount of solar radiation is higher when the PC2 value is higher (red curve), but the overall energy production is lower than that of other provinces.

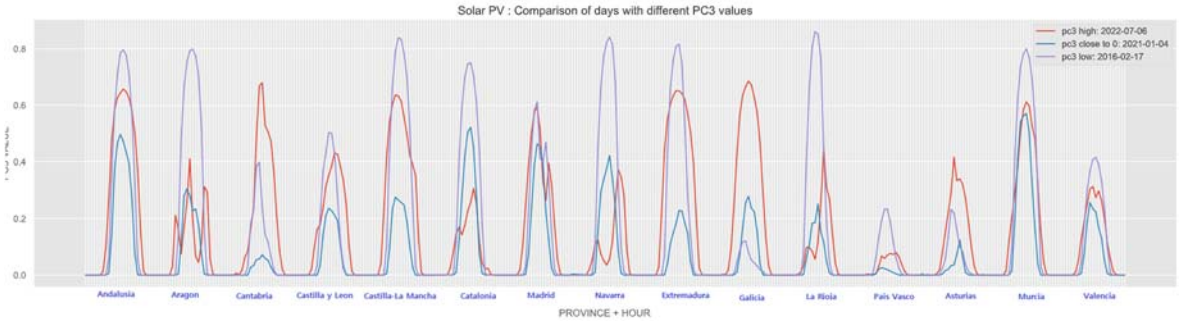

Figure 36. Three representative days according to PC3 for Solar PV generation by provinces

In Figure 36 we have selected three different days corresponding to different values of PC3. The graph shows that energy production during the day and energy production in Navarra and La Rioja regions are high when PC3 values are low (purple curve). If PC3 is high (red

curve), then it corresponds to summer days with many hours of daylight. In addition, it shows quite a fluctuating energy fish in several regions.

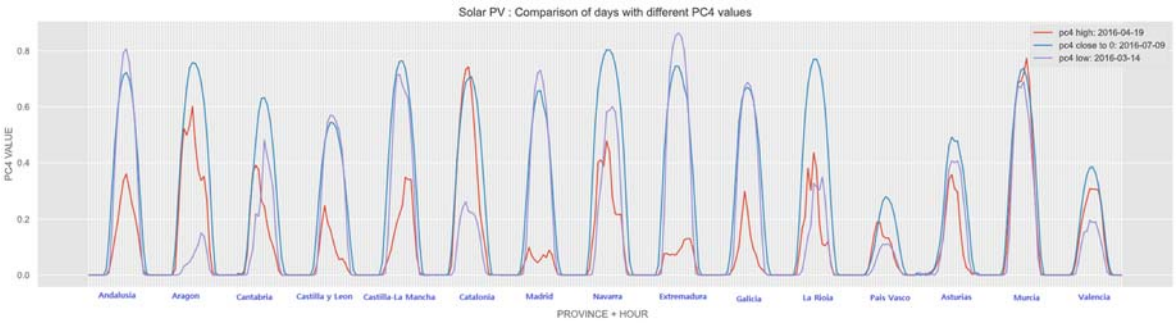

Figure 37. Three representative days according to PC4 for Solar PV generation by provinces

The fourth principal component (PC4) reveals additional insights into the solar energy data (see Figure 37), albeit with less distinct characteristics compared to other main components. The red curve in the PC4 graph depicts the variation in solar radiation production between daytime and afternoon hours. It demonstrates the temporal dynamics of solar energy generation, particularly the differences in output during these specific time intervals.

Furthermore, the intensity of solar radiation production in the western regions can be discerned from the higher values of the purple curve. This suggests a geographical influence on solar energy generation, with the western areas exhibiting specific patterns and characteristics.

Murcia, in particular, exhibits three remarkably similar curves in the PC4 analysis. This observation suggests a distinct solar radiation pattern specific to the Murcia region. The close resemblance of the curves indicates a consistent solar energy generation pattern within the area, reflecting the influence of local factors such as climate, topography, or regional solar energy policies.

The PCA analysis has provided valuable insights and initial ideas for clustering the solar energy data. By examining the principal components, particularly PC1, PC2, PC3 and PC4, we have gained a deeper understanding of the underlying patterns and relationships within the dataset. These principal components capture the major sources of variability and key

factors influencing solar energy generation. These preliminary findings serve as a foundation for clustering the data and uncovering more refined groups or clusters based on shared characteristics and patterns of solar energy generation.

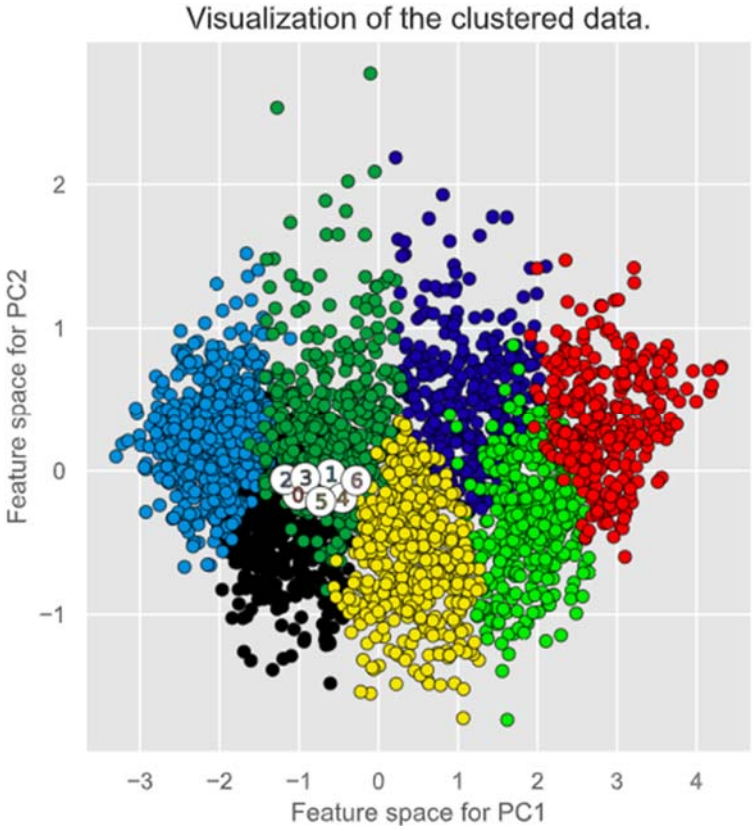

Figure 38. Clustering on data with 7 clusters for Solar PV generation by provinces

Figure 38 presents a comprehensive clustering solution by identifying seven distinct groups based on the values of PC1 and PC2. The visualization of these clusters provides valuable insights into the spatial distribution and temporal characteristics of solar radiation intensity, highlighting the differences between the southern and northern regions.

It is important to note that the boundaries between clusters in this analysis are relatively less certain compared to the clustering analysis conducted in section 4.1.2.1. This uncertainty arises due to the breakdown of data by both region and time, which introduces additional complexities and potential overlapping patterns. Nonetheless, the identified clusters offer

valuable information about the variations in solar radiation intensity and patterns within the study area, contributing to a more comprehensive understanding of solar energy dynamics.

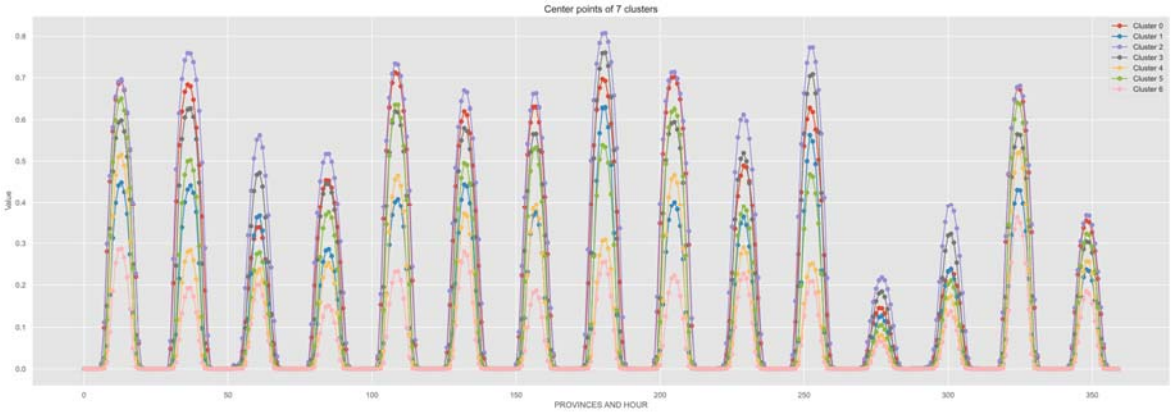

Figure 39. Center points of seven clusters for hourly Solar PV generation by provinces

Figure 39 illustrates the average values for each cluster derived from the cluster analysis. Among the identified clusters, the purple curve (Cluster 2) represents days with high solar radiation across all regions, indicating optimal conditions for solar energy generation in Spain. In contrast, the pink curve (Cluster 6) corresponds to days with the lowest solar radiation levels across all regions.

Interestingly, the province of Pais Vasco consistently exhibits significantly lower values across all clusters, suggesting a lower potential for solar radiation and solar energy generation in this region. This finding highlights the importance of considering regional variations in solar radiation intensity when analyzing and planning solar energy projects.

The figure provides valuable insights into the overall amount of solar radiation, which emerges as a critical factor in this cluster analysis. Understanding the total solar radiation potential in different clusters can aid making informed decisions regarding solar energy infrastructure development and resource allocation.

Chapter 5. WIND ENERGY GENERATION

5.1 DAILY DATA

In this chapter, an analysis of daily wind power generation data across various provinces will be undertaken. The primary objective of this study is to gain a comprehensive understanding of the temporal dynamics and regional variations in wind power production. The dataset employed for this analysis comprises 'DATE' rows and 'PROVINCES' columns, enabling us to explore and examine the intricate patterns and fluctuations in wind power generation at a localized level.

5.1.1 ENERGY GENERATION BY PROVINCES

The correlation between provinces will serve as a crucial aspect of our analysis, allowing us to identify relationships and dependencies among them in terms of wind power generation. By scrutinizing the correlation matrix, we can assess the degree of association between different provinces and determine the similarities or differences in their wind power generation patterns. This information will be instrumental in understanding the interconnectedness and potential influences between provinces concerning wind energy production.

geoname	Andalusia	Aragon	Cantabria	Castilla y Leon	Castilla-La Mancha	Catalonia	Community of Navarra	Extremadura	Galicia	La Rioja	Pais Vasco	Principado de Asturias	Region of Murcia	Valencian Community
Andalusia	1.000000	0.038997	0.149581	0.196022	0.418607	0.082785	0.028761	0.614682	0.036903	0.128400	0.340971	0.069736	0.240636	0.174750
Aragon	0.038997	1.000000	0.166839	0.463212	0.447860	0.731325	0.842946	0.114596	0.172651	0.892941	0.428917	0.164582	0.503758	0.657741
Cantabria	0.149581	0.166839	1.000000	0.530965	0.401885	0.200327	0.128934	0.360372	0.501388	0.315039	0.726743	0.463210	0.273905	0.333730
Castilla y Leon	0.196022	0.463212	0.530965	1.000000	0.698023	0.415762	0.395392	0.345431	0.716539	0.532316	0.702095	0.738570	0.538238	0.558175
Castilla-La Mancha	0.418607	0.447860	0.401885	0.698023	1.000000	0.582017	0.224156	0.479627	0.349406	0.469365	0.708917	0.497847	0.875325	0.845617
Catalonia	0.082785	0.731325	0.200327	0.415762	0.582017	1.000000	0.456077	0.109926	0.153797	0.629951	0.451684	0.292082	0.650455	0.801848
Community of Navarra	0.028761	0.842946	0.128934	0.395392	0.224156	0.456077	1.000000	0.132932	0.124659	0.858346	0.363550	0.040554	0.223836	0.339344
Extremadura	0.614682	0.114596	0.360372	0.345431	0.479627	0.109926	0.132932	1.000000	0.142254	0.251996	0.600817	0.162481	0.328867	0.248436
Galicia	0.036903	0.172651	0.501388	0.716539	0.349406	0.153797	0.124659	0.142254	1.000000	0.220077	0.424612	0.845151	0.271016	0.276408
La Rioja	0.128400	0.892941	0.315039	0.532316	0.469365	0.629951	0.858346	0.251996	0.220077	1.000000	0.560598	0.196760	0.475393	0.614099
Pais Vasco	0.340971	0.428917	0.726743	0.702095	0.708917	0.451684	0.363550	0.600817	0.424612	0.560598	1.000000	0.484686	0.547370	0.598455
Principado de Asturias	0.069736	0.164582	0.463210	0.738570	0.497847	0.292082	0.040554	0.162481	0.845151	0.196760	0.484686	1.000000	0.444247	0.441443
Region of Murcia	0.240636	0.503758	0.273905	0.538238	0.875325	0.650455	0.223836	0.328867	0.271016	0.475393	0.547370	0.444247	1.000000	0.885127
Valencian Community	0.174750	0.657741	0.333730	0.558175	0.845617	0.801848	0.339344	0.248436	0.276408	0.614099	0.598455	0.441443	0.885127	1.000000

Figure 40. Regional Correlation of Wind

Figure 40 shows the correlation matrix of daily wind generation by provinces. The analysis reveals positive correlations among all pairs of provinces. Specifically, Aragon and Rioja exhibit a correlation coefficient of 0.892941, Murcia and Valencia demonstrate a correlation of 0.885127, Castilla-La Mancha and Murcia display a correlation of 0.875325, Navarra and Rioja show a correlation of 0.858346, Castilla-La Mancha and Valencia exhibit a correlation of 0.845617, and Galicia and Asturias demonstrate a correlation of 0.845151. Additionally, there is a positive correlation of 0.842946 between Aragon and Navarra, indicating a significant association in terms of wind power generation.

These pairs of provinces exhibit geographical proximity on the map. Notably, the provinces of Navarra, La Rioja, and Aragon form a triangle, as do Castilla-La Mancha, Valencia, and Murcia. This observation highlights the spatial clustering of these provinces and suggests a potential relationship between their wind power generation dynamics.

A dendrogram analysis plays a crucial role in understanding the underlying structure of the data and guiding subsequent analyses. It visually represents the similarities and dissimilarities between data points or variables in a hierarchical manner. By examining the dendrogram, distinct clusters or groups within the data can be identified. This analysis helps in determining the optimal number of clusters or groups and provides insights into the dataset's inherent structure, which aids in variable selection, data transformation, and interpretation of outcomes in subsequent analyses such as PCA.

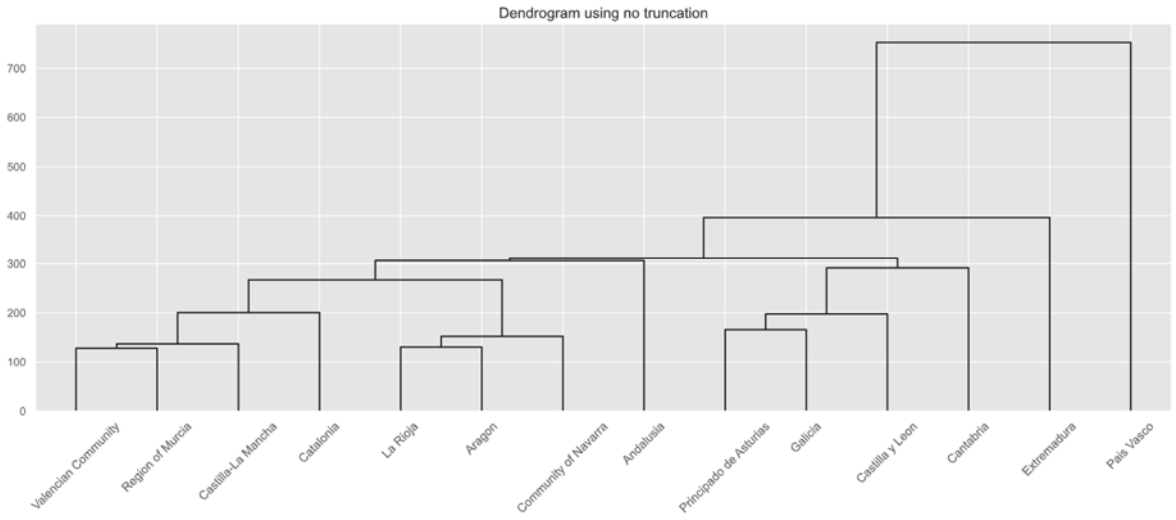

Figure 41. Dendrogram for wind generation by province with daily data

The dendrogram of Figure 41 further supports and confirms the findings observed in the correlation analysis mentioned earlier. It provides visual evidence of the clustering patterns and relationships between the variables, validating the similarities and dissimilarities identified in the correlation analysis.

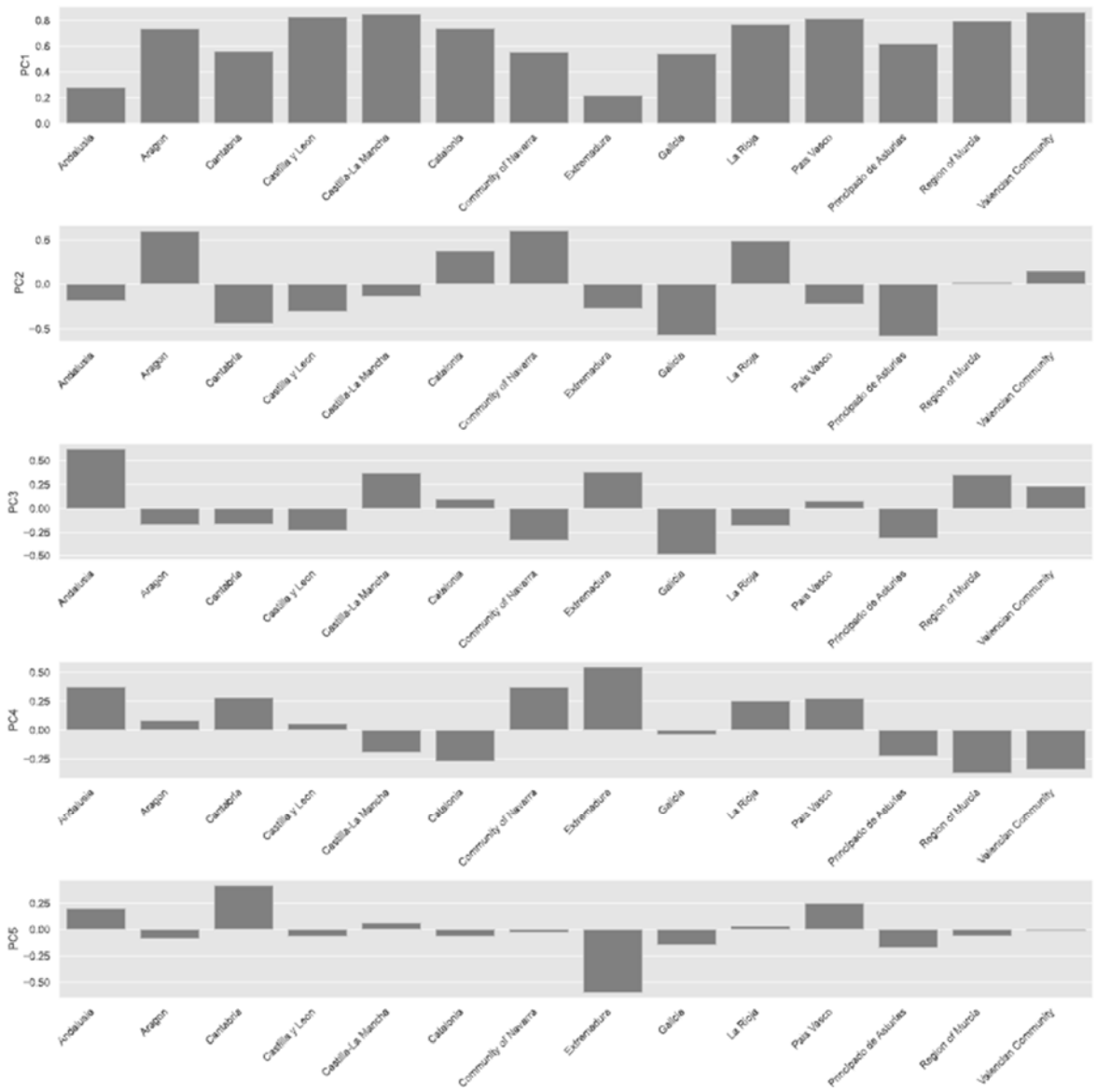

Figure 42. PCA Loadings for wind generation by provinces

Figure 42 shows the loadings of the first five main principal components. The first principal component (PC1) represents a weighted average of the wind generation, where Valencia and Castilla-La Mancha contribute the highest proportion, while Andalusia and Extremadura contribute the lowest proportion. This indicates that Valencia and Castilla-La Mancha play a significant role in determining the overall trend of solar power generation across the provinces, while Andalusia and Extremadura have relatively lesser influence in this regard.

The second principal component (PC2) in the context of wind energy represents the division of the northern regions into eastern and western parts. A positive value of PC2 indicates a stronger influence from the eastern regions, while a negative value indicates a stronger influence from the western region. This component highlights the regional variations in wind power generation between the eastern and western provinces within the northern regions.

The third principal component (PC3) in the context of wind energy analysis captures the north-south characteristics of the dataset. It assigns positive weights to the southern regions and negative weights to the northern regions, with particular emphasis on Andalusia in the south. This component highlights the regional disparities in wind energy generation between the southern and northern provinces. It shows the significance of Andalusia as a prominent contributor to the overall wind energy production in the southern regions, while also reflecting the distinctions in wind energy patterns and potential between the northern and southern parts of the analyzed area.

The fourth principal component (PC4) analysis reveals a distinct focus on the coastal areas and their relation to wind energy generation. PC4 assigns positive weights to both the southern and northern coastal regions, indicating their significance in terms of wind energy production. Conversely, negative weights are attributed to the eastern coastal regions, suggesting a relatively lower influence on wind energy patterns. This component emphasizes the role of coastal geography and its interaction with wind dynamics, highlighting the potential impact of sea proximity on wind energy generation in the analyzed areas.

The fifth principal component (PC5) analysis unveils a notable emphasis on the regions of Extremadura and Cantabria in relation to wind energy patterns. PC5 assigns a positive weight to Extremadura. On the other hand, a negative weight is attributed to Cantabria.

Overall, the PCA analysis provides a comprehensive understanding of the regional dynamics and variations in solar and wind energy generation, shedding light on the key provinces and factors shaping the energy landscape.

	Exp_variance	cum_Exp_variance
PC1	0.467727	0.467727
PC2	0.161232	0.628958
PC3	0.103392	0.732350
PC4	0.087389	0.819739
PC5	0.051343	0.871082
PC6	0.049510	0.920593
PC7	0.025591	0.946184
PC8	0.015859	0.962043
PC9	0.010420	0.972462
PC10	0.007824	0.980286
PC11	0.006949	0.987235
PC12	0.005477	0.992711
PC13	0.004457	0.997168
PC14	0.002832	1.000000

Figure 43. Cumulative explained variance of PCA for wind generation

According to Figure 43, the cumulative explained variance of the three principal components (PC1, PC2, PC3, PC4 and PC5) for wind energy is 87%. Five principal components are utilized in this analysis, exceeding the number used for solar energy analysis, due to wind power is harder to explain than solar power. Although slightly lower than solar energy, this cumulative explained variance still demonstrates the capability of these components in capturing a significant portion of the underlying variability in the wind energy dataset. The considerable cumulative explained variance highlights the usefulness of the PCA approach in revealing and representing important patterns and trends within wind energy generation. By accounting for a substantial portion of the variability, PC1, PC2, PC3, PC4 and PC5 contribute significantly to our understanding of wind energy dynamics and aid in identifying the key factors influencing wind energy generation across the analyzed provinces.

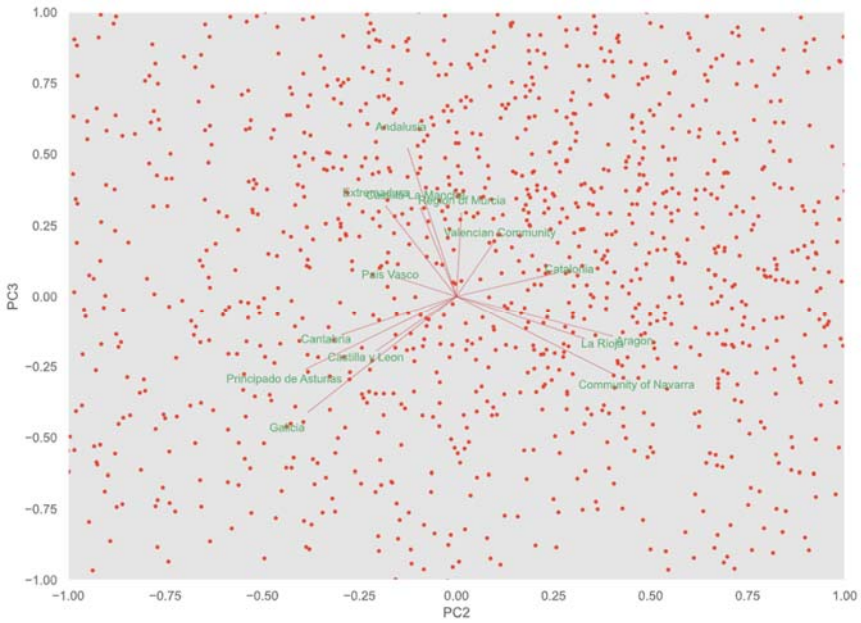

Figure 44. Biplot with PC2 and PC3 of wind generation by provinces

In the biplot representation of PC2 and PC3 (Figure 44), a symmetrical pattern emerges across the Spanish map with Galicia, Aragon, and Andalusia serving as pivotal reference points. This longitudinal symmetry indicates a consistent trend or relationship between these provinces in terms of their wind energy characteristics. The positioning of these provinces in the biplot indicates their shared attributes or similarities in relation to PC2 and PC3, highlighting the potential influence of geographic and climatic factors in shaping wind energy patterns across different regions of Spain.

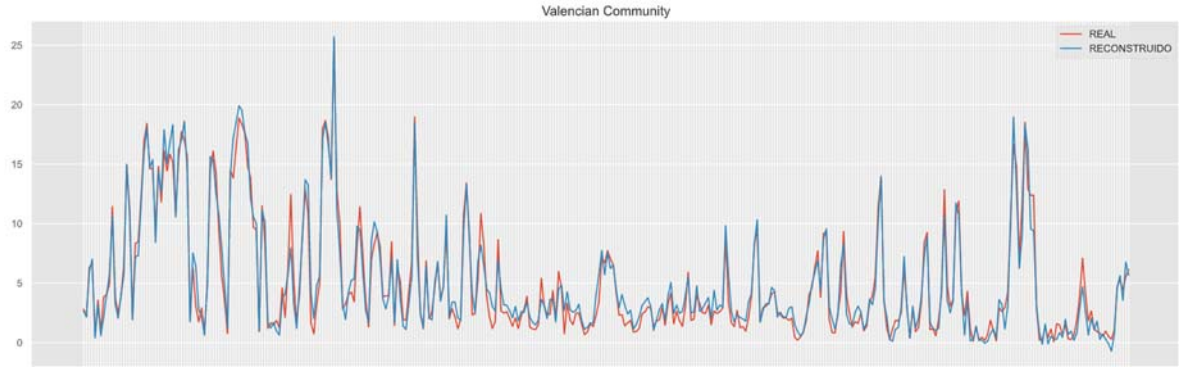

Figure 45. Reconstructed wind generation of Valencia with 5-PCs

Figure 45 illustrates the wind power generation data for Valencia throughout the entire year of 2015. The close correspondence between the reconstructed data and the actual data provides compelling evidence of the effectiveness of PCA in capturing and reproducing the underlying patterns and variability in wind power generation. This finding underscores the robustness and reliability of the PCA approach in unraveling significant insights into the dynamics of wind energy generation.

This examination of daily wind generation data provides invaluable insights into local patterns, interdependencies, and dynamic fluctuations, establishing a solid foundation for comprehensive analyses. By rigorously analyzing the daily wind power generation data, our study aims to enrich the existing knowledge in the realm of renewable energy, offering valuable insights into the temporal and regional dynamics of wind energy production.

5.2 HOURLY DATA

5.2.1 ENERGY GENERATION

In this section, the hourly data of wind energy is analyzed. PCA has been employed to identify and understand the underlying patterns and relationships in the wind energy data. The data frame used consists of the 'DATE' row and the 'HOUR' column.

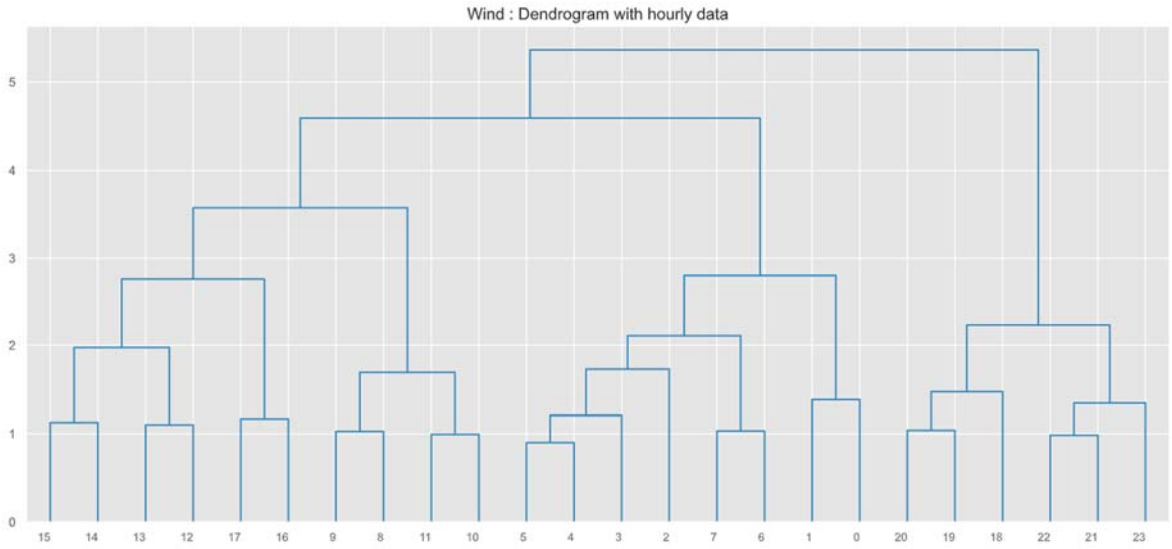

Figure 46. Dendrogram by hour for wind generation

The dendrogram shown in Figure 46 reveals a discernible differentiation between periods characterized by the presence or absence of solar radiation. This observation indicates a pronounced influence of solar radiation on wind energy dynamics. The distinct clustering patterns observed in the dendrogram highlight the interconnected relationship between solar radiation and wind energy generation. This finding underscores the significance of solar irradiance as a driving factor impacting the variability and availability of wind resources. By elucidating the connection between solar radiation and wind energy, our analysis contributes to a deeper understanding of the synergistic relationship between these renewable energy sources.

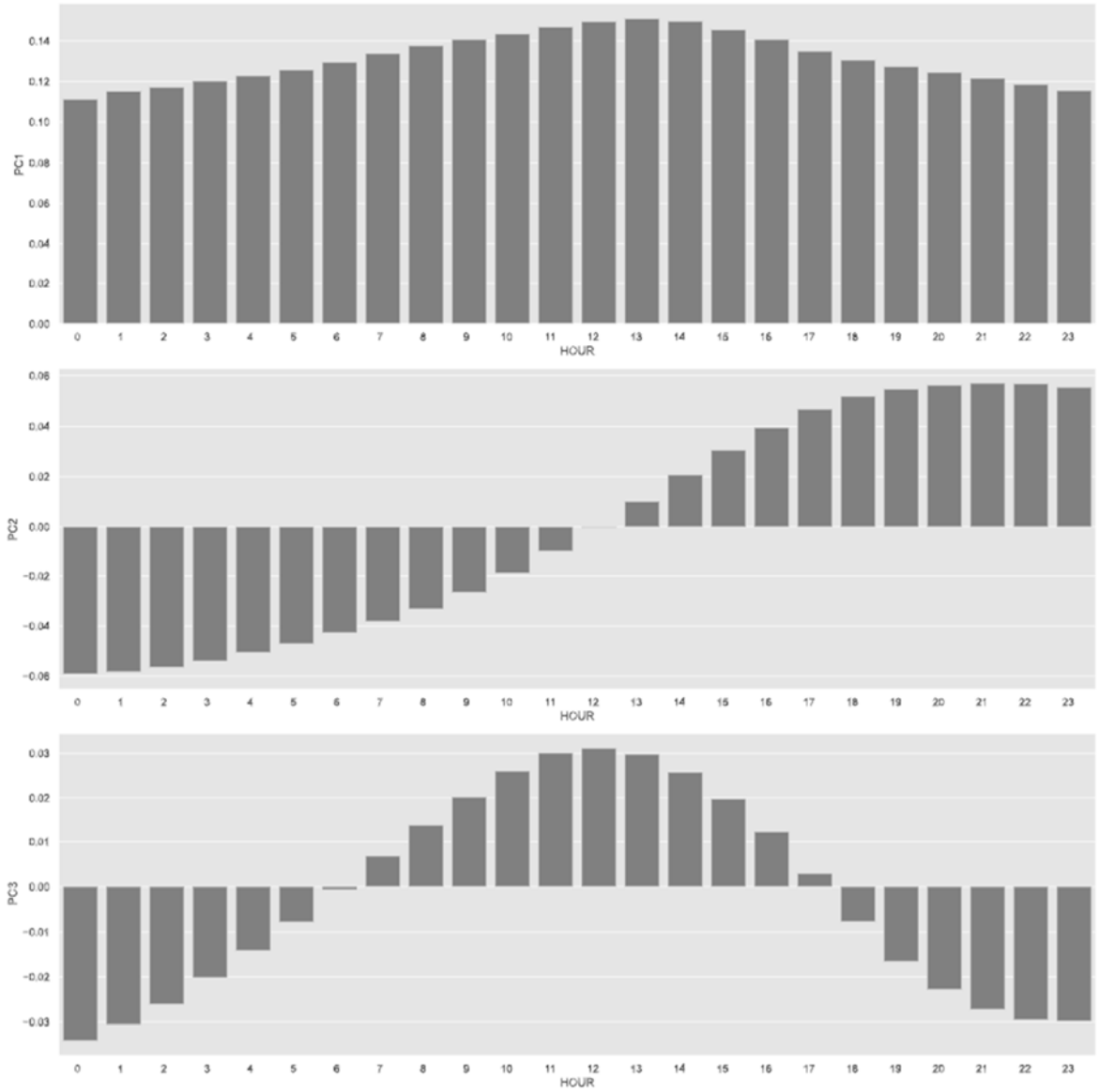

Figure 47. PCA Loadings for wind generation

Figure 47 shows the loadings of the first three principal components for the hourly wind generation. The first principal component (PC1) characterizes the weighted average of wind power generation throughout the day, with higher weights in the center of the day. A higher positive value of PC1 indicates a stronger wind on that particular day, implying a greater generation of wind energy.

The second principal component (PC2) accounts for the remaining main variability in the wind energy data that is not explained by PC1. It captures additional patterns and

relationships that are distinct from those represented by PC1. A positive value of PC2 suggests a higher wind energy generation occurring after 12 o'clock. Conversely, a negative value of PC2 suggests a stronger wind energy generation occurring before 12 o'clock.

The third principal component (PC3) captures the relationship between wind energy generation and solar energy. PC3 provides valuable insights into the temporal variations in sunlight intensity throughout the day and its impact on wind power generation. A positive value of PC3 signifies that stronger wind energy is generated during the daytime when solar radiation is at its peak. Conversely, negative values of PC3 suggest that stronger wind energy is generated during other times when solar radiation is relatively lower or null. By analyzing PC3, we gain a deeper understanding of the intricate interplay between solar energy availability and wind power generation. This knowledge enhances our ability to optimize renewable energy systems and develop strategies for maximizing the utilization of both solar and wind resources.

	Exp_variance	cum_Exp_variance	PC13	0.000306	0.998714
PC1	0.854762	0.854762	PC14	0.000248	0.998962
PC2	0.095852	0.950614	PC15	0.000192	0.999154
PC3	0.024749	0.975363	PC16	0.000159	0.999312
PC4	0.010970	0.986333	PC17	0.000131	0.999443
PC5	0.004555	0.990888	PC18	0.000113	0.999557
PC6	0.002762	0.993650	PC19	0.000096	0.999653
PC7	0.001441	0.995091	PC20	0.000083	0.999735
PC8	0.000980	0.996071	PC21	0.000078	0.999813
PC9	0.000797	0.996868	PC22	0.000069	0.999883
PC10	0.000614	0.997482	PC23	0.000062	0.999945
PC11	0.000516	0.997998	PC24	0.000055	1.000000
PC12	0.000410	0.998408			

Figure 48. Cumulative explained variance of PCA for wind generation

The cumulative explained variance of the three main principal components (PC1, PC2, and PC3) for wind generation exceeds 97% (see Figure 48). This indicates that these three components collectively capture a significant portion of the overall variability present in the

wind energy data. The high cumulative explained variance suggests that these principal components effectively represent the essential patterns and relationships within the dataset.

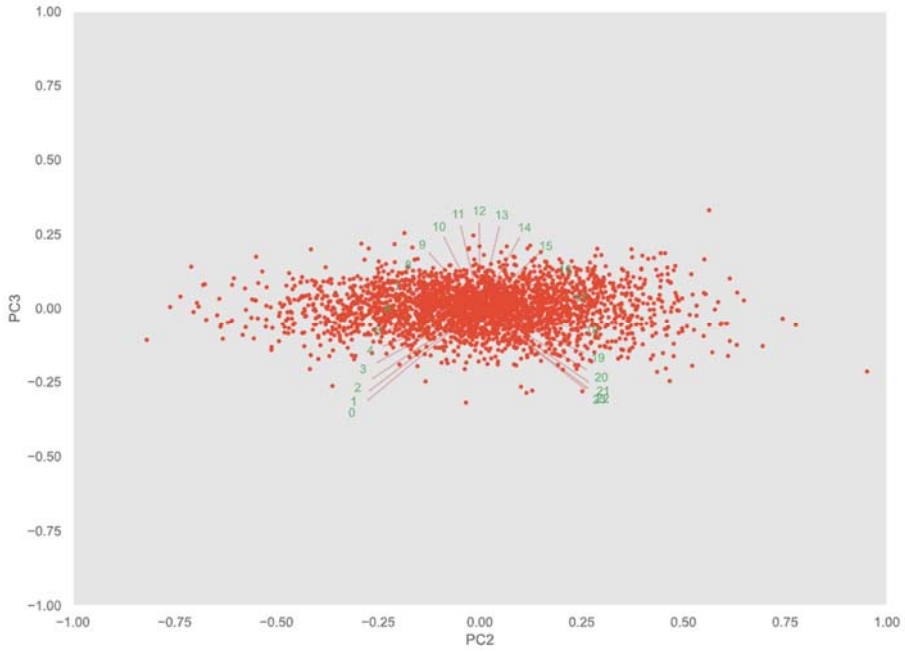

Figure 49. Biplot with PC2 and PC3 for wind generation

Based on the analysis conducted in PCA, the biplot shown in Figure 49 with PC2 and PC3 provides further insights into the relationship between wind energy generation and solar energy. The curve observed in the biplot starts at 0 and ends at 23, representing the 24-hour cycle of a day. The curve showcases the temporal variations in wind energy generation, which are influenced by the availability of solar radiation throughout the day.

The position of data points along the curve indicates the corresponding time of day when wind energy generation is relatively stronger or weaker. Data points located towards the beginning of the curve (closer to 0) represent periods when wind energy generation is higher before 12 o'clock, while points towards the end of the curve (closer to 23) represent periods of increased wind energy generation after 12 o'clock. This pattern suggests that wind energy generation tends to align with the temporal variations in solar radiation intensity, with stronger wind energy coinciding with periods of higher solar radiation. By analyzing the

biplot of PC2 and PC3, we gain a comprehensive understanding of the temporal dynamics and dependencies between solar energy and wind energy generation.

Let us look at the graph of how the characteristics of each PCA appear in real solar energy data.

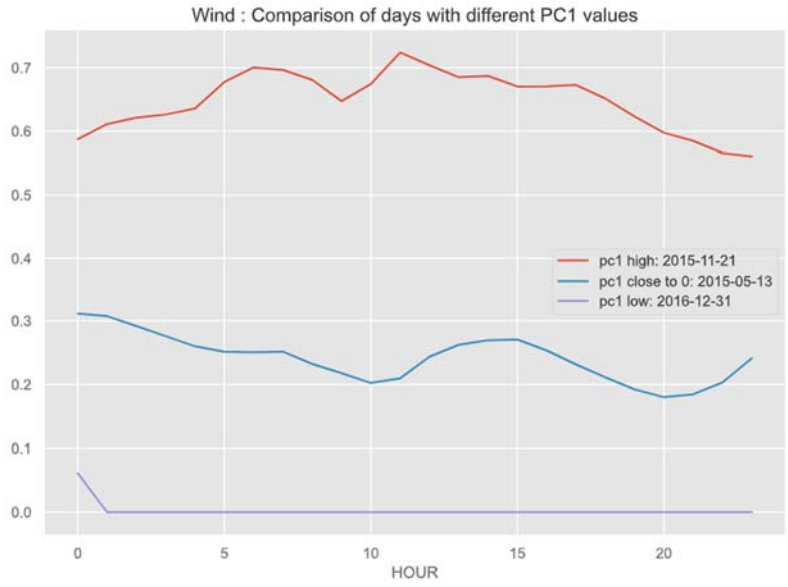

Figure 50. Three days of hourly wind generation selected according to PC1

Figure 50 shows the hourly wind generation of three distinct days, selected according to representative values of PC1. The highest red curve corresponds to the high value of PC1, indicating that the amount of wind energy generated is higher. Conversely, the lowest purple curve is observed when the value of PC1 is low, which means there is no generation of wind energy that day. The mid-curve represents the average condition when PC1 is close to zero.

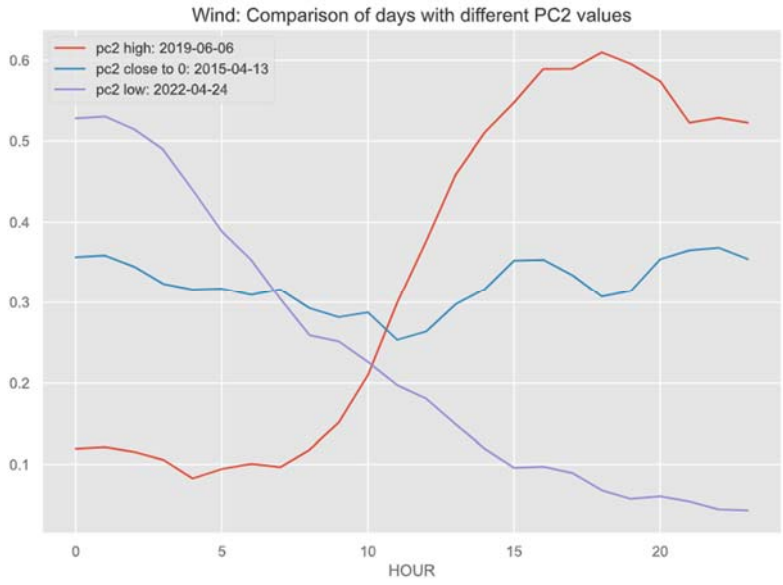

Figure 51. Three days of hourly wind generation selected according to PC2

Figure 51 illustrates three distinct curves, each representing different aspects of wind energy generation. The red curve corresponds to the characteristics of PC2 as previously discussed, indicating a higher amount of wind energy generated after 12 o'clock. Conversely, the lowest purple curve is observed when the value of PC2 is low, indicating a relatively higher amount of wind energy generated before 12 o'clock. The blue curve, corresponding to PC2 is close to zero, aligns with the average curve consisting of a more or less constant wind energy generation during the whole day.

The distinctive shapes and patterns of these curves provide valuable insights into the temporal dynamics of wind energy generation. The red curve's upward trend after 12 o'clock suggests an increase in wind energy production during the latter part of the day. Conversely, the lowest purple curve indicates a relatively higher level of wind energy generation during other time periods.

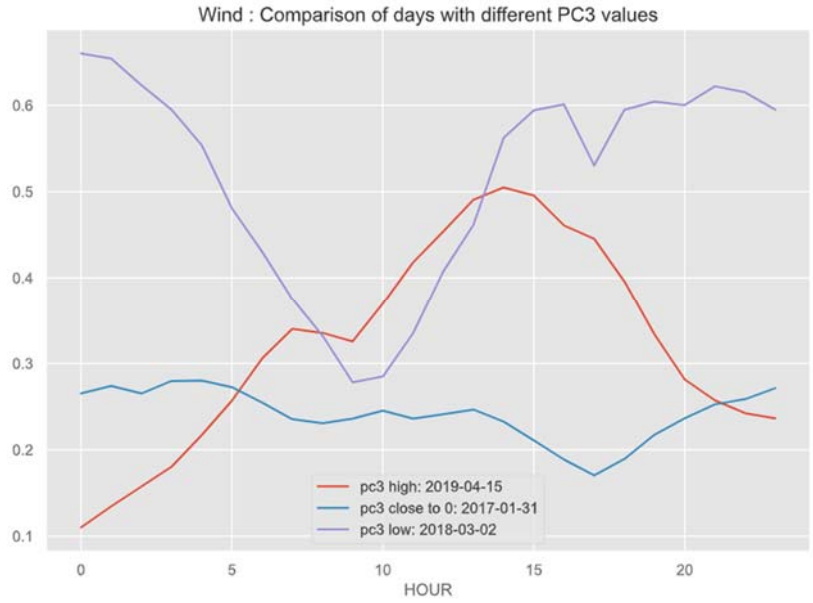

Figure 52. Three days of hourly wind generation selected according to PC3

Figure 52 shows that the red and purple curves correspond to more or less symmetrical patterns. This means that higher PC3, which is red curve generates more wind energy during the daylight hours of the day, and lower PC3, which is purple curve, generates more wind energy at other times.

By integrating the insights gained from the dendrogram analysis and PCA, we have successfully explored and reduced the dimensionality of the dataset, covering two crucial aspects of data analysis. The dendrogram analysis has offered us a deeper understanding of the hierarchical structure and potential clustering patterns within the dataset. On the other hand, the PCA analysis has enabled us to extract the most significant information and identify the key components responsible for driving the variability in the data. The combination of these techniques has provided us with a comprehensive perspective on the dataset and facilitated an initial estimation of the optimal number of clusters.

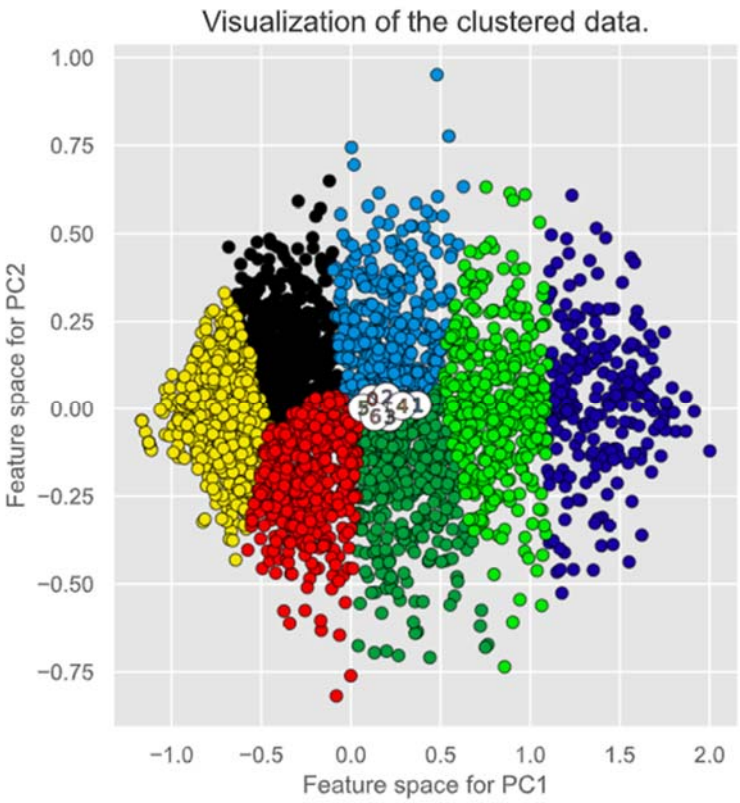

Figure 53. Clustering with 7 clusters for wind generation

Figure 53 identifies seven different groups based on the values of PC1 and PC2, indicating the appropriate clustering solution. The visual representation of these clusters shows a clear distinction between the generation intensity of wind energy and the generation of wind energy before and after 12 o'clock.

This clustering analysis facilitates the identification of distinct groups that exhibit similar patterns of wind energy generation. By applying clustering techniques to the wind energy data, inherent similarities and common characteristics in terms of wind energy generation patterns can be uncovered, leading to the grouping of data points.

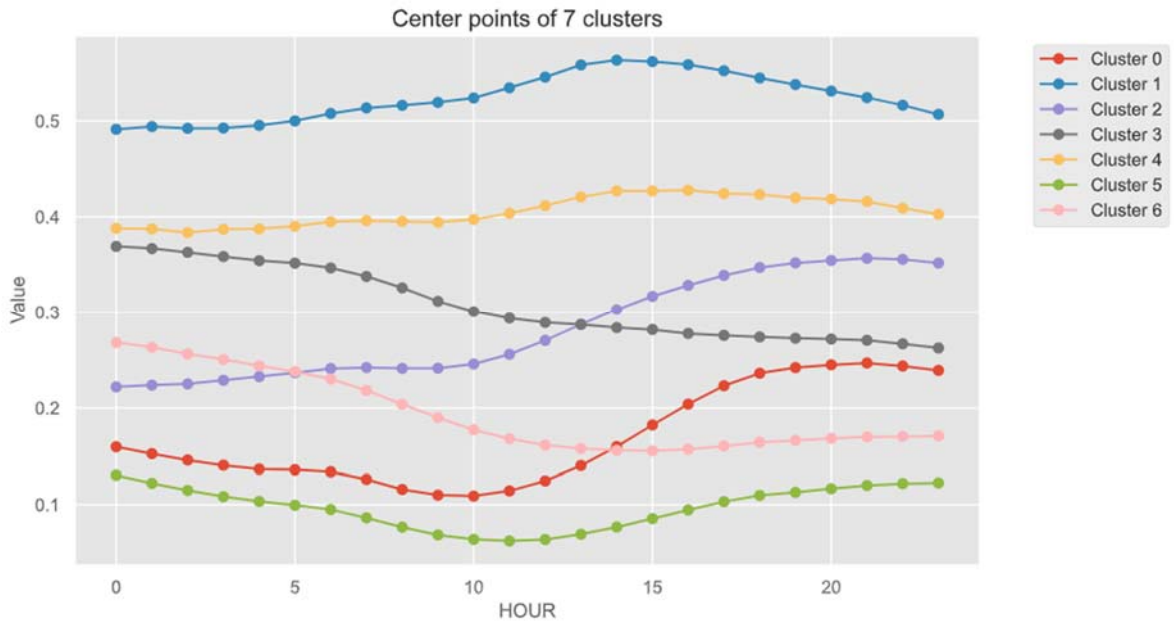

Figure 54. Main hourly patterns for wind generation

Figure 54 shows the average values for each cluster. We can see four main categories: 1 – Blue (Cluster 1), Yellow (Cluster 4) and Light Green (Cluster 5) curves, 2 – Gray (Cluster 3) and Purple (Cluster 2) curves, and 3 – Pink (Cluster 6) and Red (Cluster 0) curves. Category 1 appears when the level of wind energy generation is similar during all the hours of the day. However, depending on the amount of daily production, the corresponding cluster varies. Blue cluster generates the most wind energy, whereas light green clusters generate the least wind energy. Category 2 groups two clusters of similar total energy production, but it is distinguished by when more energy was generated, before (Cluster 3) or after 12 o'clock (Cluster 2). Category 3 has the same characteristics, but the total energy output is lower than Category 2.

This analysis provides insights into the clustering of wind energy data based on generation patterns, variations in energy production, and timing. These findings contribute to a better understanding of the dynamics and characteristics of wind energy generation, which can inform decision-making and optimization strategies in the field of renewable energy.

5.2.2 ENERGY GENERATION BY PROVINCES

In this section, an in-depth analysis is conducted on the hourly data of wind power generation energy across different provinces. This analysis aims to provide comprehensive insights into the temporal patterns and variations in wind energy generation at a regional level.

By examining this data, we can explore the dynamics of wind power generation within each province and identify any unique characteristics or trends that may emerge. This analysis will enable a deeper understanding of the factors influencing wind energy generation, such as geographical location, topography, and local wind patterns.

Through this investigation, valuable insights can be gained regarding the temporal distribution of wind energy generation across different provinces. These insights will contribute to the development of more effective strategies for optimizing wind power generation, enhancing energy planning and management, and promoting sustainable energy practices at the regional level.

In order to ensure the accuracy and reliability of the PCA analysis, a careful consideration of the dataset was undertaken. As part of this process, the decision was made to exclude the province of Extremadura from the analysis. This decision was based on the absence of installed capacity data for the period spanning from 2015 to 2019. By excluding Extremadura province, we aimed to ensure the adherence to high standards of data quality and enhances the reliability and validity of the findings.

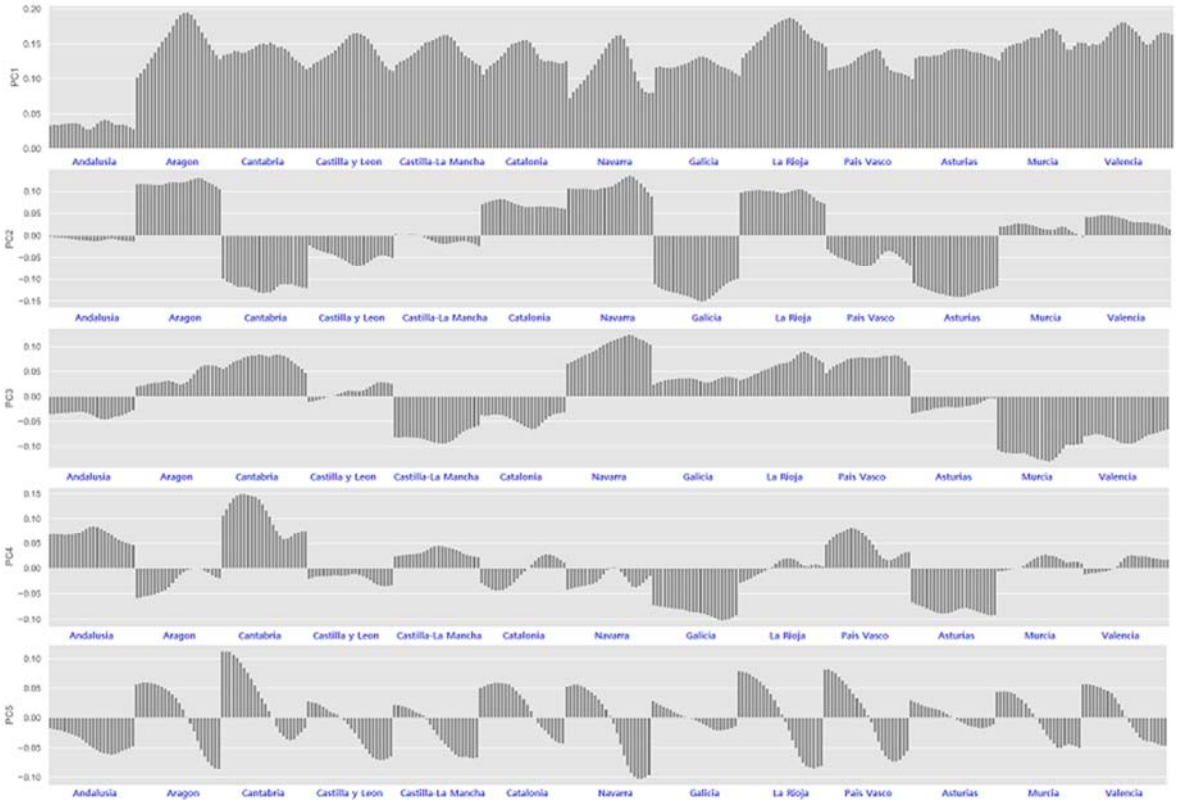

Figure 55. PCA Loadings for wind generation by provinces

According to Figure 55, the loadings defining PC1 exhibit a distinct time zone triangle pattern across multiple provinces, indicating a strong relationship between solar and wind energy. This pattern is particularly pronounced in triangular-shaped regions such as Aragon, Navarra, and La Rioja. Notably, the influence of Andalusia wind generation hours in PC1 is not prominent.

The heightened presence of the time zone triangle pattern in triangular-shaped regions, such as Aragon, Navarra, and La Rioja, suggests a geographical influence on the synchronization between solar and wind energy patterns. The unique topographical features and wind flow dynamics in these areas may amplify the correlation between solar and wind energy, leading to a more pronounced time zone triangle pattern.

Furthermore, the relatively diminished influence of Andalusia in PC1 implies a possible divergence in the relationship between solar and wind energy in this region. Factors such as geographical location, local climate conditions, and wind patterns specific to Andalusia may

contribute to a distinct energy generation profile, deviating from the overarching time zone triangle pattern observed in other provinces.

PC2 captures the divergence between the northeast and northwest regions in terms of wind energy generation. This component provides valuable insights into the regional variations and distinct patterns observed in these areas. Positive values of PC2 indicate a stronger influence from the northeast region, while negative values indicate a stronger influence from the northwest region. The presence of PC2 highlights the significance of geographical factors in shaping wind energy dynamics within the analyzed provinces.

PC3 captures the regional disparities between the northeastern and southeastern regions in terms of wind energy generation. Positive values of PC3 signify a stronger influence from the northeastern region, while negative values indicate a stronger influence from the southeastern region.

PC4 focuses on four regions: Cantabria, Pais Vasco, Galicia, and Asturias, where positive values of PC4 indicate stronger Cantabria and Pais Vasco influences, and negative values indicating stronger influences of Galicia and Asturias regions. Interestingly, note that these four areas are next to each other along the northern coast.

PC5 reveals distinctive characteristics related to the temporal distribution of wind energy generation, specifically focusing on the difference between periods before and after midnight. This principal component captures the variations in wind energy patterns and provides valuable insights into the temporal dynamics of wind power generation. A positive value of PC5 indicates a stronger influence before 12 o'clock, highlighting the significance of wind energy generation during the earlier hours of the day. Conversely, a negative value of PC5 indicates a stronger influence after 12 o'clock, emphasizing the importance of wind energy generation during the later hours of the day. The distinctive characteristics observed in PC5 of wind energy generation align closely with the findings in PC2 (Figure 47) of the Energy Generation analysis discussed earlier.

	Exp_variance	cum_Exp_variance
PC1	0.371256	0.371256
PC2	0.147664	0.518920
PC3	0.091025	0.609945
PC4	0.061729	0.671674
PC5	0.045391	0.717064
...	...	...

Figure 56. Cumulative explained variance of PCA for wind generation

The cumulative explained variance of the five main principal components (PC1, PC2, PC3, PC4 and PC5) exceeds 72%, indicating that these components capture a significant portion of the variability in the wind energy data (see Figure 56). The cumulative explained variance of the five main principal components indicates that these components capture a significant portion of the variability in the wind energy data. This underscores the significance and utility of the principal component analysis in comprehensively capturing and representing the inherent variability within the wind energy dataset.

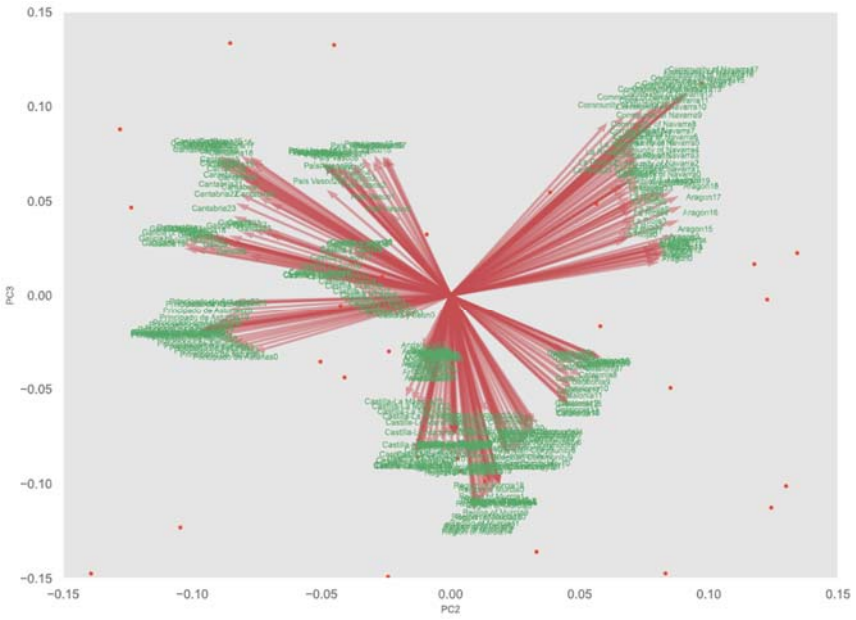

Figure 57. Biplot with PC2 and PC3 for wind generation by provinces

The biplot depicting the relationship between PC2 and PC3, shown in Figure 57, provides a spatial visualization of solar power generation across different provinces of Spain. This

insightful representation enables us to discern distinct patterns and variations in solar energy generation throughout the country. Notably, the biplot presents a rotated map of Spain, with the provinces tilted approximately 45 degrees counterclockwise, facilitating a comprehensive understanding of the geographical distribution of wind power generation.

Let us look at the graph of how the characteristics of each PCA appear in real wind energy data.

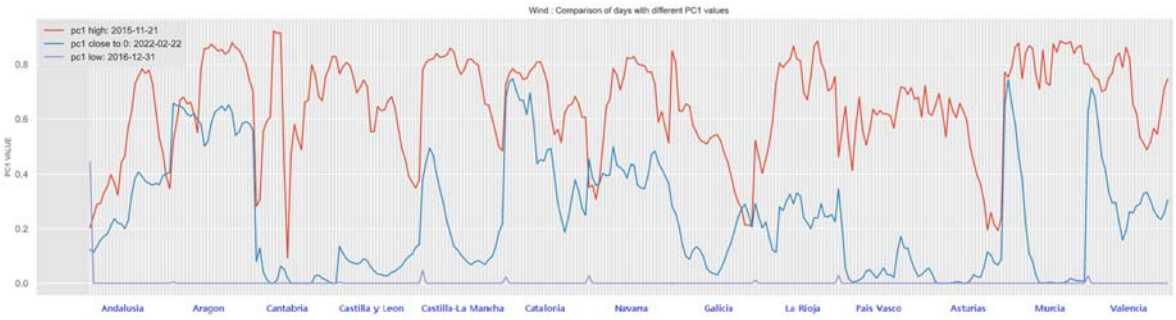

Figure 58. Three days of hourly wind generation by provinces selected according to PC1

Figure 58 shows three distinct days of wind hourly generation, showing regional differences according to three very different values of PC1. The highest red curve in all provinces corresponds to the high value of PC1, indicating a higher amount of wind energy in the day. Conversely, the lowest purple curve is observed when the value of PC1 is lower, indicating that there is no wind generation. The mid-blue curve represents the average condition when PC1 is close to zero.

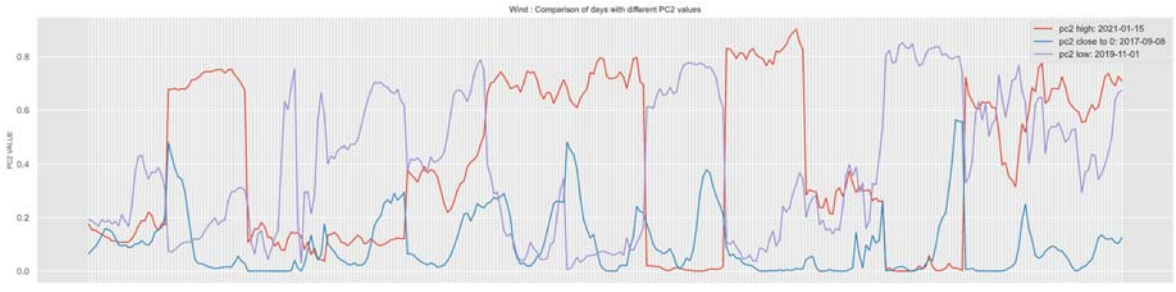

Figure 59. Three days of hourly wind generation by provinces selected according to PC2

Figure 59 shows three distinct days of wind hourly generation, selected by choosing three very different values of PC2. This graph shows three distinct and complex curves that

provide insight into wind energy patterns throughout the day, particularly with regard to the geographical characteristics of the Northeast and Northwest regions. Positive PC2 (red curve) represents energy production in the northeast region, and negative PC2 (purple) represents energy production in the northwest region.

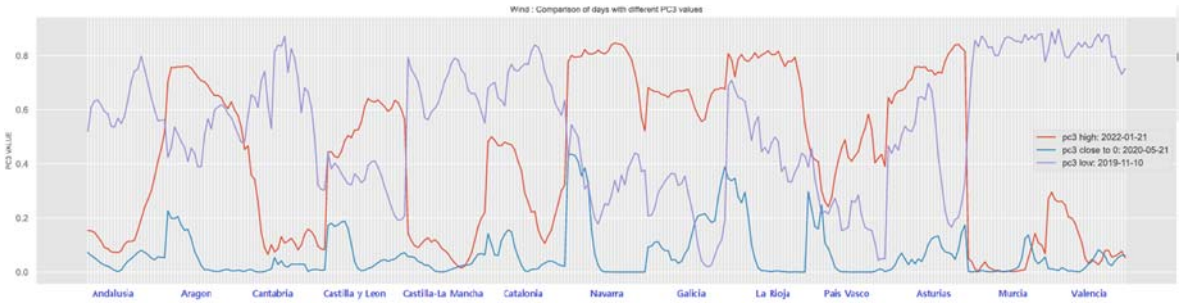

Figure 60. Three days of hourly wind generation by provinces selected according to PC3

Figure 60 shows three distinct days of wind hourly generation, showing regional differences according to three very different values of PC3. In particular, three complex curves are shown, providing insight into wind energy patterns throughout the day, particularly with regard to the geographical characteristics of the Northeast and Southeast regions. The positive PC3 (red curve) represents energy production in the northeast region, and the negative PC3 (purple) represents energy production in the southeast region.

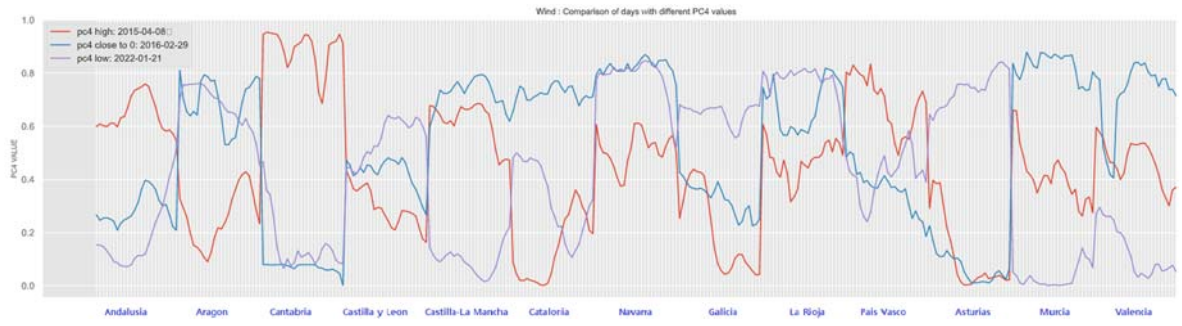

Figure 61. Three days of hourly wind generation by provinces selected according to PC4

In a similar way, Figure 61 shows three distinct days of wind hourly generation, showing regional differences according to three very different values of PC4. Three distinct and complex curves throughout the day are shown, providing insights with respect to the four provinces. The positive PC4 (red curve) represents the day when energy production is high

in Cantabria and Pais Vasco regions, and the negative PC4 (purple) represents the day when energy production is high in Galicia and Asturias regions.

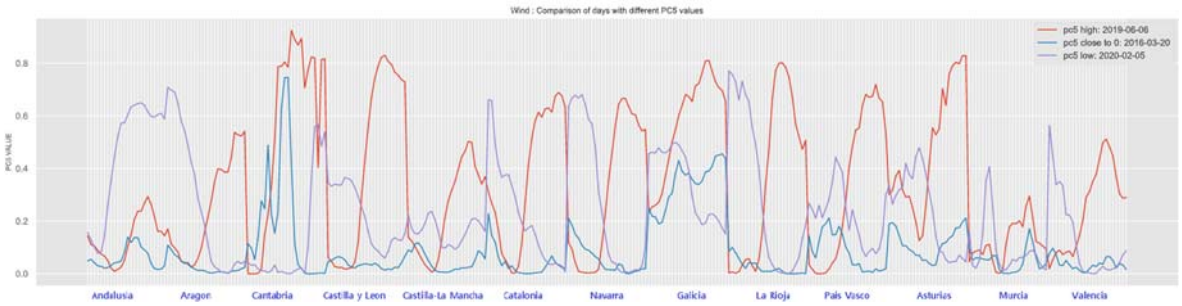

Figure 62. Three days of hourly wind generation by provinces selected according to PC5

Varying PC5 three different days have been selected (Figure 62), corresponding to three distinct and complex curves that provide insight into energy patterns associated with temporal distribution. It can be seen that the positive PC5 (red curve) and the negative PC5 (purple curve) have a symmetrical shape.

The analysis of wind energy data reveals distinct patterns and regional differences in energy generation across various provinces. This comprehensive analysis provides valuable insights into the diverse characteristics and regional variations in wind energy generation across the analyzed provinces, aiding in the development of effective strategies for renewable energy planning and management.

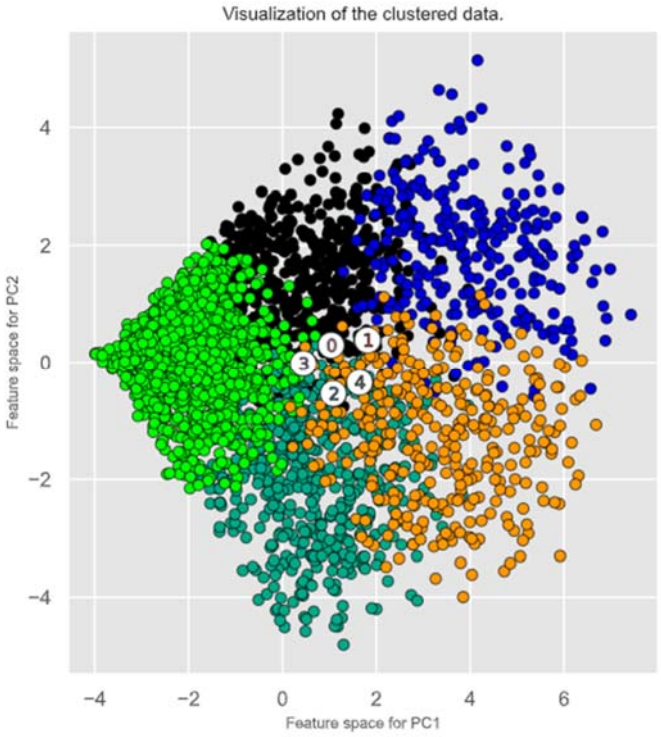

Figure 63. Clustering on data with 5 clusters for wind generation

Figure 63 presents a comprehensive clustering solution by identifying five distinct groups based on the values of PC1 and PC2. The visualization of these clusters provides valuable insights into the spatial distribution of wind energy intensity.

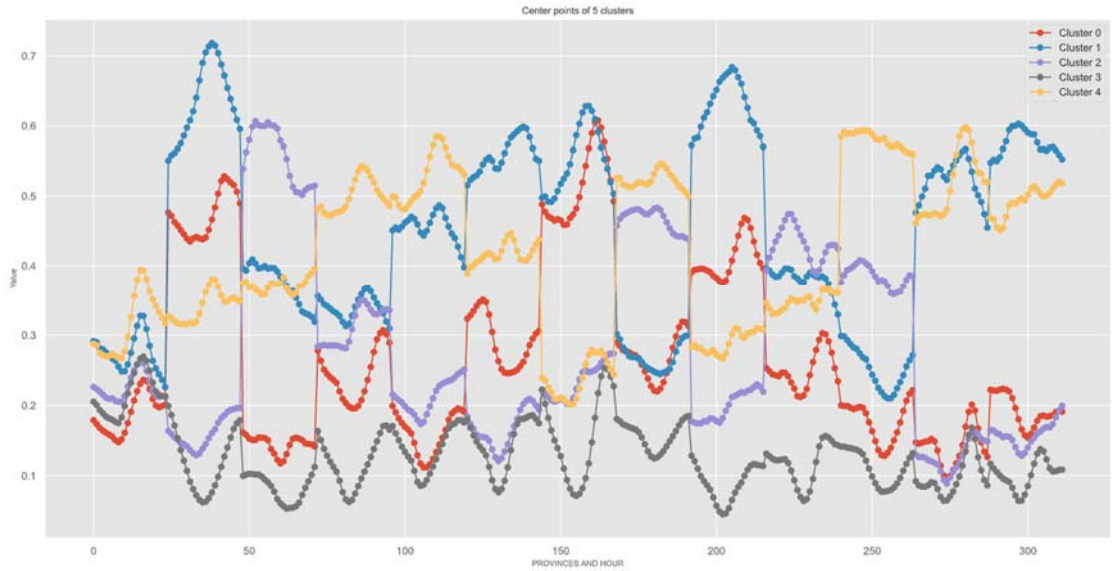

Figure 64. Main hourly patterns for wind generation by provinces

Figure 64 shows the mean values for each cluster derived from cluster analysis. For example, the red curve (Cluster 0) is a cluster of days with high wind energy production levels in the northern region, and the blue curve (Cluster 1) is a cluster of days with high wind energy production levels in the eastern region. In addition, the purple curve (Cluster 2) is a cluster of days with high energy production levels in the coastal area among the northern regions, and the purple curve (Cluster 3) is a cluster of days with low wind energy production overall.

Chapter 6. PHOTOVOLTAIC WITH WIND ENERGY GENERATION

6.1.1 RELATIONSHIP BETWEEN SOLAR PV AND WIND ENERGY

In this chapter, we will explore the relationship between solar photovoltaic (PV) and wind energy through a jointly analysis of their respective hourly data.

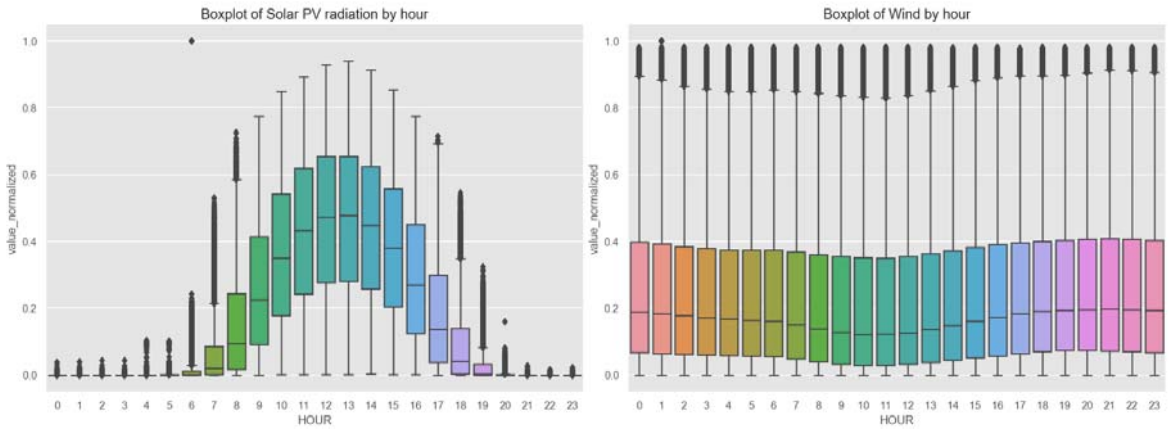

Figure 65. Box-plot of PV and Wind by hour

The boxplot analysis shown in Figure 65 reveals significant variability in solar radiation energy over time, aligning with our intuitive understanding of solar power. The peak solar radiation energy occurs at 13:00, with an average value of 0.462847. The analysis of wind energy demonstrates minimal variation over time, indicating that its generation is not significantly influenced by the time of day. The highest average value of wind energy, 0.264282.

In order to investigate the interplay between solar radiation and wind energy, we assess the association between solar radiation levels at 13:00 and wind energy measurements at 12, 13, and 14. The selection of wind energy measurements at 12, 13, and 14 is motivated by the objective of examining the temporal dynamics of wind energy generation in relation to the peak solar radiation period.

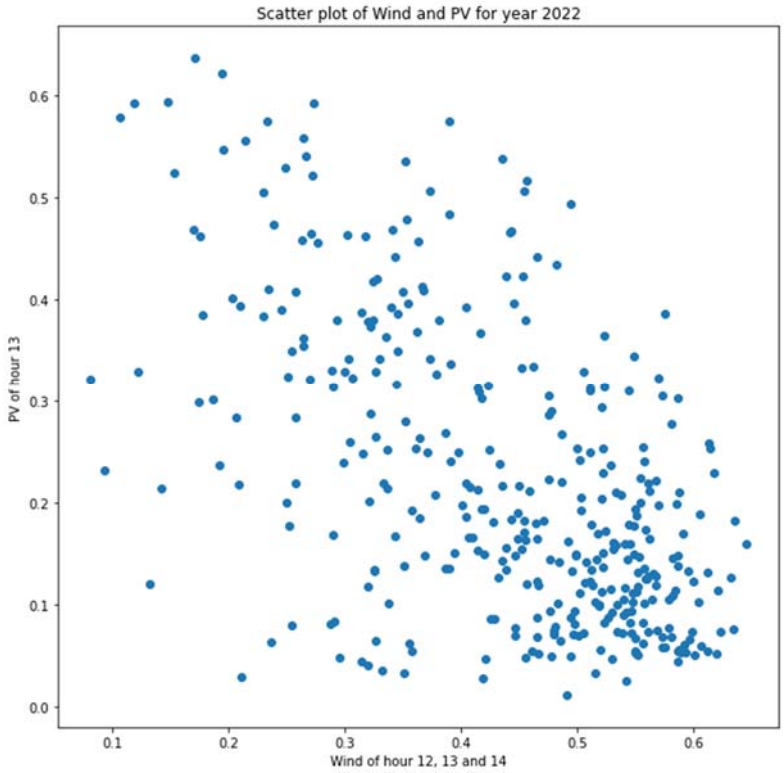

Figure 66. Scatter plot of Wind and PV for year 2022

Figure 66 presents a scatter plot illustrating the correlation between solar radiation at 13:00 and mean wind energy at 12, 13, and 14. The y-axis represents the solar radiation values, while the x-axis represents the corresponding wind energy values. The scatter plot demonstrates a negative correlation between the two variables, indicated by the downward trend of the data points. A correlation coefficient of -0.44 indicates a moderate negative correlation between the two variables. The negative sign indicates an inverse relationship, meaning that as the wind energy production increases, the solar PV generation tends to decrease, and vice versa. The magnitude of -0.44 suggests that there is a moderate strength of association between the two variables, indicating that they are somewhat related but not perfectly linearly correlated.

Though, the observed correlation coefficient of -0.44 between solar radiation at 13:00 and wind energy at 12, 13, and 14 reveals an intriguing relationship. This finding suggests that there is a tendency for increased wind energy to coincide with lower solar radiation, possibly indicating the presence of cloud cover (i.e. a "bad weather" situation). The inverse

relationship between these variables is in line with the expectation that higher wind speeds are often associated with cloudier conditions, leading to reduced solar generation. This correlation provides valuable insights into the interplay between solar radiation and wind energy generation.

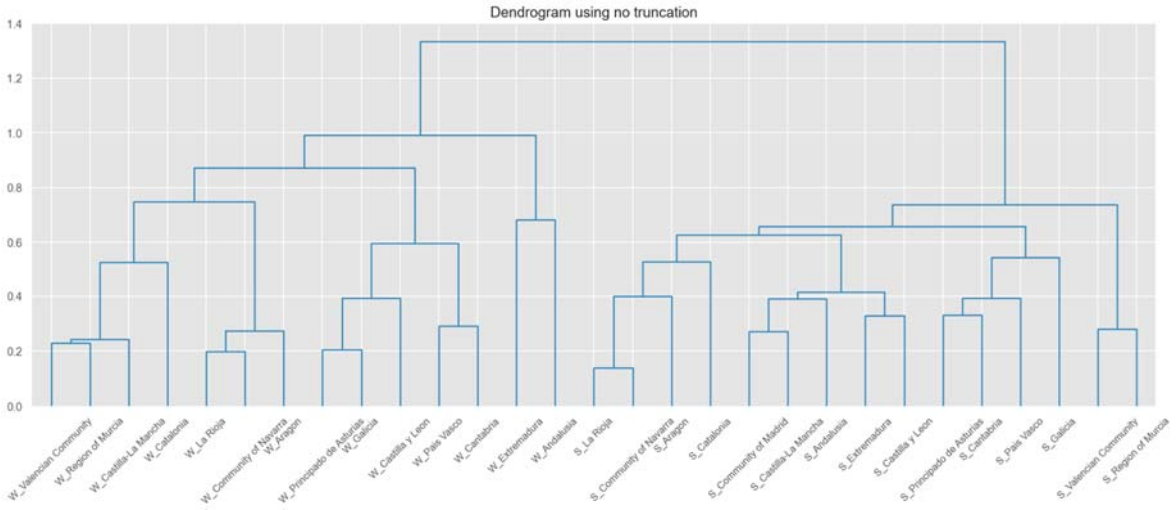

Figure 67. Dendrogram of PV and Wind

To further investigate the joint properties of solar and wind energy, we employed a dendrogram analysis. The dendrogram of Figure 67 clearly illustrates that solar and wind energy data are not grouped together within the same cluster but are instead separated. This represents the unique and independent properties of solar and wind energy generation, along with earlier findings based on correlation coefficients.

Chapter 7. CONCLUSIONS AND FUTURE WORK

The analysis conducted in this study successfully identified and characterized the daily, hourly, seasonal, and geographical changes of each renewable energy source. Specifically, the patterns and dynamics of solar, wind, and hydroelectric power generation were thoroughly investigated and understood.

Regarding solar energy, the study revealed clear patterns of sunlight intensity over various periods, highlighting regional differences and the effects of daily and seasonal changes. The findings shed light on the temporal progression of solar radiation throughout the day and the effects of cloud cover on solar energy generation. In addition, we analyze the geographical distribution of solar resources to provide insight into regional differences in solar potential in Spain.

In the case of wind energy, this study successfully revealed significant regional and hourly variations, shedding light on the temporal dynamics of wind energy generation across different time zones. The findings obtained from the analysis provided valuable insights into the temporal progression of wind energy production during the morning and afternoon periods, as well as the regional variations in energy generation. These insights contribute to a better understanding of the patterns and characteristics of wind energy in Spain, ultimately supporting informed decision-making in renewable energy planning and management.

The join analysis of solar and wind energy generation, based on hourly and daily data through the main regions of Spain has provided useful insights about the spatial and temporal correlation of these two renewable energies.

The analysis of hydro-energy was not considered in this study. This omission is primarily due to the manageable nature of hydro-energy generation, which involves factors such as reservoir management and water release strategies. As a result, hydro-energy analysis remains as a valuable area for future research the understanding the complexities and potential synergies of hydro-energy, along with solar and wind energy, can provide a more comprehensive understanding of the renewable energy landscape in Spain.

Chapter 8. REFERENCES

[1] ESIOS (System Operator's Information System). (n.d.).Glossary. ESIOS (System Operator's Information System). https://www.esios.ree.es/en/glossary

[2] G. James, D. Witten, T. Hastie & R. Tibshirani (2013). An Introduction to Statistical Learning with Applications in R. Springer (see http://www-bcf.usc.edu/~gareth/ISL/). 2nd ed. (2021)

[3] T. Hastie, R. Tibshirani & J. Friedman (2009). The Elements of Statistical Learning. Data Mining, Inference and Prediction. 2nd Ed. Springer.