**Interim Project Report**

**Topic: Data Aggregation and Analysis with Python in the field of Renewal Energy Systems**

**Introduction**

The move to a clean energy economy that prioritizes local resources, increases resiliency, creates jobs, and promotes energy independence requires high-quality renewable energy resource data and other geographic information system (GIS) data. These data are essential for making well-informed decisions on everything from policy and investment to reliable power sector planning. Data-driven decisions reflect proper ambition, maximize cost efficiency, and enable successful renewable energy investment implementation.

Planning for data-driven decision-making necessitates a number of considerations. Data, for example, might be of varying types and quality, be costly to get, and require specialised skills and resources to analyze and evaluate. This project explains the information needed to make various renewable energy decisions, as well as the tools and analysis required to turn that information into recommendations for decision-makers.

This project is designed to assist policymakers, planners, technical specialists, consultants, and researchers in bringing better data and analysis into renewable energy decision-making. There are three key sections:

• Data: renewable energy resource big data using data science

• Analysis: analytical methodologies and models

• Decisions: target setting, policymaking, investment, and power sector planning

**Background Research**

Data science is a branch of data analytics and data visualization in which unstructured or raw data is cleaned and prepared for analysis. This data is used by data scientists to obtain the necessary information for future purposes. Data science employs a variety of procedures and methods to analyse large amounts of data, which can be structured or unstructured. We obtain raw data from the internet in the form of data frames. It could be in an unstructured or semi-structured fashion. This data is subsequently filtered, cleaned, and a number of needed tasks for analysis are completed using a high-level computer language. This information is then processed and presented to help us better understand and evaluate it.

There are numerous tools available to handle the large amounts of data we have at our disposal. Data scientists extract insights from prepared data using programming languages such as Python, R, SAS, Java, Perl, and C/C++. As we are using Python in our project

As the amount of renewable energy systems grows, so does the amount of renewable energy data collect by sensors and other energy system components. So, not only can big data assist in comprehending the current state of the renewable energy business, but it can also assist in anticipating renewable energy use and output.

The following are some of the data science applications that are important in the field of renewable energy.

* **Improving present technology:** This is most commonly employed in the solar energy field. Solar panel data is acquired via sensors, and by analyzing the data pattern, we may increase the efficiency and life term of the solar panel in question.
* **Predictions for renewable energy use**: The consumption of renewable energy by customers can also be anticipated using historical data on customer energy consumption. This could be quite beneficial in the future when it comes to meeting customer needs.
* **Forecasting renewable energy output:** Solar and wind energy production can be optimized by taking weather and environmental data into account. Forecasting is simple with this information.
* **Lowering the Costs of Renewable Energy Production:** With the large amount of energy data at our disposal, we can simply forecast the cost of renewable energy production using a forecasting model. Because of the huge data and forecasting models available to us, the price of energy is falling. Renewable energy will be competitive with its traditional counterparts in terms of cost.
* **Power plants with a reliable backup system:** We can easily get the high and low power usage with the assistance of computational models, and when there is ample power, we can conserve the power that would otherwise be wasted, and vice versa, we can offer electricity with the help of our renewable energy systems when there is a shortage.

Data and analysis are essential components in making informed renewable energy decisions. This guide is divided into three sections, each of which focuses on how to use high-quality data and analytics to make informed decisions about renewable energy. The project starts with a section that explains how to make decisions about renewable energy, such as target setting, policymaking, investment, and power sector planning.

**Project Plan**

The initial step in our project will be to analyse large amounts of data from various sources. The study plan that has already been discussed will be rigorously implemented.

Following the initial step, we plan to add more machine learning elements to our project to make it a better, more efficient, and faster energy rejuvenation system.

The application of the Python programming language in the field of renewable energy is the main emphasis of this project. This language is useful not only for data analysis, but also for forecasting future scenarios in the energy sector.

The adaptability, extensive libraries, speed limits, and ease of learning are all reasons for utilising this specific language. In this project, we will be studying enormous energy data sets that cannot be easily studied in other technologies, such as python.

Python isn't just for data analytics; it's also useful in fields like artificial intelligence, machine learning, and countless additions.

So now comes the formulation and implementation phase of our project plan. After then, the results will be validated, and it will be maintained in high definition after each test.

**Summary of progress to date**

We needed raw data before we can analyze and visualize it, and we got it from a variety of open-source data websites on the internet. This information is in raw form, and also included PV solar panel sales, renewable energy consumption or production in any specific area or region where solar or wind power is more advantageous.

Global Wind and Solar Production from

1990 to 2014

Global Wind and Solar Production from

1990 to 2014

Global Wind and Solar Production from

1990 to 2014

Global Wind and Solar Production from

1990 to 2014

Global Wind and Solar Production from

1990 to 2014

From 1990 to 2014, this raw data includes imports, exports, production, consumption, transformation, and energy losses for all countries. It mostly consists of renewable energy data, as well as other types of energy data. We'll look at wind and solar data from 1990 to 2014, as well as other countries' patterns and compare them. We'll also compare the generation of wind and solar energy in the two countries.

**Methodology for data collection:** There are numerous data points in the data set, including nation, commodity transaction, year, unit, quantity, and category (Category of energy used in that specific area).

**Analytical and Statistical Issues Quantity:** foot-notes is a blank column in the data set that contains the blank cells in the data sets for which no data has been provided. As a result, that column has been removed. The data that was not accessible is represented by blank cells.

Data description of the data set of column

Data Description

country or area This column consist of most of the countries of the world.

commodity transac-

tion

In this column there are different sectors of production and

consumption according to the energy sector.

year This column consist of the year of production it ranges from

1990 to 2014.

units It consist of the unit of the different sectors of production.

quantity These are the values respective to the commodity transac-

tion column. There are zero values in most of the cells

which means the data of that particular year is not provided.

category In this column you are given with all the different types of

energy sources

Data description of the data set of column

Data Description

country or area This column consist of most of the countries of the world.

commodity transac-

tion

In this column there are different sectors of production and

consumption according to the energy sector.

year This column consist of the year of production it ranges from

1990 to 2014.

units It consist of the unit of the different sectors of production.

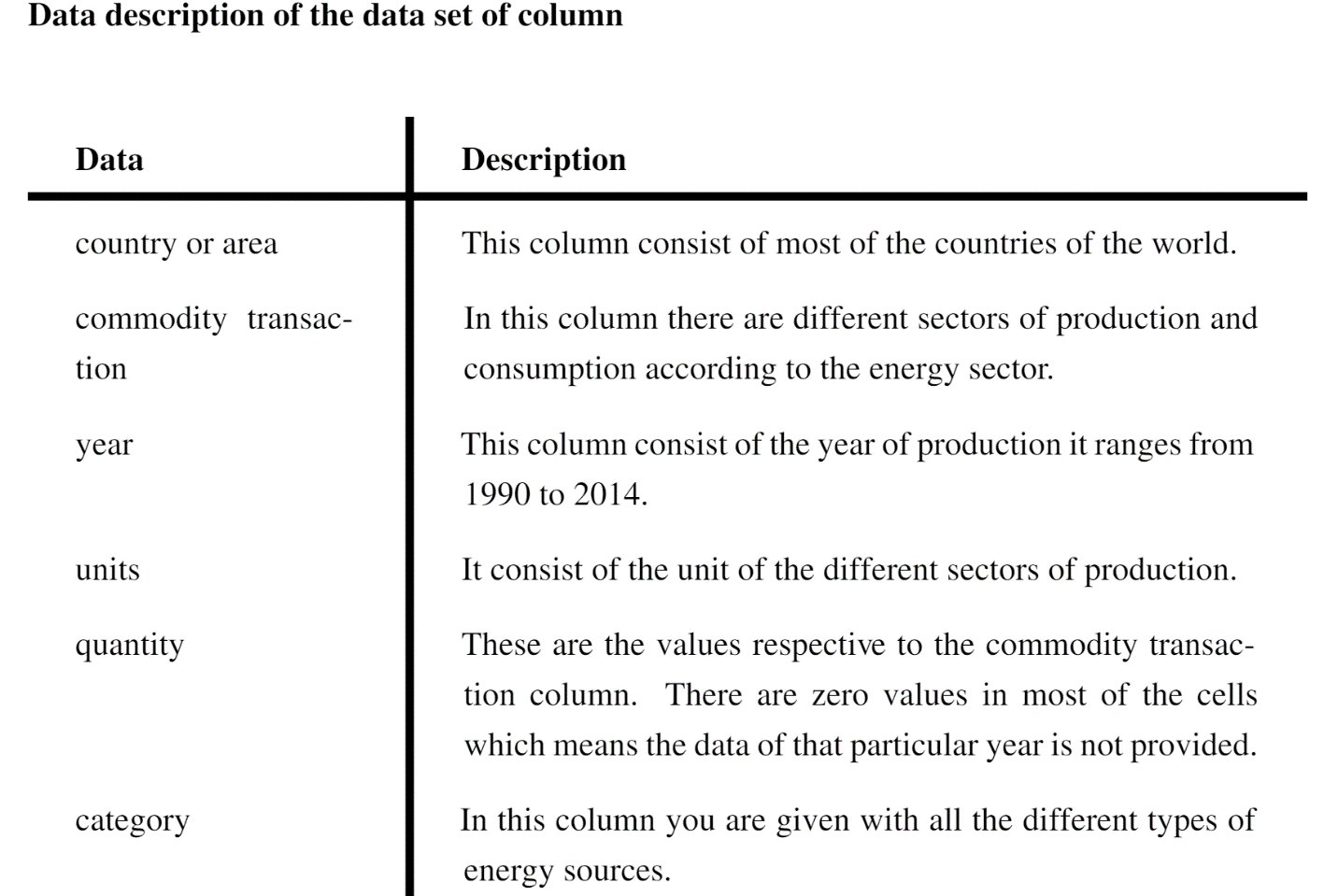
quantity These are the values respective to the commodity transac-

tion column. There are zero values in most of the cells

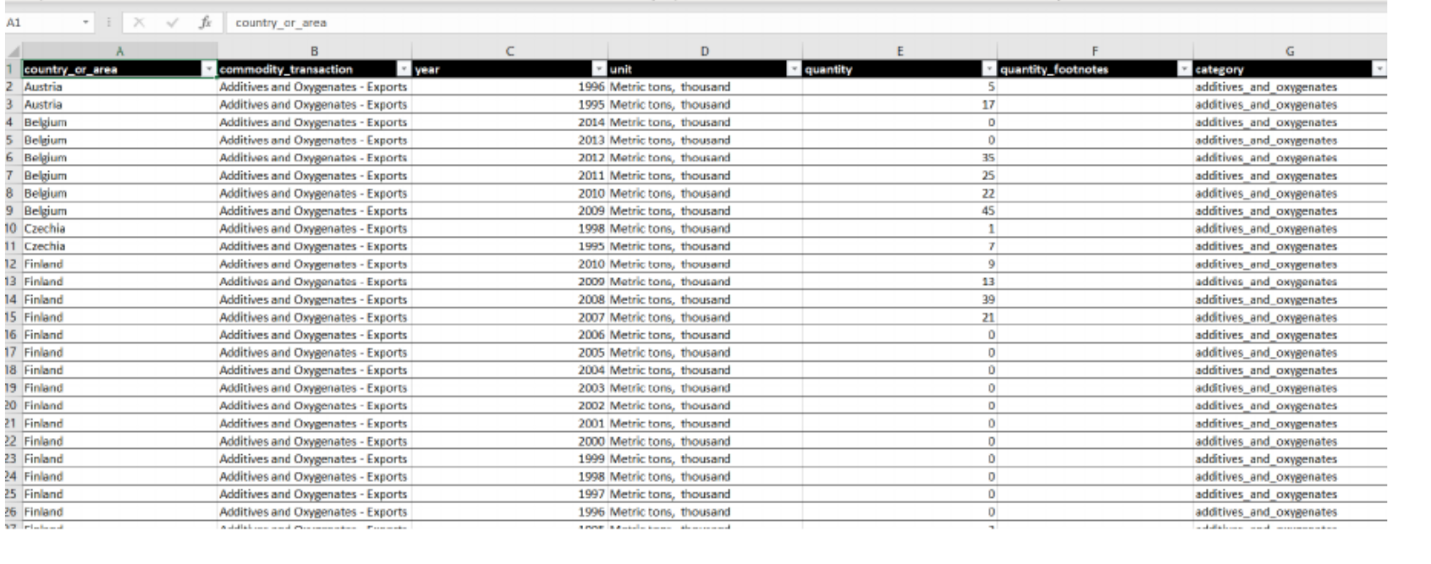
which means the data of that particular year is not provided.

category In this column you are given with all the different types of

energy sources



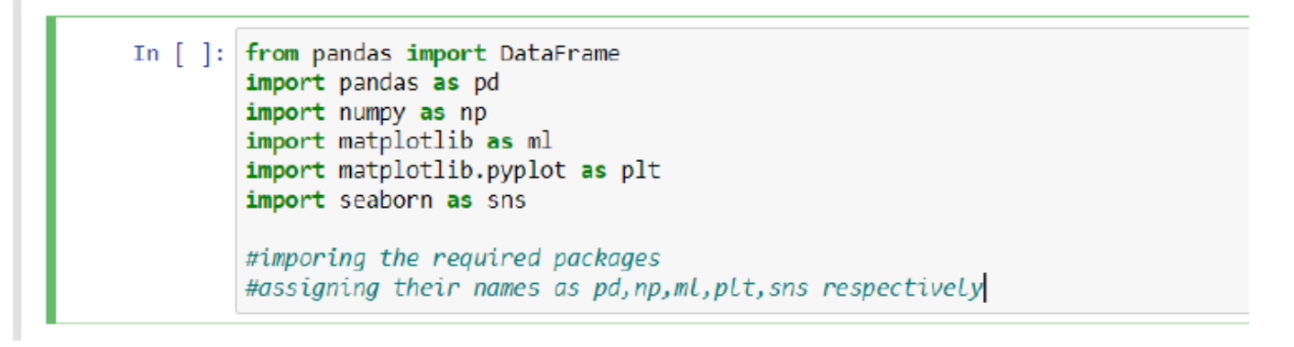
We'll use some questions to assist us analyze the data. The figure below shows the datasheet in Excel, which will give you an idea of how the data is available to us.



Giving explanations to the following set of questions will make analysis easier.

1. From 1990 to 2014, what were the trends in wind energy output in Germany, Spain, and France? Where did these three European countries produce the most wind?

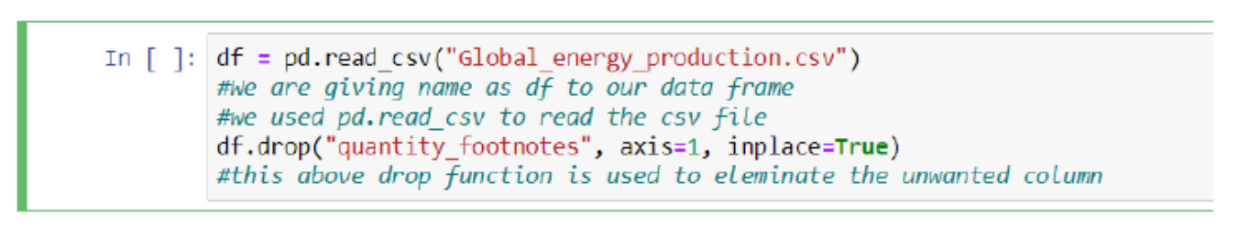
Solution: Each line of code is explained within the programmed itself. A line beginning with a hash tag is the explanation for that specific line of code.



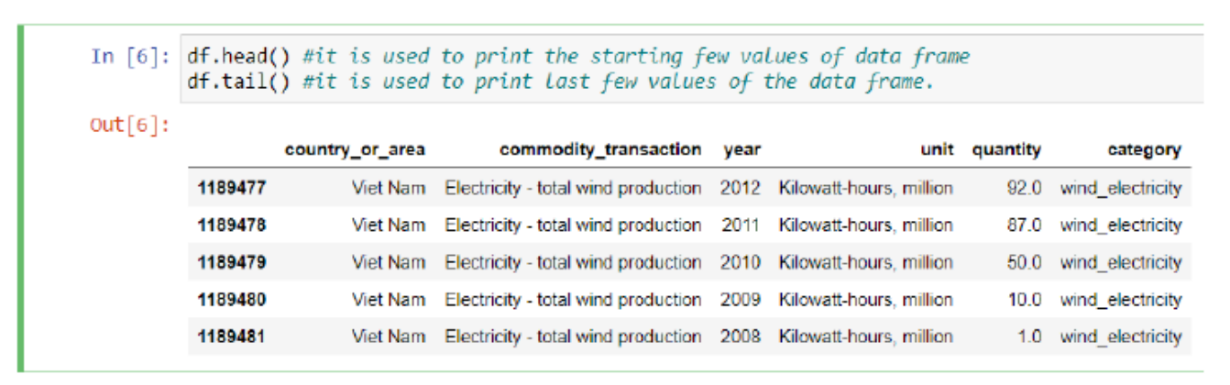
Importing the required packages



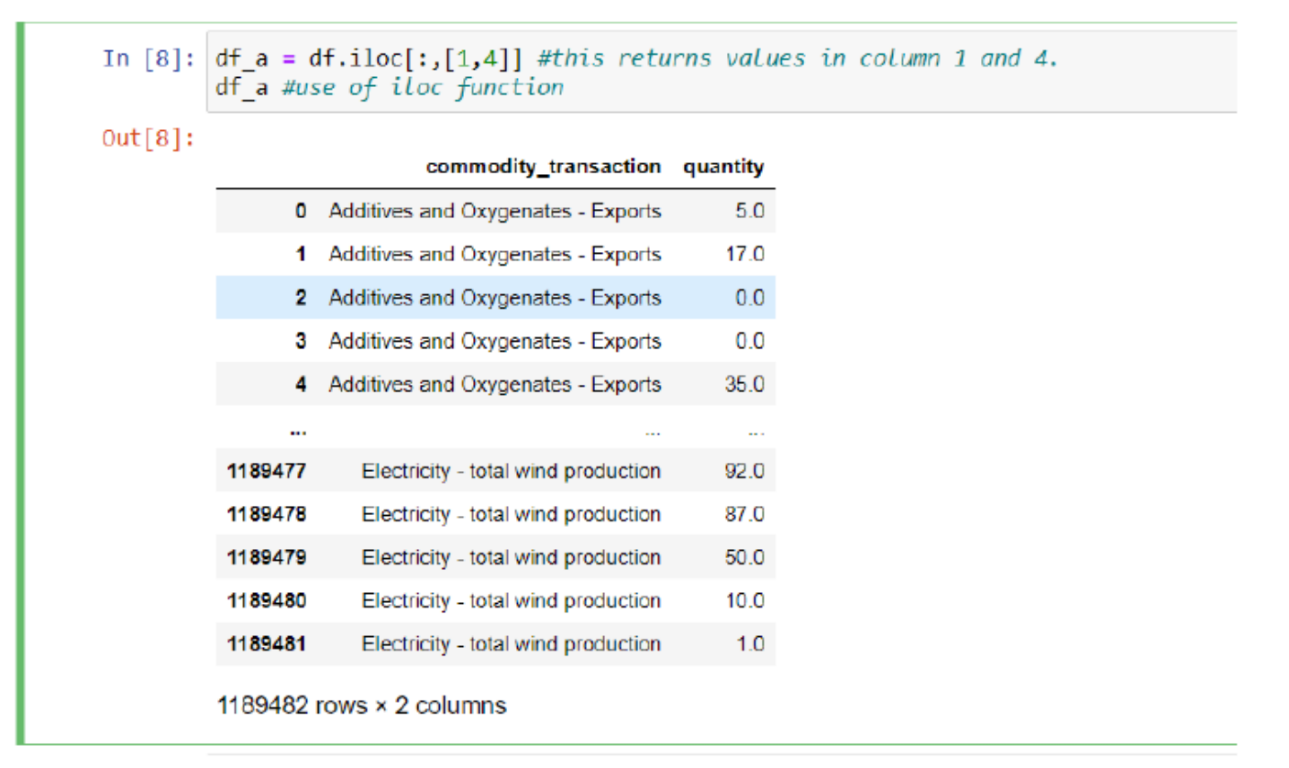
Setting up the grid style as Dark



Reading the csv file and drop functions use



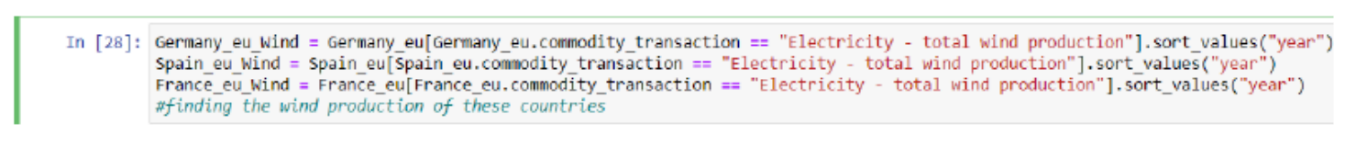
Use of head and tail function



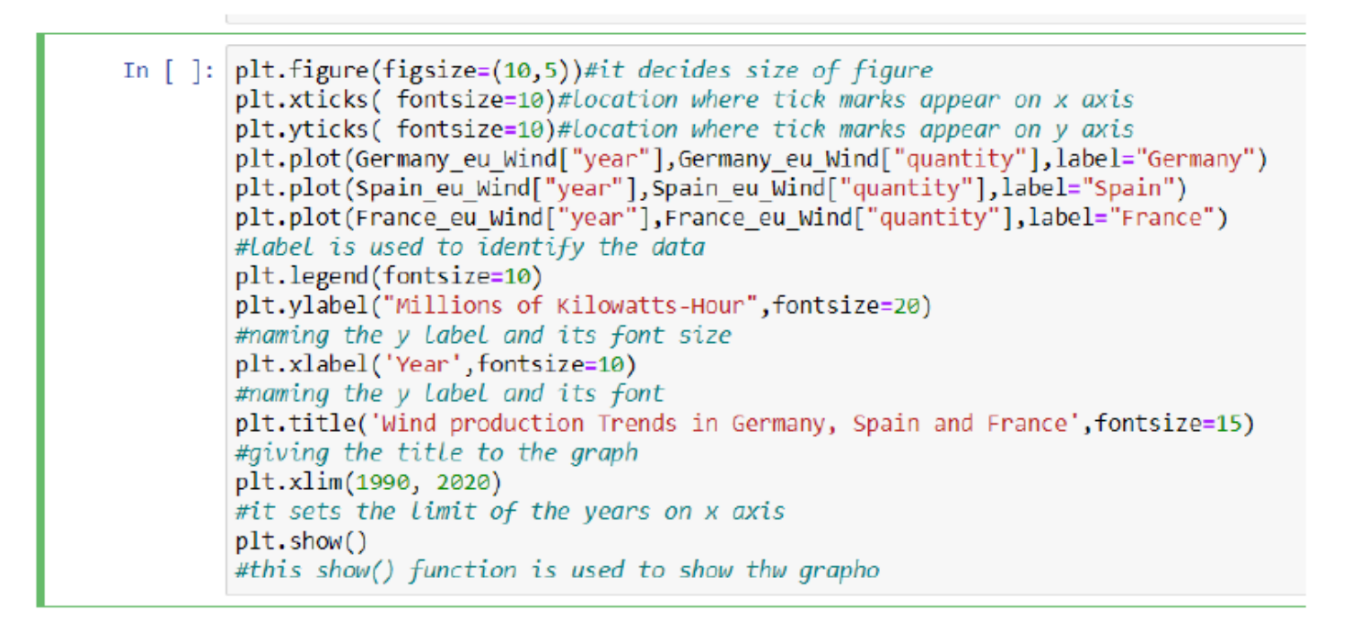
Use of iloc function to get specific column values



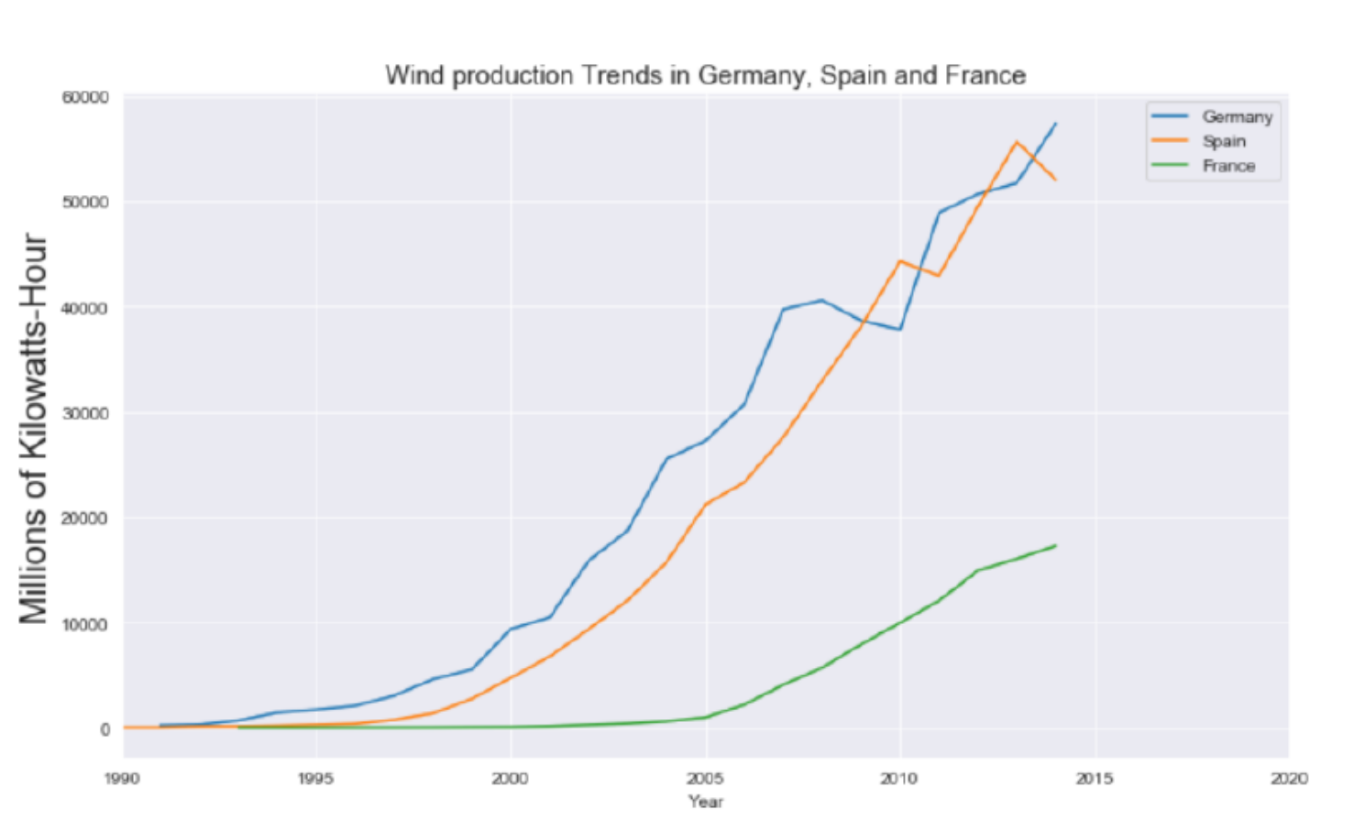
Use of isin and sort function



Again sorting and assigning for wind production



Plotting the graph



Wind energy production trends in Germany, Spain, and France

**Conclusion** - Germany's wind energy production in 2010 was lower than Spain's wind energy production. In 2010, Germany's wind energy production was less than 40000 million kilowatt hours, while Spain's was more than 40000 million kilowatt hours at the same time. This figure was three times higher than France's in 2010. However, Germany had the largest wind energy production among these three countries in 2014, with roughly 60000 million kilowatt hours.

**There are many more questions to be answered in order to collect huge data and analyse it in the same way for improved project outcomes.**

**Consideration of ethical, legal, professional and social issues:**

First, as a matter of future ethics, a substantial effort should be made to find alternate energy sources i.e., renewable energy sources. Future generations will be put at risk if we continue to rely heavily on fossil fuels and nuclear power. Second, we owe it to future generations to safeguard nonrenewable resources. Waste of resources that future generations will require, especially when we now have the technology to considerably improve energy efficiency, makes it more difficult for future generations to achieve a lifestyle comparable to ours. Finally, the ability of these renewable energy sources to be utilized is determined by the state of technology, which is the outcome of extensive research.

According to international experience, countries that use renewable energy sources on a large scale have a strategic state policy for efficient renewable energy utilization. First and foremost, a strong legislative foundation is required to promote the usage of renewable energy sources. The issue is resolved by the fact that legislators do not always take into account the main mechanisms required for the development of this technology. The applicable renewables law contains a large number of reference standards that preclude proper application of its requirements. Furthermore, there are no legal requirements that could impact the growth of renewable energy sources. Although the RES development stimulation measures have been specified, there are no mechanisms in place to implement them. It would be reasonable to establish the legal technique for stimulating their intended recipients.

We are endowed with all renewable energy sources such as hydro, wind, and geothermal, and it is also an ideal location for solar power generation. However, funds and politics are the most significant obstacles to generating electricity from renewable energy sources.

**Appendices**

This appendices has a more detailed overview of the electricity-generating technologies, with a focus on their advantages and disadvantages.

**Data structures in Python**

Data structures are a method of storing data so that we may quickly conduct various operations on it whenever we need to. When data is acquired from a data source, it is available in a variety of formats. As a result, once the data has been sorted into various data structures, it will be simple for data scientists to execute various operations on it.

Data structures are divided into two groups, which are further divided into subcategories as indicated below.

**Primitive data structures**

Basic data structures are another name for them. Simple data values are stored in this type of data structure.

* Integers- Integer data types encompass all whole numbers from negative infinity to positive infinity. For instance, 4,9, -2, -6.
* Float- Float data types include decimal figure numbers and rational numbers. 3.1,2.2,8.96, for example
* Strings- Strings are a collection of alphabets or letters. In Python, we use single or double quotes to surround the string. 'Hello' and 'bread' are two examples.
* Boolean- These are built-in data types that accept two values: 'True' and 'False.' In Python, True represents 1 and False represents 0.

**Solar Photovoltaic**

Solar photovoltaic (PV) is a possible solution for utility-scale renewable energy. Although all PV technologies convert sunlight into electricity, since the original deployment of crystalline silicon systems in the 1950s, several forms of PV technology have achieved differing levels of productivity and commercialisation.

**Wind**

Although wind, like solar, is considered a "irregular" production resource, the output of wind farms is more consistent than one might assume. Grid-connected wind farms typically have capacity factors of 25 to 40% or more (and sometimes even more) than solar farms.

**References**

1. Evidence from Multivariate Panel Data Analysis.
2. Boyle, G. (n.d.). Renewable Energy: Power for a Sustainable Future, 3rd Edition.
3. Embarak, D. O. (n.d.). *A book on data science.*
4. Jat, S. (2020). *Data Aggregation and analysis with Python in the field of Renewable Energy Systems.* Nordhausen University of Applied Science.
5. Kim, M. Y. (2020). Revisiting the Relation between Renewable Electricity and Economic Growth: A Renewable–Growth Hypothesis.
6. ManassehObib, T. R. (n.d.). *Distributed energy resource aggregation using customer-owned equipment: A review of literature and standards.*
7. Renewable Energy Data, Analysis, and Decisions: A Guide for Practitioners *Sadie Cox, Anthony Lopez, Andrea Watson, and Nick Grue National Renewable Energy Laboratory Jennifer E. Leisch United States Agency for International Development*
8. Brundtland, G. (1987). *Our Common Future*. [New York](https://www.encyclopedia.com/places/united-states-and-canada/us-political-geography/new-york): [Oxford University](https://www.encyclopedia.com/social-sciences-and-law/education/colleges-international/oxford-university) Press.
9. Bullard, R. (1993). *Confronting Environmental Racism.* Boston: South End Press.
10. Bullard, R. (1994). *Dumping in Dixie*. Boulder, CO: Westview Press.
11. Daly, H., and Cobb, J. (1989). *For the Common Good.* Boston: Beacon Press.
12. DesJardins, J. (1997). *Environmental Ethics: An Introduction to Environmental Philosophy*. Belmont, CA: Wadsworth Publishing
13. https://www.abacademies.org/articles/legal-regulation-of-renewable-energy-sources-usage-6820.html
14. Chapter 4
15. Data collection
16. Before analyzing and visualization we need the raw data and this raw data can gathered from
17. different open source data websites available on the internet. This data will be in raw form, it
18. may be the PV solar panel sales, renewable energy consumption or production in any speciﬁc
19. area or regions where solar or wind which one is more favorable. As here we are focusing on
20. the renewable energy data sets so we will be considering following websites where this data is
21. available.
22. https://www.eia.gov/
23. This website contains the energy data mostly of US. EIA is the abbreviated form of Energy
24. Information Administration.Here we have different data of prices, consumption, production,
25. exports and imports of the energy data.
26. https://www.energy.gov/
27. Energy.gov is the other website for data related to renewable energy, This Energy Department
28. is responsible to make sure USA’s Energy Future and solve the energy related problems
29. https://openei.org/
30. Open Energy Information is a website for policy makers, researchers, technology investors,
31. venture capitalists, and market professionals with energy data, information, analyzes, tools,
32. images, maps, and other resources.
33. https://data.world/
34. Here we can ﬁnd data related to each and every ﬁeld, it is the most widely used website for data
35. analysis. We can also gather energy related data from this website.
36. https://catalog.data.gov/dataset
37. Data.gov is powered by two open source applications, CKAN and WordPress, and it is de-
38. veloped publicly on GitHub.Data.gov is managed and hosted by the U.S. General Services
39. Administration, Technology Transformation Service.
40. https://www.kaggle.com/
41. The data sets available here is not speciﬁcally for renewable energy. Kaggle is general data sets
42. website, her

•Identify available data sources

•Identify if additional data sources are needed.

•Statistical analysis

•Implementation, development

•Communicate results

•Maintenance