DS5110 Project Proposal

Title: Analyzing Global Electricity Statistics

Authors: Karthik Chintamani Dileep, Shrada Chellasami, Shreya Nelakonda and Sumayya Naseem

Summary: The project involves analyzing the Monthly Electricity Statistics dataset obtained from the International Energy Agency (IEA) website which will explore the trends and patterns of electricity production and consumption in different regions of the world, and analyze the factors influencing these patterns.

By analyzing the IEA data, the project aims to gain insights into the factors that influence electricity production and consumption, and understand how these factors vary across different countries and regions. Additionally, we endeavor to establish correlations between a country's electricity production and its GDP, as well as between its electricity production and population growth. Ultimately, the project seeks to provide a comprehensive analysis of the relationship between electricity production and consumption, and contribute to a better understanding of global energy trends and patterns.

Proposed plan:

Data Gathering: The data was obtained from Monthly Electricity Statistics from the International Energy Agency (IEA) dataset which is publicly available. The data was downloaded from their website as a comma-separated text (.csv) file.

Data Cleaning and Preprocessing: To prepare the IEA's Monthly Electricity Statistics dataset for analysis, any missing values and duplicates should be removed. Column headers should be renamed for clarity, and irrelevant columns removed to simplify analysis.

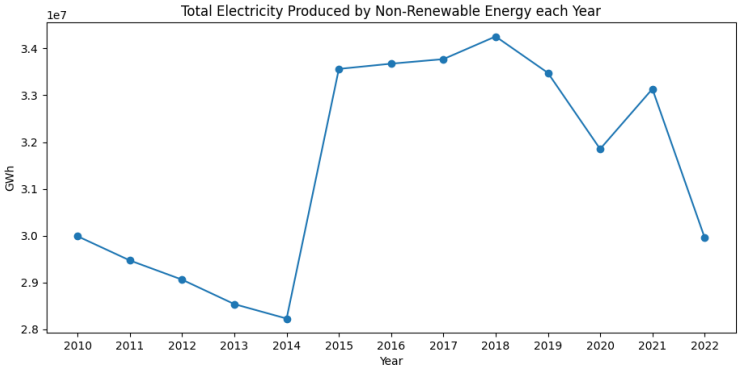
The monthly electricity data column in the IEA's Monthly Electricity Statistics dataset will be converted to multiple columns using a “pivot” function to enable easier analysis and visualization of monthly trends.

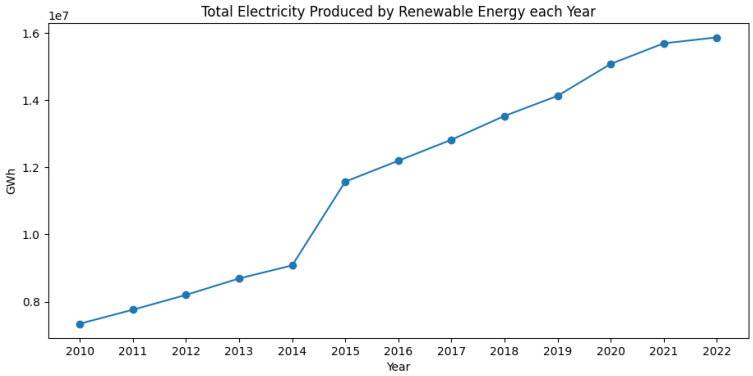
Exploratory Data Analysis: We will use the “pandas”, “matplotlib” packages in Python to design clear plots and visualizations to gain insights into the patterns and trends in the data.

Data Modeling: We plan to use “scikit-learn” package in Python to build predictive models on the data. We will use regression analysis to explore the relationship for electricity statistics for different countries and will evaluate the performance of the models using appropriate metrics and select the best model for our analysis.

Preliminary results:

We have successfully loaded and explored the dataset, and have identified some interesting patterns in the data. For example, we have done some analysis and visualized the production of electricity throughout the years using renewable and non-renewable energy resources . These preliminary results suggest that there is indeed a relationship between electricity production and energy resources, and that further analysis is warranted.





References:

Electricity consumption:

- Dataset: https://www.iea.org/data-and-statistics/data-product/monthly-electricity-statistics
- Documentation: https://iea.blob.core.windows.net/assets/18d269a6-777a-4368-b073-c407a2b8e44c/Monthlyelectricitystatistics_Documentation.pdf

World Bank - World Development Indicators. Retrieved from

- https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?end=2021&start=1960&view=chart
- https://data.worldbank.org/indicator/SP.POP.TOTL