DataVerse – T17

Renewable Energy Analysis

1985 -2021

DataVerse Team:

Batool Khanfar

Diyaa Alhariri

Omar AlTahleh

Maryam Odat

Table of Contents

01 Introduction

02 Data Collection

03 Exploratory Data Analysis

23 Model

27 Dashboard

30 DataFolio

31 Findings

32 Acknowledgement

Introduction

Business Problem

The renewable energy sector is experiencing rapid growth globally, and accurately determining the optimal capacity required to fully harness its potential is a critical challenge. This data analytic project aims to address this problem by analyzing and predicting the necessary renewable energy generation capacity while considering variables such as energy consumption, GDP, population, and renewable installed capacity plants.

The renewable energy industry is a significant player in the global economy. According to industry reports, the global renewable energy generation capacity reached 2,799 GW in 2020, with a compound annual growth rate of 7.6% between 2016 and 2020. This growth necessitates robust analysis to determine the optimal capacity needed for sustainable economic development.

Business Impact

The analysis conducted in this project will have a substantial impact on the renewable energy sector and broader economic development. By accurately determining the optimal renewable energy generation capacity, stakeholders can make data-driven decisions about investment, infrastructure development, and policy frameworks.

The analysis will help optimize resource allocation by considering variables such as energy consumption, GDP, population, and renewable installed capacity plants. For example, countries with high GDP and population growth rates may require a larger renewable energy generation capacity to support their economic development while reducing reliance on fossil fuels. By aligning the renewable energy generation capacity with these factors, stakeholders can ensure efficient and sustainable energy planning.

Furthermore, the analysis will have implications for job creation, economic growth, and environmental sustainability. By accurately estimating the necessary renewable energy generation capacity, it becomes possible to drive investments in renewable energy projects and promote the development of a skilled workforce. Studies indicate that the renewable energy sector has the potential to create millions of jobs globally while contributing to economic growth. Moreover, by reducing reliance on fossil fuels, the analysis will contribute to mitigating greenhouse gas emissions and combating climate change.

Overall, this data analytic project's findings will provide valuable insights for policymakers, investors, and stakeholders in the renewable energy sector. By leveraging data analysis techniques and considering variables such as renewable energy generation, energy consumption, GDP, population, and renewable installed capacity plants, stakeholders can make informed decisions to optimize resource allocation, drive economic growth, and promote environmental sustainability in the transition towards a cleaner and more sustainable energy future.

Data Collection

We have access to an extensive Renewable Energy dataset, covering the period from 1965 to 2022, which includes data from diverse renewable energy sources worldwide. The dataset is hosted on Kaggle and consists of 17 files, each providing valuable information about various electricity production sources, including hydro, wind, solar, biofuel, geothermal, and more. The dataset is publicly available at Renewable Energy Worldwide.

Among these files, we will specifically utilize the "modern-renewable-prod" file, which provides detailed columns for various renewable energy sources and their respective electricity production measured in Terawatt-hours (TWh), equivalent to 10^12 watt-hours.

Column Details:
1. Entity: Indicates the country or region associated with the renewable energy data.
2. Code: A three-letter code uniquely assigned to each country or region in the dataset.
3. Year: Represents the year corresponding to the renewable energy data.
4. Electricity from wind (TWh): Presents the total electricity generation in Terawatt-hours (TWh) from wind sources.
5. Electricity from hydro (TWh): Shows the total electricity generation in Terawatt-hours (TWh) from hydropower sources.
6. Electricity from solar (TWh): Displays the total electricity generation in Terawatt-hours (TWh) from solar sources.
7. Other renewables including bioenergy (TWh): Indicates the total energy generation in Terawatt-hours (TWh) from other renewables, including bioenergy.
8. Renewables (% electricity): Denotes the percentage of electricity generation derived from renewable sources.

In order to insert new features to the dataset, we require additional features such as the area of the country, population, and GDP. We obtained this information from external resources.
For the country area, we utilized https://countrycode.org and connected the website to an Excel file, which was then processed using pandas to clean and extract data under the "code country" column.

For population data, we retrieved it from the World Bank through the following link: https://data.worldbank.org/indicator/SP.POP.TOTL. Furthermore, we acquired GDP data from the dataset found at https://ourworldindata.org/grapher/gdp-world-regions-stacked-area. Specifically, we utilized the "GDP" column from this dataset.

The data from the GDP dataset was initially in a pivot format, so we performed an unpivot operation to reshape the data from wide to long format. We leveraged Python, pandas, and pandasql to execute SQL code within pandas and merged the data using the country code as a key.
After merging the datasets, we obtained a new dataset with 5252 rows and 17 columns. The time span of the data ranges from 1985 to 2021. To standardize the energy production values, we converted them from TWh to gigawatt-hours (GWh).

Exploratory Data Analysis

Through our analysis of global renewable energy production, energy consumption, GDP, and population data, our analysis has provided valuable insights into energy practices worldwide. This section presents the findings of our exploratory data analysis (EDA) on global renewable energy production, energy consumption, GDP, and population data. It offers essential perspectives on energy practices for policymakers, energy planners, and stakeholders. The analysis emphasizes significant trends, correlations, and opportunities, thereby guiding informed decision-making towards a greener and more sustainable future.

- **Statistical Summary:**

The information gives a quick and simple description of the data, which is include Count, Mean, Standard Deviation, median, mode, minimum value, maximum value, range, standard deviation, etc.

Table 1. Descriptive Statistical Analysis Table.

	count	mean	std	min	25%	50%	75%	max
Unnamed: 0	5253.0	2.626000e+03	1.516555e+03	0.000000e+00	1.313000e+03	2.626000e+03	3.939000e+03	5.252000e+03
Year	5253.0	2.006182e+03	9.200711e+00	1.985000e+03	2.000000e+03	2.007000e+03	2.014000e+03	2.021000e+03
Electricity from wind (TWh)	5253.0	2.461154e+00	2.000582e+01	0.000000e+00	0.000000e+00	0.000000e+00	4.000000e-02	6.556000e+02
Electricity_from_hydro_TWh	5252.0	2.005887e+01	7.368310e+01	0.000000e+00	2.000000e-02	1.503000e+00	9.702500e+00	1.321710e+03
Electricity_from_solar_TWh	5253.0	8.980895e-01	8.993691e+00	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e-02	3.270000e+02
Other renewables_including_bioenergy_TWh	5253.0	2.133536e+00	8.446564e+00	0.000000e+00	0.000000e+00	1.000000e-02	5.000000e-01	1.699316e+02
Total_enewable_energy_production_TWh	5253.0	2.554784e+01	9.987509e+01	0.000000e+00	7.000000e-02	2.220000e+00	1.288000e+01	2.452532e+03
Renewables_percentage_electricity	5253.0	3.031623e+01	3.236187e+01	0.000000e+00	1.515151e+00	1.666667e+01	5.476191e+01	1.000000e+02
AREA_KM2	5027.0	8.396666e+05	2.182110e+06	2.100000e+01	2.874800e+04	1.632700e+05	6.003700e+05	1.710000e+07
Population	5253.0	4.168255e+07	1.455070e+08	1.023300e+04	2.140215e+06	8.746084e+06	2.869868e+07	1.412360e+09
GDP	5093.0	3.465304e+11	1.386961e+12	4.756284e+07	6.979155e+09	3.300020e+10	1.770000e+11	2.330000e+13
Primary_energy_consumption_TWh	5253.0	8.317954e+02	3.038302e+03	1.240929e-01	1.839642e+01	1.022910e+02	4.795935e+02	4.379089e+04
Primary_energy_consumption_GWh	5253.0	8.317954e+05	3.038302e+06	1.240929e+02	1.839642e+04	1.022910e+05	4.795935e+05	4.379090e+07
Electricity_from_wind_GWh	5253.0	2.461154e+03	2.000582e+04	0.000000e+00	0.000000e+00	0.000000e+00	4.000000e+01	6.556000e+05
Electricity_from_hydro_GWh	5252.0	2.005887e+04	7.368310e+04	0.000000e+00	2.000000e+01	1.503000e+03	9.702500e+03	1.321710e+06
Electricity_from_solar_GWh	5253.0	8.980895e+02	8.993691e+03	0.000000e+00	0.000000e+00	0.000000e+00	1.000000e+01	3.270000e+05
Other_renewables_including_bioenergy_GWh	5253.0	2.133536e+03	8.446564e+03	0.000000e+00	0.000000e+00	1.000000e+01	5.000000e+02	1.699316e+05
Total_Renewable_energy_production_GWh	5253.0	2.554784e+04	9.987509e+04	0.000000e+00	7.000000e+01	2.220000e+03	1.288000e+04	2.452532e+06

- **Correlation :**

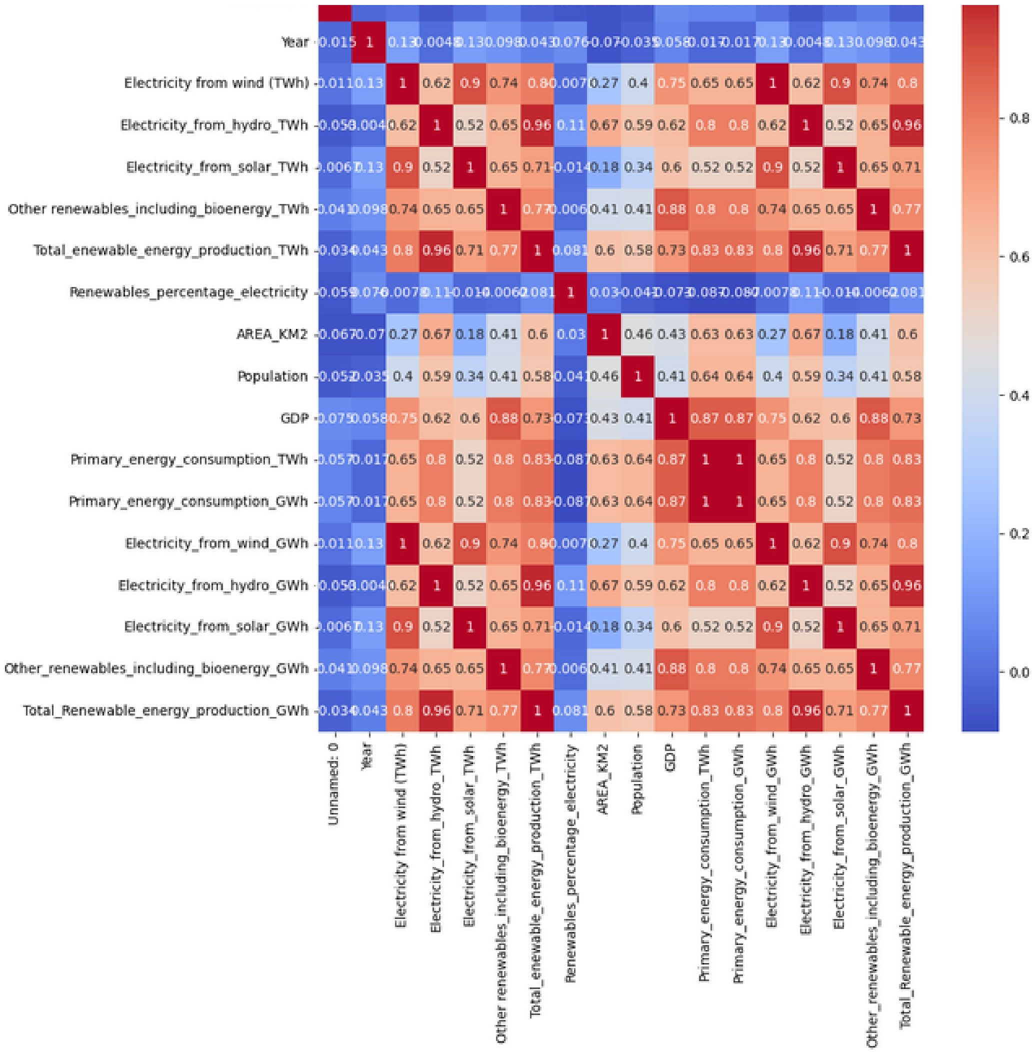

Figure 1. Correlation Heat Map.

The exploratory data analysis (EDA) conducted on global renewable energy production, energy consumption, GDP, and population data reveals significant correlations between various variables. The heatmap visualizes these correlations, with darker shades of red indicating strong positive correlations and darker shades of blue indicating strong negative correlations. The correlation coefficient, ranging from -1 to 1, provides insights into the strength and direction of the relationships between the variables.

Positive Correlations:

1. Electricity generation from wind (Electricity_from_wind_GWh) exhibits a robust positive correlation of 0.9 with electricity generation from solar (Electricity_from_solar_GWh). This finding suggests that an increase in wind electricity generation corresponds to an increase in solar electricity generation, and vice versa, indicating a harmonious coexistence between these renewable energy sources.

2. Electricity generation from wind (Electricity_from_wind_GWh) is positively correlated (0.75) with Gross Domestic Product (GDP), indicating a potential link between wind electricity generation and the economic performance of the region represented by GDP. This insight underscores the economic significance of investing in wind energy infrastructure.
3. Electricity generation from hydro sources (Electricity_from_hydro_GWh) shows a positive correlation of 0.8 with primary energy consumption (Primary_energy_consumption_TWh). This suggests that an increase in hydro electricity generation is linked to higher primary energy consumption, showcasing the role of hydroelectric power in meeting energy demands.
4. Electricity generation from other renewables (including bioenergy) (Other_renewables_including_bioenergy_GWh) demonstrates a strong positive correlation of 0.88 with GDP, indicating that the generation of electricity from other renewable sources aligns with higher economic performance represented by GDP. This emphasizes the economic benefits of diversifying the renewable energy mix.
5. Electricity generation from other renewables (including bioenergy) (Other_renewables_including_bioenergy_GWh) is positively correlated (0.8) with primary energy consumption (Primary_energy_consumption_TWh), signifying that increased electricity generation from these sources is associated with higher primary energy consumption, supporting the idea of using various renewables to meet energy demands sustainably.
6. Primary energy consumption (Primary_energy_consumption_TWh) is positively correlated (0.87) with GDP, suggesting that higher energy consumption tends to be associated with stronger economic performance represented by GDP. This highlights the integral role of energy consumption in driving economic growth.

Negative Correlations:
1. Electricity generation from wind (Electricity_from_wind_GWh) exhibits a negative correlation of -0.27 with the area in square kilometers (AREA_KM2). This indicates that as the area decreases, electricity generation from wind sources tends to increase. This finding emphasizes the potential benefits of wind energy in regions with limited space for large-scale installations.
2. Electricity generation from solar (Electricity_from_solar_GWh) shows a negative correlation of -0.18 with the area in square kilometers (AREA_KM2). As the area decreases, electricity generation from solar sources tends to increase. This underscores the adaptability and efficiency of solar energy installations in confined spaces.

In conclusion, the analysis sheds light on the interconnections between renewable energy production, energy consumption, GDP, and population. The positive correlations highlight opportunities for sustainable energy planning and economic growth, while the negative correlations provide valuable insights for optimizing renewable energy deployment in spatially constrained regions. These findings serve as a valuable foundation for informed decision-making and pave the way for a greener and more sustainable future.

- **Electricity Production From Wind By Continent:**

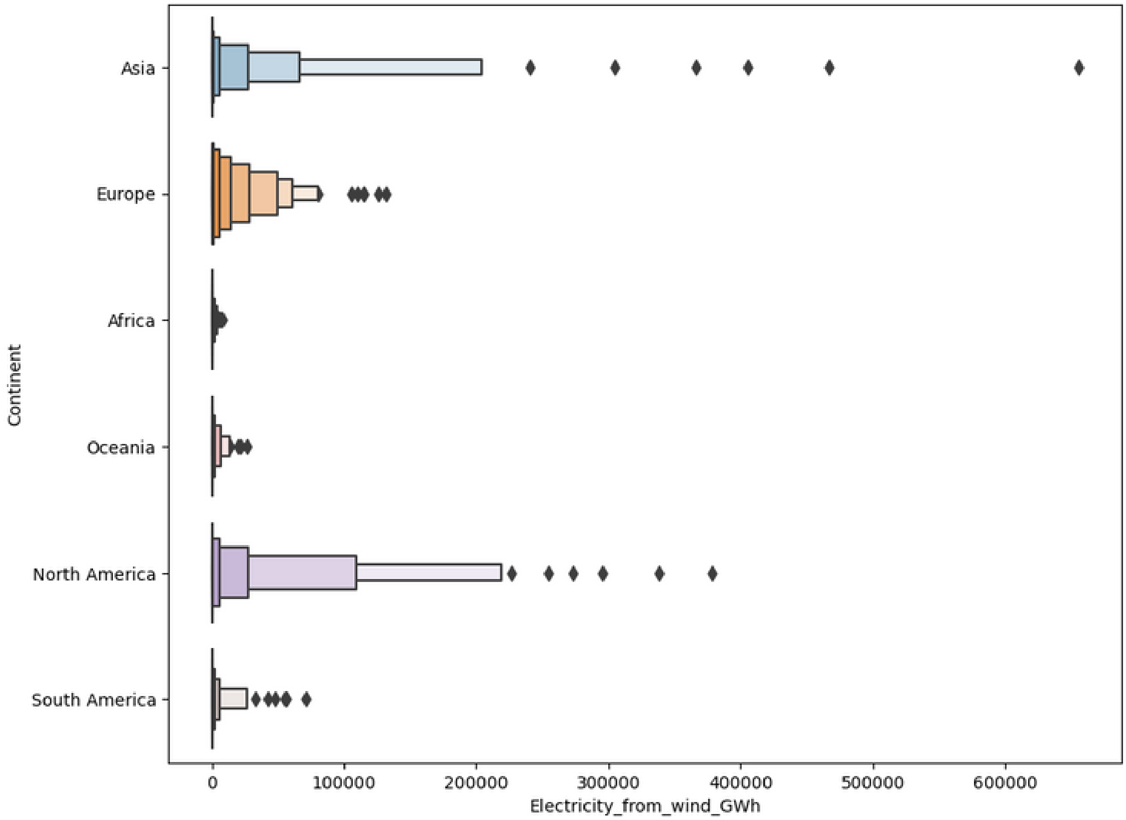

Chart 1. Electricity Production From Wind By Continent.

From the Boxplot for Electricity from wind , we can have below observations:

1. North America has the highest median 'Electricity_from_wind_GWh':

This could be attributed to the availability of favorable geographical locations and wind patterns in North America, which are conducive to harnessing wind energy efficiently. Additionally, North America may have invested significantly in wind energy infrastructure and technology, leading to higher wind energy generation.

2. Asia has the second-highest median 'Electricity_from_wind_GWh':

Asia is a vast and diverse continent with countries like China and India, which have been actively adopting renewable energy, including wind power, to meet their growing energy demands and reduce carbon emissions.

3. Europe has the third-highest median 'Electricity_from_wind_GWh':

Europe has been a pioneer in renewable energy adoption, and many European countries have implemented policies and incentives to promote wind energy development. and Public awareness and support for renewable energy, including wind power, are generally high in many European countries. Citizens often express a willingness to support and participate in renewable energy projects, such as community wind farms, through crowd-funding initiatives. This positive attitude towards renewable energy fosters a conducive environment for the development and expansion of wind energy projects in Europe.This has resulted in a substantial contribution to wind energy generation in the continent.

4. South America has the fourth-highest median 'Electricity_from_wind_GWh':

Countries in South America, such as Brazil and Chile, have been increasing their wind energy capacity, driven by environmental concerns and a desire to diversify their energy mix. Many regions in South America have access to favorable wind resources, especially in coastal areas and high-altitude regions. The presence of strong and consistent winds makes these areas suitable for wind energy development, contributing to higher wind energy generation. These efforts have led to a notable increase in wind energy generation.

5. Oceania has the fifth-highest median 'Electricity_from_wind_GWh':

Oceania includes countries like Australia and New Zealand, which have been actively investing in renewable energy sources, including wind power. The presence of strong wind resources in certain regions of Oceania contributes to its wind energy generation. Oceania is surrounded by vast coastlines, offering significant potential for offshore wind energy development. Offshore wind farms in Oceania benefit from stronger and more consistent winds compared to onshore locations, contributing to higher wind energy generation.

6. Africa has the lowest median 'Electricity_from_wind_GWh':

While wind energy is gradually gaining traction in some African countries, the continent's wind energy potential is relatively underexploited compared to other continents. Factors like:

- Geographical Challenges:

Africa's vast and diverse geography presents challenges for wind energy development. Suitable wind energy sites may be located in remote or hard-to-access areas, making it difficult and costly to establish and maintain wind farms.

- Energy Infrastructure Limitations:

Many regions in Africa face limitations in their overall energy infrastructure. The lack of a well-developed power grid can hinder the transmission and distribution of wind energy to areas where it is needed, resulting in lower wind energy generation.

- Financial and Investment Constraints:

The development of wind energy projects requires significant upfront investments. In many African countries, financial constraints and a lack of access to capital can impede the scaling up of wind energy projects.

- Access to Technology and Equipment:

Limited access to specialized wind energy technology and equipment may hinder the expansion of wind energy infrastructure in Africa.

- Dependency on Traditional Energy Sources:

Some African countries heavily rely on traditional energy sources such as fossil fuels or biomass for their energy needs. Transitioning to renewable energy sources like wind may require overcoming resistance and ensuring energy security during the transition period. contribute to lower wind energy generation in Africa.

Electricity Production From Hydro By Continent:

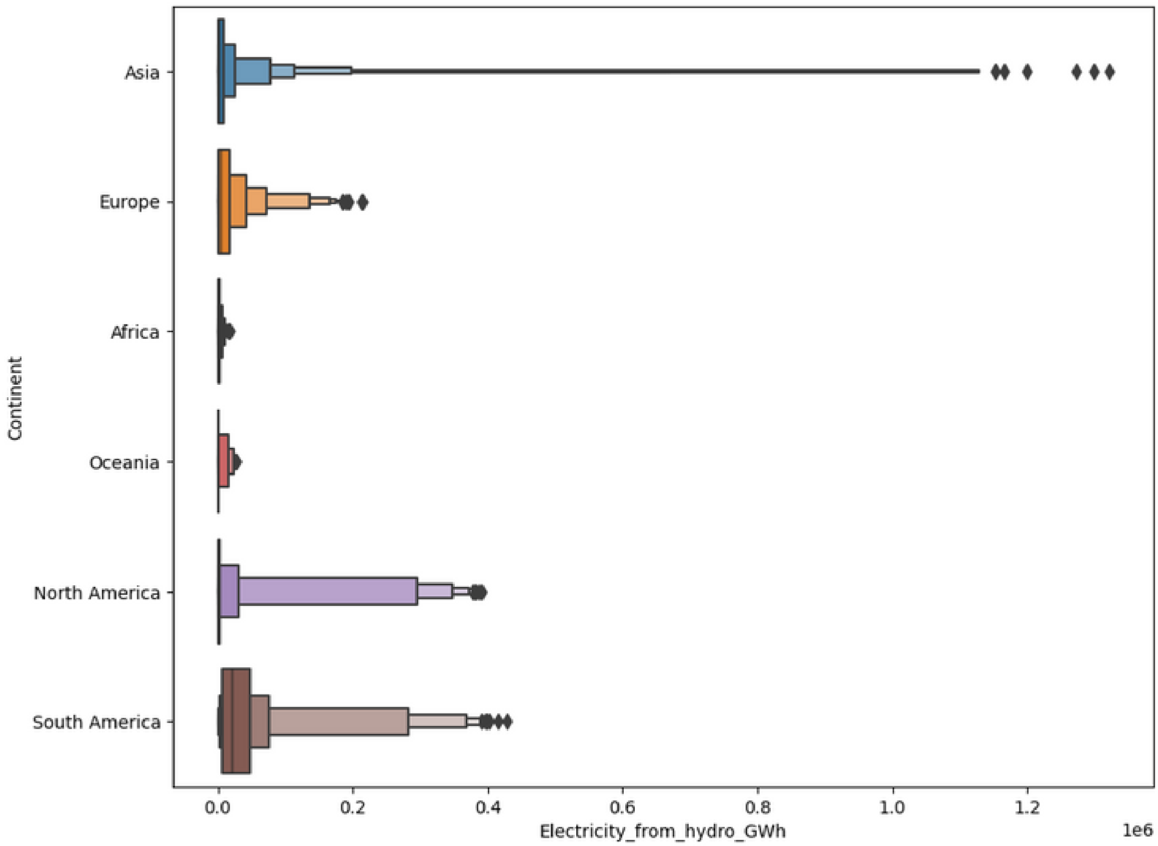

Chart 2. Electricity Production From Hydro By Continent.

From the Boxplot for Electricity from Hydro , we can have below observations:

1. **North America has the highest median 'Electricity_from_hydro_GWh':**

This could be attributed to the presence of large rivers and abundant water resources in North America, offering significant hydroelectric power generation potential. North America has invested in the development of hydroelectric projects, taking advantage of its favorable geographical features, such as the Columbia River and Niagara Falls.

2. **South America has the second-highest median 'Electricity_from_hydro_GWh':**

South America is home to some of the world's largest rivers, such as the Amazon and Paraná rivers, providing ample opportunities for hydroelectric power generation. Countries like Brazil and Venezuela have heavily invested in hydropower infrastructure, contributing to a notable increase in hydroelectric generation in the continent.

3. **Asia has the third-highest median 'Electricity_from_hydro_GWh':**
Asia is geographically diverse, with several large rivers like the Yangtze, Mekong, and Ganges-Brahmaputra, providing extensive potential for hydroelectric power generation. Countries like China and India have undertaken significant hydroelectric projects, adding to the overall hydroelectric capacity in the region.

4. **Europe has the fourth-highest median 'Electricity_from_hydro_GWh':**
Europe has developed a long history of harnessing hydropower, with countries like Norway and Sweden heavily reliant on hydroelectric generation. European nations have invested in upgrading and modernizing existing hydroelectric facilities to optimize electricity generation.

5. **Oceania has the fifth-highest median 'Electricity_from_hydro_GWh':**
Oceania includes countries like New Zealand and Papua New Guinea, which have been investing in hydroelectric projects to meet their energy demands and reduce reliance on fossil fuels. Oceania's topography, including mountainous regions, contributes to the availability of suitable sites for hydroelectric power plants.

6. **Africa has the lowest median 'Electricity_from_hydro_GWh':**
While Africa has substantial hydroelectric potential, a large portion of it remains untapped due to various challenges. Factors like limited infrastructure, funding constraints, and political instability have hindered the development of large-scale hydroelectric projects in Africa.

Electricity Production From Solar By Continent:

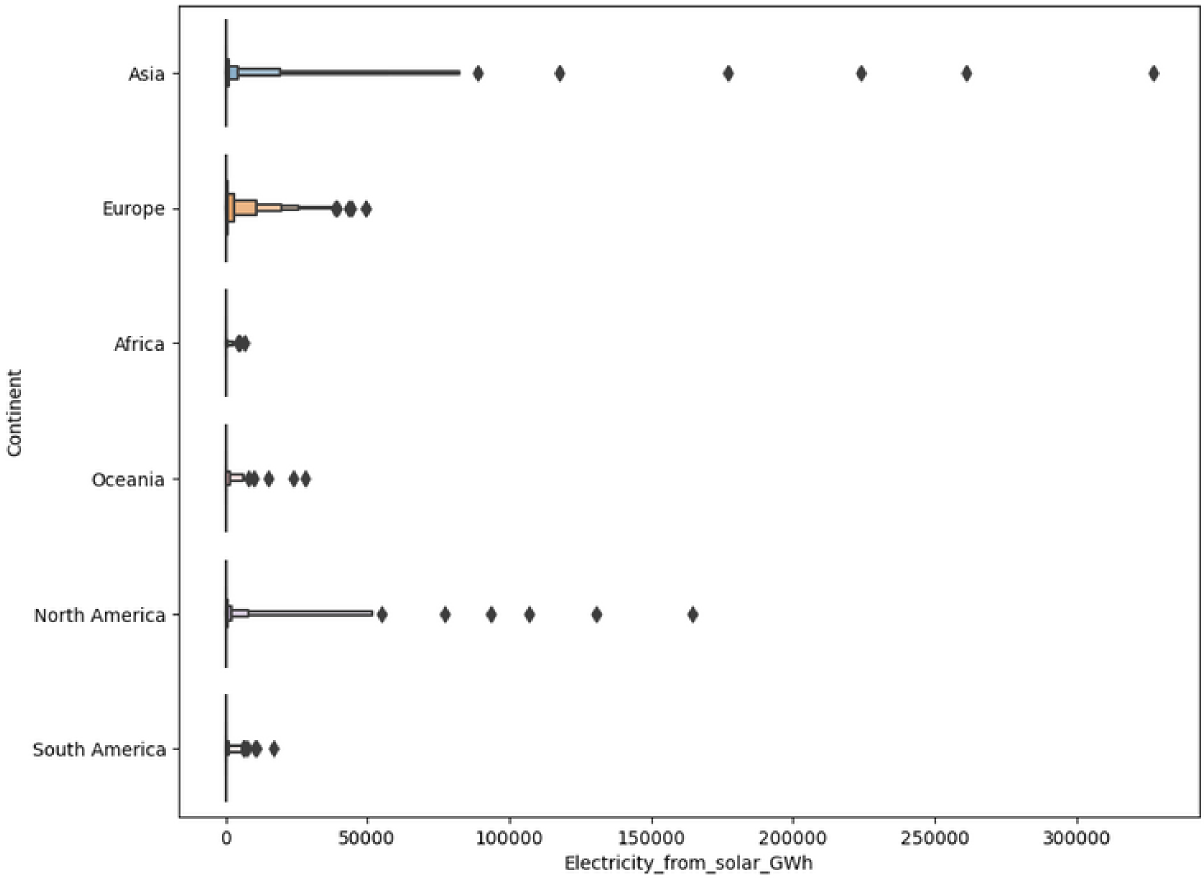

Chart 3. Electricity Production From Solar By Continent.

From the Boxplot for Electricity from Solar , we can have below observations:

1. **Asia has the highest median 'Electricity_from_solar_GWh':**
Abundant Solar Resources Asia has a diverse climate and a large landmass with ample sunlight in many regions. Countries located closer to the equator, such as India, receive substantial solar irradiance, making solar energy a viable and attractive option for electricity generation.

2. **Europe has the second-highest median 'Electricity_from_solar_GWh':**
European countries have implemented various supportive policies, feed-in tariffs, and renewable energy targets to promote solar energy adoption. These policies have incentivized investments in solar power generation. Europe has been a leader in solar energy research and development. Technological advancements have increased the efficiency of solar panels and reduced the cost of solar installations, making solar energy more competitive.

3. **North America has the third-highest median 'Electricity_from_solar_GWh':**
The United States has been a key driver of solar energy growth in North America. The country's commitment to clean energy, combined with decreasing solar panel costs, has led to increased solar energy adoption.

4. **Oceania has the fourth-highest median 'Electricity_from_solar_GWh':**
Island Nations and Remote Areas: Oceania includes numerous island nations and remote regions with limited access to traditional energy sources. Solar energy offers a sustainable and decentralized solution to meet their electricity needs.

5. **South America has the fifth-highest median 'Electricity_from_solar_GWh':**
Solar energy is still an emerging sector in many South American countries. While the region has vast solar potential, the pace of solar energy adoption may be slower compared to other continents due to various economic and policy factors.

6. **Africa has the lowest median 'Electricity_from_solar_GWh':**
Energy Infrastructure Challenges: Some African countries face challenges in energy infrastructure development and access to financing, which can hinder large-scale solar energy projects.. Africa's energy mix heavily relies on traditional sources like fossil fuels and biomass. Solar energy may not be the primary focus for some countries at present, but this is gradually changing as awareness of renewable energy benefits increases.

Electricity Production From Bioenergy By Continent:

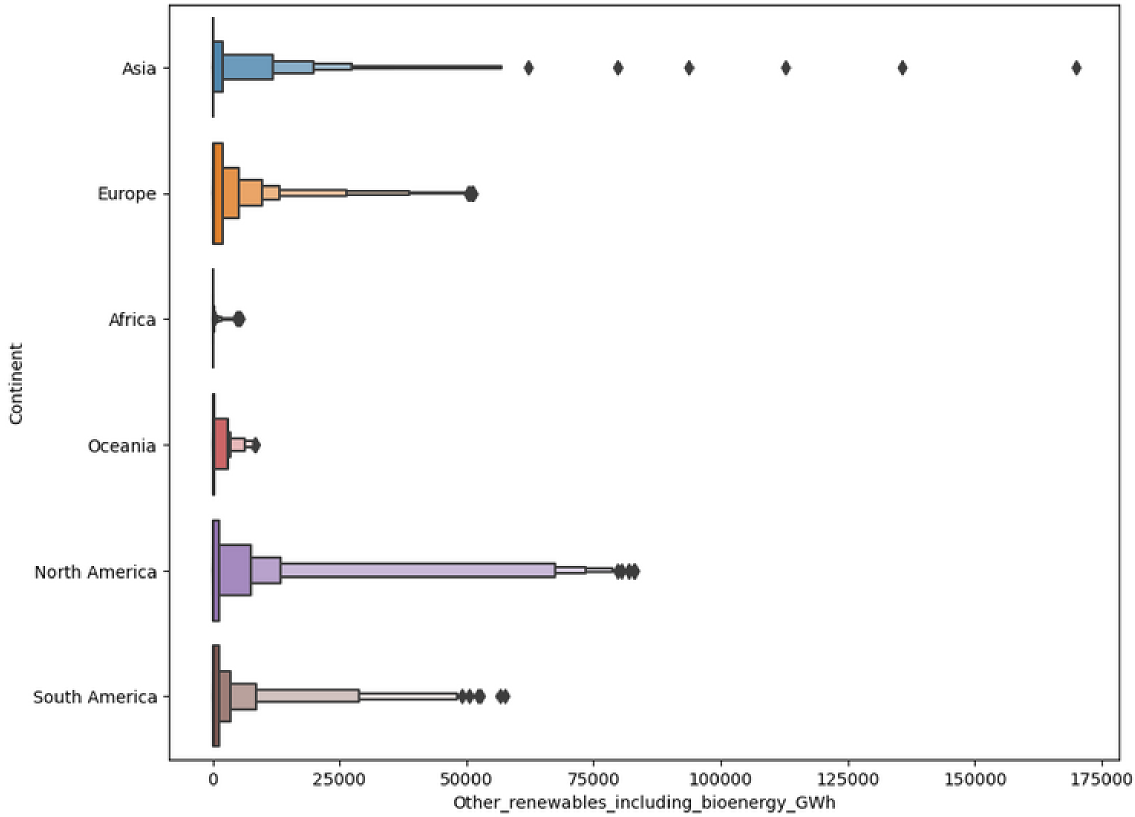

Chart 4. Electricity Production From Bioenergy By Continent.

From the Boxplot for Electricity from Bioenergy , we can have below observations:

1. **North America has the highest median 'Electricity_from_Bioenergy_GWh':**
Abundance of Biomass Resources: North America possesses significant biomass resources, including agricultural residues, forest products, and organic waste. These resources serve as feedstock for bioenergy production, contributing to higher bioenergy generation.

2. **Asia has the second-highest median 'Electricity_from_Bioenergy_GWh':**
Agricultural and Forestry Sector: Asia has large agricultural and forestry industries that generate significant biomass residues. Utilizing these residues for bioenergy production provides an attractive solution for sustainable energy generation.

3. **Europe has the third-highest median 'Electricity_from_Bioenergy_GWh':**
Biomass and Waste-to-Energy Policies: European countries have implemented policies to promote the use of biomass and waste-to-energy technologies for bioenergy production. These policies support the growth of the bioenergy sector.

4. **South America has the fourth-highest median 'Electricity_from_Bioenergy_GWh':**
Biomass-rich Environment: The region has vast biomass resources from agricultural activities and forestry. Utilizing these resources for bioenergy production has contributed to higher bioenergy generation.

5. **Oceania has the fifth-highest median 'Electricity_from_Bioenergy_GWh':**
Biomass Utilization: Oceania countries have been exploring biomass utilization for bioenergy generation. The presence of biomass resources, such as agricultural residues and organic waste, contributes to bioenergy production.

6. **Africa has the lowest median 'Electricity_from_Bioenergy_GWh':**
Limited Infrastructure: Some African countries face challenges in developing the necessary infrastructure for large-scale bioenergy projects, which affects overall bioenergy generation. Economic Constraints: Economic constraints may hinder investments in bioenergy projects, limiting the growth of the bioenergy sector in Africa.

Production of Renewable Energy And The Primary Energy Consumption.:

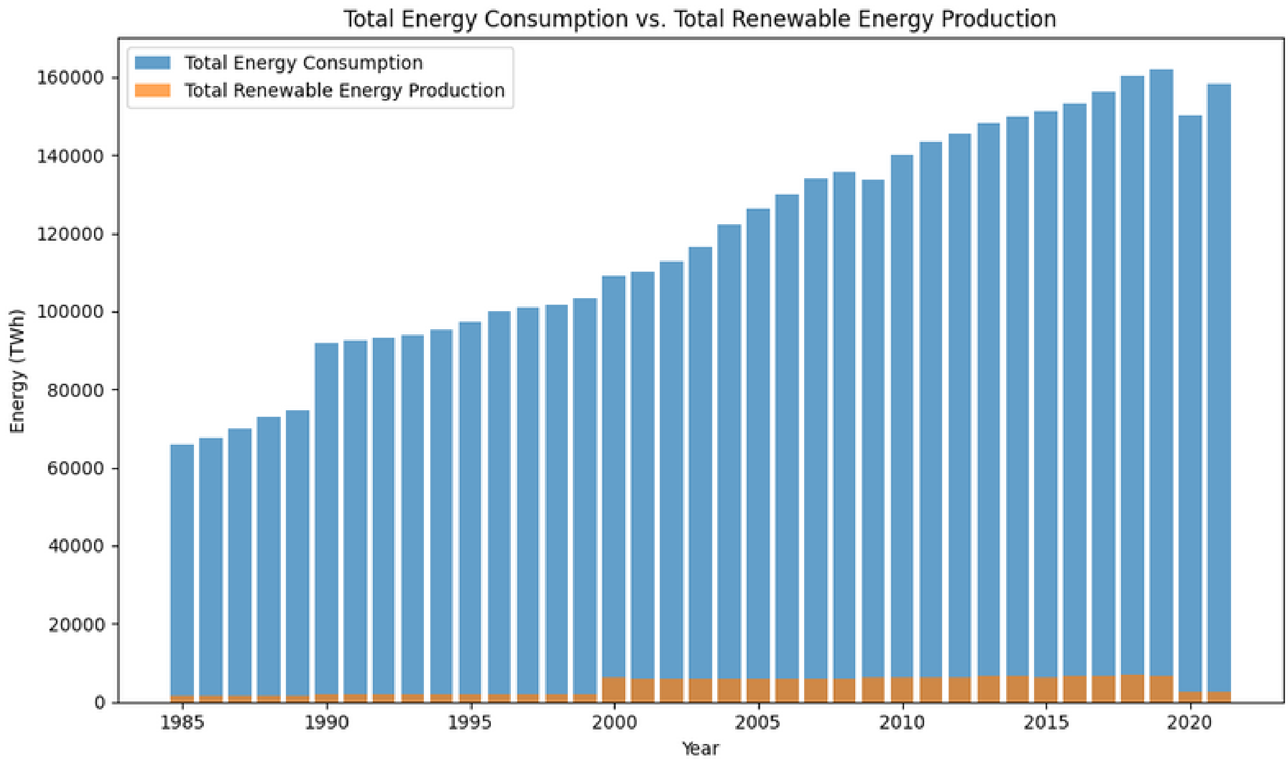

Chart 4. Production of Renewable Energy And The Primary Energy Consumption.1985 - 2021.

The bar chart represents the total renewable energy generation and total energy consumption on a global scale. The chart clearly illustrates that the global energy consumption is significantly larger than the total renewable energy production. The stark difference in the bar lengths highlights the considerable gap between the two figures, emphasizing the world's heavy reliance on non-renewable energy sources. This visual representation underscores the urgent need for sustainable energy practices and the transition towards greater renewable energy adoption to address the growing energy demands while mitigating environmental impacts.

- **Production of Renewable Energy Sources:**

The examination of trends in electricity generation from Wind, Hydro, Solar, and Bioenergy from 1985 to 2021 enables us to discern the ranking of the top renewable energy sources based on their respective electricity generation capacities.

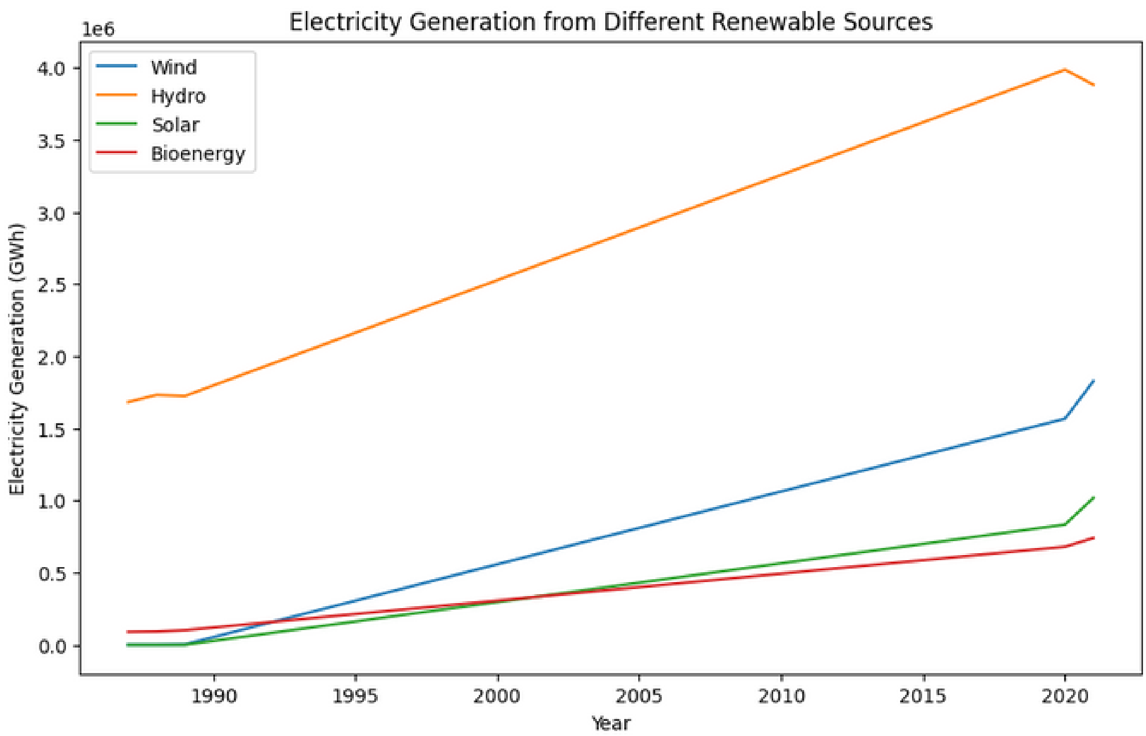

Chart 5. Energy Production Of The Renewable Energy Sources 1985 - 2021.

The line chart above displays the electricity generation from four renewable energy sources - Hydro, Wind, Solar, and Bioenergy - from 1985 to 2021. Overall, we can observe a consistent increase in electricity production from these sources over the years. Notably, starting from 2020, there is a significant surge in electricity generation from Wind, Solar, and Bioenergy, while Hydroelectric generation experiences a noticeable decrease during the same period.

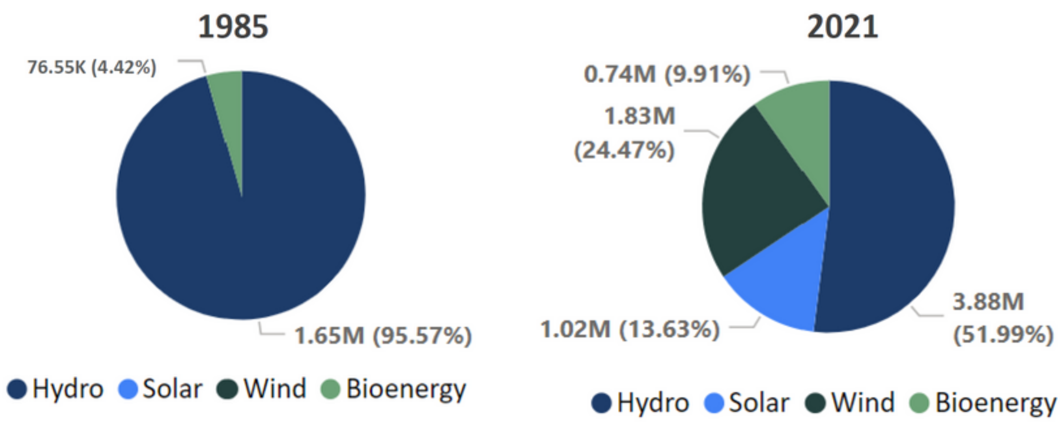

Chart 6. Energy Production Of The Renewable Energy Sources 1985 & 2021.

51.99% hydroelectric is the leading renewable energy source in 2021.

The two pie charts illustrate the production of renewable energy sources in the years 1985 and 2021. In 1985, hydroelectric production accounted for a significant 95.57% of the renewable energy output, representing one of the earliest and most established renewable energy sources in history. Bioenergy contributed 4.42%, while the other sources had minimal contributions.

Fast forward to 2021, and the landscape has evolved. Hydroelectric production decreased to 51.99%, but it still remains the leading renewable energy source, generating 3.88 million units of energy. Wind power follows as the second-highest contributor at 24.47%, producing 1.83 million units, while solar energy makes up 13.63% and contributes 1.02 million units. In contrast, bioenergy shows a slight growth, generating 0.74 million units and accounting for 9.91% of the total renewable energy production in 2021. This trend demonstrates the growing importance of various renewable sources and the increasing diversity in the global energy mix.

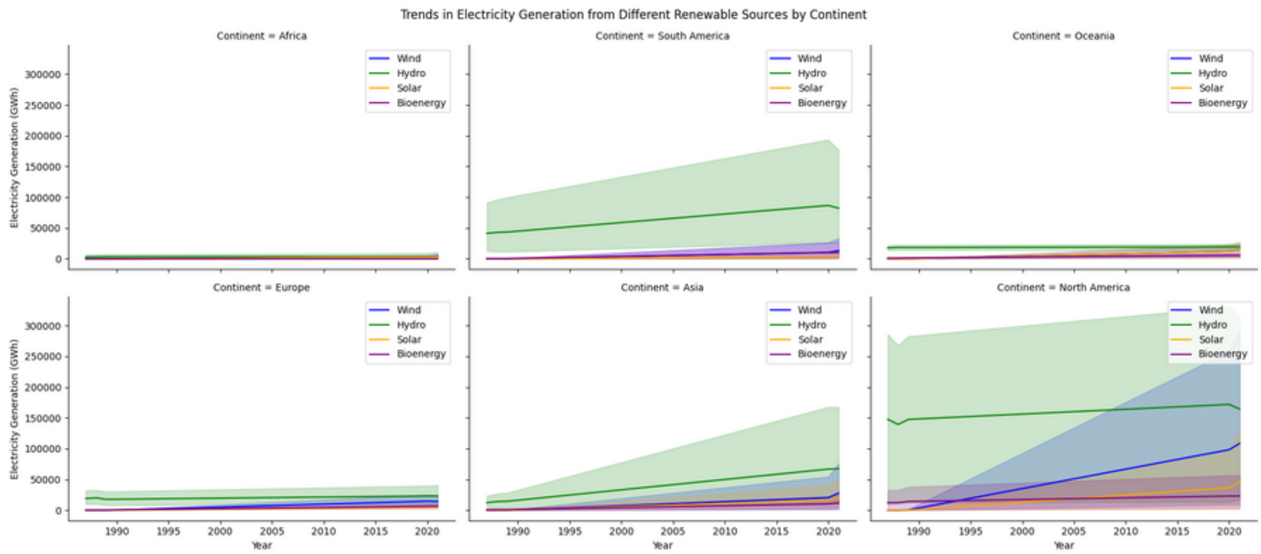

Chart 7. Energy Production Of The Renewable Energy Sources 1985 – 2021 By Continent.

1) **Africa** having lower values in electricity generation from different renewable energy sources (wind, hydro, solar, bioenergy) compared to other regions can be attributed to several factors:

1. Infrastructure and Investment Challenges:
2. Many African countries face challenges in developing and maintaining energy infrastructure for renewable energy projects. Lack of investment and funding constraints can hinder the growth of renewable energy installations.
3. Limited Technological Adoption:
4. The adoption of advanced renewable energy technologies might be slower in some African regions due to limited access to modern technologies, research, and development.
5. Dependency on Traditional Energy Sources:
6. Many African countries still heavily rely on traditional energy sources like biomass and fossil fuels for their energy needs. The transition to renewable energy may be hindered by existing energy dependencies.
7. Economic Constraints:
8. Economic factors can play a significant role in the implementation of renewable energy projects. High upfront costs and financing challenges can be barriers to the development of renewable energy infrastructure.

2) **Oceania**

1. Isolated Geography:
2. Oceania comprises a vast area of scattered islands and remote regions, which may pose logistical challenges for large-scale renewable energy projects and grid connectivity.
3. Limited Population Density:
4. Some areas in Oceania have low population densities, which can lead to lower energy demand and, in turn, less emphasis on extensive renewable energy development.

Energy Imports:

Some Oceania countries might import a significant portion of their energy, which can reduce the incentive to develop renewable energy projects locally.

3) Europe

1. Europe has historically relied on a mix of energy sources, including nuclear and fossil fuels, which were well-established before the widespread adoption of renewable energy technologies.
2. Energy Infrastructure: Europe's existing energy infrastructure and grid systems were built to accommodate conventional energy sources. Transitioning to renewable energy may require significant upgrades and investments in the grid.
3. Land Constraints: Some European countries have limited available land for large-scale renewable energy projects, such as solar farms or wind farms, which can impact the expansion of renewable energy capacity.
4. Intermittency Challenges: Renewable energy sources like wind and solar can be intermittent and dependent on weather conditions. Balancing this variability with stable energy demand can be a challenge.

3) **Asia**

- Industrialization and Urbanization: Asia has experienced rapid industrialization and urbanization, leading to increased energy demand from conventional sources like coal and natural gas. The transition to renewable energy may be slower due to the existing energy infrastructure and demand.
- Energy Mix Dominated by Coal: Many Asian countries heavily rely on coal as a primary source of electricity generation due to its abundance and affordability. Shifting away from coal to renewable energy requires substantial investment and policy changes.
- Population and Energy Demand: Asia is home to some of the world's most populous countries, resulting in significant energy demand. Meeting this demand through renewable energy sources requires substantial capacity and infrastructure development.

Asia's high electricity generation from hydroelectric power can be attributed to several factors: 1) Abundant Water Resources: Asia is home to many large rivers and water bodies, providing abundant water resources suitable for hydroelectric power generation.

2)Energy Security and Independence: Hydropower offers a reliable and locally available energy source, reducing dependence on imported fossil fuels.

3)Infrastructure Development: Asia's rapid economic growth and urbanization have driven the need for increased electricity supply, and hydropower has been developed to meet this demand.

4) **South America**

The high hydroelectric power generation in South America can be attributed to several favourable factors:

Abundant Water Resources: South America is rich in rivers, including the Amazon River, Paraná River, and Orinoco River, among others. These rivers provide abundant water resources that are suitable for hydroelectric power generation.

Geographical Features: The continent's diverse topography includes numerous mountain ranges and highland areas, creating ideal conditions for the development of large-scale hydroelectric projects.

Renewable Energy Emphasis: Many South American countries have placed a strong emphasis on renewable energy development, and hydropower plays a crucial role in meeting sustainability goals.

Export Potential: Some South American countries export surplus hydroelectric power to neighboring regions, contributing to energy trade and cooperation in the continent.

5) **North America**

The high electricity generation from hydro and wind sources in North America can be attributed to several favorable factors:

1. Abundant Renewable Resources: North America is blessed with vast and diverse renewable energy resources. In particular, it has numerous rivers and water bodies suitable for hydroelectric power generation, as well as expansive land areas with strong and consistent wind resources.
2. Hydropower Potential: North America has several major rivers, such as the Mississippi, Colorado, and Columbia, which provide significant hydropower potential for electricity generation. There are also opportunities for the development of smaller-scale hydroelectric projects.
3. Wind Energy Potential: Many regions in North America experience consistent and strong winds, especially in areas like the Great Plains and along coastlines, making it conducive for wind energy development.
4. Energy Diversity Goals: Many North American regions aim to diversify their energy mix to reduce dependence on fossil fuels and enhance energy security. Hydro and wind power offer stable and indigenous energy sources.

Renewable Energy Production By Continent:

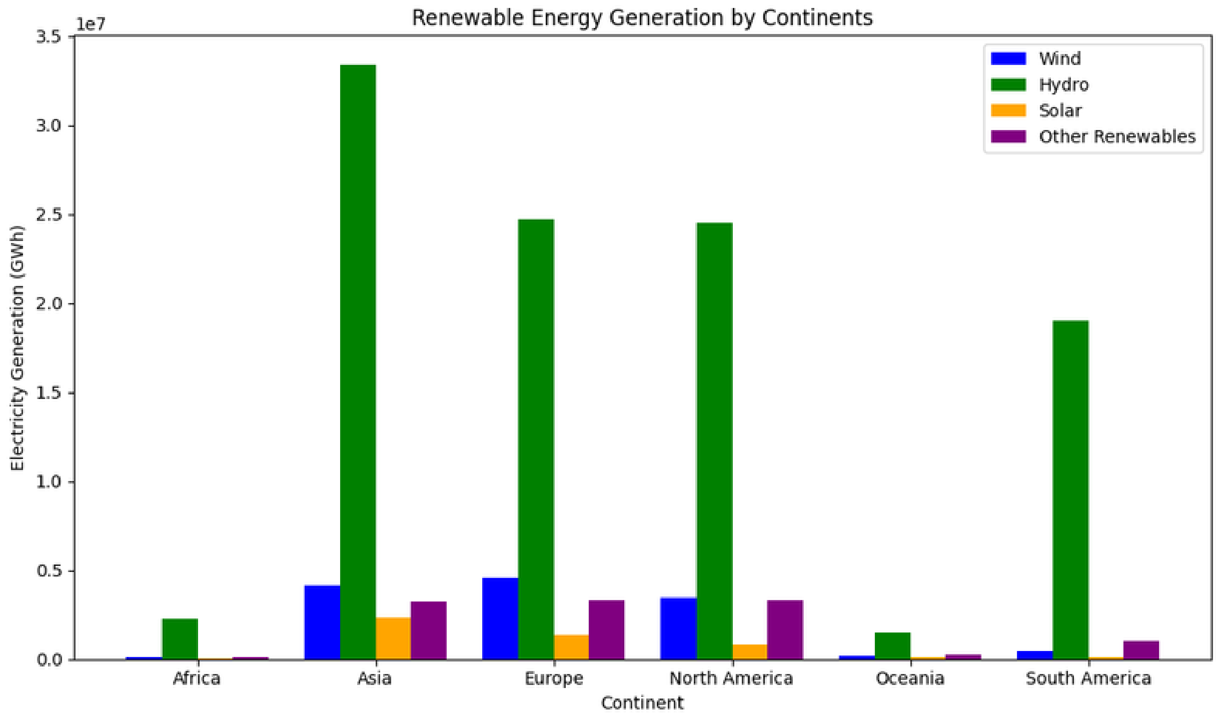

Chart 8. Renewable Energy Production By Continent.

The bar chart provides a visual representation of renewable energy production by continent, revealing significant variations among regions. Oceania emerges as the continent with the lowest renewable energy generation capacity, while Asia stands out as the highest producer.

Hydroelectric energy takes the lead as the predominant renewable energy source across all continents, signifying its widespread utilization and potential. Conversely, solar energy demonstrates the lowest contribution to renewable energy production, suggesting room for further development and adoption.

Top 10 Countries In Renewable Energy sources Production:

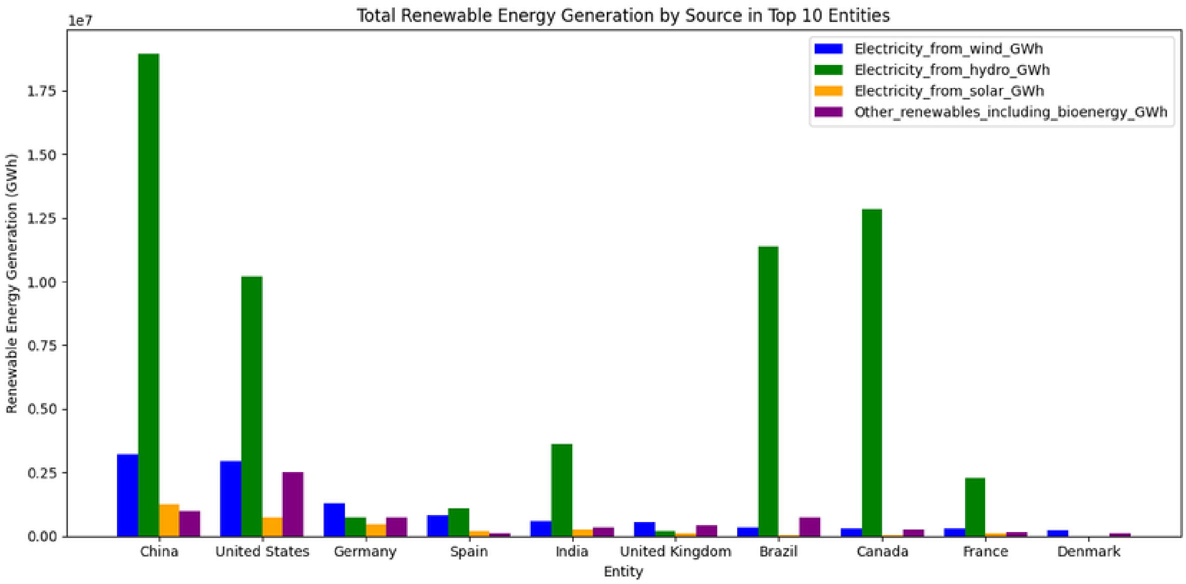

Chart 9. Top 10 Countries In Renewable Energy sources Production.

The bar chart presents a comprehensive view of the top 10 countries leading in renewable energy generation. China emerges as the global leader, followed by the United States, with Denmark occupying the 10th position in renewable energy production.

Hydroelectric energy stands out as the primary renewable energy source in the majority of these countries, signifying its widespread adoption and significant contribution to their energy portfolios. However, notable exceptions are observed in Germany, the United Kingdom, and Denmark, where hydro is not the dominant renewable energy source.

On the other hand, solar and bioenergy demonstrate comparatively lower contributions to renewable energy generation across most of the countries in the top 10 list. This highlights the potential for further development and utilization of solar and bioenergy technologies in these nations.

Renewables Sources across (AREA_KM2, Population , GDP):

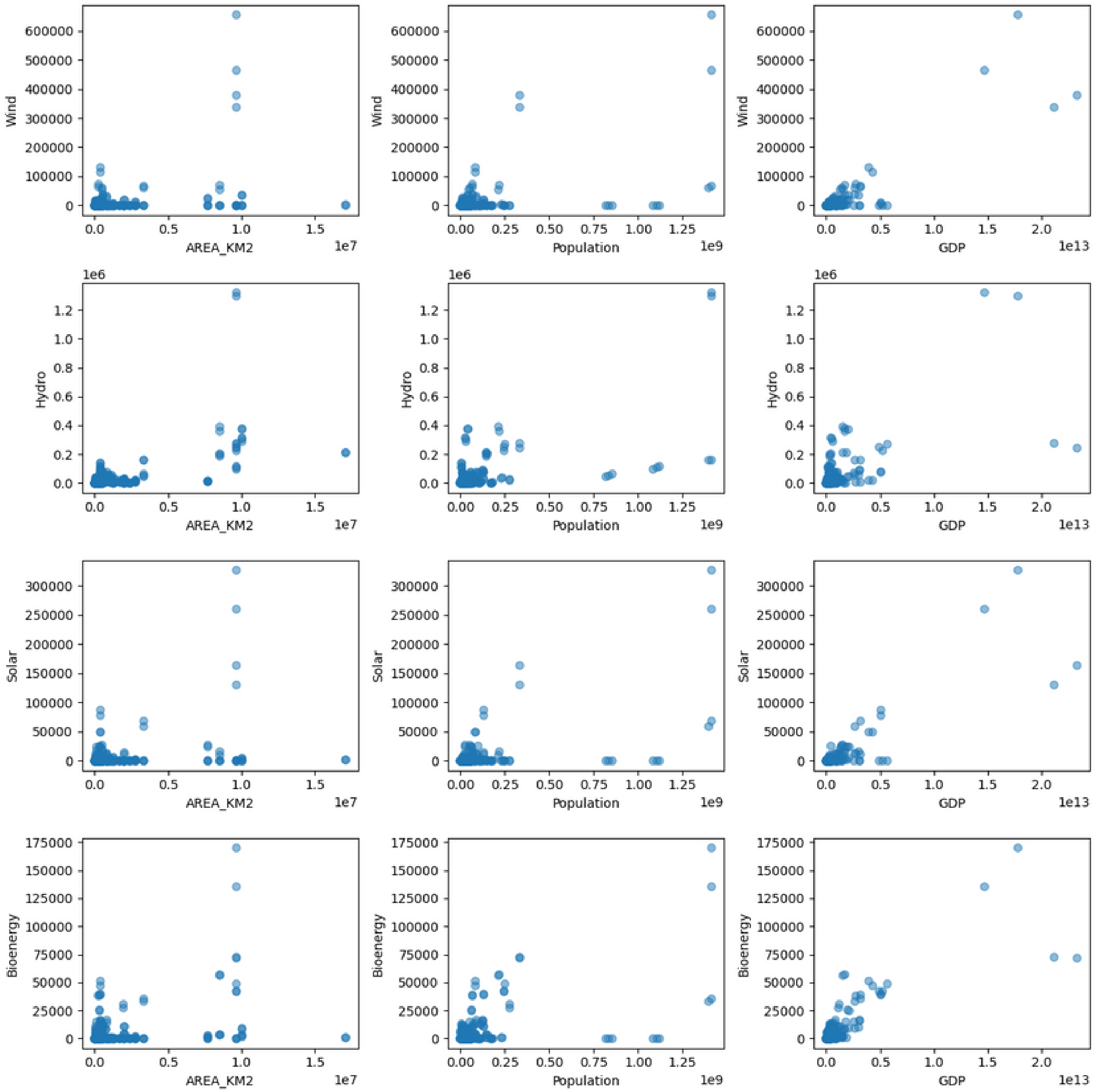

Chart 10. Renewables Sources across (AREA_KM2, Population , GDP).

From the Scatterplot, we can have below observations:

1)The scatter plot depicts the relationship between (wind energy generation) and three factors: GDP, Population, and Area_KM2. Upon analysis, it becomes evident that the primary factor that influences wind energy generation is GDP. The scatter plot points exhibit a discernible upward trend, indicating a positive correlation between wind energy generation and GDP. As GDP increases, so does the generation of wind energy. However, the scatter plot for Population and Area_KM2 reveals a relatively scattered distribution with no significant pattern, suggesting that wind energy generation does not have a substantial correlation with population density or geographical area. Consequently, the results underscore the importance of economic development (as reflected by GDP) in driving the growth and utilization of wind energy in the studied region.

2) The scatter plot illustrates the relationship between 'Electricity_from_solar_GWh' (solar energy generation) and three factors: GDP, Population, and Area_KM2. Similarly to wind energy, the analysis demonstrates that the primary factor influencing solar energy generation is GDP. The scatter plot points show a noticeable upward trend, indicating a positive correlation between solar energy generation and GDP. As GDP increases, solar energy generation also tends to increase. However, the scatter plots for Population and Area_KM2 exhibit a dispersed distribution with no significant discernible pattern, suggesting that solar energy generation does not have a strong correlation with population density or geographical area. Thus, the results reinforce the significance of economic development, represented by GDP, in driving the expansion and adoption of solar energy in the studied region.

3) The scatter plot represents the correlation between 'Electricity_from_hydro_GWh' (hydroelectric energy generation) and three factors: GDP, Population, and Area_KM2. Analogous to wind and solar energy, the analysis indicates that the most influential factor impacting hydroelectric energy generation is GDP. The scatter plot points exhibit a visible upward trend, signifying a positive relationship between hydroelectric energy generation and GDP. As GDP increases, hydroelectric energy generation tends to increase as well. Conversely, the scatter plots for Population and Area_KM2 demonstrate a scattered distribution without a clear pattern, indicating that hydroelectric energy generation does not display strong correlations with population density or geographical area. Thus, the findings emphasize the importance of economic development, represented by GDP, in driving the expansion and utilization of hydroelectric energy in the studied region.

4) The scatter plot portrays the relationship between 'Other_ renewables_ including_ bioenergy_ GWh' (bioenergy generation) and three factors: GDP, Population, and Area_KM2. Similarly to wind, solar, and hydro energy, the analysis highlights GDP as the primary factor influencing bioenergy generation. The scatter plot points exhibit a noticeable upward trend, indicating a positive correlation between bioenergy generation and GDP. As GDP increases, bioenergy generation also tends to increase. However, similar to the other renewable energy sources, the scatter plots for Population and Area_KM2 reveal a dispersed distribution with no significant evident pattern, suggesting that bioenergy generation does not have a strong correlation with population density or geographical area. Consequently, the results reinforce the significance of economic development, represented by GDP, in driving the expansion and adoption of bioenergy in the studied region.

Gross Demotic Product (GDP) By Continent:

This data visualization sheds light on the economic landscape and regional economic performance.

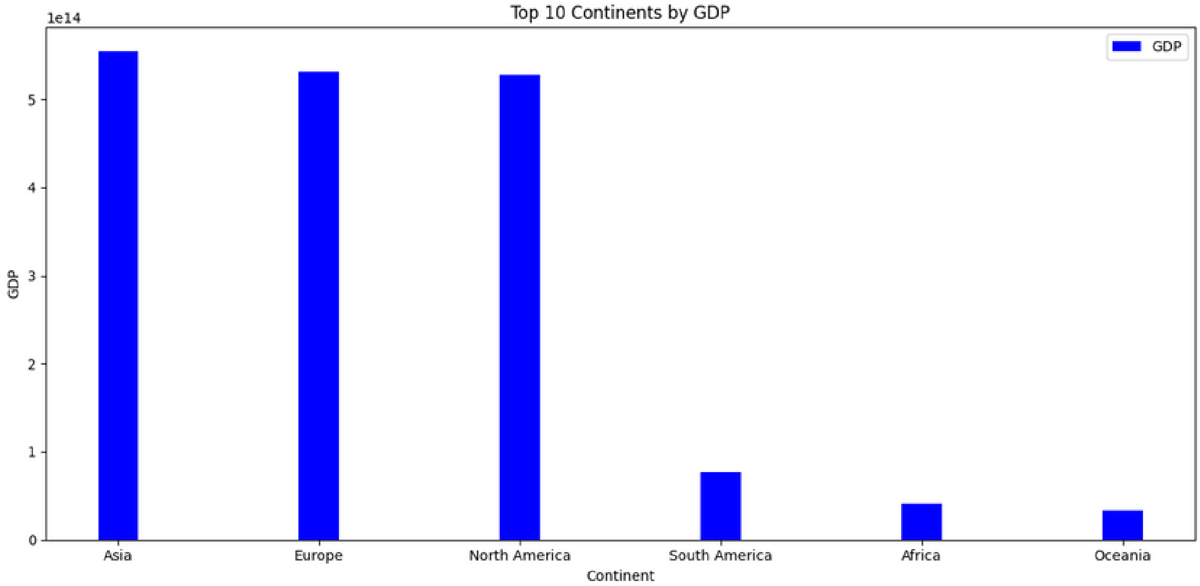

Chart 11. Gross Demotic Product (GDP) By Continent.

The bar chart provides a visual representation of the Gross Domestic Product (GDP) by continent, revealing significant disparities in economic output across different regions. Asia emerges as the top continent, boasting the highest GDP among all the continents, showcasing its economic prominence and robust growth. On the other end of the spectrum, Oceania records the least GDP, reflecting its relatively smaller economic size compared to other continents.

Top 10 Countries By Area:

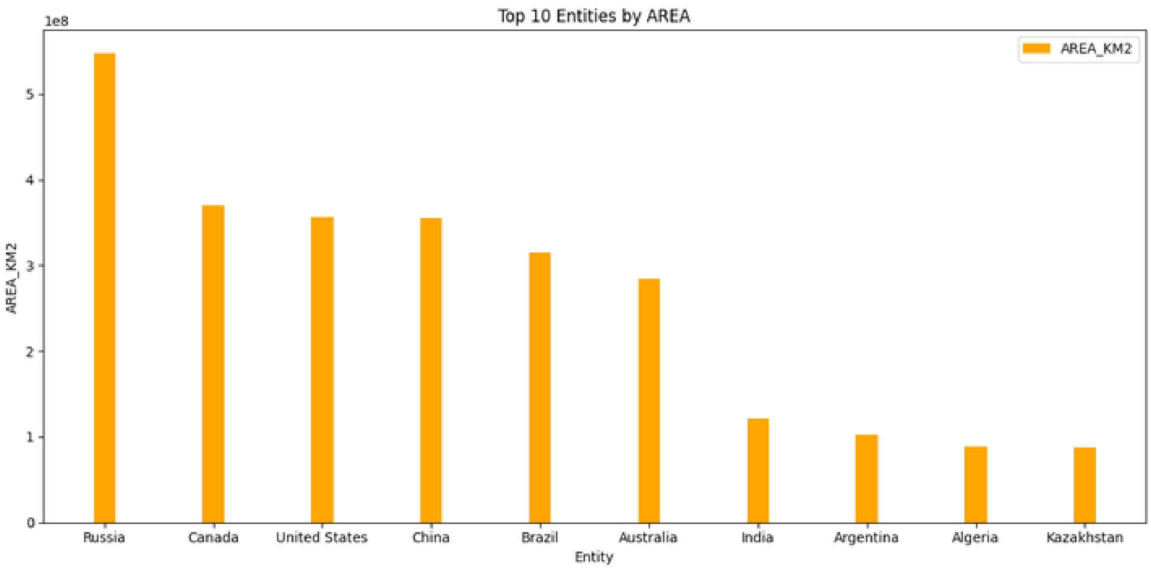

Chart 12. Top 10 Countries By Area.

The presented chart highlights the top 10 countries ranked by land size, providing a clear depiction of their relative geographical expanses. At the forefront, Russia emerges as the largest country in terms of land area, boasting an extensive territory that surpasses all others. Following closely, Canada and the United States secure their positions as the second and third-largest countries, respectively, in terms of land size.

The chart also showcases Kazakhstan, securing its place as the 10th largest country by land area.

Top 10 Countries By Population:

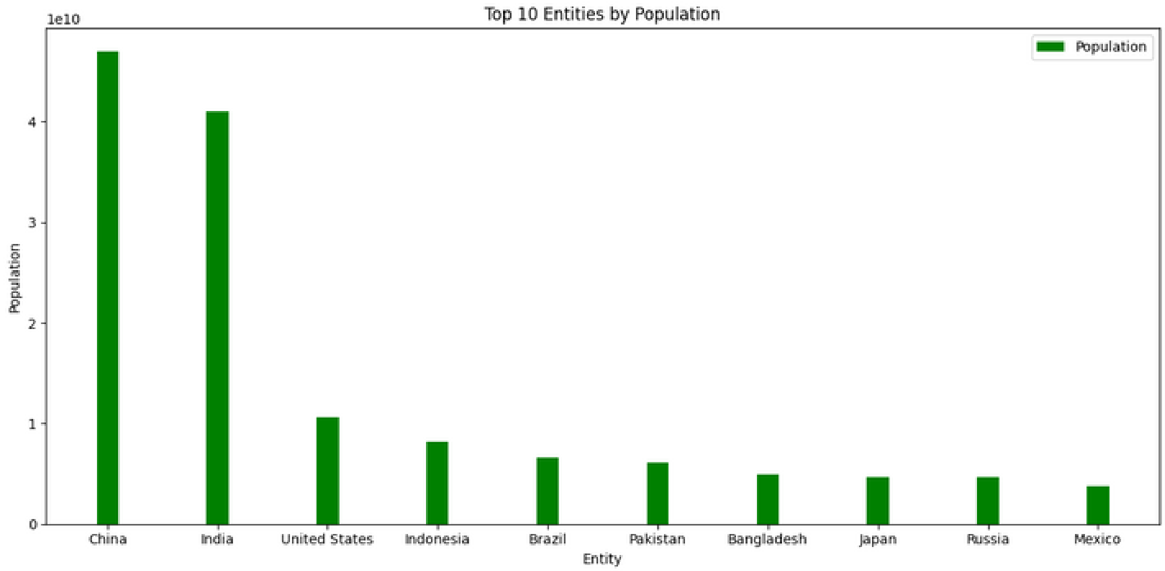

Chart 13. Top 10 Countries By Population.

The presented chart provides a comprehensive view of the top 10 countries ranked by population, offering valuable insights into the global distribution of human inhabitants. Evidently, China takes the lead as the most populous country, showcasing its significant demographic presence on the world stage. Following closely, India and the United States secure their positions as the second and third most populous countries, respectively.

Furthermore, the chart highlights Mexico's standing as the 10th most populous country, rounding out the list with its substantial population.

Model

Our objective was to create a predictive model to estimate the percentage of renewable energy in each country. To achieve this, we utilized specific features that are closely related to the country's characteristics. These features include electricity production from wind, electricity production from solar, electricity production from hydro, energy consumption, population, GDP per capita, and the country's area in square kilometers

To forecast the percentage of renewable energy, we employed multiple regression techniques, such as Linear Regression, Lasso, and Ridge. After conducting thorough experiments and evaluating the models, it became apparent that Linear Regression outperformed the other techniques

For model training and evaluation, we employed a dataset divided into two subsets: 70% of the data was utilized for training, and the remaining 30% was allocated for testing the model's performance.

The model was trained on a set of training data without applying normalization.

We get the weight table :

Table 2. weight table.

	Coffiecent
Renewables_percentage_electricity	1.000000e+00
AREA_KM2	-3.660518e-19
Population	-1.481619e-20
GDP	-2.500919e-25
Electricity_from_wind_GWH	7.453108e-19
Electricity_from_wind_GWh	9.965074e-18
Electricity_from_hydro_GWh	3.502318e-18
Electricity_from_solar_GWh	-3.921346e-17
Other_renewables_including_bioenergy_GWh	-7.425910e-17
Africa	-1.104369e-15
Asia	-9.267677e-16
Europe	-8.586826e-16
North America	-9.331996e-16
Oceania	-1.108713e-15
South America	4.939435e-15

We get The Error Matrix is:

Table 3. Error Matrix.

Metrix Error	Value
the mean squared error	2.633663244586887 e-24
R-square score of the model	1.0
The mean absolute error	6.920576502272 437e-13

After this highest of errors, we decide to make normalization then trained the model and below you will find the weight table for model:

Table 4. Model weight Table.

	Coffiecent
AREA_KM2	-0.121064
Population	0.107515
GDP	0.596917
Electricity_from_wind_GWH	-2.768613
Electricity_from_wind_GWh	-0.218587
Electricity_from_hydro_GWh	2.889116
Electricity_from_solar_GWh	-1.237201
Other_renewables_including_bioenergy_GWh	1.079725
Africa	0.079220
Asia	-0.105146
Europe	-0.025573
North America	-0.094127
Oceania	-0.078662
South America	0.224288

We get The Error Matrix is:

Table 5. Error Matrix.

Metrix Error	Value
The mean squared error	0.09
R-square score of the model	0.17
The mean absolute error	0.24

We used a different algorithm, specifically Ridge and Lasso regression, to make predictions, and interestingly, we obtained identical results. Upon examining the weight table, we observed that the influence of the EU and Africa continents is significantly close to zero. Therefore, we made the decision to delete them from the features.

After that we get good result:

Table 6. Model weight Table.

	Coffiecent
AREA_KM2	-0.089077
Population	0.134084
GDP	0.565906
Electricity_from_wind_GWH	-2.784037
Electricity_from_wind_GWh	-0.181909
Electricity_from_hydro_GWh	2.798660
Electricity_from_solar_GWh	-1.160713
Other_renewables_including_bioenergy_GWh	1.020743
Asia	-0.127813
North America	-0.113837
Oceania	-0.101904
South America	0.203155

The performance metrics for the best Linear Regression model are as follows:

Table 7. Error Matrix.

Metrix Error	Value
The mean squared error	0.09
R-square score of the model	0.16
The mean absolute error	0.24

So , We decided to adopt the model.

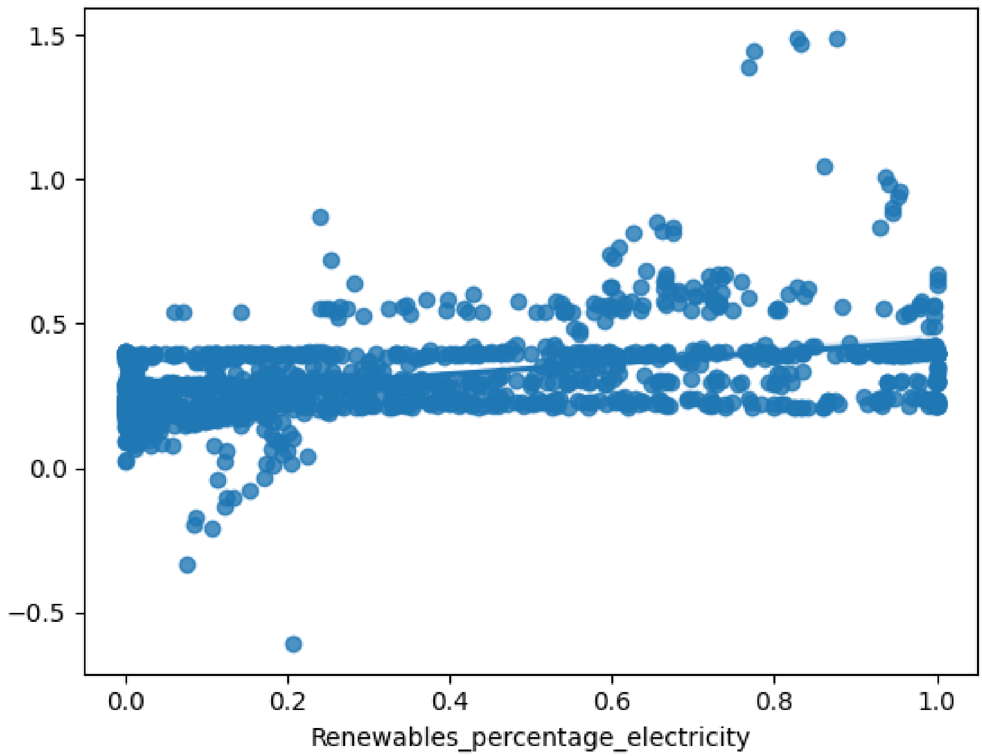

Chart 14. True vs. Predicted Renewable Energy Production: Regression Plot

Dashboard

The Global Energy Production dashboard created using Tableau Public. This interactive dashboard comprises three informative pages:

- **Main page:**

At the outset, this page showcases a group of key performance indicators (KPIs) represented through card visualizations. The displayed KPIs include total Energy Consumption, total GDP, total energy production, Number of Countries, and total population.

The page also features four charts:

1. Map: This map illustrates energy production by countries, providing an overview of global energy production distribution.
2. Pie Chart (1): Here, you can see a breakdown of energy production in six continents, with Asia having the highest value, reaching 43,058.
3. Pie Chart (2): This chart presents energy consumption in six continents, with Asia having the highest consumption value, totaling 1,760,684.
4. Dual Combination Chart: This chart compares energy consumption versus energy production from the years 1985 to 2021. The bars represent energy consumption, while the line represents energy production.

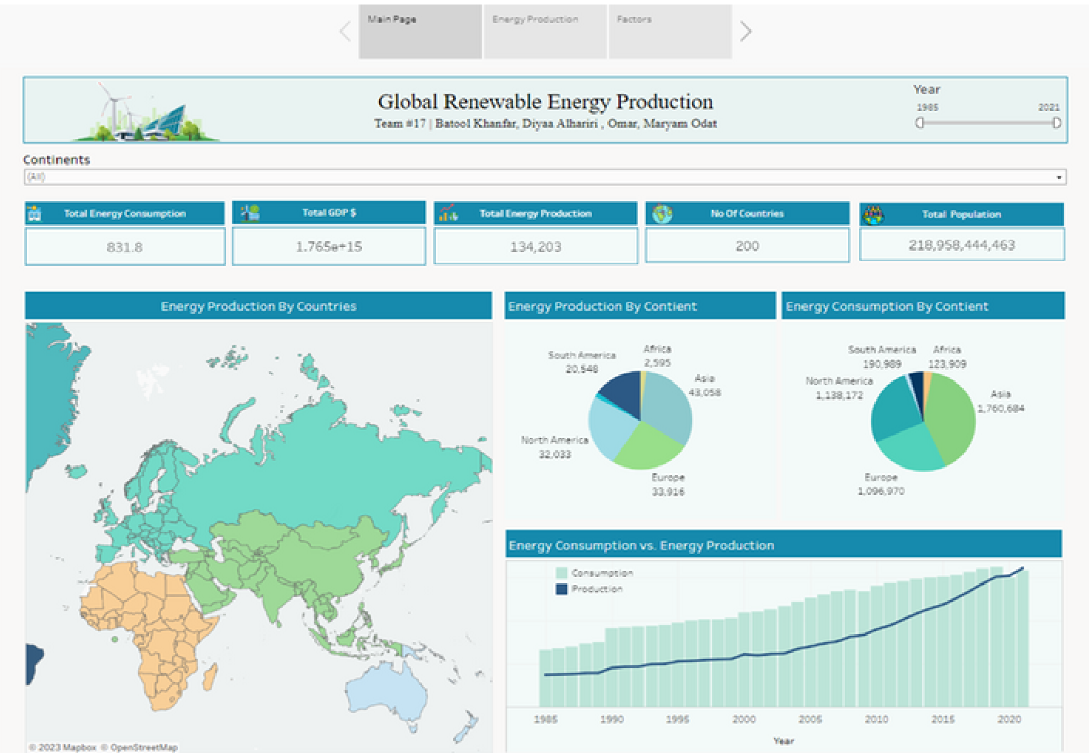

Figure 2. Main Page of Dashboard.

Two interactive filters are available for this page, allowing you to select specific years and continents.

- **Energy production page:**

On this page, you will find comprehensive details of energy production presented in various ways, featuring three distinct charts:

1.Area Chart (Continuous): This chart depicts the trend of energy sources (solar, wind, hydro, bioenergy) across different continents from 1985 to 2021.
2. Horizontal Bars: The chart showcases the Top 5 Countries in terms of energy sources (solar, wind, hydro, bioenergy).
3. Line Chart (Continuous): This chart visualizes energy source trends (solar, wind, hydro, bioenergy) across the years 1985 to 2021.

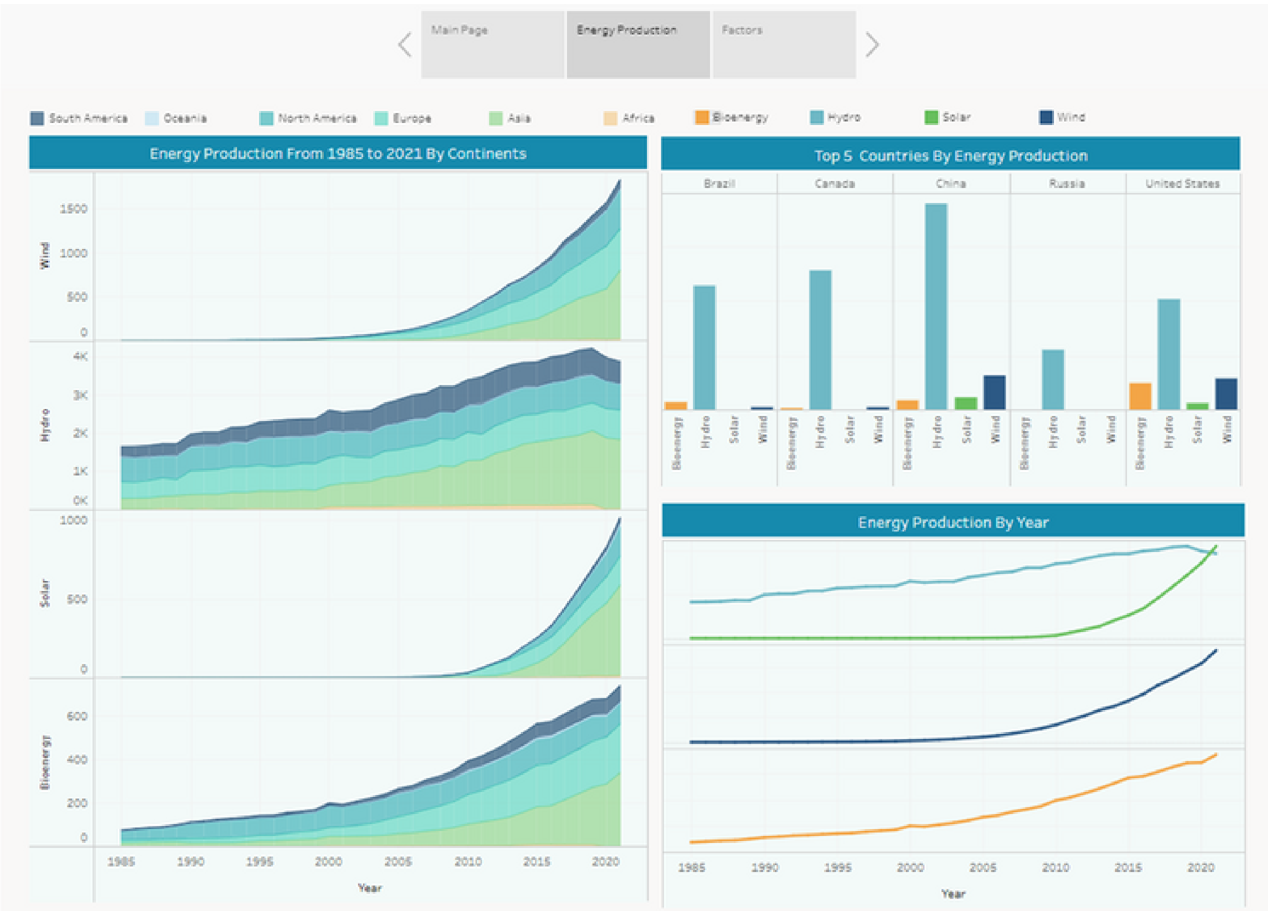

Figure 3. Energy Production Page.

- **Factors page:**

The third page provides detailed insights into various factors concerning continents and countries, using six informative bar charts:

1.GDP Values in Continents: This chart displays the GDP values for each continent, providing a comparison between them.
2.Population Values in Continents: This chart represents population values for each continent, offering a clear perspective on population distribution.
3.Area Values in Continents: This chart showcases the area values of each continent, highlighting the variations in landmass.
4.Top 10 Countries with GDP Values: Here, you can find a comparison of the top 10 countries based on their GDP values.
5.Top 10 Countries with Population Values: This chart presents the top 10 countries with the highest population values.
6.Top 10 Countries with Area Values: The final chart ranks the top 10 countries according to their land area values.

Figure 4. Factors Page.

DataFolio

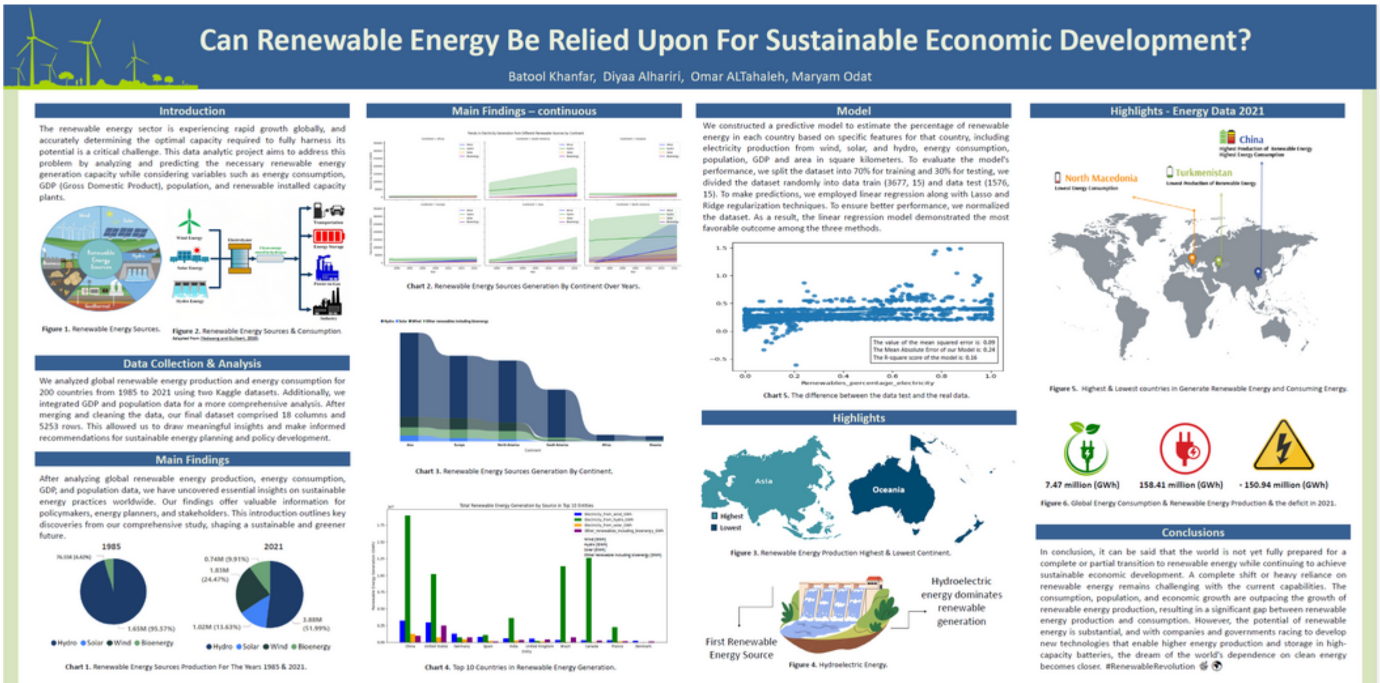

Figure 5. Renewable Energy Production DataFolio

The DataFolio presents a comprehensive analysis of global renewable energy production, consumption, GDP, population, and other relevant factors. In the Introduction section, the aim is to explore the potential capacity for sustainable economic development through renewable energy sources. The Data Collection and Analysis section details the process of gathering and cleaning data from two Kaggle datasets, incorporating GDP and population data, resulting in a rich dataset for in-depth examination.

In the Main Findings section, valuable insights emerge from the analysis, shedding light on trends, regional disparities, and dominant renewable energy sources across continents and countries. The Model section describes the construction of a predictive model using linear regression, Lasso, and Ridge techniques, allowing for informed estimations of renewable energy percentages in different countries.

Highlights offer key takeaways from the analysis, emphasizing critical observations and implications for stakeholders. Finally, the Conclusions section summarizes the findings, highlighting the challenges in transitioning to renewable energy and the potential for future developments in sustainable energy practices. The DataFolio provides essential information for policymakers, investors, and energy planners seeking to optimize resource allocation and foster a greener, more sustainable energy future globally.

Findings

In 2021, global renewable energy production reached 7.47 million GWh, while global energy consumption amounted to 158.41 million GWh, resulting in a substantial deficit of -150.94 million GWh. Renewable energy sources accounted for a mere 4.71% of our total energy consumption, indicating that the world is still significantly distant from meeting its energy demands solely through renewables. The data underscores the pressing need for further progress and widespread adoption of renewable energy solutions to bridge this considerable gap and advance towards a more sustainable energy future.

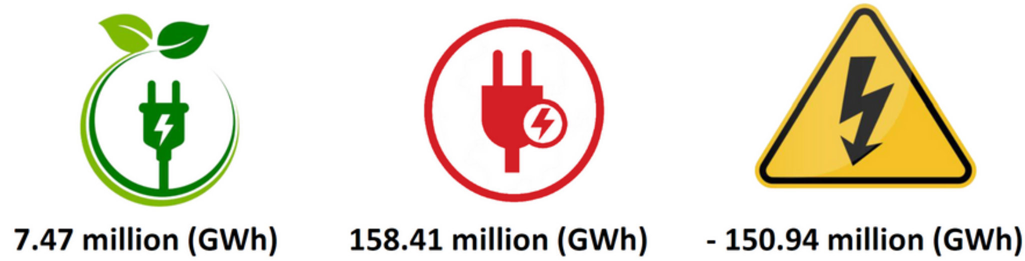

Figure 6. Global Energy Consumption & Renewable Energy Production & the deficit in 2021.

158.41 Million The Global Energy Consumption in 2021.

In summary, the world currently faces challenges in achieving a full or partial transition to renewable energy while sustaining economic development. The rapid growth in energy consumption, population, and economic activities has outpaced the expansion of renewable energy production, leading to a noticeable gap between supply and demand. Despite these hurdles, the potential of renewable energy is vast.

Efforts by both private enterprises and governments to advance technology, particularly in energy production and storage through high-capacity batteries, bring us closer to the realization of a cleaner and more sustainable energy future. While the journey towards a renewable energy-dominant world is ongoing, continuous innovation and investments offer promising prospects for attaining our goal of greater dependence on clean energy sources.

Future Analysis Topics:

1. Financial Incentives & Renewable Energy: Study the impact of financial incentives on renewable energy adoption.
2. Renewable Energy & Climate Change: Explore renewables' role in mitigating climate change.
3. Social & Economic Impact of Renewable Projects: Analyze the effects of large-scale renewable projects on communities.

Acknowledgements

We extend our heartfelt gratitude to the Correlation One program for providing us with invaluable learning opportunities and resources to enhance our data analysis skills. The program has been instrumental in shaping our analytical capabilities and fostering our passion for data science.

We would like to express our sincere appreciation to Jana Marzouk, our Teacher Assistant (TA), for her continuous support, guidance, and dedication throughout the course. Her expertise and encouragement have been crucial in our learning journey.

Special thanks to Sana Ali for her assistance and contributions, which have significantly enriched our understanding of the subject matter.

We are also deeply thankful to Dr. Sherif Halawani and Dr. Mohammad Daoud for their mentorship and for sharing their knowledge and expertise. Their valuable insights have been instrumental in shaping our understanding of data analysis and its real-world applications.

Lastly, we would like to acknowledge the support of our fellow learners and teammates, whose collaboration has been an essential part of our growth and success in this program.

Together, these individuals and institutions have played a pivotal role in our development as data analysts, and we are sincerely grateful for their support and encouragement.