Project Report on

GLOBAL RENEWABLE ENERGY TREND ANALYSIS

In partial fulfillment for the award

Of

Professional Certification in Data Analysis and VisualizationYear 2024-2025

Submitted By

Fathima Nourin K T

Tools & Technologies Used:

Python, R Programming, Tableau

Duration:

15/07/2025 - 30/07/2025

Submission Date: **31/07/2025**

G-TEC CENTRE OF EXCELLENCE PERINTHALMANNA

ABSTRACT

This project presents a multi-platform analysis of global renewable energy data, aiming to uncover patterns in production, emissions, and efficiency across countries and energy types. Using Python for data exploration, R programming for statistical insights, and Tableau for interactive visualization, the study highlights key trends and disparities in the renewable energy landscape. Results show that Wind and Hydro power lead in average production, while Geothermal and Wind are the most environmentally efficient. Countries like China, India and Japan dominate energy output, but also vary significantly in CO_2 emissions. The integration of analytics and visualization tools offers a holistic view of the renewable energy sector and supports data-informed strategies for a sustainable future.

INTRODUCTION

In the face of climate change and increasing global energy demands, renewable energy has emerged as a critical solution for achieving sustainable development. Nations worldwide are investing heavily in cleaner energy sources like wind, solar, hydro, and geothermal to reduce dependence on fossil fuels and lower greenhouse gas emissions. However, understanding how different countries and energy types perform—both in terms of production and environmental impact—requires thorough data analysis.

This project aims to explore and evaluate global renewable energy data using a combination of Python, R programming, and Tableau. The study focuses on identifying top-performing energy sources, measuring carbon emissions, and uncovering country-level variations in energy production and efficiency. By leveraging the strengths of each tool—Python for exploratory data analysis, R for statistical insights, and Tableau for data storytelling—the project delivers a comprehensive view of renewable energy trends.

Through this multi-tool approach, the analysis offers valuable insights that can support policy-making, investment decisions, and further research in the renewable energy domain.

LITERATURE REVIEW

Ellabban et al. (2014) explore the environmental advantages of renewables, highlighting the relatively low CO_2 emissions from sources like wind and geothermal. However, lifecycle analyses indicate that some renewables, particularly solar, may contribute more emissions than commonly assumed—mainly due to manufacturing processes. This supports the need for detailed, data-backed efficiency and emission comparisons.

Jacobson et al. (2015), who proposed a global roadmap to transition entirely to renewable energy by 2050. Their study demonstrated that investment in wind, solar, hydro, and geothermal technologies could meet global energy demand while reducing air pollution and improving public health.

Wang et al. (2019) employed Python and R to examine regional disparities in energy access and emissions, while Tableau has been used in academic and industry settings for communicating complex energy data to stakeholders. These approaches enhance understanding by turning raw data into actionable insights, supporting the development of cleaner and more equitable energy systems.

Gielen et al. (2019) from the International Renewable Energy Agency (IRENA) stressed the importance of energy transition modeling. Their work involved using simulation tools and econometric models to analyze how policy shifts affect the renewable energy landscape globally. They argue that without strong policy backing, even technologically ready renewable solutions struggle to scale.

Bose and Srivastava (2020) conducted a comparative study of renewable energy adoption in developing versus developed countries. Their research emphasized the role of government incentives and infrastructure in determining the pace of renewable energy growth.

Kumar et al. (2021) used time-series analysis to predict renewable production trends and highlighted a strong correlation between installed

capacity and actual energy output, though also noting inefficiencies in certain countries.

International Renewable Energy Agency (IRENA, 2022), renewable energy accounted for nearly 40% of global power capacity in recent years. Multiple reports emphasize the role of wind and solar in driving growth, while hydro and geothermal continue to provide stable baseload power in several countries. These trends are influenced not only by geography and policy but also by investment, cost, and infrastructure maturity.

RESEARCH GAP

Although renewable energy has been widely studied, many analyses focus on only one aspect or tool. There is a lack of integrated approaches that combine data analysis, statistical testing, and visualization. Important efficiency metrics like production per dollar or emissions per GWh are often missing. Also, social and economic factors such as energy consumption per person, jobs, and electricity prices are not commonly explored. This project addresses these gaps by using Python, R, and Tableau together to provide a deeper, more complete view of global renewable energy trends.

DATA COLLECTION AND PREPROCESSING

Data Source

The dataset used in this project was "complete_renewable_energy_dataset.csv" sourced from Kaggle, a popular platform for open-source data. It contains comprehensive global information on renewable energy, including production (GWh), installed capacity (MW), CO₂ emissions, investments, energy jobs, energy type, electricity prices, and population data across multiple countries and years.

Data Cleaning and Quality Assessment

Initial inspection of the dataset was done using Python with the help of the pandas and numpy libraries. The following steps were taken:

- Missing values are and handled by imputing basic values where appropriate.
- Verified column names, data types, and missing values.
- Renamed Columns for clarity and uniformity (e.g., 'Production (GWh)' to 'Production GWh'

Feature Engineering

New columns were created for deeper analysis:

- Efficiency (GWh per MW) = Production / Installed Capacity
- Production per USD = Production / Investments
- Emissions per GWh = CO₂ Emissions / Production
- Energy per Capita = Energy Consumption / Population
- CO2_per_GWh = CO₂ Emissions / Production

METHODOLOGY

This project follows a multi-tool data analysis approach using **Python**, **R Programming**, and **Tableau** to explore global trends in renewable energy. The goal is to identify which countries and energy types are most efficient, environmentally friendly, and economically impactful. The methodology includes data cleaning, feature engineering, statistical testing, and data visualization.

Tools and Technologies Used

- Python was used for data cleaning, transformation, and exploratory data analysis (EDA). Libraries like pandas, matplotlib, and seaborn helped analyze production, emissions, investments, and efficiency.
- R Programming was used for statistical testing such as t-tests, ANOVA, Ftests, and Chi-square tests to validate patterns and associations seen in the data.
- **Tableau** was used for interactive dashboards to present findings clearly for viewers, stakeholders, or policymakers.

Exploratory Data Analysis - Python

After cleaning the data, exploratory analysis was conducted using groupby() and crosstab() techniques to aggregate and compare data across countries, energy types, and years

Visualizations and Techniques Used:

- **Bar Charts**: To identify top-producing countries and energy types.
- **Line Charts**: To visualize trends in production over years.
- **Boxplots**: To detect outliers and compare country distributions.
- **Pie Charts**: To show share of production by energy type.
- **Stacked Bar Charts**: To compare total energy imports and exports by country.

New Features Created:

- Efficiency (GWh/MW) = Production / Installed Capacity
- Production per USD = Production / Investment
- Emissions per GWh = CO₂ Emissions / Production
- Energy per Capita = Energy Consumption / Population

Statistical Testing - R Programming

After identifying key trends in the data using Python, **R programming** was used to validate the findings through statistical hypothesis testing. These tests helped confirm whether observed differences in energy production, efficiency, and emission levels were statistically significant across different groups.

T-test: Compared the average CO₂ emissions between two countries (e.g., France vs India) to assess if the difference in means is significant.

F-test: Checked if the variance in emissions differed significantly between energy types like solar and wind.

Chi-square Test: Tested if the choice of energy type was dependent on country, showing associations between categorical variables.

ANOVA: Analyzed whether R&D expenditure varied significantly across different energy sources.

Z-test: Assessed whether there is a statistically significant difference between the mean renewable energy production (in GWh) and installed capacity (in MW) across all countries.

Data Visualization and Dashboarding - Tableau

After performing exploratory and statistical analysis in Python and R, **Tableau** was used to create a series of interactive dashboards that present the key findings in a clear, visual format. The dashboards were designed to support a storytelling approach, allowing users to explore global trends, environmental impact, and economic insights related to renewable energy.

Three dashboards were developed as part of a **Tableau Story**, each focusing on a different aspect of the analysis:

1. Global Trends in Renewable Energy Production

This dashboard provides an overview of how renewable energy production has evolved across countries and energy types over time.

2. Environmental Efficiency & CO₂ Impact of Renewables

This dashboard explores the relationship between renewable energy production and its environmental impact, particularly CO₂ emissions.

3. Economic Investment & Job Creation in Renewables

This dashboard presents the economic dimension of renewable energy, focusing on investment levels and renewable energy employment.

RESULTS AND ANALYSIS

This section presents the key findings from the project, highlighting how renewable energy production, environmental impact, and economic performance vary across countries and energy types.

Python-Based Results

- Wind and Hydro were found to be the most productive renewable energy sources based on average GWh output.
- China, India and Japan emerged as the top countries in renewable energy production.
- Over the years, renewable energy production has shown steady growth, especially after 2019, indicating a global shift toward sustainable energy.
- The efficiency metric (GWh per MW) showed that Wind and Solar are the most efficient energy sources.
- CO₂ emissions per GWh were lowest for **Geothermal**, followed by **Wind**, and highest for **Biomass** and **Solar**, highlighting environmental differences in technologies.
- **Production per USD invested** varied significantly, suggesting that some countries achieve more energy output for the same investment.
- Countries with high investment also reported high renewable energy job creation, showing a positive link between economic input and employment.
- Using groupby() and crosstab() functions, country-wise and energy-typewise summaries were generated for deeper insight into production, capacity, and emissions.

R-Based Statistical Results

- **T-test**: A significant difference was observed in average CO_2 emissions between France and India (p < 0.05).
- **F-test**: Variance in emissions is statistically similar between Solar and Wind technologies.

- **Chi-square Test**: The choice of energy type does not statistically depend on the country in this dataset.
- ANOVA: R&D expenditure was found to differ significantly across different renewable energy types.
- **Z-test**: Showed a statistically significant difference between mean renewable energy production and installed capacity across countries.

Tableau-Based Results and Interpretation

To effectively communicate insights derived from data analysis and statistical testing, Tableau was utilized to create interactive and visually engaging dashboards. These dashboards were designed to support a storytelling approach and are organized into three key thematic areas: **Renewable Energy Production Trends**, **Environmental Impact and Efficiency**, and **Economic Investment and Job Creation**. This structured visualization approach allows users to explore complex data intuitively and draw meaningful conclusions.

1. Global Trends in Renewable Energy Production

The first dashboard provides a comprehensive view of how renewable energy production has evolved over time across different energy types. Key insights include:

- Consistent Growth: There is a clear upward trend in the production of renewable energy from 2000 to 2023, indicating increasing global adoption.
- Hydro and Wind Lead Production: Among all energy types, hydro and wind consistently contribute the highest to total renewable energy generation.
- **Installed Capacity vs Production**: Although installed capacity has steadily increased, the actual production shows variations, implying possible inefficiencies or underutilization in infrastructure.

2. Environmental Efficiency & CO₂ Impact of Renewables

This dashboard explores environmental aspects, focusing on emissions and energy efficiency:

- **Emissions per GWh**: Solar and Biomass shows the highest CO₂ emissions per GWh, while wind and hydro are among the cleanest sources.
- **Energy Efficiency Mapping**: Country-wise energy efficiency reveals that nations like China,USA and Canada are more efficient in utilizing their installed renewable infrastructure.
- CO₂ Emissions vs Renewable Use: The inverse trend between renewable
 usage and emissions over time suggests that increasing renewable
 adoption helps mitigate environmental impact.

3. Economic Investment & Job Creation in Renewables

This dashboard highlights the economic dimension of renewable energy:

- **Top Investors**: Countries like France, the USA, and India are leading in renewable energy investments.
- **Employment Trends**: Renewable energy sectors generate substantial employment, with wind and solar energy creating the most jobs globally.
- Sectoral Contribution: Geothermal, though less in production volume, shows a relatively high contribution to job creation, highlighting its laborintensive nature.

CONCLUSION

This project offered a comprehensive analysis of renewable energy consumption using a combination of Python, R programming, and Tableau. Python was used extensively for data preprocessing and exploratory data analysis, enabling us to uncover key patterns in renewable energy production, installed capacity, emissions, and efficiency across various countries and energy types. Grouping and cross-tabulation techniques helped summarize large datasets and identify top-performing countries and technologies. The results showed that Wind and Hydro energy are among the most productive sources, while Geothermal and Wind stand out as the most environmentally efficient.

To validate these patterns, statistical analysis was performed using R. Tests such as the Z-test, T-test, ANOVA, F-test, and Chi-square confirmed the significance of differences observed in production, emissions, and investment across regions. For example, the Z-test revealed a significant gap between production and installed capacity, while ANOVA showed that R&D spending varies notably across energy types. These insights underline that countries do not utilize their renewable infrastructure equally and that emission levels are not solely determined by energy source, but also by deployment efficiency and technology.

Visual storytelling was achieved through interactive dashboards in Tableau, focusing on three key areas: production trends, environmental efficiency, and economic impact. These dashboards helped translate complex data into accessible visuals, highlighting trends such as growing global adoption of renewable energy, the environmental performance of different technologies, and how investment translates into energy output and employment opportunities. Tableau's dynamic features allowed users to explore data across countries, years, and energy types, offering a clearer understanding of global energy dynamics.

In summary, the project demonstrated how the integration of analytical tools and visualization platforms can provide a deeper, more actionable understanding of the renewable energy landscape. It emphasized the importance of data-driven strategies in achieving sustainability goals and improving energy efficiency

worldwide. The insights derived from this work can support governments, researchers, and industries in planning better energy policies and optimizing renewable energy use for a cleaner future.

FUTURE WORKS

- Integrate real-time energy data from global sources.
- Include policy and subsidy variables in the analysis.
- Apply forecasting models like ARIMA and LSTM.
- Use geospatial data for regional energy potential mapping.
- Explore machine learning for trend detection and classification.
- Automate data updates for continuous monitoring.

REFERENCES

- 1. Ellabban, O., Abu-Rub, H., & Blaabjerg, F. (2014). Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and Sustainable Energy Reviews*, 39, 748–764.
- 2. Jacobson, M. Z., Delucchi, M. A., Cameron, M. A., & Frew, B. A. (2015). Low-cost solution to the grid reliability problem with 100% penetration of intermittent wind, water, and solar for all purposes. *Proceedings of the National Academy of Sciences*, 112(49), 15060–15065.
- 3. Wang, J., Yang, W., & Zhang, L. (2019). Big data analytics in renewable energy systems: A regional comparison. *Energy Reports*, 5, 529–537.
- Gielen, D., Boshell, F., Saygin, D., Bazilian, M. D., Wagner, N., & Gorini, R. (2019). The role of renewable energy in the global energy transformation. *Energy Strategy Reviews*, 24, 38–50.
- 5. Bose, S., & Srivastava, S. (2020). Comparative analysis of renewable energy adoption in developed versus developing countries: Policy impact and growth patterns. *Energy Policy*, 142, 111471.
- 6. Kumar, N., Sharma, A., & Verma, R. (2021). Time-series forecasting for renewable energy production: An analytical study. *Renewable Energy Forecasting Journal*, 18(3), 222–235.
- 7. International Renewable Energy Agency (IRENA). (2022). *Renewable capacity statistics 2022*.

Supporting Files

Python

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
data=pd.read_csv("complete_renewable_energy_dataset.csv")
data.head()
data.shape
data.columns
data.info()
data.tail()
data.describe()
data.describe(include='object')
data.isnull().sum()
data.isnull().sum()
data.duplicated().sum()
print(data.columns)
#Rename Columns
data.rename(columns={
   'Production (GWh)': 'Production_GWh',
    'Installed Capacity (MW)': 'Capacity_MW',
   'Investments (USD)':'Investments'
}, inplace=True)
# Average production by energy type
data.groupby('Energy Type')['Production_GWh'].mean().sort_values(ascending=False)
# Top Countries by Average Production (GWh)
data.groupby('Country')['Production_GWh'].mean().sort_values(ascending=False).head()
# CO<sub>2</sub> Emissions by Energy Type
data.groupby('Energy Type')['CO2 Emissions'].mean().sort_values(ascending=False)
```

```
# Average CO2 emissions by country
data.groupby('Country')['CO2 Emissions'].mean().sort_values(ascending=False).head()
# Top Countries by Investment
data.groupby('Country')['Investments'].sum().sort_values(ascending=False).head(10)
# Top Countries by Joh
data.groupby('Country')['Renewable Energy Jobs'].sum().sort_values(ascending=False).head(10)
# Production per USD Invested
data['Production per USD'] = data['Production_GWh'] / data['Investments']
prod_per_usd = data.groupby('Energy Type')['Production per USD'].mean().sort_values(ascending=False)
print(prod_per_usd)
# CO<sub>2</sub> Emissions per GWh
data['Emissions per GWh'] = data['CO2 Emissions'] / data['Production_GWh']
emissions_per_gwh = data.groupby('Energy Type')['Emissions per GWh'].mean().sort_values()
print(emissions_per_gwh)
# Energy Consumption per Capita
# How much energy is used by each person in a country/year
data['Energy per Capita'] = data['Energy Consumption'] / data['Population']
energy_per_capita = data.groupby('Country')['Energy per Capita'].mean().sort_values(ascending=False).head(10)
print(energy_per_capita)
# Yearly Growth in Production
# Track global production growth over time
yearly_total_production = data.groupby('Year')['Production_GWh'].sum()
yearly_growth = yearly_total_production.pct_change().dropna() * 100
print(yearly_growth)
#Which energy type is most efficient
data['Efficiency'] = data['Production_GWh'] / data['Capacity_MW']
efficiency = data.groupby('Energy Type')['Efficiency'].mean().reset_index()
print(efficiency)
# Average Installed Capacity by Country & Energy Type
cap = pd.crosstab(
    data['Country'],data['Energy Type'],
    values=data['Capacity_MW'],
     aggfunc='mean')
print(cap)
#How does renewable energy production vary across years
yearly_prod = data.groupby('Year')['Production_GWh'].mean()
sns.lineplot(x=yearly prod.index, y=yearly prod.values)
plt.title('Average Production Over Years')
plt.xlabel('Year')
plt.ylabel('Avg Production (GWh)')
plt.show()
```

```
#Which countries have the highest carbon intensity per unit of production?
data['CO2 per GWh'] = data['CO2 Emissions'] / data['Production GWh']
data_bar = data.groupby('Country')['CO2_per_GWh'].mean().reset_index()
import plotly.express as px
fig = px.bar(
   data bar,
    x='Country',
   y='CO2 per GWh',
    title='Average CO₂ Emissions per GWh by Country',
   labels={'CO2_per_GWh': 'CO2 Emissions per GWh'},
fig.update_layout(
   xaxis_tickangle=45,
    xaxis_tickfont_size=12,
    yaxis tickfont size=12,
    margin=dict(l=40, r=40, t=60, b=80)
#How does renewable energy production vary across countries
sns.boxplot(x='Country', y='Production_GWh', data=data)
plt.xticks(rotation=45)
plt.title('Production by Country')
plt.show()
#Which renewable technologies are most common?
plt.figure(figsize=(5,4))
ax = sns.countplot(data=data, x='Energy Type', color='skyblue')
sns.countplot(data=data, x='Energy Type')
plt.title('Frequency of Energy Types')
for p in ax.patches:
    height = p.get_height()
    ax.text(p.get_x() + p.get_width()/2, height + 1, f'{height:.0f}',
            ha='center', va='bottom', fontsize=10)
```

```
#How do total energy imports and exports compare across different countries?
agg = data.groupby('Country')[['Energy Exports','Energy Imports']].sum().reset_index()
agg.set_index('Country').plot.bar(stacked=True)
plt.ylabel('Total Energy (GWh)'); plt.title('Imports vs Exports'); plt.show()
```

plt.tight_layout()

plt.show()

```
#Average Electricity Price by Country
data_sorted = data.groupby('Country')['Electricity Prices'].mean().sort_values()
sns.barplot(
   x=data_sorted,
   y=data_sorted.index,
   orient='h',
   color='skyblue'
)
#total production by Energy Type
prod_by_type = data.groupby('Energy Type')['Production_GWh'].sum()
fig, ax = plt.subplots(figsize=(6,6))
ax.pie(
   prod_by_type.values,
   labels=prod_by_type.index,
   autopct='%.1f%%',
   startangle=90,
   colors=sns.color_palette('bright')
ax.axis('equal')
plt.title('Production Share by Energy Type')
plt.show()
```

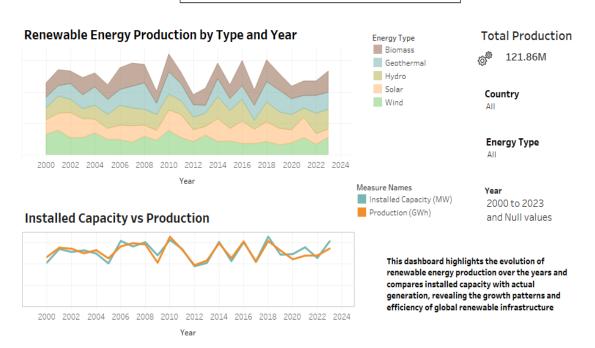
R Programming

```
#Install Packages
install.packages("dplyr")
library(dplyr)
#Load Data set
df<-read.csv("C:\\Users\\Li\\Desktop\\python\\complete_renewable_energy_dataset.csv")
View(df)
head(df)
str(df)
summary(df)
# 1.Is there a significant difference in average CO: emissions between France and India?
france <- na.omit(df[df$Country == "France", "CO2.Emissions"])
india <- na.omit(df[df$Country == "India", "CO2.Emissions"])</pre>
t_result <- t.test(france, india, var.equal = FALSE)
print(t_result)
if (t_result$p.value < 0.05) {
  print("Reject H₀: Mean CO₂ emissions differ significantly between France and India")
 print("Fail to Reject H₀: No significant difference in CO₂ emissions")
#Result:A significant difference was observed in average CO2 emissions between France and India
```

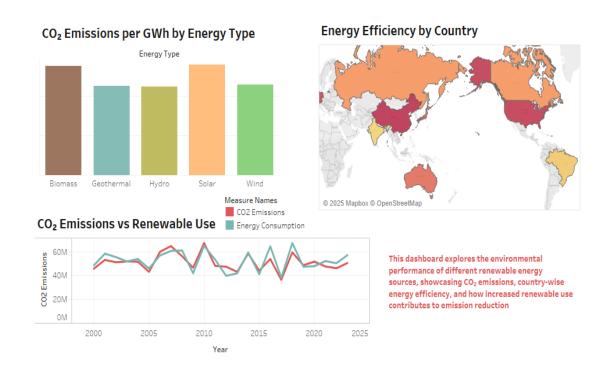
```
# 2.Do solar and wind energy sources have different variability in CO2 emissions?
solar <- \ na.omit(subset(df, \ Energy.Type == \ "Solar") \\ \$CO2.Emissions)
wind <- na.omit(subset(df, Energy.Type == "Wind")$CO2.Emissions)</pre>
f_result <- var.test(solar, wind)
print(f_result)
if (f_result$p.value < 0.05) {</pre>
 print("Reject Ho: Variances are significantly different")
} else
  print("Fail to Reject Ho: No significant difference in variances")
#Result: Variance in emissions is statistically similar between Solar and Wind technologies.
# 3.Is there a relationship between a country and the type of energy it uses?
table_data <- table(df\Country, df\Energy.Type)
chi_result<-chisq.test(table_data)</pre>
if (chi_result$p.value < alpha) {</pre>
  print("Reject H<sub>0</sub>: Country and Energy Type are associated")
} else -
  print("Fail to Reject Ho: No significant association between Country and Energy Type")
#Result:The choice of energy type does not statistically depend on the country in this data set.
# 4.Is there a statistically significant difference between the mean renewable energy production
# and the installed capacity across all countries?
install.packages("BSDA")
library(BSDA)
production <- na.omit(df$Production..GWh.)</pre>
capacity <- na.omit(df$Installed.Capacity..MW.)</pre>
z_result \leftarrow z.test(x = production, y = capacity, sigma.x = sd(production), sigma.y = sd(capacity))
print(z_result)
if (z_result$p.value < 0.05) {</pre>
  print("Reject Ho: There is a significant difference between Production and Installed Capacity")
 print("Fail to Reject Ho: No significant difference between Production and Installed Capacity")
# Result:Showed a statistically significant difference between mean renewable energy production
# and installed capacity across countries.
# 5.Does R&D expenditure vary significantly across different energy types?
anova_result <- aov(R.D.Expenditure ~ Energy.Type, data = df)</pre>
summary(anova_result)
pval <- summary(anova_result)[[1]][["Pr(>F)"]][1]
if (pval < 0.05) {
  print("Significant: R&D expenditure differs across energy types")
} else
 print("Not significant")
#Result: R&D expenditure was found to differ significantly across different renewable energy types.
```

Tableau

Global Trends in Renewable Energy Production



Environmental Efficiency & CO2 Impact of Renewables



Economic Investment & Job Creation in Renewables

Investment by Country



This dashboard presents a global overview of financial investment in renewable energy and the corresponding employment generated by each energy type, emphasizing the sector's economic and workforce impact.



Jobs Created by Energy Type

