

Achievement 6.1

Table of Content

1. DATA.....	2
1a. Data Source:.....	2
1b. Data Collection:.....	2
1c. Data Contents:.....	2
1d. Data Profile:.....	2
1d. Limitations and Ethics:.....	4
1e. Data Relevancy:.....	5
1e. Questions to explore:.....	5
2. Hypothesis.....	5

Global Sustainable Energy from 1996 to 2016

1. DATA

1a. Data Source:

This data is publicly available open-source data. It was downloaded from the website [kaggle.com](https://www.kaggle.com). This dataset can use under Creative Common licenses: Attribution 4.0 International (CC BY 4.0). The information sources and retrieved by the World Bank according to the source More details: [Data Set](#)

1b. Data Collection:

The data collection method is not specified but the initial source of the data is the world bank, a reputed government organization worldwide.

1c. Data Contents:

Data consist of information on accessing electricity from urban and rural areas, renewable energy (RE) generation, and the proportion of Renewable Energy share of total electricity.

1d. Data Profile:

- The data has 15 columns and 5444 rows. After the Data Quality check the dataset remains the same.
- A few corrections were made in Excel for easiness. Such as removing irrelevant data rows.
- Using Python change the column position, handle duplicates and missing values, outliers in all columns, remove the impossible negative values, and correct as necessary.

Index	Columns name	Description	Time Variant/ Invariant	Data type
1	Country Name	Name of the country	Time invariant	Qualitative
2	Country Code	Code of the country	Time invariant	Qualitative

3	Time	Duration of data collected (Year)	Time Variant	Quantitative
4	Time Code	Code generated from the year	Time Variant	Quantitative
5	Access to Clean Fuels and Technologies for cooking (% of total population) [2.1_ACCESS.CFT.TOT]	Percentage of the total population with access to Clean Fuels and Technologies for cooking	Time invariant	Quantitative
6	Access to electricity (% of rural population with access) [1.2_ACCESS.ELECTRICITY.RURAL]	Percentage of rural population with access to electricity	Time invariant	Quantitative
	Access to electricity (% of total population) [1.1_ACCESS.ELECTRICITY.TOT]	Percentage of the total population with access to electricity	Time invariant	Quantitative
7	Access to electricity (% of urban population with access)	Percentage of urban population with access to electricity	Time invariant	Quantitative
8	Energy intensity level of primary energy (MJ/2011 USD PPP)	A ratio between energy supply and gross domestic product is measured at purchasing power parity. Energy intensity indicates how much energy is used to produce one unit of economic output. A lower ratio indicates that less energy is used to create one output unit.	Time invariant	Quantitative
9	Renewable electricity output (GWh) [4.1.2_REN.ELECTRICITY.OUTPUT]	Electric output (GWh) of power plants using renewable resources, including wind, solar PV, solar thermal, hydro, marine, geothermal, solid biofuels, renewable municipal waste, liquid biofuels, and biogas. Electricity production from	Time invariant	Quantitative

		hydro-pumped storage is excluded.		
10	Renewable electricity share of total electricity output (%) [4.1_SHARE.RE.IN.ELECTRICITY]	Electricity generated by power plants using renewable resources as a share of total electricity output.	Time invariant	Quantitative
11	Renewable energy consumption (TJ) [3.1_RE.CONSUMPTION]	This indicator includes energy consumption from all renewable resources: hydro, solid biofuels, wind, solar, liquid biofuels, biogas, geothermal, marine, and waste	Time invariant	Quantitative
12	Renewable energy share of TFEC (%) [2.1_SHARE.TOTAL.RE.IN.TFEC]	Share of renewable energy in total final energy consumption	Time invariant	Quantitative
13	Total electricity output (GWh) [4.1.1_TOTAL.ELECTRICITY.OUTPUT]	Total number of GWh generated by all power plants	Time invariant	Quantitative
14	Total final energy consumption (TFEC) (TJ) [1.1_TOTAL.FINAL.ENERGY.CONSUM]	This indicator is derived from energy balances statistics and is equivalent to total final consumption excluding non-energy use	Time invariant	Quantitative

1d. Limitations and Ethics:

1. Limitation: The data with a lot of missing values. The Data contains typos with the Country name. Data can be used in accordance with the creative commons license and use the data with copyright policies and manipulation mentioned on the [website](#) and [CC License](#).

2. Ethical issue: The data does not contain any personal information. The data set contains only the geographical location of consumers. Therefore no way of identifying an individual. No action took against any concern related to ethics.

1e. Data Relevancy:

This dataset is very relevant to the project scope as it enables us to go through the variables that impact on project's final objectives.

1e. Questions to explore:

1. Which country shows the highest renewable energy output for the total population?
2. Which country covers the most availability coverage of electricity in rural and urban areas?
3. What does the relationship between RE and total population accessibility?
4. Which country is more efficient in electricity coverage with both energy sources and in both rural and urban coverage?
5. Forecasting about the next decade's RE consumption spread all over the world.

Only 3rd question can be answered -

What does the relationship between REoutput and total population accessibility?

- When it comes to the relationship between Renewable Energy output and total population accessibility it shows a strong relationship to each other.
- Depending on the Energy the total accessibility population of all countries can be categorized into four categories
- When you consider the accessibility of the population, a large portion of the population is highly accessible to RE, and there is a tiny portion with no access.

2. Hypothesis

All sources of energy output and consumption optimization will directly predict the sustainable energy supply to a larger population worldwide.