**World sustainability data**

* ***1a. Data Source:***

This data is publicly available open-source data. It was downloaded from Kaggle.com [link](https://www.kaggle.com/datasets/truecue/worldsustainabilitydataset?select=WorldSustainabilityDataset.csv)

* ***1b. Data Collection***:

The dataset was collected for TrueCue Women+Data Hackathon. The dataset were generated from several merged data sources. There are 54 fields included in the data set. The wide nature of the dataset could give rise to a range of analytical possibilities to answer a varieties of questions.

* ***Why this data set***

I choose this data because with the world moving towards sustainability. It is imperative for countries to know their standing relative to their commitment to achieving the SDG.

* ***1c. Data Contents:***

The dataset includes various globally agreed indicators to track the performance of 173 countries against a range of sustainability metrics over 19-year period.

* ***1d. Data Profile:***

The data has 54 columns and 3287 rows. After the Data Quality check the dataset.

* ***Data Cleaning and Data Consistency Checks***

● Checked for missing data:

● Check for duplicate data. No duplicate found

● Check for mixed data type. Converted all mixed data types

**Column Details**

|  |  |  |  |
| --- | --- | --- | --- |
| ***Column*** | ***Column description*** | ***Data type*** | ***Time Variant*** |
| ***Year*** |  | Quantitative | Yes |
| ***Country code*** |  | categorical variables | NO |
| ***Country name*** | Name of country | categorical variables | NO |
| ***Access to electricity (% of population)*** | Percentage of population with access to electricity. Electrification data are collected from industry, national surveys and international sources. | Quantitative | NO |
| ***Adjusted net national income per capita (annual % growth)*** | Percentage yearly growth of adjusted net national income divided by number of people in country | Quantitative | NO |
| ***Adjusted net savings, excluding particulate emission damage (% of GNI)*** | Measure that monitors whether savings and investment in a country compensate for depreciation of natural and physical capital | Quantitative | NO |
| ***Adjusted savings: carbon dioxide damage (% of GNI)*** | Measure that monitors whether savings and investment in a country compensate for depreciation of natural and physical capital | Quantitative | NO |
| ***Adjusted savings: natural resources depletion (% of GNI)*** | Measure that monitors whether savings and investment in a country compensate for depreciation of natural and physical capital | Quantitative | NO |
| ***Adjusted savings: net forest depletion (% of GNI)*** | Measure, calculated as the product of unit resource rents and the excess of roundwood harvest over natural growth. | Quantitative | NO |
| ***Adjusted savings: particulate emission damage (% of GNI)*** | Measure that monitors whether savings and investment in a country compensate for depreciation of natural and physical capital | Quantitative | NO |
| ***Automated teller machines (ATMs) (per 100,000 adults)*** | Number of physical ATMs that allow clients to make financial transactions in a public place. | Quantitative | NO |
| ***Children out of school (% of primary school age)*** | Percentage of primary-school-age children who are not enrolled in primary or secondary school. Children in the official primary age group that are in preprimary education should be considered out of school. | Quantitative | NO |
| ***Compulsory education, duration (years)*** | Duration of compulsory education is the number of years that children are legally obliged to attend school. | Quantitative | NO |
| ***Continent*** | Continent classification according to UN Convention | categorical variables | NO |
| ***Cost of business start-up procedures, female (% of GNI per capita)*** | Cost to register a business is normalized by presenting it as a percentage of the gross national income (GNI) per capita. | Quantitative | NO |
| ***Cost of business start-up procedures, male (% of GNI per capita)*** | Cost to register a business is normalized by presenting it as a percentage of the gross national income (GNI) per capita. | Quantitative | NO |
| ***Exports of goods and services (% of GDP)*** | Exports of goods and services represent the value of all goods and other market services provided to the rest of the world | Quantitative | NO |
| ***Final consumption expenditure (% of GDP)*** | Sum of household final consumption expenditure (private consumption) and general government final consumption expenditure (general government consumption). | Quantitative | NO |
| ***GDP (Current US$)*** | Current value (in USD) of total goods and services produced within a country. | Quantitative | NO |
| ***General government final consumption expenditure (% of GDP)*** | Final consumption expenditure (formerly general government consumption) includes all government current expenditures for purchases of goods and services (including compensation of employees). | Quantitative | NO |
| ***Gini index (World Bank estimate)*** | Score awarded based on how equitably income is dispersed within a country | Quantitative | NO |
| ***Gross national expenditure (% of GDP)*** | Sum of household final consumption expenditure (formerly private consumption), general government final consumption expenditure (formerly general government consumption), and gross capital formation (formerly gross domestic investment). | Quantitative | NO |
| ***Gross savings (% of GDP)*** | Calculated as gross national income less total consumption, plus net transfers | Quantitative | NO |
| ***Imports of goods and services (% of GDP)*** | Value of all goods and other market services received from the rest of the world | Quantitative | NO |
| ***Incidence of HIV, all (per 1,000 uninfected population*** | HIV rates per 1000 people | Quantitative | NO |
| ***Income Classification (World Bank Definition)*** | World Bank assigns the world’s economies to four income groups—low, lower-middle, upper-middle, and high-income countries. | categorical variables | NO |
| ***Individuals using the Internet (% of population)*** | Individuals who have used the Internet (from any location) in the last 3 months | Quantitative | NO |
| ***Individuals using the Internet (% of population)*** | % of countrys total population that has access to working internet | Quantitative | NO |
| ***Inflation, consumer prices (annual %)*** | Rate at which prices of goods and services brought by conusmers rise or fall. Estimated by using consumer price indices. | Quantitative | NO |
| ***Life expectancy at birth, total (years)*** | Value which captures the expectated age a newborn baby will live to | Quantitative | NO |
| ***Life expectancy at birth, total (years)*** | Average life expectancy for newborn babies | Quantitative | NO |
| ***Population, total*** | Total population | Quantitative | NO |
| ***Prevalence of undernourishment (%)*** | Percentage of the population whose habitual food consumption is insufficient to provide the dietary energy levels that are required to maintain a normal active and healthy life. | Quantitative | NO |
| ***Primary completion rate, total (% of relevant age group)*** | Number of new entrants (enrollments minus repeaters) in the last grade of primary education, regardless of age, divided by the population at the entrance age for the last grade of primary education. | Quantitative | NO |
| ***Proportion of population below international poverty line (%)*** | Percentage of the population living on less than $1.90 a day at 2011 international prices. The international poverty line is currently set at $1.90 a day at 2011 international prices. | Quantitative | NO |
| ***Proportion of population covered by at least a 2G mobile network (%)*** | Percentage of inhabitants living within range of a mobile-cellular signal, irrespective of whether or not they are mobile phone subscribers or users. | Quantitative | NO |
| ***Proportion of population covered by at least a 3G mobile network (%)*** | Percentage of inhabitants living within range of a mobile-cellular signal, irrespective of whether or not they are mobile phone subscribers or users. | Quantitative | NO |
| ***Proportion of population using basic drinking water services, by location (%)*** | Percentage of the population who have sustained access to basic water drinking services. | Quantitative | NO |
| ***Proportion of seats held by women in national parliaments (%)*** | Percentage of parliamentary seats in a single or lower chamber held by women. | Quantitative | NO |
| ***Pupil-teacher ratio, primary*** | Average number of pupils per teacher in primary school. | Quantitative | NO |
| ***Regime Type (RoW Measure Definition)*** | Regime classified considering the competitiveness of access to power as well as existence of liberal principles. Regime type is a weighted index that considers a variety of political metrics. | categorical variables | NO |
| ***Renewable electricity output (% of total electricity output)*** | Share of electrity generated by renewable power plants in total electricity generated by all types of plants. | Quantitative | NO |
| ***Renewable energy consumption (% of total final energy consumption)*** | Share of renewables energy in total final energy consumption. | Quantitative | NO |
| ***Rural population (% of total population)*** | Percentage of a countries total population that occupy rural areas. | Quantitative | NO |
| ***School enrollment, preprimary (% gross) -*** | Ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown | Quantitative | NO |
| ***School enrollment, primary (% gross)*** | Ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown | Quantitative | NO |
| ***School enrollment, secondary (% gross)*** | Ratio of total enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown | Quantitative | NO |
| ***School enrollment, secondary, female (% gross)*** | Ratio of female enrollment, regardless of age, to the population of the age group that officially corresponds to the level of education shown | Quantitative | NO |
| ***Total greenhouse gas emissions including land use change and forestry, measured in million tonnes of carbon dioxide-equivalents*** |  | Quantitative | NO |
| ***Total natural resources rents (% of GDP)*** | % Income accrued from natural resources (e.g. Exports) as a percentage of GDP | Quantitative | NO |
| ***Trade (% of GDP) -*** | Proportion of total GDP that is comprised by Trade activity. | Quantitative | NO |
| ***Unemployment rate, male (%)*** | Share of labour force without work but available and seeking employment (male) | Quantitative | NO |
| ***Unemployment rate, women (%)*** | Share of labour force without work but available and seeking employment (female) | Quantitative | NO |
| ***Urban population (% of total population)*** | What proportion of total population live in urban areas? | Quantitative | NO |
| ***Women Business and the Law Index Score (scale 1-100)*** | Index measures how laws and regulations affect women’s economic opportunity. Overall scores are calculated by taking the average score of each of the eight areas (Going Places, Starting a Job, Getting Paid, Getting Married, Having Children, Running a Business, Managing Assets and Getting a Pension), with 100 representing the highest possible score. | Quantitative | NO |
| ***World Regions (UN SDG Definition)*** | World Region as classified by UN. | categorical variables | NO |

* ***Limitations and Ethics:***

The data is real world data set and contains a lot of missing data. This presents all the challenges of working with data in professional world, such as missing values. For instance, a given country may not have data available for every field and every year.Dealing with missing values is quite challenging, as the data contains a lot of typing errors and unavailable information.

In terms of ethics, the data does not contain any personal information that could be used to trace any one.

***Questions to explore:***

1) Which countries still has the least access to clean water and sanitization?

2) Which continent has the least number of female enrollment in school?

3) Which countries has the highest proportion of population below the international poverty line ($1.25), therefore are behind in SDG 1?

4) Country with the highest unemployment rate