Project Description

## Name: Rasil Kodah

## Problem Statement:

As a data analyst at the United Nations working in the department of economic and social affairs, your primary responsibility is to report and evaluate the global progress on the Sustainable Development Goals (SDGs) focusing on identifying trends and disparities, presenting actionable insights.

You are specialized in sustainable development and global affairs and have been tasked to monitor the SDG progress of individual countries in the following deliverables:

1. **Top Performing Nations**: Identifying countries that show exceptional progress, specifying the goals they are excelling in.
2. **Struggling Nations**: Recognizing countries that are lagging behind, pinpointing the goals where they are struggling compared with the first category.
3. **Reasons for Disparities**: Investigating factors contributing to their success or struggles, such as policies, economic stability, geographic challenges, or technological or educational advancements.

A visual representation of the findings is required to compare the leading countries in SDG and the struggling countries by means of the nation’s resources, economy, policies, and overall prioritization of sustainable development.

## Possible Impact of Your Analysis:

Investigating this dataset will help allocate resources more effectively by identifying patterns and trends. This will allow for machine-learning based time-series forecasting which relies on the historical records and the performed prediction analysis. Exploring the data will also define where the strategies succeed in some regions and fail in other regions, giving an opportunity for more customized policies.

Cultivating these results will aid in providing a more thorough and specified action plan for countries to progress in their development goals depending on their own conditions and circumstances which will encourage cooperation and coordination between nations to achieve the shared objective for a more sustainable environment.

## Dataset(s):

The dataset represents the Sustainable Development Report (SDR) of 2023 which contains extensive data on the progress towards the Sustainable Development Goals (SDG) index scores for each individual country. Each entry shows the sustainability score for these countries as well as regional classification and performance on each of the 17 SDG’s.

### **Dataset Details**

The dataset itself is divided into two complementary datasets: Primary Dataset (sustainable\_development\_report\_2023.csv), and Supplementary Dataset (sdg\_index\_2000-2022.csv).

* **Size**: 167 rows for primary, 4,140 rows for supplementary, and 21 columns each.
* **Column labels for the dataset (primary and supplementary)**:
  1. country\_code: Country code.
  2. country: Country name.
  3. region: the region in which the country belongs to
  4. overall\_score: The overall sustainability (SDG index) score of the country.
  5. year: Year of the data.
  6. sdg\_index\_score: Overall SDG index score.
  7. goal\_1\_score to goal\_17\_score: Individual scores for each of the 17 SDGs.
* **Data Types:**

1. **Categorical**: country\_code, country.
2. **Numerical**: year, sdg\_index\_score, and scores for all 17 goals (goal\_1\_score to goal\_17\_score).

Source: <https://www.kaggle.com/datasets/sazidthe1/sustainable-development-report>

Project Scoping Document

## Name

Rasil Kodah

## Business Problem

The 2030 Agenda for Sustainable Development is a global plan adopted by all UN Member States in 2015 and outlines 17 Sustainable Development Goals (SDGs) aimed at achieving peace, prosperity, and environmental sustainability for all regions over the world.

The SDGs are built upon decades of prior work and research, including Agenda 21 (1992), a comprehensive plan for sustainable development adopted at the Earth Summit, Johannesburg Declaration (2002), which consists of reaffirmed commitments to poverty eradication and environmental protection, Rio+20 (2012), which launched the process to develop the SDGs, and Millennium Development Goals (MDGs) which contains eight goals to reduce extreme poverty by 2015.

As a data analyst at the United Nations working in the department of economic and social affairs, your primary responsibility is to report and evaluate the global progress on the Sustainable Development Goals (SDGs) focusing on identifying trends and disparities, presenting actionable insights.

You are specialized in sustainable development and global affairs and have been tasked to monitor the SDG progress of individual countries in the following deliverables:

1. **Top Performing Nations**: Identifying countries that show exceptional progress, specifying the goals they are excelling in.
2. **Struggling Nations**: Recognizing countries that are lagging behind, pinpointing the goals where they are struggling compared with the first category.
3. **Reasons for Disparities**: Investigating factors contributing to their success or struggles, such as policies, economic stability, geographic challenges, or technological or educational advancements.

A visual representation of the findings is required to compare the leading countries in SDG and the struggling countries by means of the nation’s resources, economy, policies, and overall prioritization of sustainable development. By analyzing this data, the UN will be able to designate each country’s needs and priorities so they can advance in the sustainable goal for cleaner energy and a safer future.

## Business Impact

Investigating this dataset will help allocate resources more effectively by identifying patterns and trends. This will allow for machine-learning based time-series forecasting which relies on the historical records and the performed prediction analysis. Exploring the data will also define where the strategies succeed in some regions and fail in other regions, giving an opportunity for more customized policies.

Cultivating these results will aid in providing a more thorough and specified action plan for countries to progress in their development goals depending on their own conditions and circumstances which will encourage cooperation and coordination between nations to achieve the shared objective for a more sustainable environment.

## Dataset(s)

The dataset represents the Sustainable Development Report (SDR) of 2023 which contains extensive data on the progress towards the Sustainable Development Goals (SDG) index scores for each individual country. Each entry shows the sustainability score for these countries as well as regional classification and performance on each of the 17 SDG’s.

The dataset itself is divided into two complementary datasets: Primary Dataset (sustainable\_development\_report\_2023.csv), and Supplementary Dataset (sdg\_index\_2000-2022.csv).

* **Size**: 167 rows for primary, 4,140 rows for supplementary, and 21 columns each.
* **Column labels for the dataset (primary and supplementary)**:
  1. country\_code: Country code.
  2. country: Country name.
  3. region: the region in which the country belongs to
  4. overall\_score: The overall sustainability (SDG index) score of the country.
  5. year: Year of the data.
  6. sdg\_index\_score: Overall SDG index score.
  7. goal\_1\_score to goal\_17\_score: Individual scores for each of the 17 SDGs.
* **Data Types:**

1. **Categorical**: country\_code, country.
2. **Numerical**: year, sdg\_index\_score, and scores for all 17 goals (goal\_1\_score to goal\_17\_score).

Source: <https://www.kaggle.com/datasets/sazidthe1/sustainable-development-report>

## Methods

The SDG scores represent this dataset’s variables which are dependent on each individual country names; each country contains a distinct SDG score depending on its performance and progress on the SDG plans.

Using Microsoft Excel, I intend to clean and organize the data as necessary; searching for null values and performing EDA, to which the dataset is to be potentially transferred to either SQL or Python for further analysis. After all the data has been prepared and the expected insights have been gathered, I will then use Tableau to visualize the findings and create a dashboard that will contain a bird-eye view as well as a detailed view as the main tabs and story board as a complementary tab.

## Dashboard

The dashboard will primarily have a bird eye view and a detailed view and a storyboard as a secondary tab view. It will contain each individual country's SDG index score for each of the 17 goals as well as the overall score for each country. In the center of the dashboard, a bubble map will show each of the SDG indicators for each country pinned at their capital and potentially showing the resources, energy consumption, environmental state, and other potential parameters that might be added should an additional dataset will be needed. Date picker will be placed on the top of the dashboard to be able to scroll through the years (2000-2023) and observe how the data interacts with each other as the timeframe changes.



## Milestones

During this project, it is expected to achieve the following milestone should everything go according to plan:

* Perform data wrangling, data cleaning, and EDA on the selected datasets.
* Transfer the data set into a SQL or python database or continue to work on Microsoft Excel should the process be difficult or time consuming.
* Identify key insights through statistical calculations (median, mean, mode)
* Obtain the necessary correlations and tables.
* Visualize the findings using tableau.
* Documenting the steps and notating them in the final report.

## Timeline

The following table represents the timeline of the datafolio project. Each week contains a different deliverable that contributes to the completion of the project, should everything progress according to the intended milestones:

|  |  |  |
| --- | --- | --- |
| Week | Due Date | Tasks |
| Week 1 | January 14 | Project Description + Project Scoping |
| Week 2 | January 21 | Data Curation |
| Week 3 | January 28 | Exploratory Data Analysis |
| Week 4 | February 4 | Datafolio + Dashboard |
| Week 5 | February 18 | Dashboard + Final Report |