# CRISP-DM Template for SDG Dashboard Design

This template will guide you through the CRISP-DM process while designing a dashboard focusing on one of the Sustainable Development Goals (SDG) indicators. Each section corresponds to a step in the CRISP-DM cycle. Provide detailed explanations and documentation of your work at each stage.

## 1. Business Understanding

In this section, define the objectives of your dashboard, which is the data-driven research question. Identify the specific SDG indicator you will focus on and explain the importance of this indicator to the broader SDG goals. Clearly state what you aim to achieve with the dashboard.

Despite global efforts, about 759 million people around the world still lack affordable and clean energy. Slow progress in energy efficiency, lack of electricity access for millions, health risks from inadequate cooking solutions, and disparities in sustainable energy access contribute to this challenge, limiting economic development and quality of life.

Based on this problem statement, I formulated the research question: "What are the key factors hindering the adoption of affordable and clean energy solutions in developing countries?". With this dashboard, which will focus on Goal 7: Affordable and Clean Energy, and will cover several indicators, namely 7.1, 7.1.2, 7.2.1, 7.a.1, I hope to provide a clear insight into the key factors hindering the adoption of affordable and clean energy solutions in developing countries.

Due to time constraints, I will only cover four of these key factors in my dashboard.

Key Questions to Address:

* - What is the purpose of the dashboard?
* - Which SDG indicator are you focusing on?
* - What insights do you hope to provide with this dashboard?

## 2. Data Understanding

Describe the data you are using to build the dashboard. Explain the sources of the data, the variables included, and any initial findings from exploring the data. Highlight any data quality issues or unusual patterns.

I have collected various data tables from various sources (referenced below), which are of high quality, causing only a few minor data quality issues that have been resolved.

From exploring this data, I have gained many initial insights into each of the LDCs. To give a few examples: the level of access to electricity and clean and renewable energy, the efforts each country has made and is making to improve its infrastructure and access to electricity, the stability of each LDC, and the international flow into some of these countries to help them adopt clean and affordable energy.

Sources:

[Our World in Data](https://ourworldindata.org/data?topics=Energy+and+Environment~Renewable+Energy) – Source where I have collected most of the data about the electricity of the LDC’s

[World Bank Group](https://www.worldbank.org/ext/en/home) – Source where I have collected most of the data about the development of the LDCs

[Conflict and the identification of the Least Developed Countries: Theoretical and statistical considerations](https://www.un.org/development/desa/dpad/wp-content/uploads/sites/45/publication/CDP-bp-2012-13.pdf) – Source where I have collected data about the current conflict and non-conflict LDCs

Variables:   
Country Name

Country Code

Year

Proportion of population with primary reliance on clean fuels and technologies for cooking

Access to electricity

International flows to developing countries for clean energy

Sum of GDP, PPP per Country (LDC’s)

International flows to developing countries for and spent on infrastructure (LDC’s)

Quality of overall infrastructure (LDC’s )

Total infrastructure investments (LDC’s)

Total-oda-for-infrastructure-by-recipient

Key Questions to Address:

* - What are the sources of your data?
* - What variables are included in the dataset?
* - Are there any data quality issues?
* - What initial insights did you gain?

## 3. Data Preparation

Document the steps you took to clean, transform, and prepare the data for analysis. This may include removing duplicates, handling missing values, and creating new variables that will be useful for the dashboard.

All the data I downloaded came from trustworthy websites that usually publish high quality data. This meant that I didn't have to do much to clean, transform or prepare the data for analysis, as it was mostly already prepared. Only for a few data tables I had to change data types or create a new unique identifier per row to connect different data tables.

The only new variables I had to create were unique identifiers per country and year to link different data tables, and the list of conflict and non-conflict LDCs.

Key Questions to Address:

* - What data cleaning processes did you perform?
* - How did you handle missing values?
* - Did you create any new variables or features?

## 4. Modeling

Explain any models or algorithms you used to analyze the data. If your dashboard includes predictive elements, describe the models and how they were trained. If not, explain how you structured the data to create meaningful visualizations.

Because all the data I downloaded came from trusted sources, I didn't have to do much to ensure the accuracy of the data tables. When I first explored the downloaded data tables, I noticed that many of the data tables contained extensive information for countries that are not classified as LDCs. As these are outside the scope of this research, I created a spreadsheet called 'List of LDC's'. Using this table I was able to display all the data relating to LDCs only.

I have not used any algorithms in this Power Bi dashboard.

Key Questions to Address:

* - What models or algorithms did you use?
* - How did you ensure the model’s accuracy?
* - How did you structure the data for visualization?

## 5. Evaluation

Evaluate the results of your modeling or visualization efforts. Discuss whether the dashboard meets the original objectives set out in the business understanding phase. Identify any limitations and how they might impact the dashboard's effectiveness.

At the outset of this research, I attempted to cover all of the key factors hindering the adoption of clean and renewable energy solutions in third world countries, but due to time constraints and the realization that there are many reasons, I decided to cover only four in this research and dashboard.

Using this dashboard and it’s visualizations, I have been able to greatly support my presentation about the key factors hindering the adoption of clean and renewable energy solutions for third world countries.

Although this dashboard is very clear when I walk through it with my presentation, I think it would be hard to understand if you just saw this dashboard without any presentation. This dashboard could possibly be improved by adding some guidance for those who haven't seen or remembered my presentation.

Key Questions to Address:

* - Does the dashboard meet the business objectives?
* - What are the limitations of your analysis?
* - How could the dashboard be improved?

## 6. Deployment

Outline the steps required to deploy the dashboard for use. This may include publishing it to a platform like Power BI, sharing it with stakeholders, and ensuring it is regularly updated with new data.

I will be hosting my dashboard on my GitHub repository. By doing this, it will not only allow me to easily share the project with my mentor(s), peers, and the client for feedback, but it will also enable version control, making it easier to track changes and improvements over time.

If for whatever reason I need to keep my dashboard actively updated, I will manually check every few weeks to see if any new data has been uploaded to any of my previous sources. For example, I currently have data on the GDP of Haiti but that data only goes up to the year 2020, so if it were to be updated to any later years, I could add that to my dashboard.

Key Questions to Address:

* - How will you deploy the dashboard?
* - Who is the target audience for the dashboard?
* - How will you ensure the dashboard stays up to date?

## 7. Future Research

List down the steps that might address the shortcomings in your research. These could be availability of data, outdated research data.

Most of the data I have used so far only goes up to 2020. Some even go back to 2008-2009. Although this data is still useful, I could give a clearer picture if the data were more up to date. At the time of writing, new data on LDCs may be collected and published soon. With access to this newer data, I'd be able to give a clearer insight into the four key factors hindering the adoption of clean and renewable energy solutions for third world countries.

Another minor issue is that in my dashboard I try to cover data from all LDCs, but not all countries have data for every year. This means that some visualizations may be missing some of the countries.

All in all, Together with my presentation and the data I have collected, I hope to give the client a clearer picture of what is preventing these LDCs from adopting clean and renewable energy solutions.

Key Questions to Address:

* Are there emerging techniques or tools that could provide deeper insights?- Who is the target audience for the dashboard?
* How might the research impact different stakeholders or communities?

This template is designed to help you systematically document your work at each stage of the CRISP-DM process, ensuring a comprehensive and well-structured approach to designing your SDG dashboard.