GENERAL SIR JOHN KOTELAWALA DEFENCE UNIVERSITY
INTAKE 41

Faculty of Management, Social Sciences, and Humanities
Department of Languages

BSc in Applied Data Science Communication

Advanced SQL and Cloud Databases - LB 2224
Assignment 1 – Task 03
Year 02 – Semester 02
14.12.2025

Assessment Title: Exploring Real-World Datasets Using Power BI, SQL Databases, and Cloud Databases

Group Members:

- **P.R.N.D. Jayalath - D/ADC/24/0002**
- **W.P.C.N.S Perera - D/ADC/24/0046**
- **JDT Jayawickrama - C/ADC/24/6854**

Content

Task 03

1. Introduction

The study focuses on global progress towards Sustainable Development Goal 7 (Affordable and Clean Energy) and Goal 13 (Climate Action) using climate-related indicators which published by the United Nations. Nowadays, sustainable development has become a critical global concern. Therefore, it is essential for evaluating the trends across the countries regarding carbon emission, renewable energy adoption, and greenhouse gas levels.

The study aims to provide a clear understanding regarding the above SDG indicators through the detailed process of data acquisition, preparation, transformation, and visualization. With the integration of SQL Server and Power BI allowed for a more efficient analytical workflow. It provides a clear understanding about global climate performance, regional disparities, and long-term sustainability trends.

2. Dataset

The dataset used for the analysis obtained from the United Nations Data Portal focusing on indicators related to Sustainable Development Goal 7 (Affordable and Clean Energy) and Goal 13 (Climate Action). The final energy consumption from renewable sources such as solar, wind, hydro, and biomass showing country's total energy use comes from clean and renewable energy. There are 03 datasets including renewable energy consumption, greenhouse gas emissions generated from agricultural and non-agricultural industrial processes. The dataset of greenhouse gas emissions generated from agricultural activities and industrial processes captures emissions from livestock, fertilizer use, crop production, manufacturing, and other industrial activities. The other dataset of greenhouse gas emissions from sectors outside agriculture mainly focused on the emission sources of fossil fuel, waste disposal. And industrial energy use.

3. Challenges Faced During Data Import, Cleaning, and Modelling

3.1 Dataset Acquisition and Preparation

As the first step of the project, the climate-related datasets were obtained from the official UN Data Portal regarding Sustainable Development Goals. The data sources mainly include data on greenhouse gas emissions, CO2 emissions, and renewable energy usage. Several excel sheets are downloaded which contained both simple and complex tables. As the initial review of the datasets, manually opened each excel file to check its structure. Some sheets contained clear column names such as Country, Year but some sheets used wide format for each year as separate columns. Some sheets had blank header fields which leads to confusing labels such as column46, column 47 etc. when imported to SQL. There were some numerical columns which contained text characters, commas, or symbols while indicating the data cleaning required.

3.2 Importing data into SQL Server

After the data preparation process of the files, a SQL database names SDG_Climate_DB was created in SQL Server Management Studio and import the files to the database via SSMS import flat files.

There were several complications during import,

a. **Inconsistent column names and formats**
   There were some columns in the tables with generic names like column46, column 47 etc. making difficult to understand immediately what the data represented. Therefore, the metadata and documentation have to reference for correct identification of relevant columns.

b. **Large volume of data**
   The dataset consists of multiple tables with large number of rows. Without losing the data there should be a careful resource management to import them efficiently.

c. **Mixed data types**
   Some columns which intended to store numerical values, were stored as text or mixed with non-numeric characters. Therefore, it is required to identify those columns and convert them into appropriate data types for analysis.

d. **Import failures**
   At the first attempt some tables did not load correctly due to merged header cells in excel and unrecognized column formats. To fix that, have to reopened and fix the unnecessary formatting and save it again.
   After that the tables of,

- EN_ATM_GHGT_NAIP (Non-Annex I)
- EN_ATM_GHGT_AIP (Annex I)
- SYB67_310_202411_Carbon Dioxide Emission Estimates

Were successfully imported and verified the each import using below code.

```
--check columns
SELECT TOP 5 * FROM EG_FEC_RNEW;
SELECT TOP 5 * FROM EN_ATM_GHGT_AIP;
SELECT TOP 5 * FROM EN_ATM_GHGT_NAIP;
```

3.3 Initial Data Assessment and Understanding

This step was essential before cleaning to determine how much cleaning was required and which transformations were necessary. The below commands were used for that exploratory assessment.

```
--get column names
SELECT
    TABLE_NAME,
    COLUMN_NAME,
    DATA_TYPE
FROM
    INFORMATION_SCHEMA.COLUMNS
WHERE
    TABLE_NAME IN ('EG_FEC_RNEW', 'EN_ATM_GHGT_AIP', 'EN_ATM_GHGT_NAIP')
ORDER BY
    TABLE_NAME, COLUMN_NAME;
```

```
SELECT COUNT(*) AS TotalRows FROM EN_ATM_GHGT_NAIP;
```

As the findings after executing above queries below findings got.

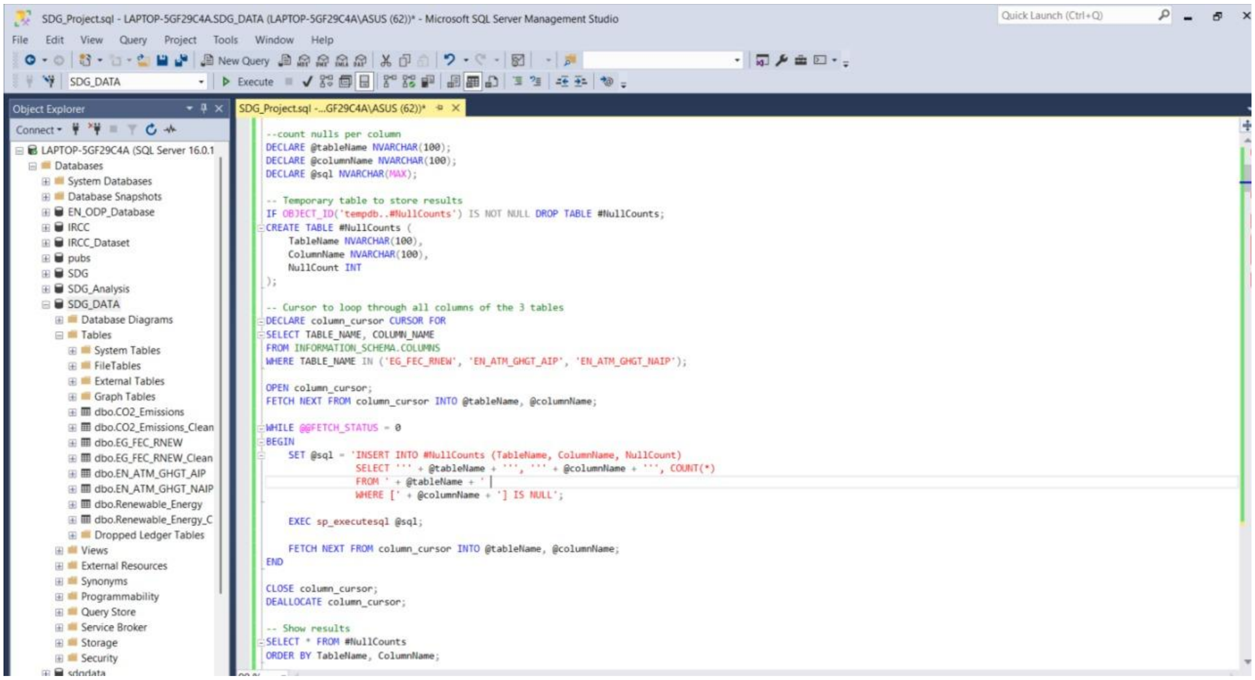
There were a lot of tables that were very wide, and a few of them had more than 50 columns, some tables that had columns full of null values and some tables used non-standard country names.

3.4 Data Cleaning Process

There were several SQL techniques to do for the data cleaning process. Those were as follows,

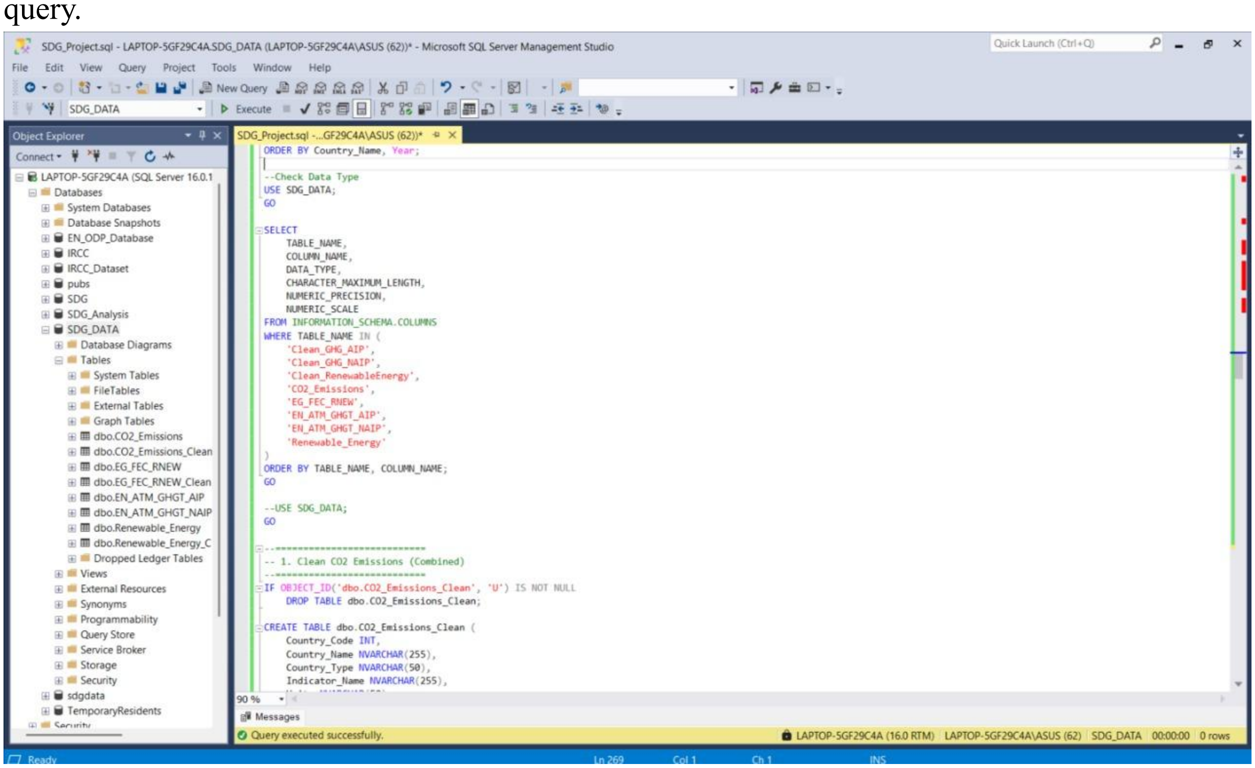
a. **Identifying and Removing 100% Null Columns**

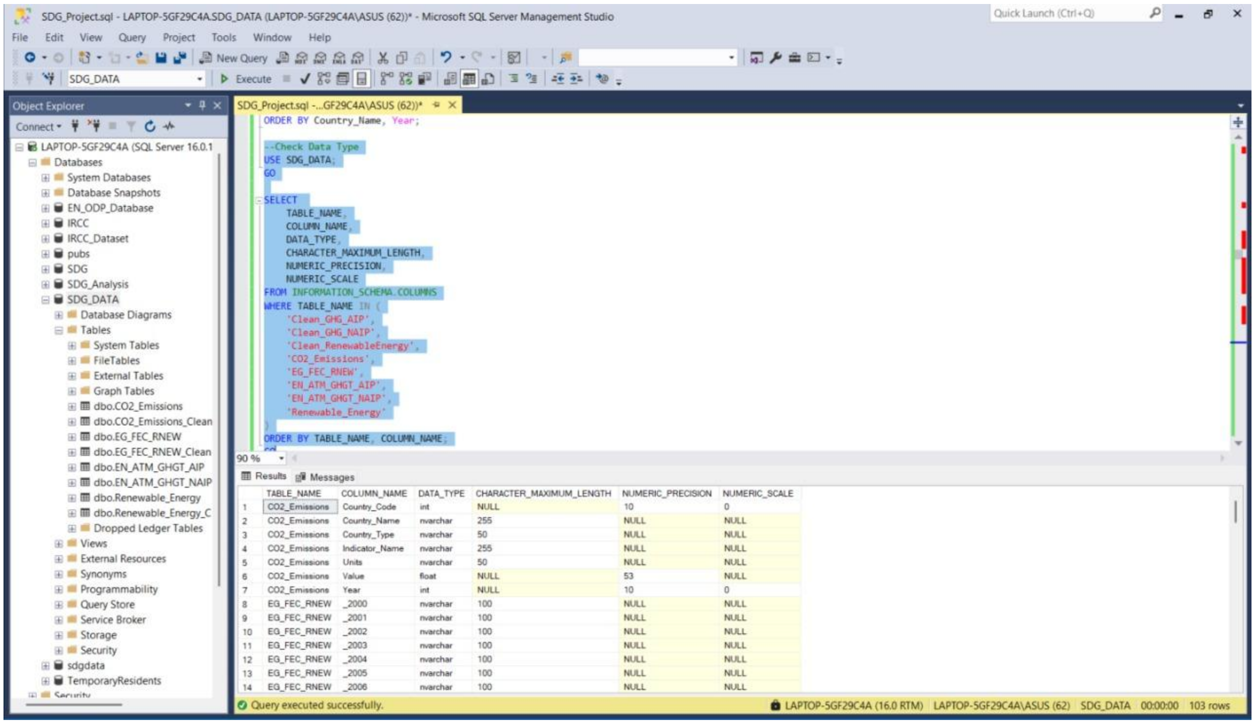
There were several columns such as column46 to column52 which contained as empty data and those columns had been Imported from blank cells in Excel.



b. **Fixing data types**

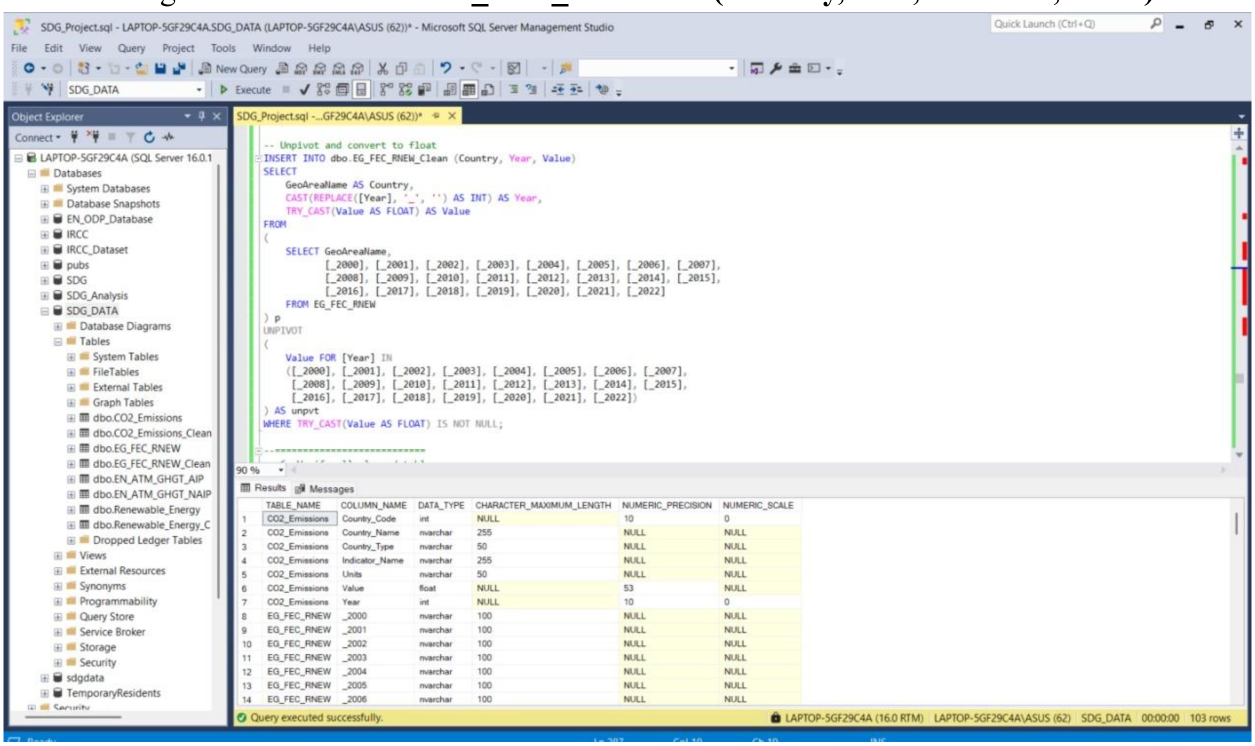
The data types were checked to maintain the standardization of the values by using below query.





c. Unpivoting Wide Format Tables

The UNPIVOT was used to transform the long formats into a structure which suitable for Power BI modelling. It converted the wide messy tables into clean analytical tables by standardizing the table named **Clean_CO2_Emissions (Country, Year, Indicator, Value)**



3.5 Data Modelling and Integration

After cleaning all the tables, the tables combined into a unified modelling structure by merging Annex I and Non-Annex I emissions tables into a single table by adding a new column named CountryType while removing intermediate tables. It ensured the consistent among units of measurement, indicator codes, country names, and year formats. Then the final table of Clean_CO2_Emissions exported to Power BI for visualization.

4. How Connecting Power BI to SQL Server Improved the Analytical Workflow

The SDG related CO_2 emissions, renewable energy indicators, and greenhouse gas metrics datasets were stored in the SDG_Climate_DB SQL Server database. After completing the steps of the SQL Server, Power BI was used as the main platform for the visualization and reporting. By connecting SQL Server directly to the Power BI, it ensures Power BI always accessed the most accurate and consistently structured datasets without cleaning data repeatedly inside Power BI.

4.1 Establishing the SQL Server Connection

To establish the connection process, first I opened the Power BI Desktop and navigated to Get Data → SQL Server and entered the SQL server name and selected the relevant database which contained all the prepared cleaned tables in SSMS.

Because the dataset size is manageable, Import Mode was used for the project to ensure the fast navigation, filtering, and visual rendering within Power BI. After the connection was established, selected only the tables which cleaned and prepared already as standardized, filtered, and unpivoted in SQL Server which suitable for analytical use.

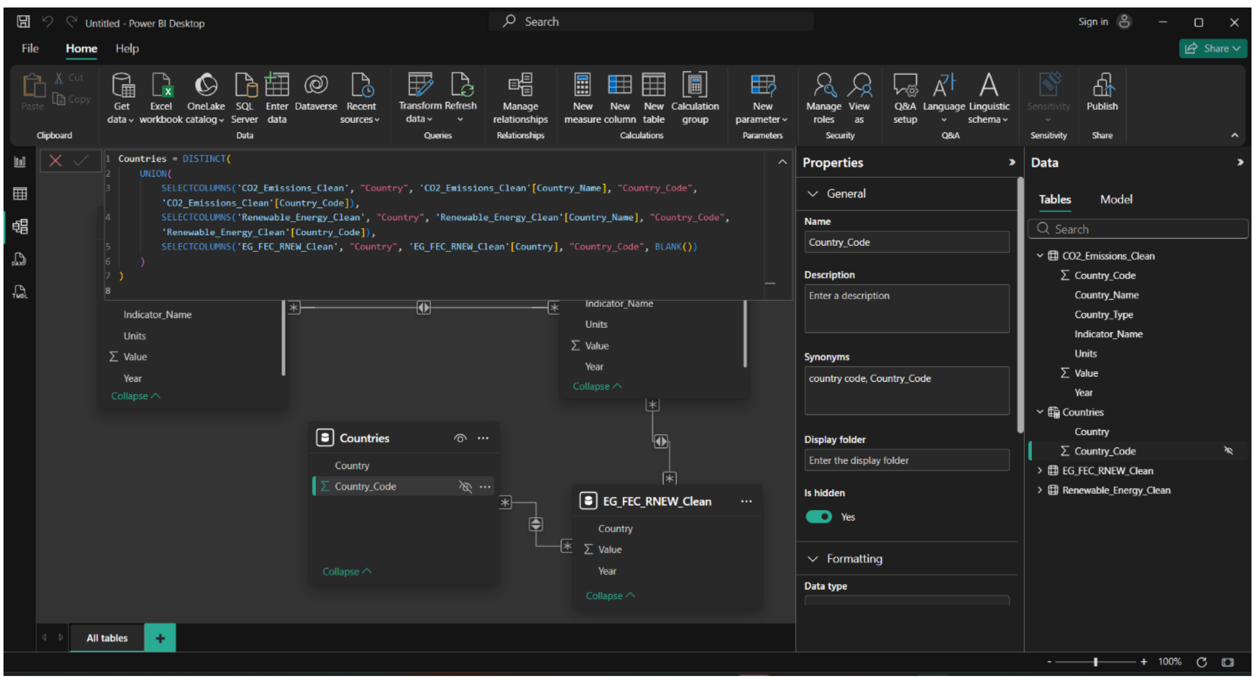
4.2 Selecting and Transforming Data in Power Query

Then opened the Power Query Editor after loading the tables to perform the additional validation and minor adjustments. Because the most cleaning part was done in the SQL Power Query mainly focused on structural checks such as removing unnecessary columns where exists duplicated metadata when importing, ensuring numeric year columns, verifying value columns were float. The final tables of CO2_Emissions_Clean, Renewable_Energy_Clean, and EG_FEC_RNEW_Clean formed the foundation of the Power BI model.

4.3 Creating Relationships in Power BI

After all the tables imported, then the data model was created to form a structure that allowed to seamless filtering of all indicators based on the relevant shared attributes. Using the Model View in Power BI, verified each table contained a properly cleaned Year column. Then the Year column dragged from each fact table and connected it to a newly created master Year table. It allowed slicers and charts based on Year to update all the visuals at same time.

As a major challenge, there were different structures for country identifiers. Country codes did not always match and duplicate country names appeared across tables. To solve this, a single master Country Lookup Table using DAX. It avoids the incorrect joins and duplicate country values across fact tables, making country slicers filter all visuals correctly while maintaining a clean and normalized data model.



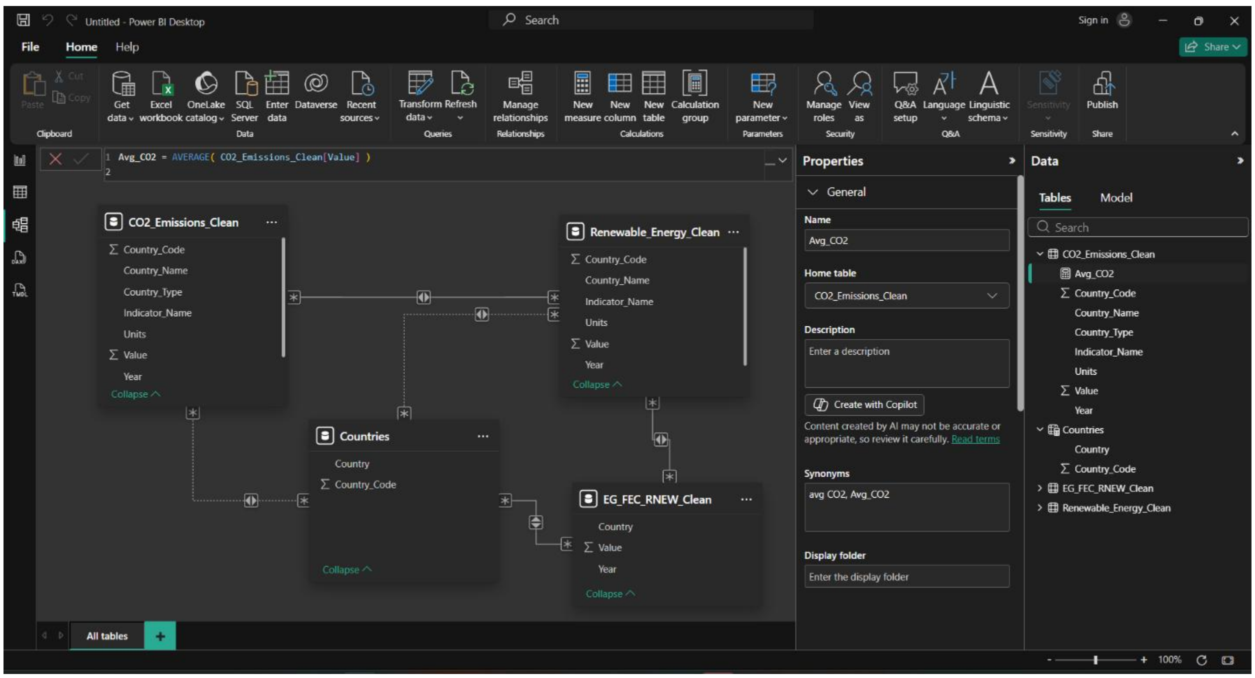
5. Application of Advanced Power BI Features

5.1 Uses of DAX Measures

There are different Power BI features that can be applied to enhance the analytical capability of the SDG dashboard. Among those key advanced features, DAX was required although the most preprocessing tasks were completed in SQL server. For aggregations, comparisons, ranking, and progress tracking Dax was created.

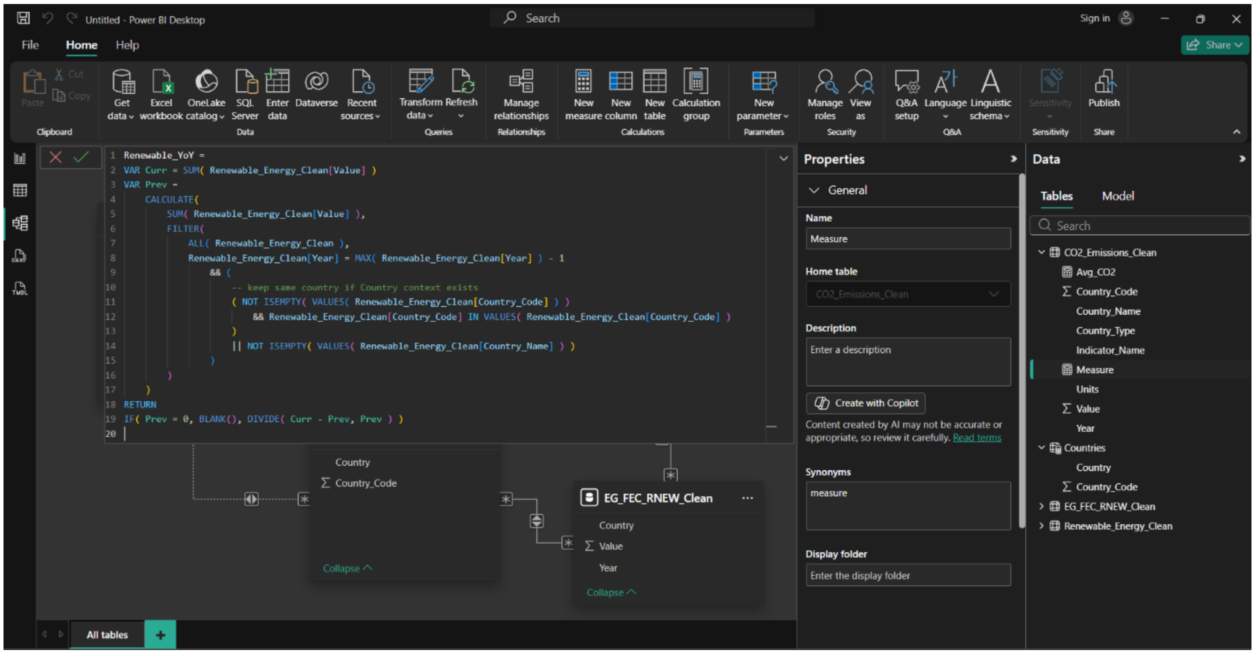
a. Average CO_2 Emissions

The below DAX was used to calculates the average CO_2 emissions for any selected country, year or indicator. That measurement allows for easy comparison between different regions and time periods. It used across cards, KPI visuals, and comparison charts to show overall emission intensity.



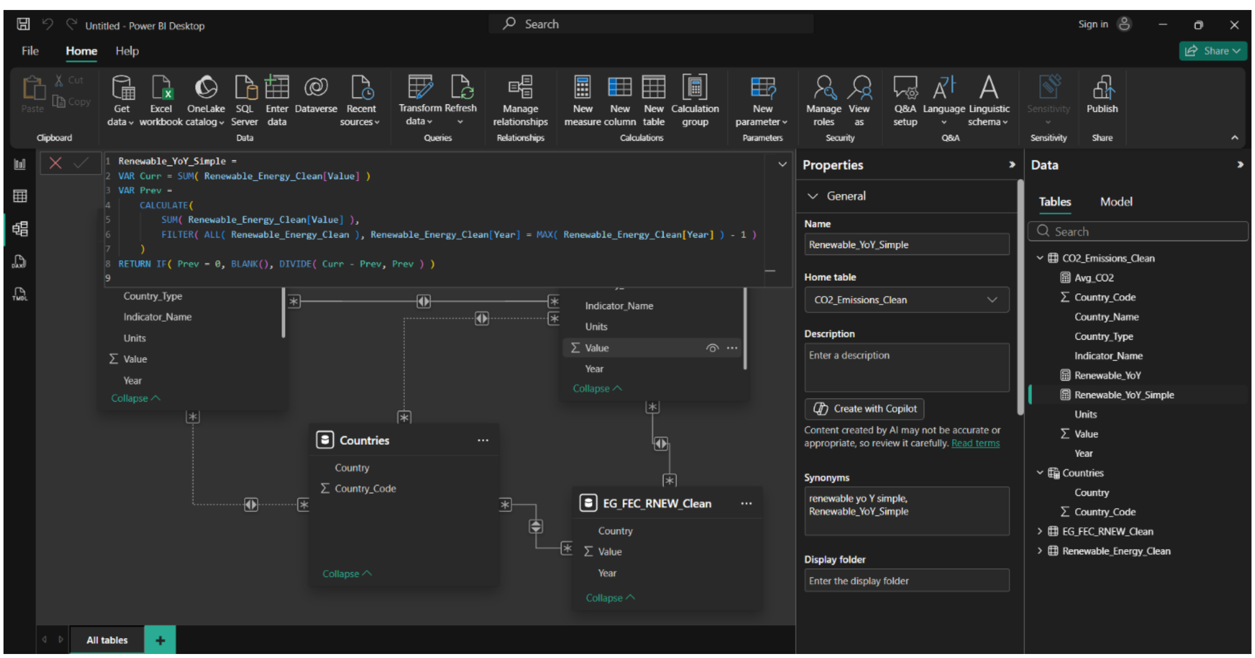
b. Total Renewable Energy

This measure allows to understand how much renewable energy each country contributed within a selected time range. It also used in bar charts, summary KPI cards, and trend analysis visuals by providing quick insight into which countries are progressing in renewable energy adaption.



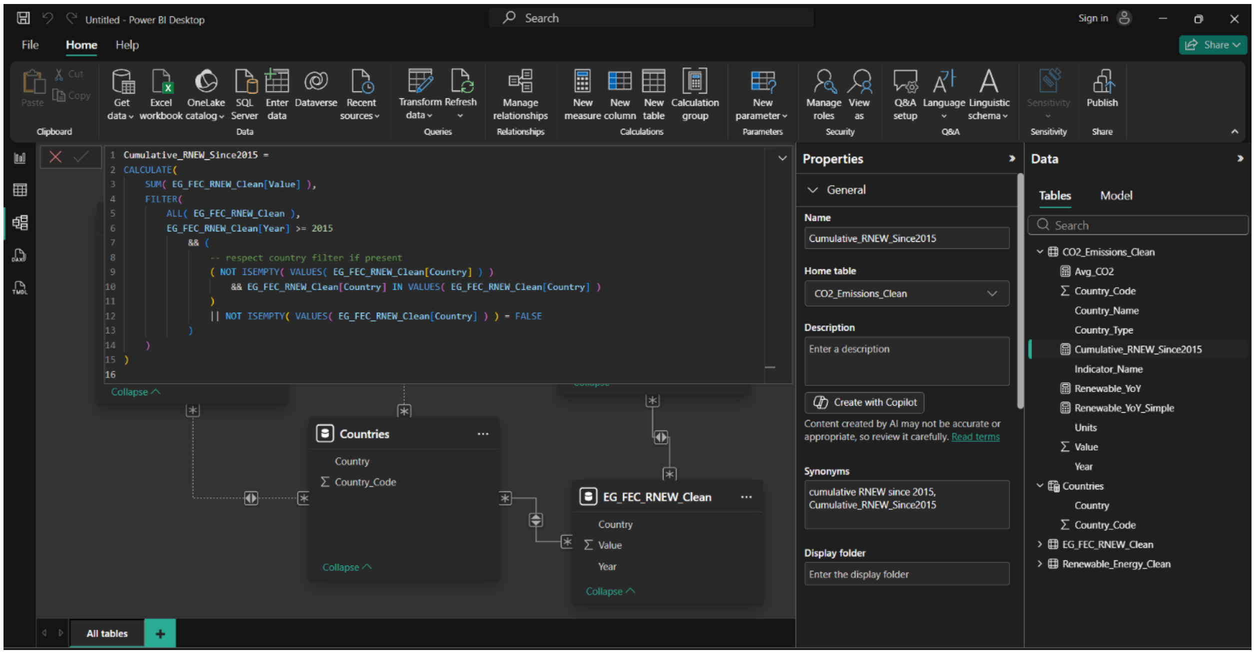
c. Year-over-Year (YoY) Change

Below DAX used to calculate the percentage increase or decrease in renewable energy compared to the previous year.



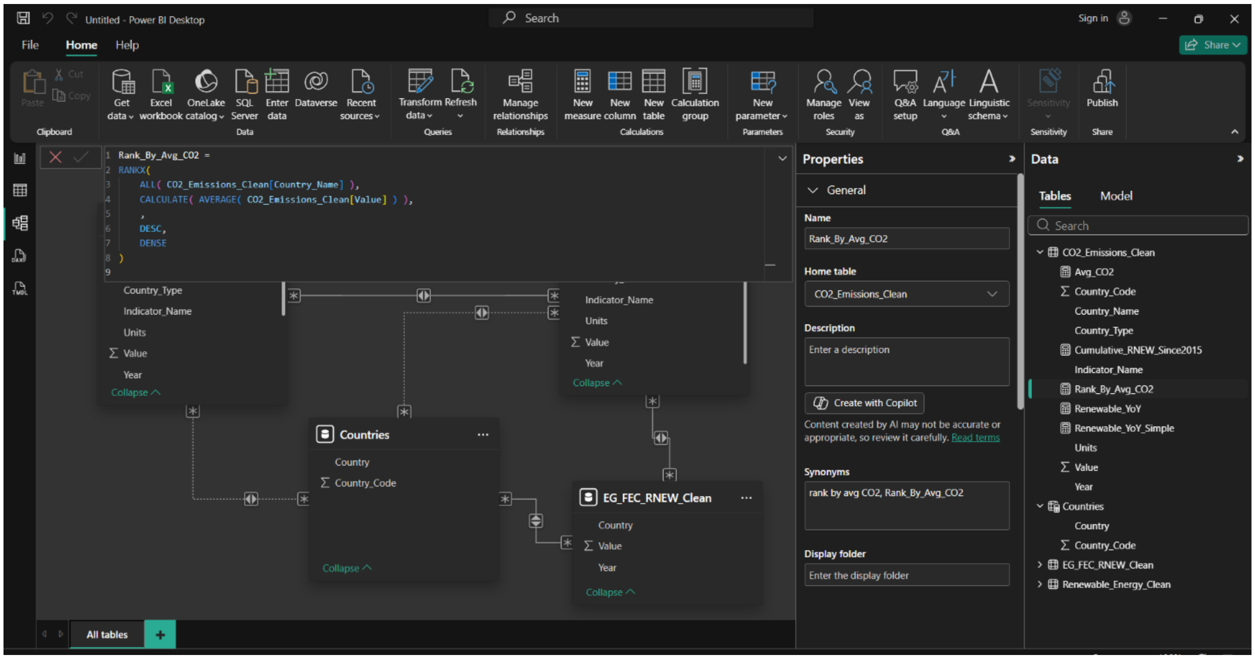
d. Cumulative SDG Progress Since 2015

It measures the total renewable energy since 2015 which is the official SDG tracking start year. It allows to shows the long-term progress used in cumulative line charts.



e. Ranking Countries by Average CO_2 Emissions

The ranking measure is useful when comparing countries based on emission intensity. It allows the dashboard to show top emi



tters, lowest emitters, and regional comparisons.

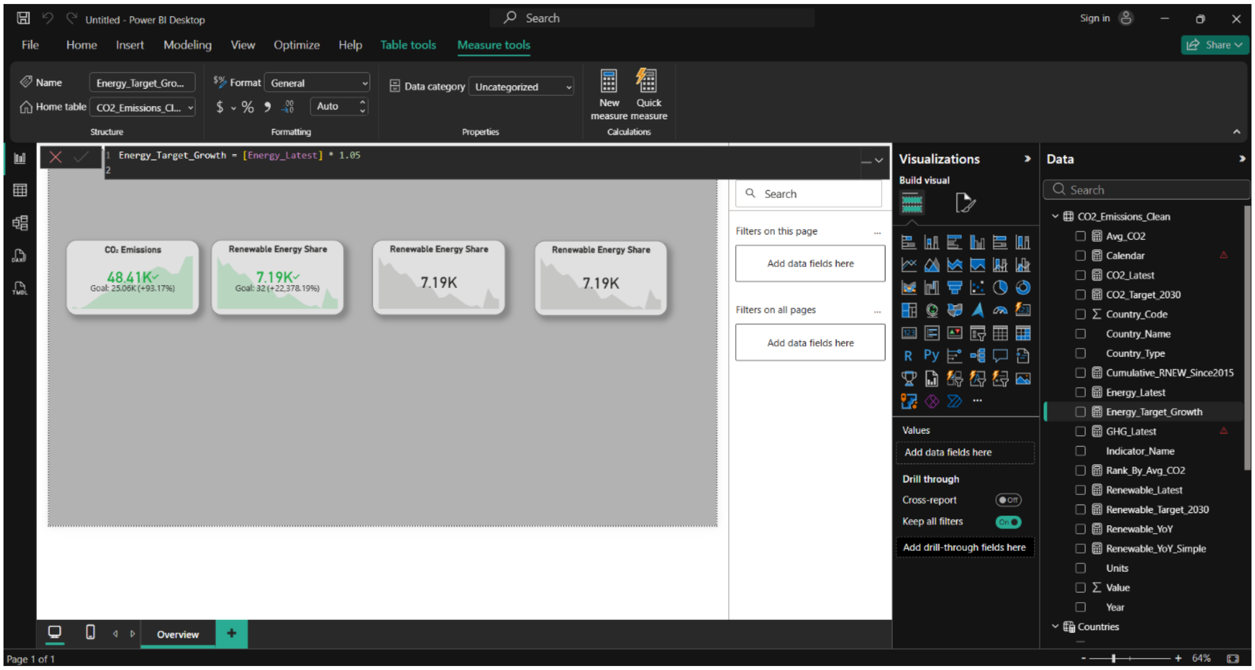
5.2 Custom Visuals

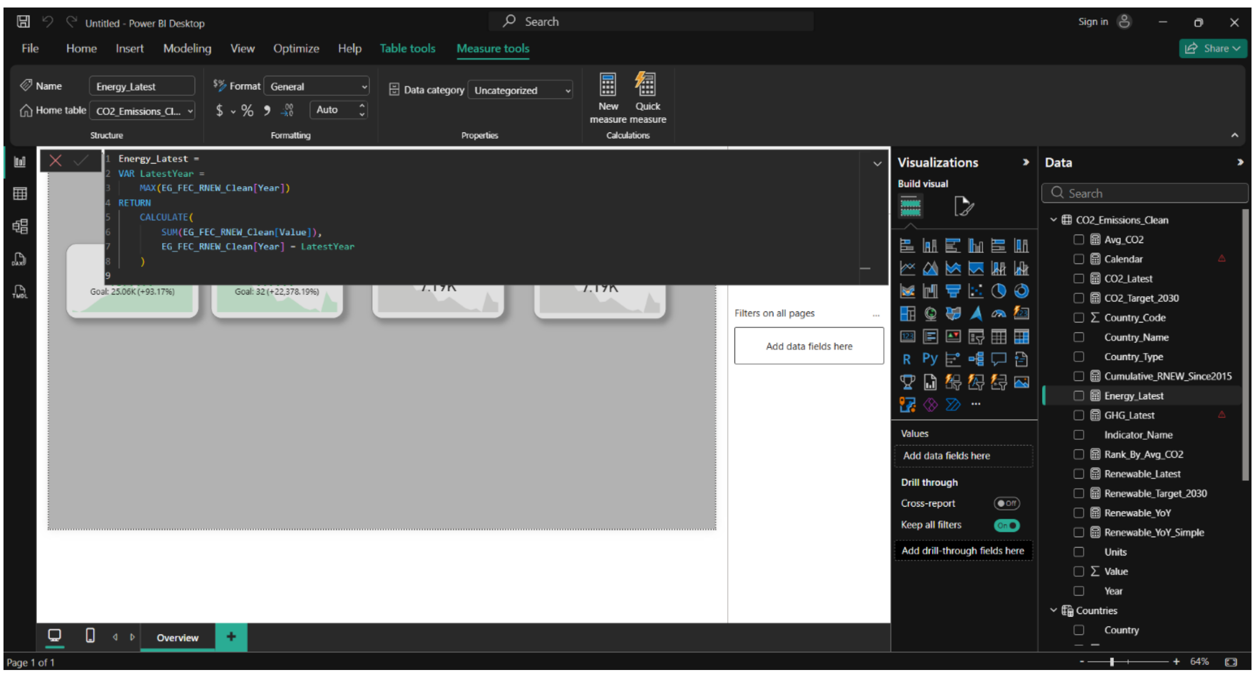
The custom visuals were used to track trends easily, compare countries, and evaluate progress toward global sustainability objectives. So for that below custom visuals were used.

KPI Cards

To provide a high-level summary metrics Key Performance Indicator (KPI) visuals were used to understand at a glance. KPIs allows to SDG progress monitoring of different trends as follows.

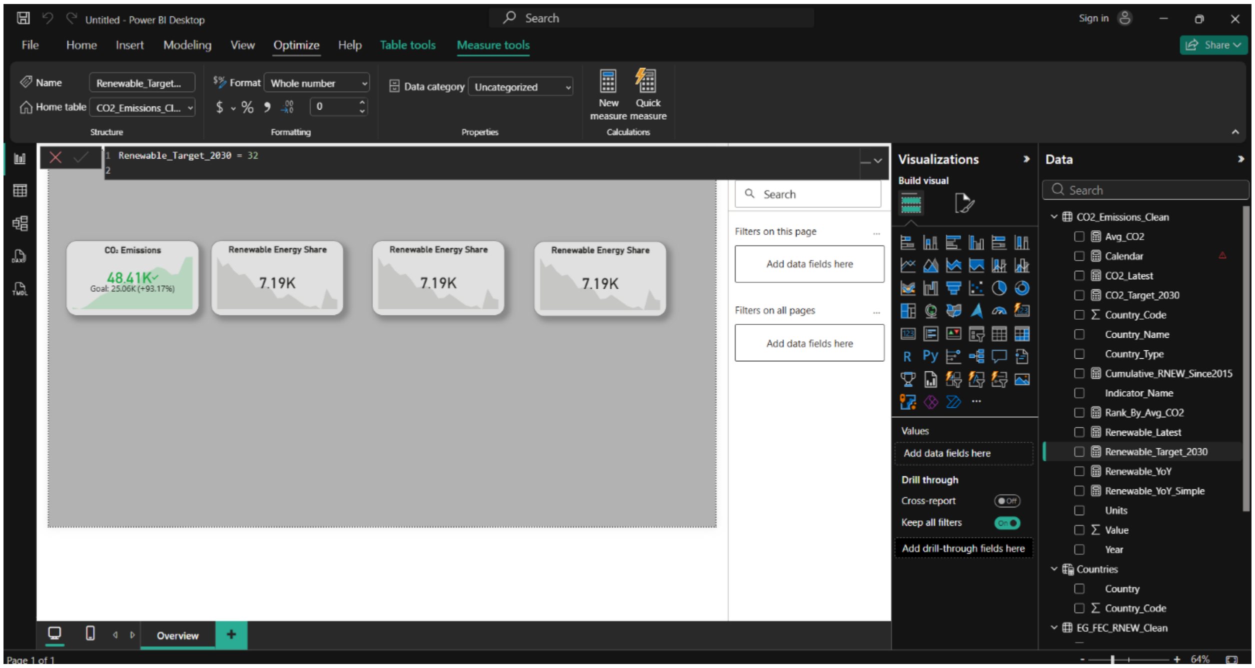
a. **Energy Consumption Trend**
The energy consumption KPI gives a quick overview regarding how a country's energy is changing over time. A DAX measure was used to present the latest year's energy consumption because Power BI cannot automatically limit a KPI to the latest year.

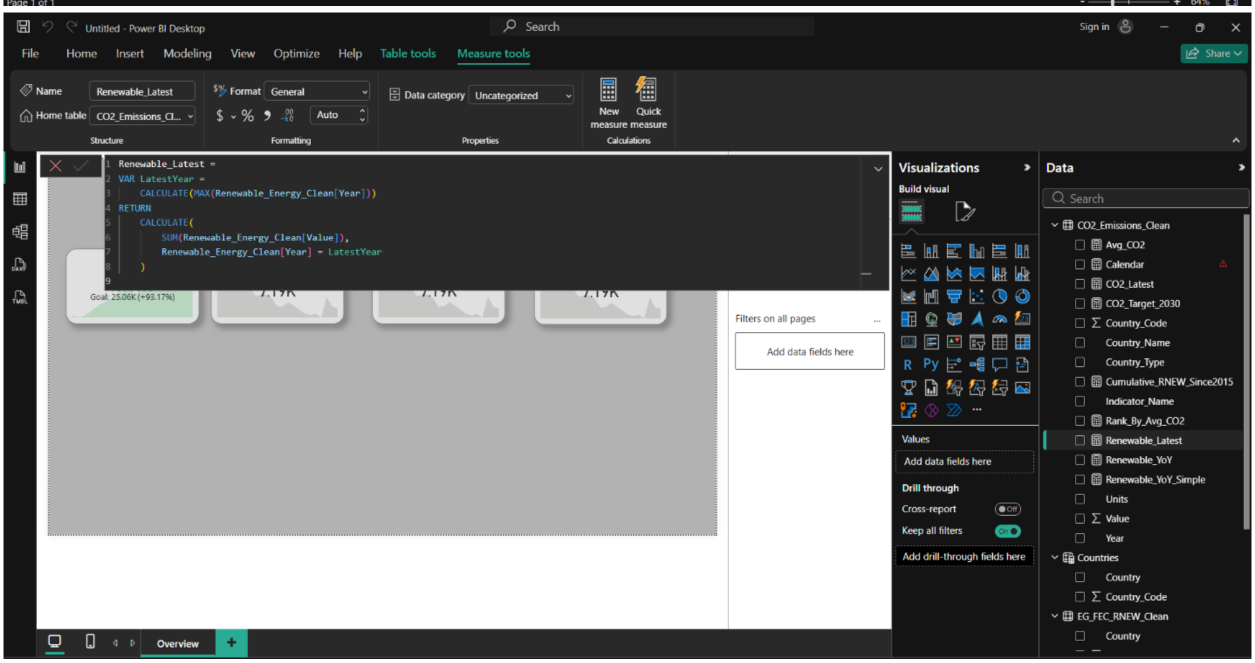




b. Renewable Energy Share

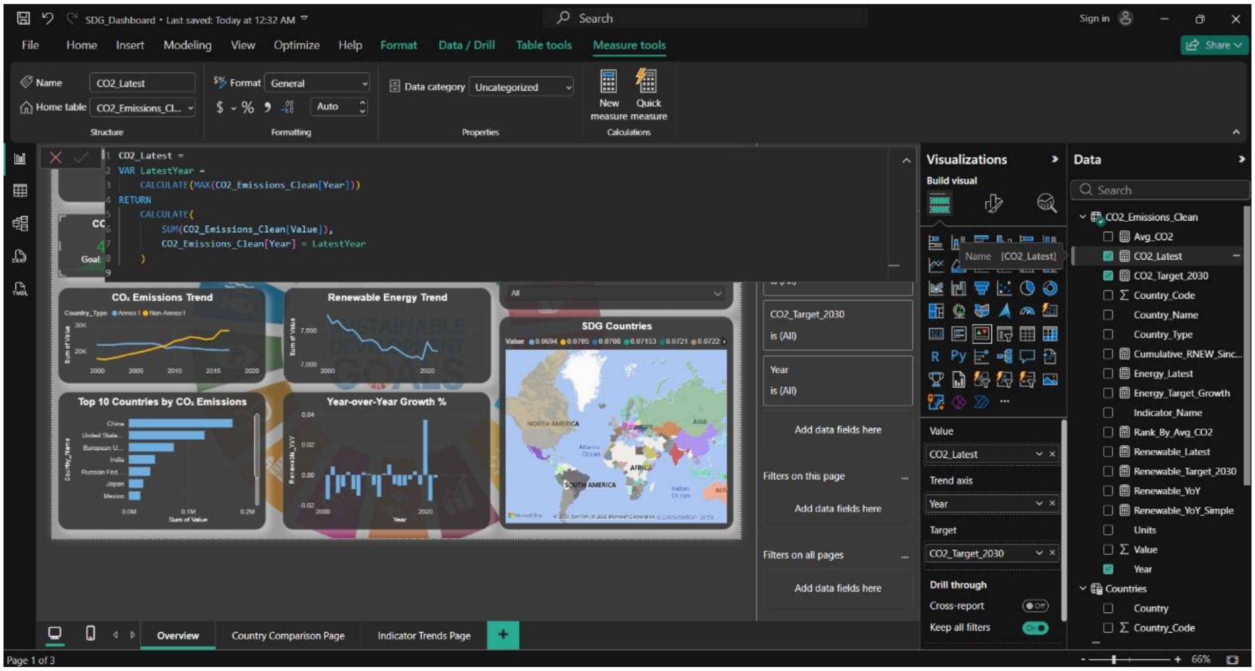
The renewable energy KPI provides a simple and immediate insight about a country's renewable energy performance. A DAX measures were created to display the most recent and accurate metric which shows the latest available value by filtering to a specific year or evaluating dynamic targets. The DAX allows the renewable energy share KPI to make more accurate, reliable and meaningful SDG reporting.

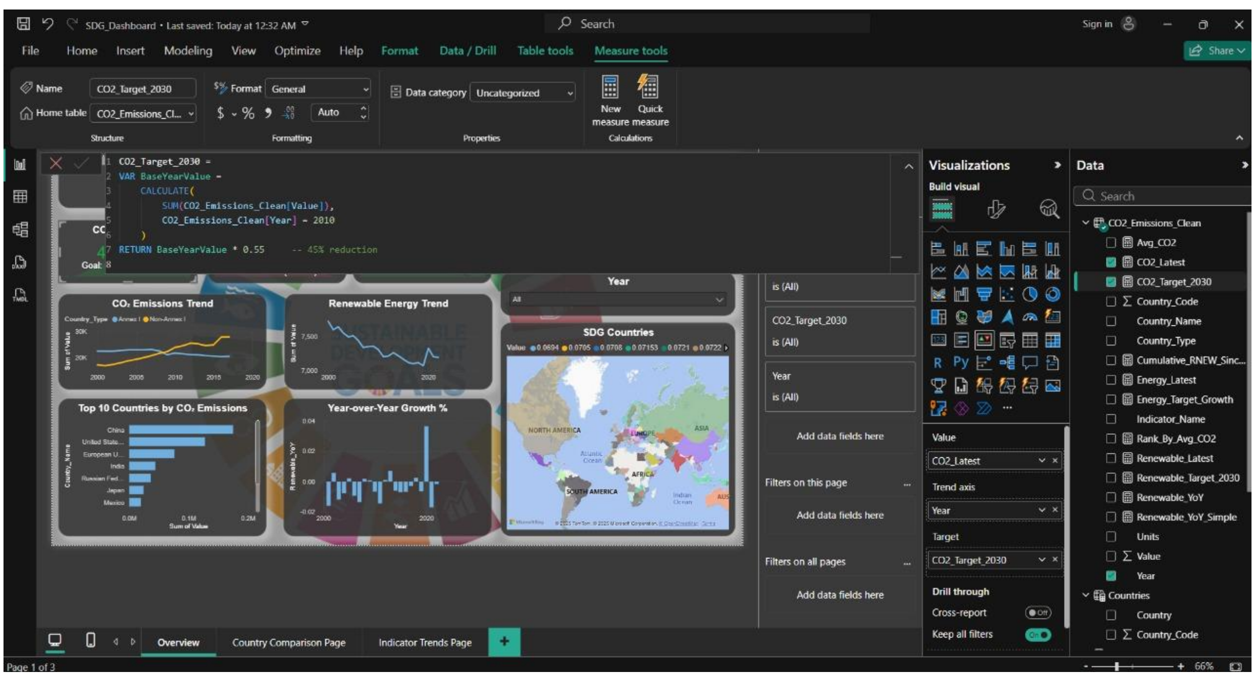




c. CO_2 Emissions

The KPI allows users to understand the climate performance without reading detailed emission chart. Because the CO_2 emission is vary every year, it is essential for analysis having most relevant insight. Therefore, a custom DAX is used to extract the most recent CO_2 emission level.


SDG_Dashboard • Last saved: Today at 12:32 AM
File
Home
Insert
Modeling
View
Optimize
Help
Format
Data / Drill
Table tools
Measure tools
Name
CO2_Latest
Home table
Format
General
Data category
Uncategorized
New measure
Quick measure
CO2_Latest =
VAR LatestYear =
CALCULATE(MAX(CO2_Emissions_Clean[Year]))
RETURN
CALCULATE(
SUM(CO2_Emissions_Clean[Value]),
CO2_Emissions_Clean[Year] = LatestYear
)
CO₂ Emissions Trend
Renewable Energy Trend
SDG Countries
Top 10 Countries by CO₂ Emissions
Year-over-Year Growth %
Visualizations
Build visual
Data
CO2_Emissions_Clean
Avg_CO2
CO2_Latest
CO2_Target_2030
Country_Code
Country_Name
Country_Type
Energy_Latest
Energy_Target_Growth
Indicator_Name
Rank_By_Avg_CO2
Renewable_Latest
Renewable_Target_2030
Renewable_YoY
Renewable_YoY_Simple
Units
Value
Year
Countries
Country
Value
Trend axis
Target
Drill through
Cross-report
Keep all filters
Filters on this page
Filters on all pages
Add data fields here
Overview
Country Comparison Page
Indicator Trends Page
Page 1 of 3

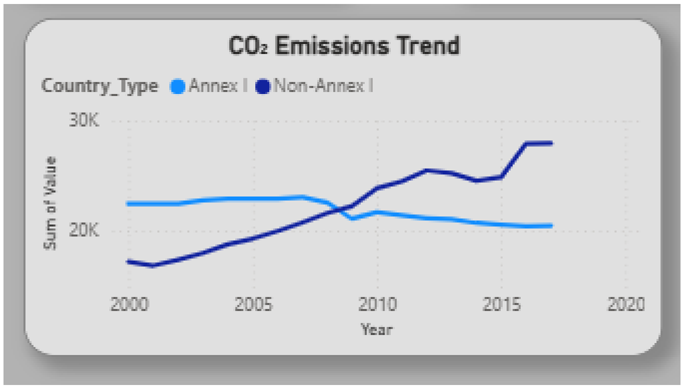

SDG_Dashboard • Last saved: Today at 12:32 AM
Name
CO2_Target_2030
Home table
Format
General
Data category
Uncategorized
New measure
Quick measure
CO2_Target_2030 =
VAR BaseYearValue =
CALCULATE(
SUM(CO2_Emissions_Clean[Value]),
CO2_Emissions_Clean[Year] = 2010
)
RETURN BaseYearValue * 0.55 -- 45% reduction
CO₂ Emissions Trend
Renewable Energy Trend
SDG Countries
Top 10 Countries by CO₂ Emissions
Year-over-Year Growth %
Visualizations
Data
Overview
Country Comparison Page
Indicator Trends Page
Page 1 of 3

Line Charts

To visualize the changes of CO_2 emissions and renewable energy year-by-year line charts were used. Line charts can be used to trend analysis to reveal long-term patterns, peaks, and transitions.

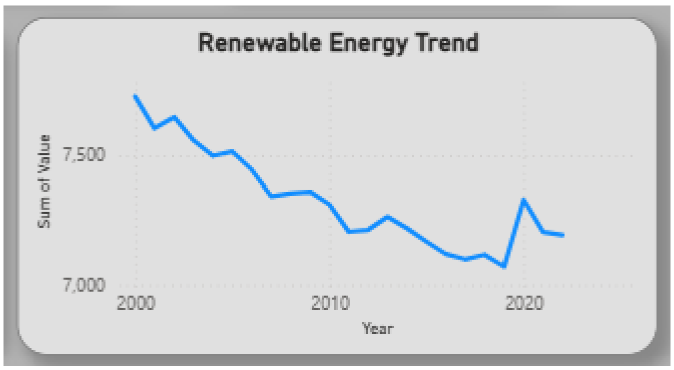
a. CO_2 emission trend line chart

Because CO_2 emissions act as time dependent trends, line chart was used to reveal long-term climate impact, years with significant rise or impact, and country-to-country performance differences. It reveals the patterns and anomalies that tables cannot show easily when climate analysis.



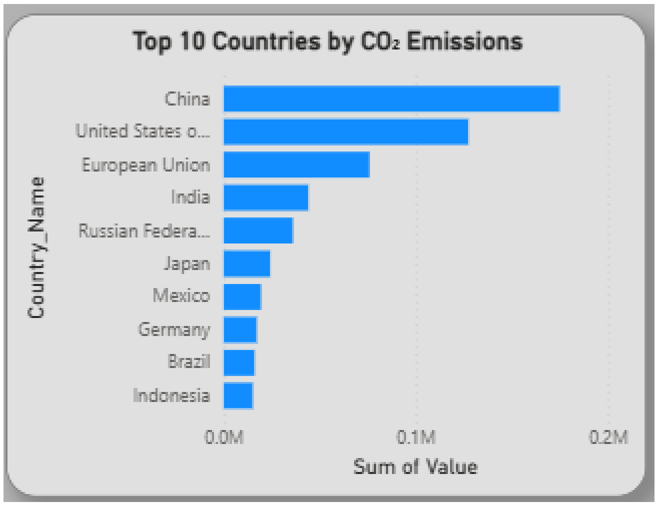
b. Renewable energy line chart

The line chart shows how renewable energy changed year over year while helping to analyze how fast renewable energy is growing, the growth consistency, and the growth difference between countries. It captures the rate of transition toward clean energy while supporting the energy policy evaluation.



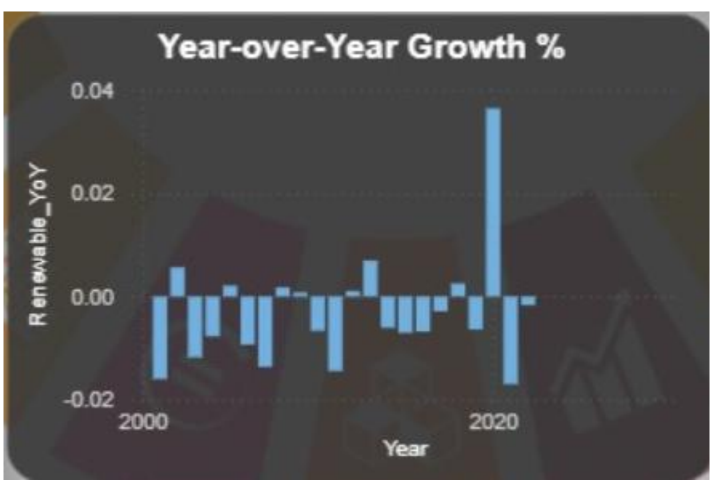
Bar Chart

Bar chart was used to highlight the highest and lowest performers in emissions across countries while understanding the leadership and inequality among those countries. It allows policymakers and researches to understand and determine where emission reduction efforts are most needed.



Column Chart

Column chart used to represent the speed of the improvement of the renewable energy. It shows the percentage growth from one year to the next by allowing comparing which years had the strongest improvement, which countries made rapid progress and the policy impact of them by highlighting the growth years and decline years.



Map Visual

The map visual used to showcase the global view of all SDG-reporting countries which included in the dataset. Map visuals act as a high-level visualization by covering global regions without selecting just countries. It allows users to identify the regional coverage of the SDG reporting.

6. Insights and Trends Identified from the Dashboard

6.1 Insights from Dashboard - Page 1

The dashboard provides data-driven insights about SDG 7 & SDG 13 Global Sustainability. It provides the understanding of how different countries adopting to clean energy, controlling carbon emission and the overall sustainability performance. When examining all the charts, KPIs and maps as a whole, it provides essential global patterns regarding energy use, climate impact and long-term development strategies. The KPI which shows the CO_2 emissions trend represent that the global emission has increased continuously especially among industrial countries. The manufacturing countries like China, India, United States indicate a noticeable upward because of the large-scale manufacturing, fossil-fuel-based electricity generation and high population density. Because of the many countries not yet enough to shift toward green industries at a large scale, still they heavily depend on high-carbon activities. The top 10 countries of CO_2 emitters emphasize the evenly distribution of the emissions around the world. As an important insight, the global targets remain unachievable unless major emitters are existing strongly while coordinating reduction strategies.

The renewable energy trend showcases the overall improvement of the renewable energy across many countries which gradually investing in solar, wind, hydroelectric and other clean sources. Sometimes there are fluctuations of the trend due to changes of the political leadership, renewable subsidies, international energy prices or disruptions like natural disasters or financial crisis. But it provides a clear long-term insight on how much renewable energy contribute to total consumption of energy. When considering the share of the renewable energy, it remains low while highlighting the dominance of fossil fuels in many countries. It represents the overall slow transition to sustainable energy. From the insights of Year-over-Year growth, some countries experience positive growth while others experiencing negative growth due to technical reversals or technical constraints.

The dashboard provides a critical insight that the world is moving two opposite directions. It means, renewable energy capacity and CO_2 emissions both are increasing. The renewable energy is not enough fast to cover the fossil fuel consumption. Though the global transition is happening, they are not so quick enough to meet climate changes, It is the most significant trend which the dashboard demonstrates.

6.2 Insights from Dashboard - Page 2

The second page of the dashboard clearly shows the continuous CO_2 emission increase year by year globally. When consider about the developed and developing countries via the Annex I and Non-Annex I respectively, in developed countries the emissions level is high due to industrial activities. But in developing countries the emissions are growing as economies expand and energy demand increases. The top 10 CO_2 emitting countries, highlight the countries which should target for the biggest impact regarding the global climate actions. The CO_2 trend over time shows a steady over the years in many countries. Few countries show a small decline emphasizing improved energy efficiency and climate policies.

When consider about the renewable energy performance, there is a clear upward trend in renewable energy production and consumption in worldwide. The top 10 countries of renewable energy outperform regarding the global renewable energy average significantly. The top countries are consists of strong policy frameworks, high public acceptance of renewables, financial capacity to invest new infrastructure and the flow of natural resources unlike other countries. Yearly fluctuations of renewable energy demonstrate the delays or changes, seasonal and climatic variations or transition challenges from old energy systems to new systems. Through the maximum renewable energy value, and maximum CO_2 emission value demonstrate the which year achieved the highest renewable output, the biggest polluter in the data model respectively. The map shows the CO_2 emission trends across the countries.

7. How the Analysis Supports Policymakers and NGOs

With the comprehensive insights that the dashboard provides regarding global energy use and climate performance on SDG 7 and SDG 13, can directly support policymakers, climate agencies, and NGOs in several impactful ways. Because the CO_2 emission continue to rise among the dashboard insights, it allows NGOs to advocate for sector-specific reforms and push them for faster decarbonization in critical regions. Time-series line charts help to guide policymakers by allowing them to comparing past strategies succeeded or failed within periods of policy success or setbacks. The top 10 emitters and renewable performers help NGOs to prepare grand proposals for high priority regions for equitable climate financing. The renewable energy insights allow NGOs to use them as evidence for establishing stable long-term renewable energy policies where the countries showing irregular growth. The comparisons about developed and developing countries can be used by policymakers to create differentiated climate strategies according to the tailored solutions based on country needs. Through the map visualization of spatial insights, decision makers allow to targeted actions rather than general ineffective approaches. As a summary, the dashboard converts raw data into actionable insights while enabling stakeholders to make informed, strategic, and impactful climate actions.