Unearthing the Environmental Impact of Human Activity:A Global CO2 Emission Analysis.

OVERVIEW :

Global warming is one of the biggest challenges currently being faced by the human race, although correlation is not causation, a likely cause of global warming is due to increased atmospheric carbon dioxide from human activities. **CO2 Emission** refers to the Carbon Dioxide emitted throughout the world. For this analysis we will be focusing on CO2 Emissions and its effect on the world we live in as well as some key factors and stats that may play a role in the emission of CO2 globally. Fossil fuel use is the primary source of CO2. The data throws light onto how much fossil fuels are burnt, per year per nation, which amounts to an increase in CO2 every year. This will help researchers and environment experts to predict global warming. So countries should set a goal to decrease this amount yearly.

Analysing Global Co2 Emission across countries from 1975 to 2020. This dataset contains a record of Co2 Emission by each Country and Region of Earth, here we are going to analyse and visualise Country wise, Region wise and Overall Co2 Emission on Earth.

INTRODUCTION

- Define Problem / Problem
  Understandingo
  Specify the
  business problem
  - Business requirements
  - Literature Survey
  - Social or Business Impact.
- Data Collection & Extraction from Database
  - Collect the dataset,
  - Storing Data in DB
  - Perform SQL Operations
  - Connect DB with Tableau
- Data Preparation
  - Prepare the Data for Visualization
- Data Visualizations
  - No of Unique Visualizations
- Dashboard
  - Responsive and Design of Dashboard
- Story
  - No of Scenes of Story
- Performance Testing
  - Amount of Data Rendered to DB ‘
  - Utilization of Data Filters
  - No of Calculation Fields
  - No of Visualizations/ Graphs
- Web Integration
  - Dashboard and Story embed with UI With Flask
- Project Demonstration & Documentation
  - Project Documentation-Step by step project development procedure

Define Problem / Problem Understanding :

Specify the business problem :

Refer Project Description

Business requirements :

The business requirements for analysing the Co2 Emission Globally over time, identifying affecting factors, creating interactive dashboards and reports, identifying areas for improvement, making data-driven decisions, comparing to countries average and creating forecasting models for future performance. The ultimate goal is to gain insights and reduce the emission through data visualization techniques.

Literature Survey :

A literature survey is a method of researching existing literature and studies related to a specific topic. In the context of analyzing the Global Co2 Emission, a literature survey would involve reviewing studies and articles that have been published on the topic of Emission, as well as studies specific to Co2.The literature survey would include sources such as academic journals, industry reports, and online articles. It would aim to identify different internal and external factors that are responsible and commonly used to determine Co2 Emission, as well as any best practices or strategies that have been identified for reducing emission.The literature survey would also explore any existing research on Co2 Emission specifically, and would aim to identify any challenges or opportunities that the Countries can opt to reduce emission.

Social or Business Impact. :

Social Impact: Carbon dioxide emissions are the primary driver of global climate change. It's widely recognised that to avoid the worst impacts of climate change, the world needs to urgently reduce emissions.

Business Model/Impact: By conducting an analysis the countries can identify areas for improvement and take steps to reduce factors that are responsible for Co2 Emission for environmental sustainability by improving the efficiency and transitioning to low carbon alternatives.

Data Collection & Extraction from Database :

Data collection is the process of gathering and measuring information on variables of interest, in an established systematic fashion that enables one to answer stated research questions, test hypotheses, and evaluate outcomes and generate insights from the data.

Collecting the data'S

Please use the link to download the dataset:
https://drive.google.com/file/d/1n764uDPT_ZF7kzGFLtpxkwBBsDBScbWm/view?usp=sharing

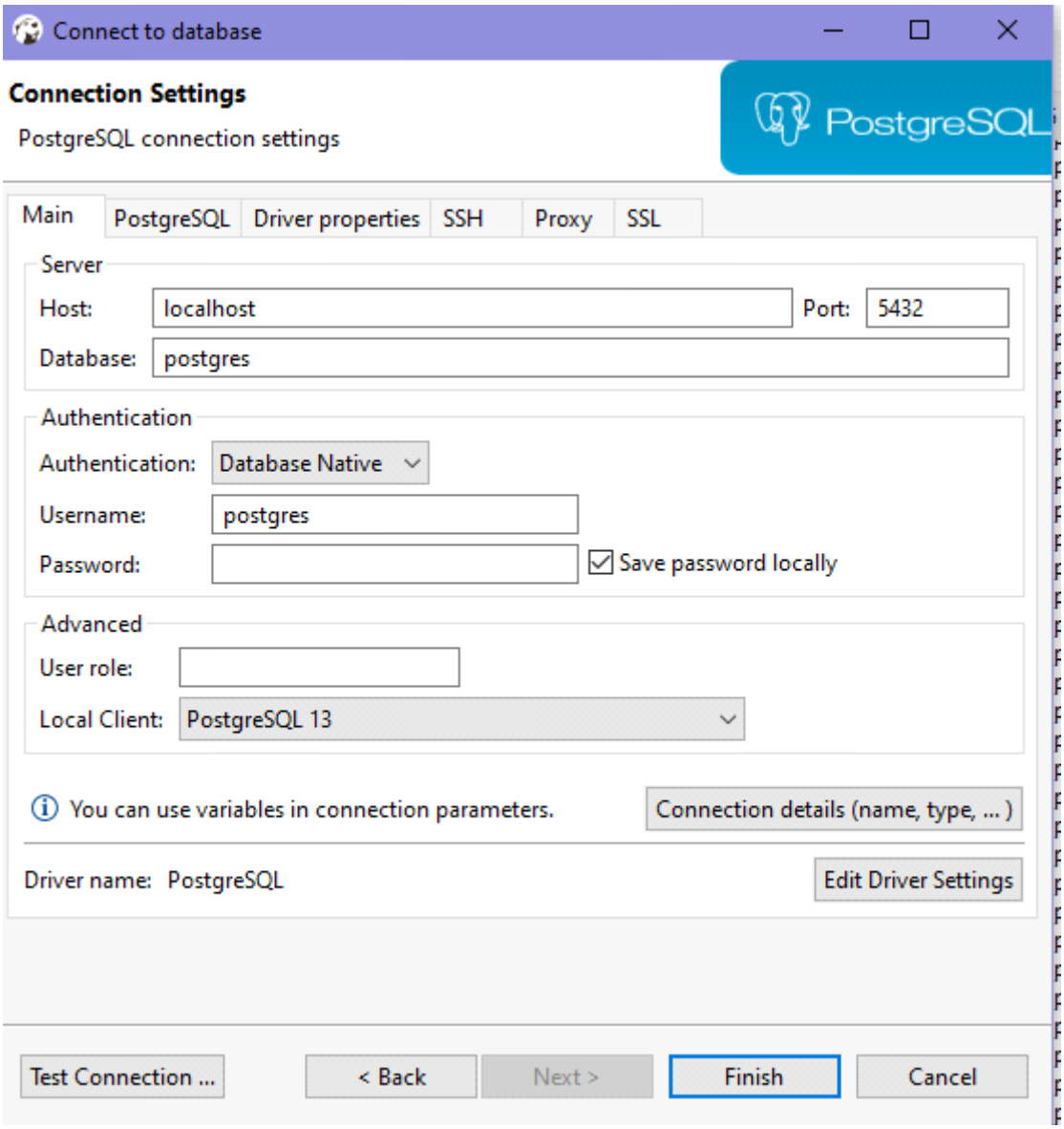
Understand the data :

Dataset consists CO2 emissions in metric ton per capita of every country around the world. The data is collected from 1975 to 2020. In this dataset Countries and regionsare included. Data is initially pre-processed using excel.

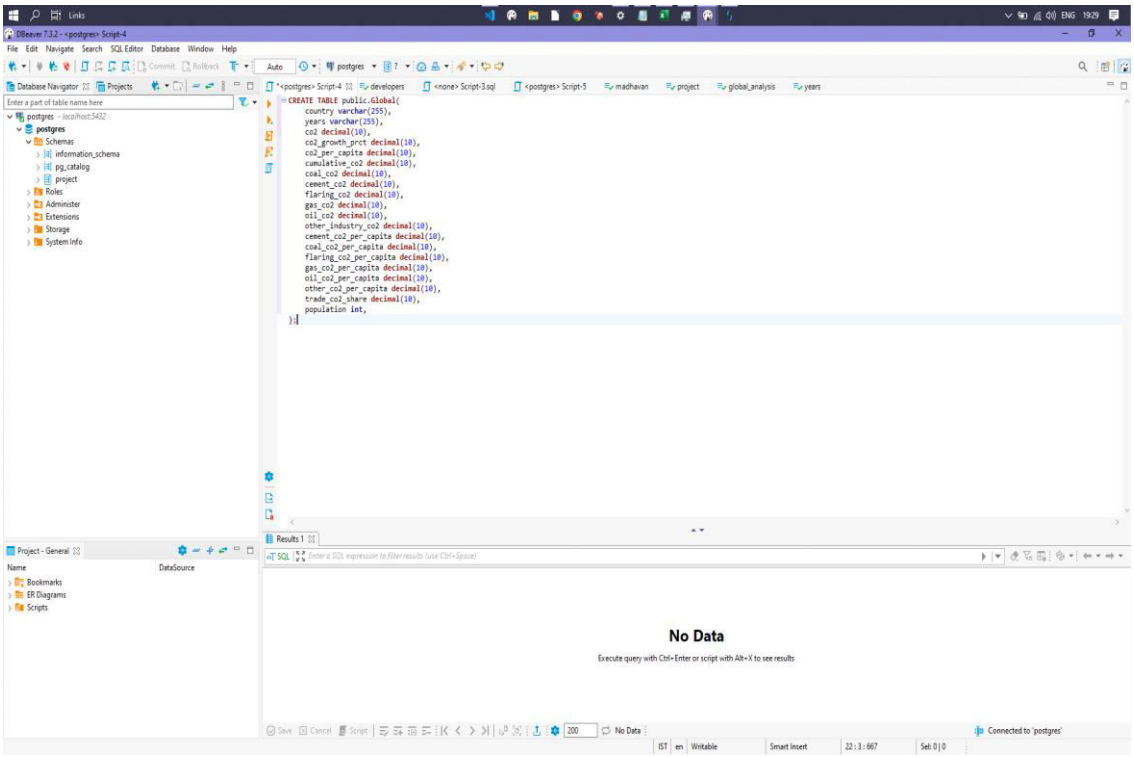
The dataset contains

- Country- Country for which Co2 is Recorded
- Year- Year the data was recorded
- Co2 Emission (In Million Metric Tons)
- Co2 Growth per Capita
- Co2 Per Capita
- Cumulative Co2
- Several Fossil Fuels rate of Emission

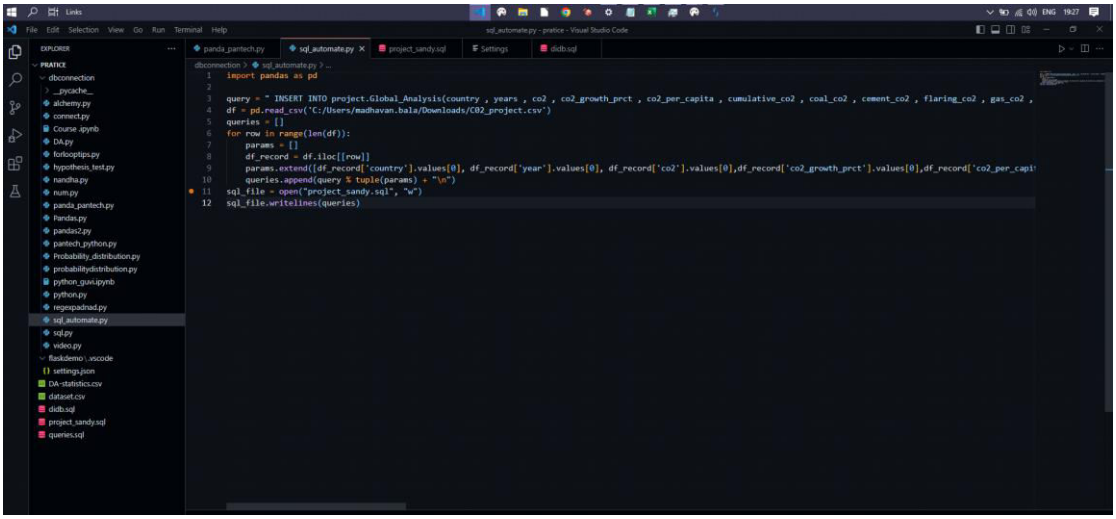
Creating Database to transfer data's :



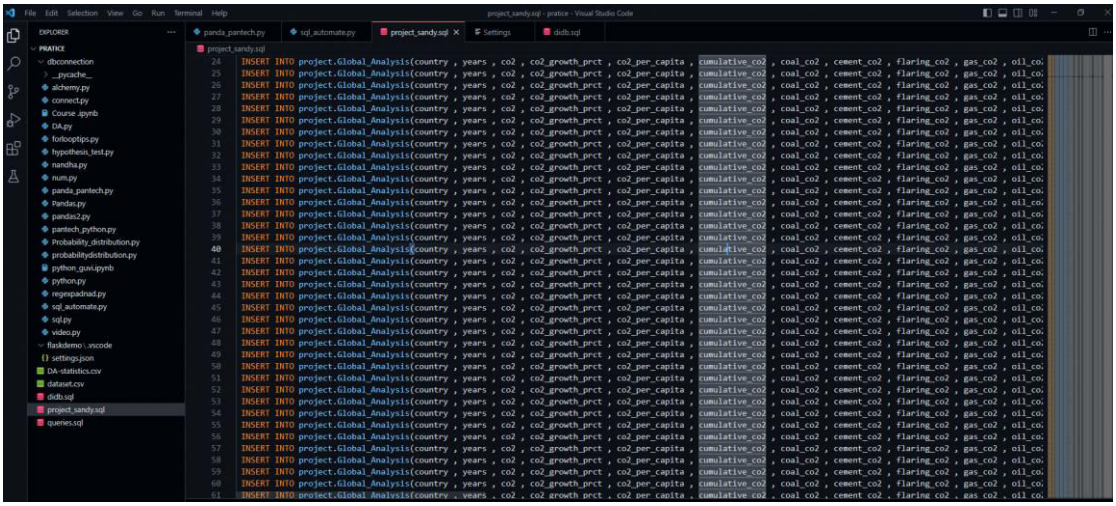
Creating Table in Database :



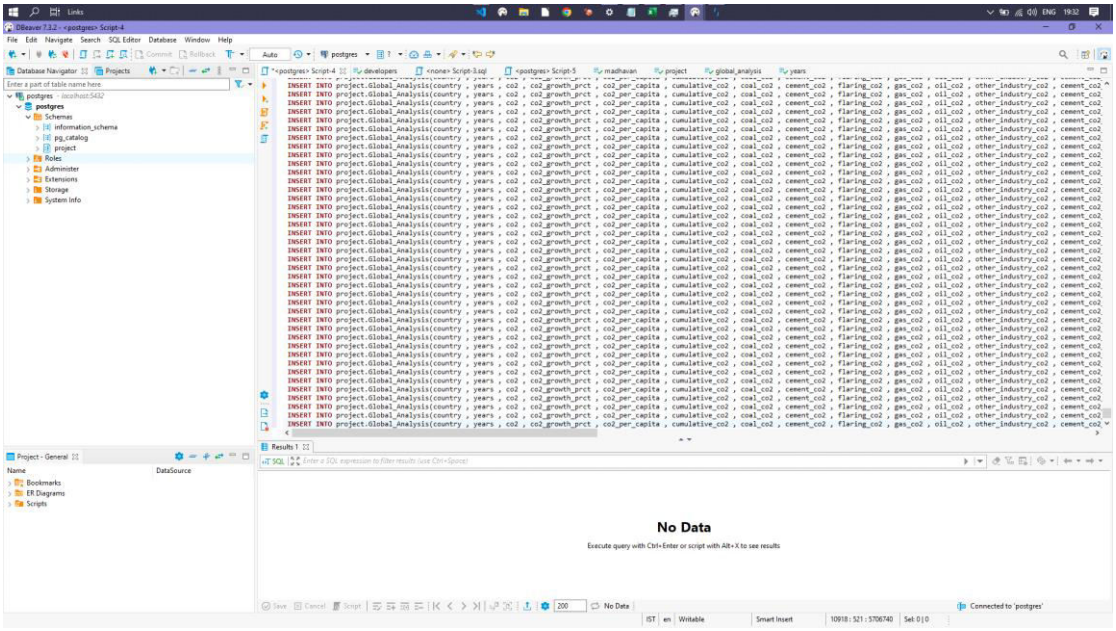
Automate a query using python language :



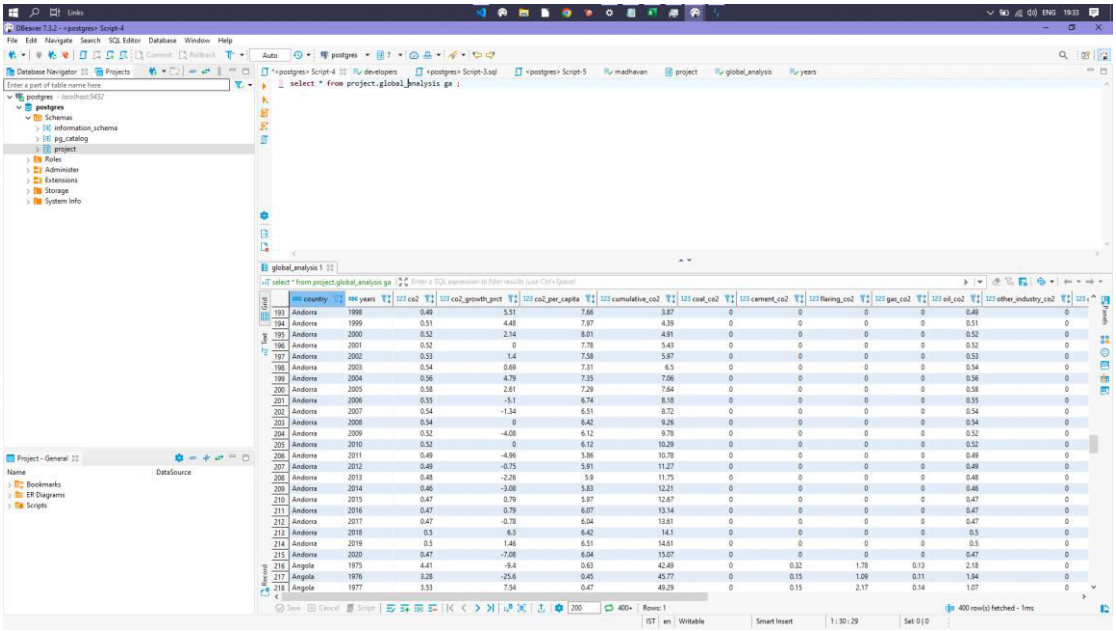
Stores the queries into .CSV format :



Copy and paste the queries in DB to give values in the table :



Using select * query to check the table values



Python code to Automate SQL Query
[Please refer the above images for better understanding]

```
import pandas as pd

query = " INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s','%s');"
df = pd.read_csv('C:/Users/madhavan.bala/Downloads/C02_project.csv')
queries = []
for row in range(len(df)):
    params = []
    df_record = df.iloc[[row]]
    params.extend([df_record['country'].values[0], df_record['year'].values[0], df_record['co2'].values[0],df_record['co2_growth_prct'].values[0],df_record['co2_per_capita'].values[0],df_record['cumulative_co2'].values[0],df_record['coal_co2'].values[0],df_record['cement_co2'].values[0],df_record['flaring_co2'].values[0],df_record['gas_co2'].values[0],df_record['oil_co2'].values[0],df_record['other_industry_co2'].values[0],df_record['cement_co2_per_capita'].values[0],df_record['coal_co2_per_capita'].values[0],df_record['flaring_co2_per_capita'].values[0],df_record['gas_co2_per_capita'].values[0],df_record['oil_co2_per_capita'].values[0],df_record['other_co2_per_capita'].values[0],df_record['trade_co2_share'].values[0],df_record['population'].values[0]])
    queries.append(query % tuple(params) + "\n")
sql_file = open("project_sandy.sql", "w")
sql_file.writelines(queries)
```

Sample query from using the automation

INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1977','7320.281','0.77','10.697','246465.042','2968.749','167.369','43.606','995.146','3145.412','0.0','0.245','4.338','0.064','1.454','4.596','0.0','0.0','684625058');
INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1978','7552.39','3.17','10.983','254017.432','2986.386','170.0','47.225','1045.356','3303.423','0.0','0.247','4.343','0.069','1.52','4.804','0.0','0.0','687959112');
INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1979','7732.355','2.38','11.193','261749.787','3076.822','167.743','35.216','1109.399','3343.175','0.0','0.243','4.454','0.051','1.606','4.839','0.0','0.0','691165026');
INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1980','7735.701','0.04','11.148','269485.489','3164.58','166.719','30.445','1139.153','3234.805','0.0','0.24','4.56','0.044','1.642','4.662','0.0','0.0','694253884');
INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1981','7453.972','-3.64','10.696','276939.46','3028.981','160.601','29.53','1180.789','3054.07','0.0','0.23','4.347','0.042','1.694','4.383','0.0','0.0','697202619');
INSERT INTO project.Global_Analysis(country , years , co2 , co2_growth_prct , co2_per_capita , cumulative_co2 , coal_co2 , cement_co2 , flaring_co2 , gas_co2 , oil_co2 , other_industry_co2 , cement_co2_per_capita , coal_co2_per_capita , flaring_co2_per_capita , gas_co2_per_capita , oil_co2_per_capita , other_co2_per_capita , trade_co2_share , population)VALUES ('Europe','1982','7394.331','-0.8','10.568','284333.791','3039.17','156.424','27.797','1195.424','2975.515','0.0','0.224','4.344','0.04','1.708','4.253','0.0','0.0','700021734');

Data Preparation :

Prepare the Data for Visualization :

Preparing the data for visualization involves cleaning the data to remove irrelevant or missingdata, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data, preparing the data for visualization software, and ensuring the data is accurate and complete. Since the Data is initially pre-processed we can skip this step. Basically this process helps to make the data easily understandable and ready for creating visualizations to gain insights into the performance and efficiency.

Data Visualization :

Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data setsmore accessible, intuitive, and easier to interpret. By using visual elements such as charts, graphs, and maps, data visualizations can help people quickly identify patterns, trends, and outliers in the data.

No of Unique Visualizations :

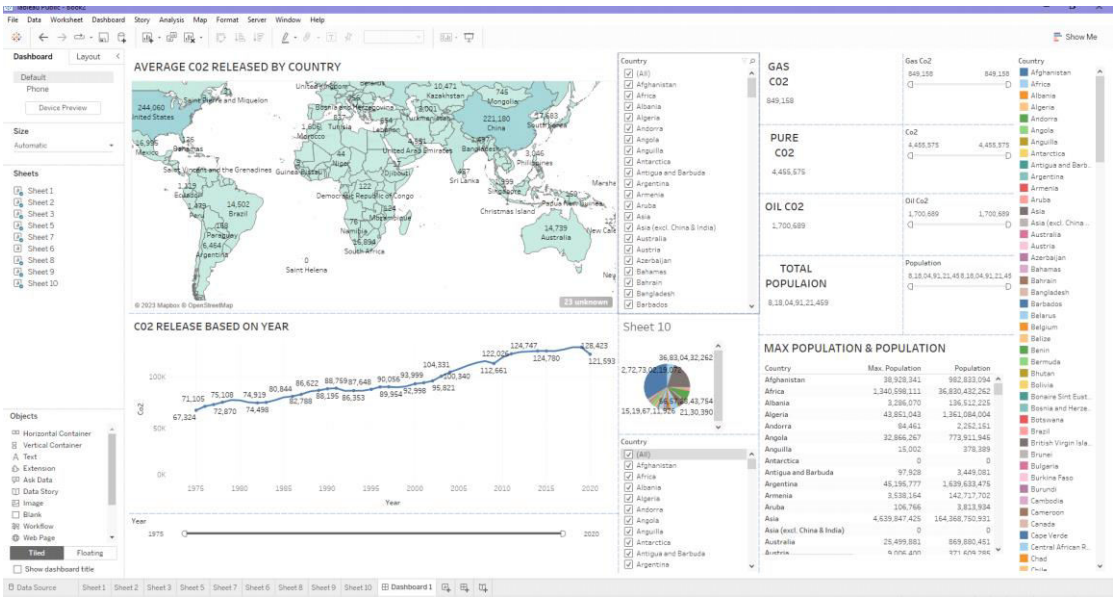
The number of unique visualizations that can be created with a given dataset. Some common types of visualizations that can be used to analyze the Co2 Emission include bar charts, line charts, Tree Map, scatter plots, pie charts, Maps etc. These visualizations can be used to compare performance, track changes over time, show Emission, and relationships between variables, breakdown of factors and emission by countries and continenrts.

Dashboard :

A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data,and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data inthe form of charts, graphs, and tables.

Responsive and Design of Dashboard :

The responsiveness and design of a dashboard for analyzing the globally Co2 Emission. It is crucial to ensure that the information is easily understandable and actionable. Key considerations for designing a responsive and effective dashboard include user-centred design, clear and concise information, interactivity, data-driven approach, accessibility, customization. Once you have created views on different sheets in Tableau, you can pull them into a dashboard.

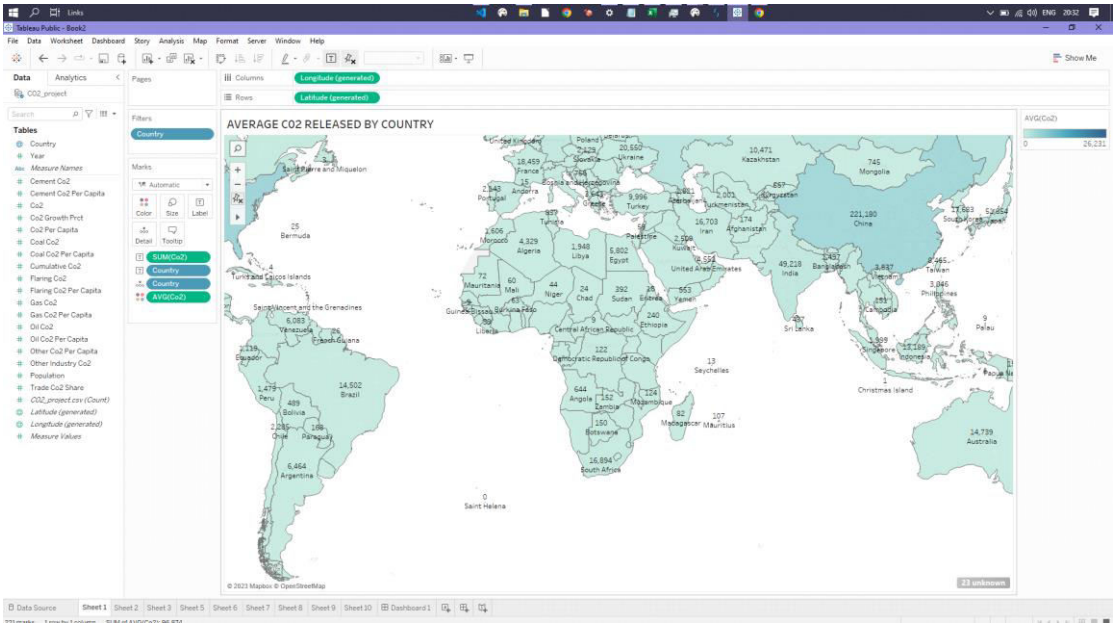


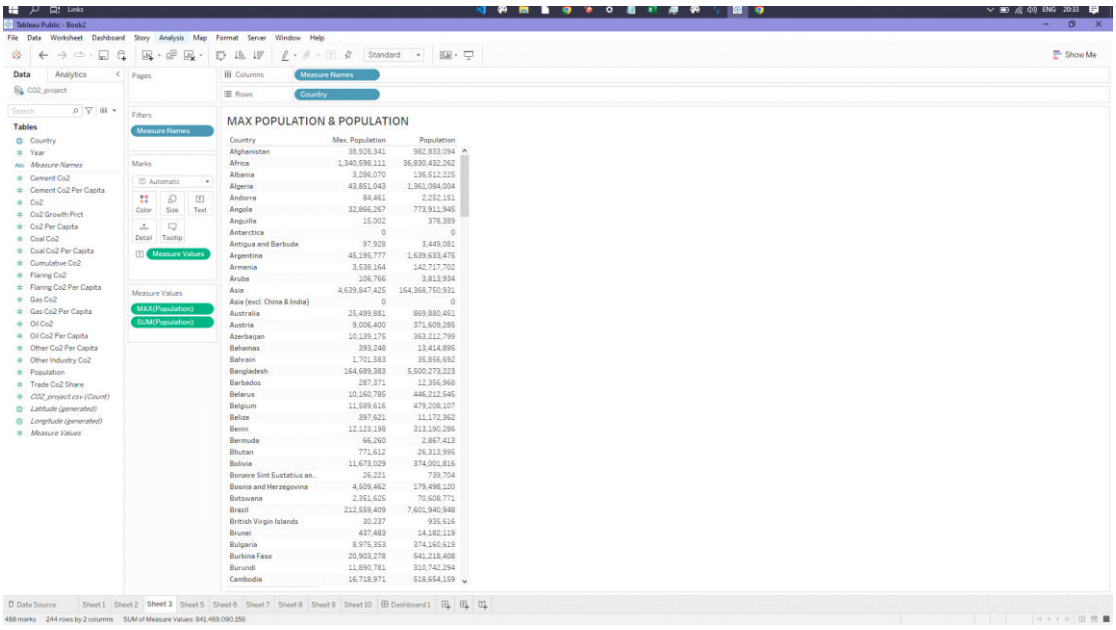
Story :

A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. A data story typically includes a clear introduction that sets the stage and explains the context for the data, a body that presents the data and analysis in a logical and systematic way, and a conclusion that summarizes the key findings and highlights their implications. Data stories can be told using a variety of mediums, such as reports, presentations, interactive visualizations, and videos.

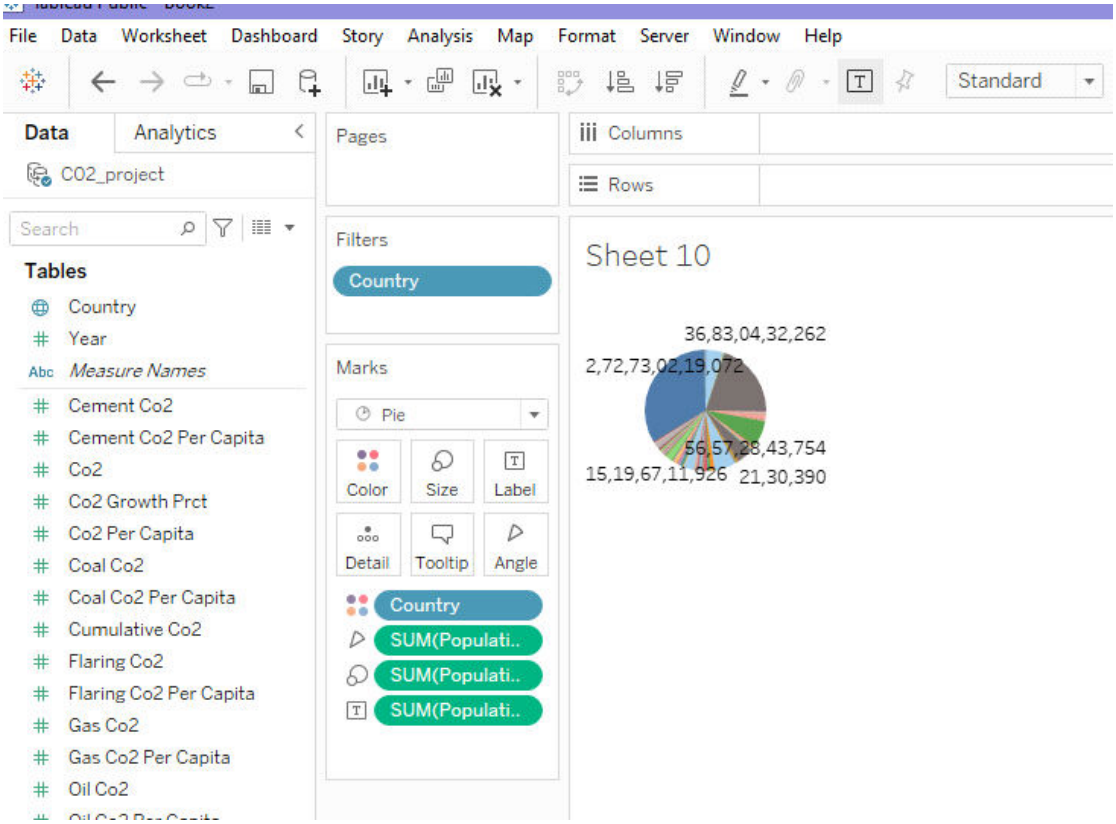
No of Scenes of Story :

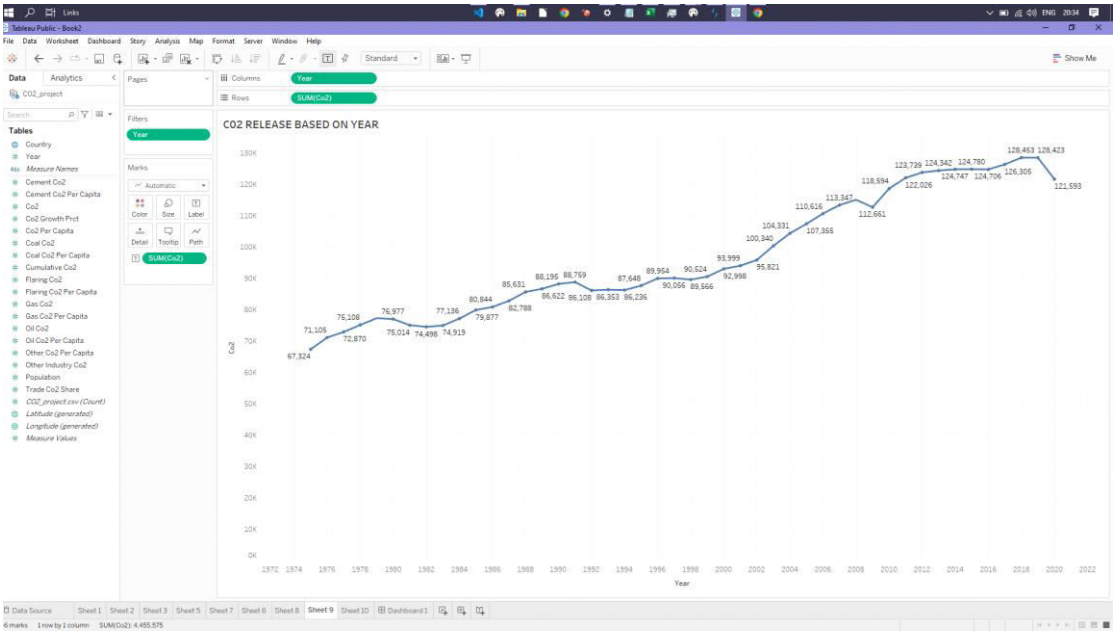
The number of scenes in a storyboard for a data visualization analysis of the Co2 Emission will depend on the complexity of the analysis and the specific insights that are trying to be conveyed. A storyboard is a visual representation of the data analysis process and it breaks down the analysis into a series of steps or scenes.





Country	Max. Population	Population
Afghanistan	38,928,341	982,833,094
Africa	1,340,598,111	36,830,432,262
Albania	3,286,070	136,612,225
Algeria	43,851,043	1,361,084,004
Andorra	84,461	2,262,151
Angola	32,866,267	773,911,945
Anguilla	15,002	378,389
Antarctica	0	0
Antigua and Barbuda	97,928	3,449,081
Argentina	45,196,777	1,639,633,475
Armenia	3,538,164	142,717,702
Aruba	106,766	3,813,934
Asia	4,639,847,425	164,368,750,931
Asia (excl. China & India)	0	0
Australia	25,499,881	869,880,451
Austria	9,006,400	371,609,285
Azerbaijan	10,139,176	363,212,799
Bahamas	393,248	13,414,895
Bahrain	1,701,583	35,856,692
Bangladesh	164,689,383	5,500,273,223
Barbados	287,371	12,356,968
Belarus	10,160,785	446,212,545
Belgium	11,589,616	479,208,107
Belize	397,621	11,172,362
Benin	12,123,198	313,190,286
Bermuda	66,260	2,867,413
Bhutan	771,612	26,313,995
Bolivia	11,673,029	374,001,816
Bonaire Sint Eustatius an..	26,221	739,704
Bosnia and Herzegovina	4,509,462	179,498,120
Botswana	2,351,625	70,608,771
Brazil	212,559,409	7,601,940,948
British Virgin Islands	30,237	935,616
Brunei	437,483	14,182,119
Bulgaria	8,975,353	374,160,619
Burkina Faso	20,903,278	641,218,408
Burundi	11,890,781	310,742,294
Cambodia	16,718,971	518,654,159



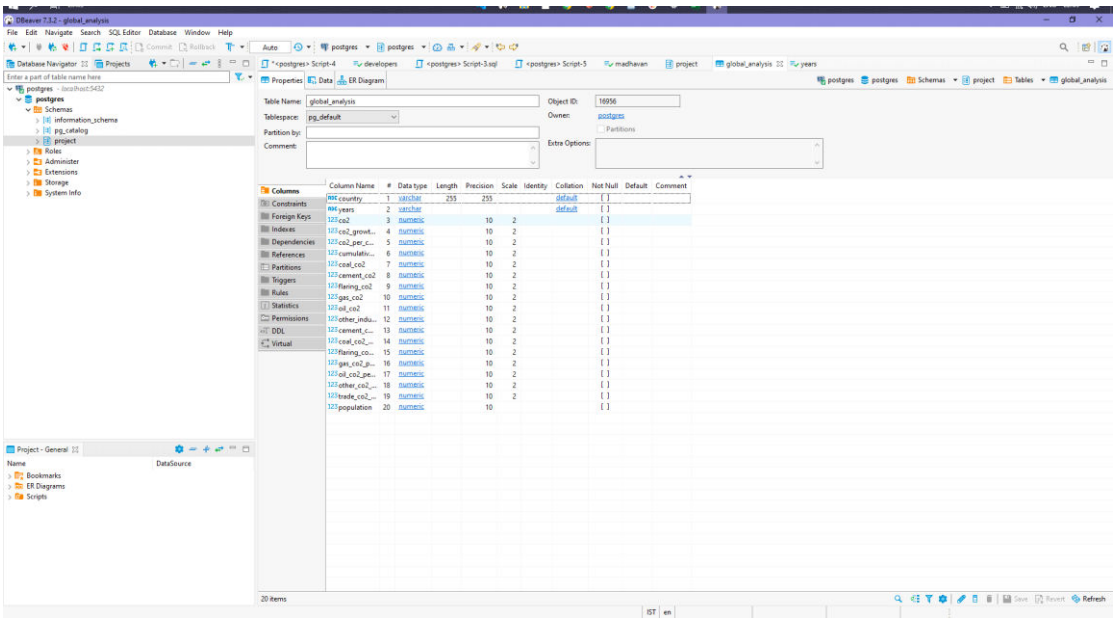


Performance Testing

Amount of Data Rendered to DB

- The amount of data that is rendered to a database depends on the size of thedataset and the capacity of the database to store and retrieve data.

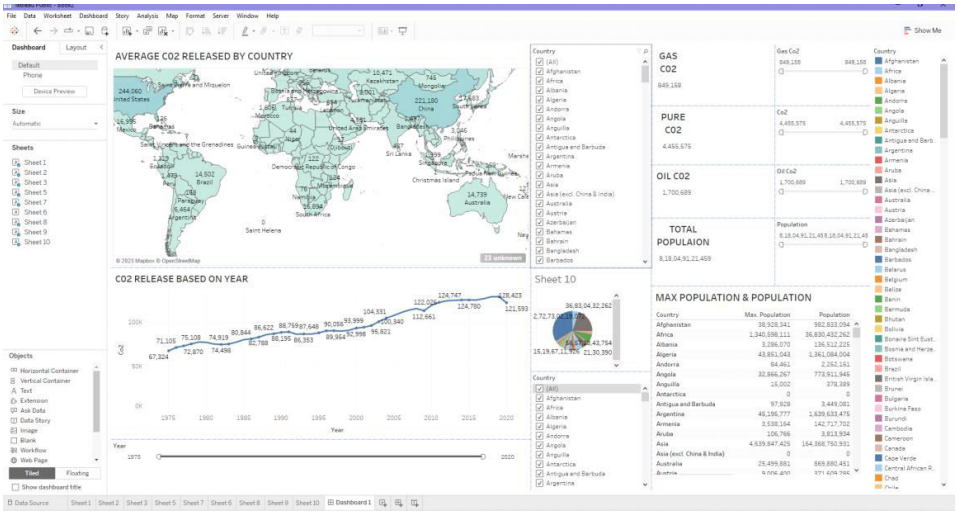
- Open the SQL Server Management Studio, go to the database then click toexpand the tables, select the table and Right click and select Properties

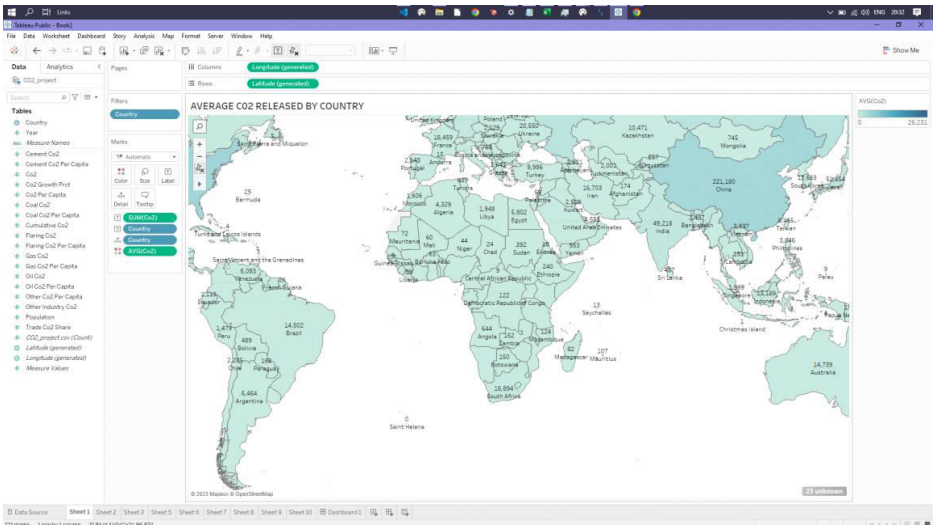


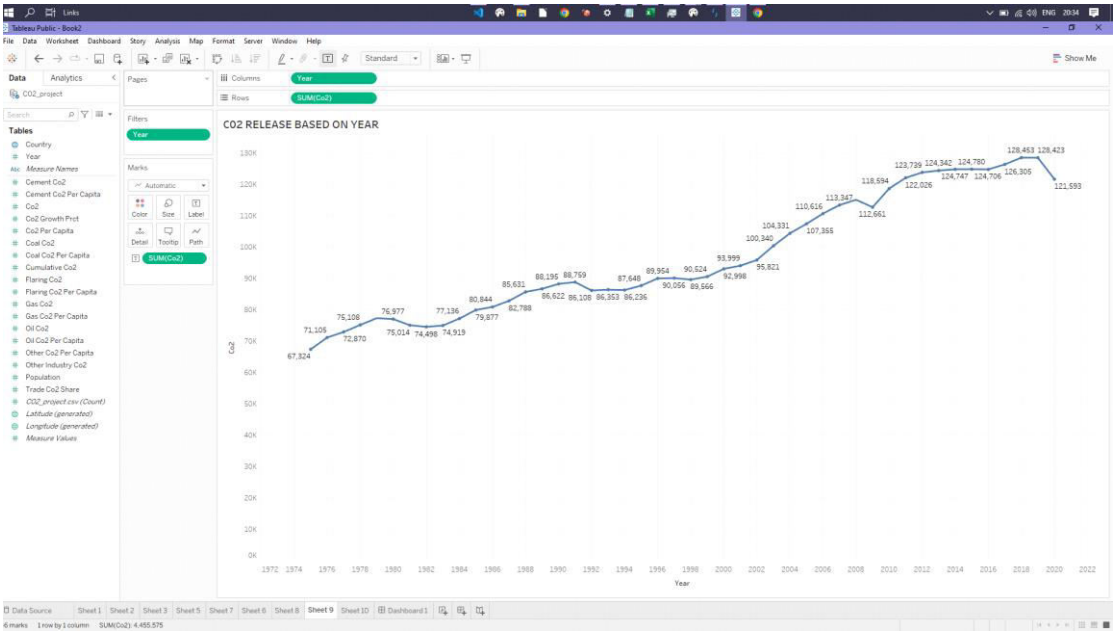
No of Visualizations/ Graphs

- Top World Emission
- Top Emitting Countries
- Co2 Emission over Time
- Co2 Emission India vs USA
- Total Emission by Continents
- Co2 Emission per Capita
- Co2 Emission by International Factors
- Emission Rate over Years
- Donut Charts-Coal Co2,Cement Co2,Gas Co2,Oil Co2
- Co2 Emission over past 10 years
- Continent Contribution in Co2 Emission
- Cumulative Co2 and Co2 per Capita
- Co2 Emission in 2020
- China vs India Co2 emission due to internal factors
- Overall Contribution by China in Co2 Emission

Dashboard and Story embed :


AVERAGE C02 RELEASED BY COUNTRY
C02 RELEASE BASED ON YEAR
GAS C02
PURE C02
OIL C02
TOTAL POPULAION
MAX POPULATION & POPULATION


AVERAGE C02 RELEASED BY COUNTRY


C02 RELEASE BASED ON YEAR

Project Documentation :

Below mentioned deliverables to be submitted along with other deliverables

Project Documentation-Step by step project developmentprocedure.