**TABLE OF CONTENT**

# 1.0 BACKGROUND PROJECT 1-3

1.1 Problem need to be solved 4

1.2 Objective 5

1.3 Data schema 6-9

# 2.0 ARCHITECTURE AND ETL PIPELINE

2.1 Pipeline structure 10-11

2.2 Data Warehouse Architecture 12

2.3 ETL Pipeline 13

2.3.1 Step to Extract and Transform (Phyton)

2.3.2 Step to Load Data into Database (PostgreSQL)

2.3.3 Step to Load OLAP Data into Visualization Tool (Tableau)

# 3.0 DATABASE 14-15 4.0 RESULT AND DATA ANALYSIS

4.1 OLAP Operations

4.1.1 Slicing 16-17

4.1.2 Pivot 18-19

4.1.3 Roll Up 20-21

4.1.4 Dicing 22-23

4.2 Visualization

4.2.1 Average Yield Value for Each Food Item Based on Average

Wind Speed 24

4.2.2 Top 10 Highest Country Based on Average Wind Speed 25

4.2.3 Top 10 Average Wind Speed Effect the Food Supply 26-27

4.2.4 Total Port of Vessels for each Country 28

# 5.0 CONCLUSION 29 6.0 REFERENCES 30

# 7.0 APPENDICE 31-45

iii

## 1.0 Background Project

The Sustainable Development Goal of Zero Hunger (SDG 2) is one of the 17 global goals by the United Nations in 2015 as part of the 2030 Agenda for Sustainable Development. SDG 2 aims to defeat hunger and develop sustainable agriculture worldwide. For our project, the title we have chosen is the Food Supply Chain Monitoring and Optimization System which is closely related to the Sustainable Development Goal of Zero Hunger (SDG 2). This project recognizes the urgent need for establishing an efficient and sustainable food supply chain in order to address issues with food security.

Our project's main goal is to create a complete system that analyses and improves the food supply chain. To do this, we make use of a variety of datasets that offer crucial data for efficient monitoring and optimization. The worldwide nation dataset, which acts as the project's fact table, is one important dataset used in this project. Among the attributes in this dataset are the nation (primary key), the region, the population, and the area (sq). We may learn important things about the expanding populations and how they affect food consumption in various nations by utilising this dataset. Understanding these dynamics enables us to predict food demand and make informed plans to ensure everyone has access to food more accurately.

Additionally, we make use of the global food consumption dataset, which includes information on the country, year, population, and particular food consumption values (such as those for beef(kg/capita), poultry(kg/capita), sheep (kg/capita), soy, maize (1000 tonnes), rice (1000 tonnes), soy(1000 tonnes) and wheat (1000 tonnes). We may examine consumption trends and preferences across nations using this dataset. We can make sure there is a sufficient range of healthy food alternatives available to fulfill the various nutritional demands of different locations by detecting these trends and monitoring changes in food preferences. We can address any shortfalls in food availability and make wise decisions to improve food security thanks to this study.

We also pay close attention to the global food supply port transportation dataset, which offers important details on the transportation component of the food supply chain. Data on nations, port names, ships in port, typical departure and arrival times, type, region local coverage, and area global coverage are all included in this dataset. We can find transportation system inefficiencies and boost the overall effectiveness of food distribution by analyzing this dataset. By streamlining the transport procedure, we lower waste and minimize damage, guaranteeing that food is delivered effectively and safely.

Additionally, the information on global weather conditions is crucial for agricultural planning for keeping an eye on the food supply chain. The features of this dataset contain information about the nation, city, latitude, longitude, average air temperature, average wind direction (in degrees), average wind speed (in km/h), and average sea level air pressure (in hPa). By taking into account these variables, we can determine how weather conditions affect crop development and guarantee a steady supply of food. Understanding weather patterns enables us to foresee problems and take the necessary action to lessen their impact on agricultural productivity, food supply and monitor the transportation condition at the same time.

The worldwide crop yield dataset, which contains data on parameters including country, food code, food item, year, unit, and yield value, is the last one we use. We can examine production levels and trends across multiple countries and categories of food thanks to this information. We can analyze agricultural performance, pinpoint low-productivity areas, and improve crop production methods by analyzing this data. These observations help to improve overall productivity, decrease food waste, and increase the effectiveness of the food supply chain.

In conclusion, the project Food Supply Chain Monitoring and Optimization System is created to solve the issues of food security. Our initiative intends to improve the efficiency, sustainability, and responsiveness of the whole food supply chain by utilizing data-driven insights and applying optimization methodologies. We aim to significantly affect the eradication of hunger and the establishment of a more sustainable and secure global food system by helping to accomplish the Sustainable Development Goals, particularly SDG 2 of Zero Hunger.

## 1.1 Problem need to be solved

The Food Supply Chain Monitoring and Optimization System project aims to address food security by developing an approach that leverages data-driven insights and optimization strategies to enhance the effectiveness, sustainability, and adaptability of the entire food supply chain. The project aims to make significant contributions towards achieving the Sustainable Development Goal of Zero Hunger (SDG 2).

The problem to be solved:

1. How can consumption patterns and preferences of different countries be analyzed to gain insights into demand trends and optimize the food supply chain accordingly?
2. What measures can be implemented to streamline the distribution process, reduce losses, and improve overall supply chain performance in order to enhance efficiency in food distribution systems?
3. How can weather data be utilized to understand the impact of weather conditions on crop transportation of food item worldwide to ensure a reliable and resilient food supply in the face of changing weather patterns and extreme events?

These problems highlight the need for data analysis and optimization strategies to improve various aspects of the food supply chain, including crop selection, distribution systems, weather impact, and understanding consumption patterns. By addressing these challenges, the project aims to contribute towards achieving SDG 2 and promoting sustainable and efficient food supply chains globally.

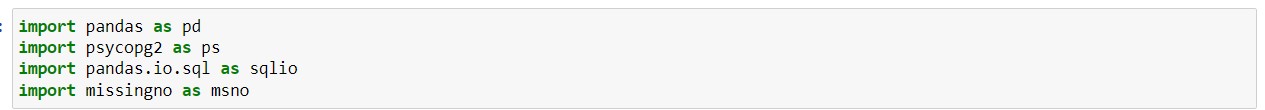
## 1.2 Objective

The objective of this project are:

1. To improve food supply chain operations based on consumption patterns.
2. To enhance efficiency in food distribution systems.
3. To utilize country weather data for reliable food supply.

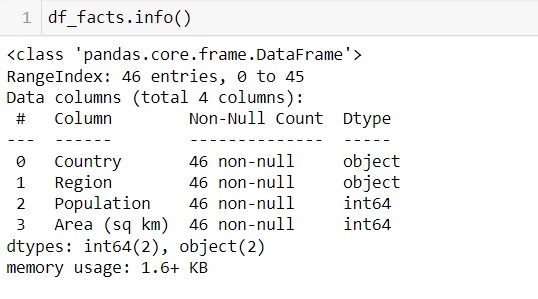
## 1.3 Data Schemas

A database schema defines how data is organized within a relational database including table names, fields, data types, and the relationships between these entities. A schema describes the organization and storage of data in a database and defines the relationship between various tables. This dataset of this project contains five tables which are Global Country acts as facts table and the other datasets are Global Crop Yield, Global Food consumption, Global Weather Condition, and Global Food Supply Port Transportation. In this project, we used 4 library packages which are pandas to display the data schema with python code in Jupyter as shown in Figure 1.3.1.



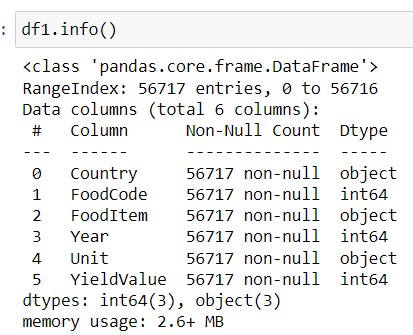
## Figure 1.3.1: Libraries were used to find data schema

The figure below shows the



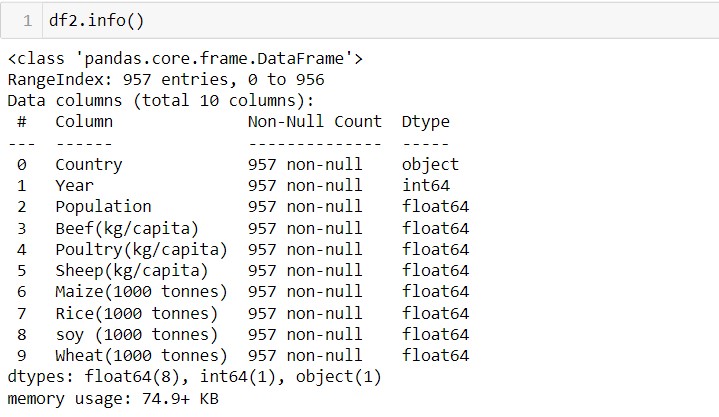
### 1.3.2: Data Schema Global Country Tables

The Figure 1.3.2 shows the Global Crop Yield Table from the Food Supply Monitoring and Optimization System. This table has 56717 rows and 6 columns in total such as Country, FoodCode, FoodItem, Year, Unit and YieldValue. It shows that there are 3 integer (int64) columns and 3 columns with object data type which are typically represented as string.



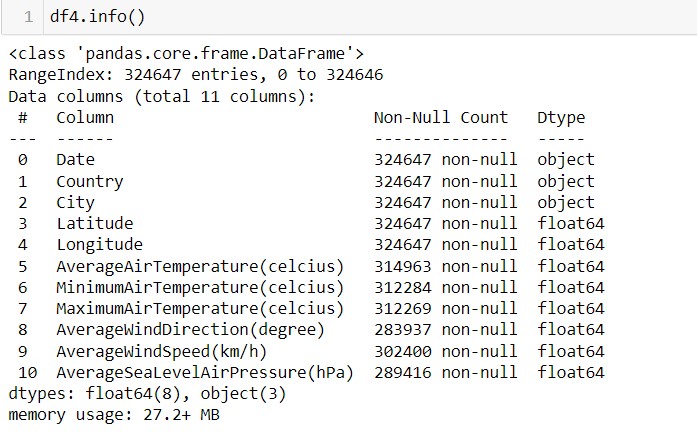
**Figure 1.3.3: Data Schema of Global Crop Yield Table.**

Additional, Global Food Consumption Table is shown in Figure 1.3.4. Based on the result, this table consists of 10 columns in total which are Country, Year, Population, Beef (kg/capita), Poultry(kg/capita), Sheep(kg/capita), Maize(1000 tones), Rice (1000 tones), soy (1000 tonnes), and lastly is Wheat (1000 tonnes). There are 3 datatypes in this data schema. These data types are string, integer and float. From the figure below, we have 8 float attributes, and 1 for each integer and string attributes.



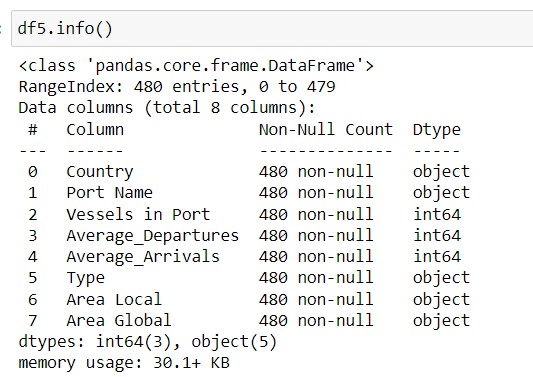
### 1.3.4: Data Schema Global Food Consumption Tables

Furthermore, the Global Weather Condition Table is another table for our project. This table contains 10 columns altogether. The names for each column are Date, Country, City, Latitude, Longitude, AverageAirTemperature(celcius), MinimumAirTemperature(celcius), MaximumAirTemperature(celcius), AverageWindDirection(celcius), AverageWindSpeed(km/h) and AverageSeaLevelAirPressure (hPa). This table consists of only 2 data types such as string and float. Based on the result from figure below, this table has 3 string attributes and 8 float attributes as shown in Figure 1.3.5.



### 1.3.5: Data Schema Global Weather Condition Tables

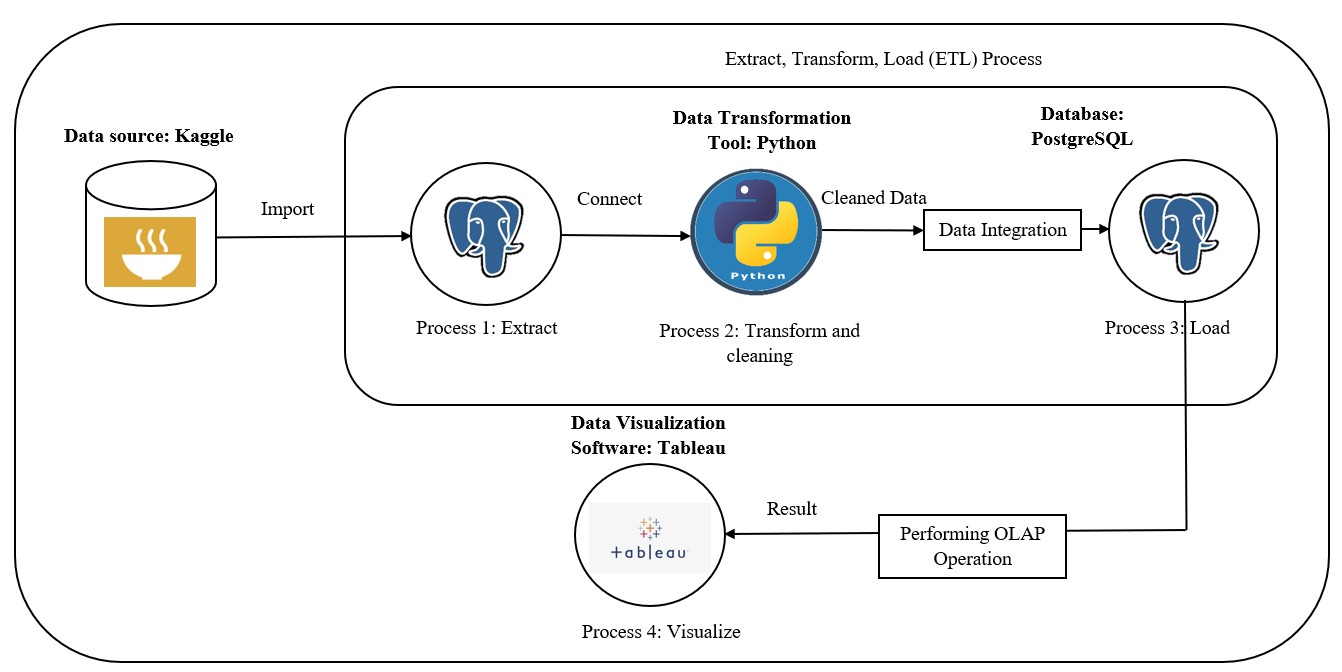
Lastly, the last table for our database is Food Supply Port Transportation in our project. This data frame contains a total of 8 columns. Among these, 3 columns have the data type 'int64', representing integer values, while the remaining 5 columns have the data type 'object', indicating string or categorical values.



**1.3.6: Data Schema Food Supply Port Transportation Tables**

**2.0 Architecture and ETL Pipeline**

#### 2.1 Pipeline structure



## Figure 2.1

Based on Figure 2.1, for data collection and preparation, utilize Kaggle as a platform to access relevant datasets related to Global Country, Global Crop Yield, Global Food Consumption, Global Weather and Food Supply Port Transportation. Use Python to download and preprocess the datasets. Python provides libraries such as Pandas and NumPy that facilitate data manipulation, cleaning, and transformation.

For database design and management, install and set up PgAdmin to design the database schema based on the ERD. Create the five tables Global Country, Global Crop Yield, Global Food Consumption, Global Weather and Food Supply Port Transportation with the appropriate fields and relationships. Use SQL statements to define the tables, primary keys, foreign keys, and constraints.

For data loading and transformation, use Python to load the preprocessed data into the PostgreSQL database using the psycopg2 library. Convert the datasets into appropriate formats for example and insert them into the respective tables using SQL queries. Perform any necessary data transformations or manipulations within the database using SQL statements. This can include joining tables, aggregating data, and creating derived columns.

For data warehouse construction, once the data is loaded and transformed, the PostgreSQL database serves as the foundation for the data warehouse. Design and implement appropriate indexes, views, and materialized views in the database to optimize data retrieval and analysis.

For data visualization and analysis, connect Tableau to the PostgreSQL database. Tableau supports various data connectors, including PostgreSQL, and can easily access and visualize data from the data warehouse. Create interactive dashboards and visualizations in Tableau to explore the data across the five tables. We can leverage Tableau's drag-and-drop interface to create insightful visualizations, such as charts, graphs, and maps, to gain actionable insights and communicate them effectively.

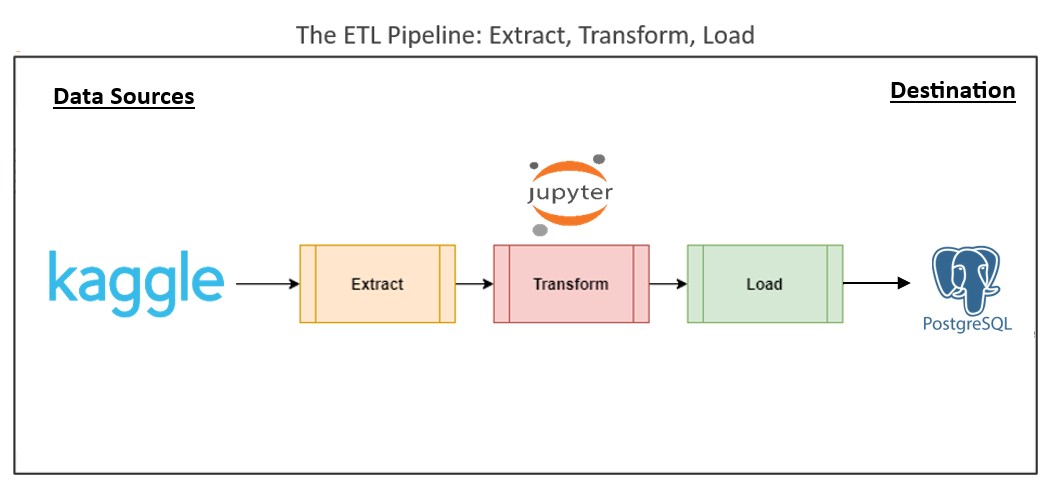
### 2.2 Data Warehouse Architecture

The Architecture of our project is close to Kimball’s apporach. Our group is starting the project by understanding and documenting the most critical problem from one of the 17 Sustainable Development Goals (SDGs), we extract the idea about the project needs, and project questions being asked. Zero Hunger is chosen as our main focus and the idea is about defeating hunger by developing a system to ensure a constantly sufficient food supply worldwide.

Then we document all data sources available from the online resource like kaggle throughout the project. After gathering all the required data ,we build ETL pipelines that extract from kaggle, transform by jupyter through python, and load data from the data sources into a denormalized data model in postgresql. The dimensional model is built as a star schema with central fact tables surrounded by dimension tables as shown in figure 3.1 of our database part.Lastly ,we proceed by enabling user-friendly data visualization and OLAP operation analysis.

All these aspects align with the Kimball approach, which focuses on delivering user-centric data warehouse solutions with bottom-up procedure. Kimball’s approach focuses on a bottom-up procedure.

### 2.3 ETL Pipeline



## Figure 2.3

The dataset is downloaded from Kaggle. The dataset is composed of five tables which are Global Country, Global Crop Yield, Global Food Consumption, Global Weather and Food Supply Port Transportation. The table has been extracted and the process data cleaning and data integrating using Jupyter Notebook. After the cleaned data has been provided, the data will transform to a new csv file before loaded to PostgreSQL database to create visualization using Tableau.

**2.3.1 Step to Extract and Transform (Python)**

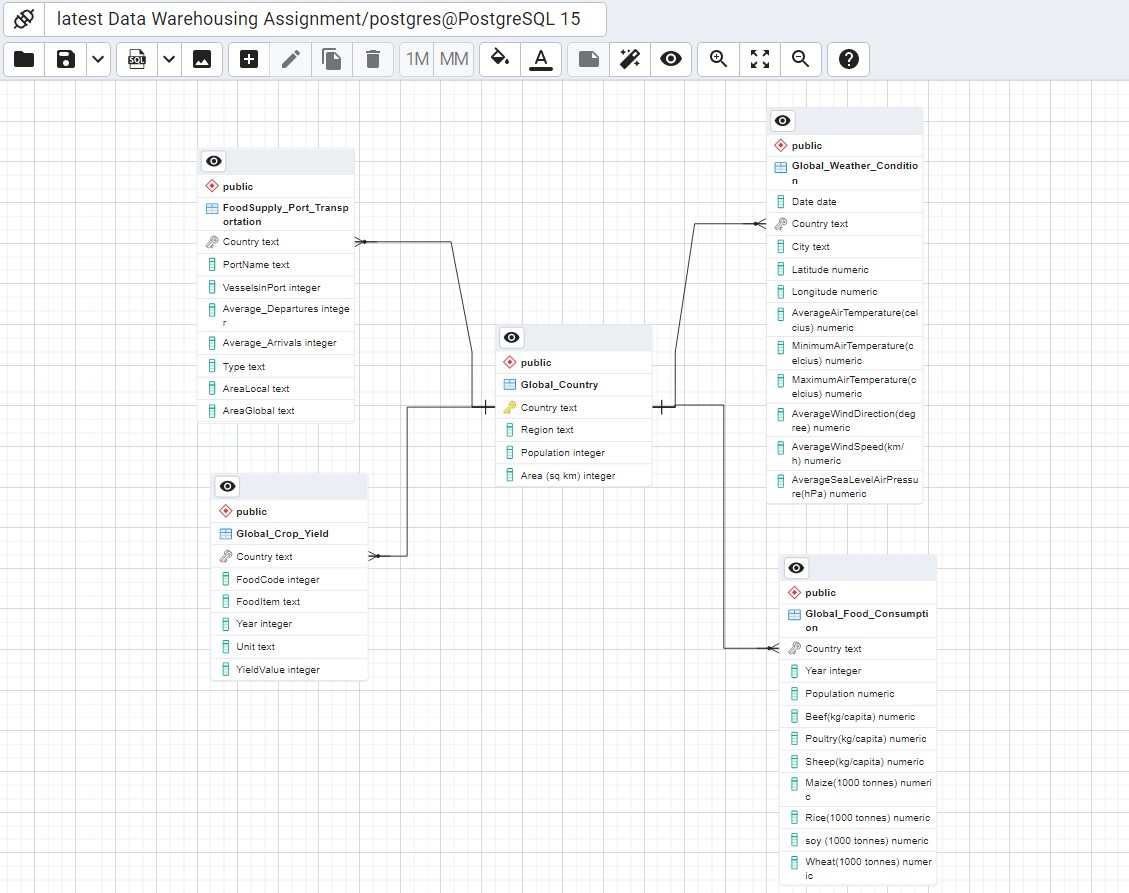
Refer To Appendices

**2.3.2 Step to Load Data into Database (PostgreSQL)** Refer To Appendices

**2.3.3 Step to Load OLAP Data into Visualization Tool (Tableau)**  Refer To Appendices

### 3.0 Database

**Figure 3.1 : Data schema from PostgreSQL**



The figure 3.1 showed the relational model of our database by loading the 5 tables including Global\_Country, Global\_Crop\_Yield, Global\_Food\_Consumption, Global\_Weather\_Condition and FoodSupply\_Port\_Transportation.The Entity Relationship Diagram(ERD) is clearly showing that our data warehouse schema is a star schema .Star schema is the fundamental schema among the data mart schema and it is simplest.

The data model is said to be star schema as its physical model resembles to the star shape having Global fact table at its center and the dimension tables like Global\_Crop\_Yield, Global\_Food\_Consumption, Global\_Weather\_Condition and FoodSupply\_Port\_Transportation at its peripheral representing the star’s points. Based on Figure 3.1, the fact table in a star schema has one to many relationships with dimension tables. The relationship between the Global\_Country(fact table) and Global\_Crop\_Yield(dimension table) is one-to- many relationship with foreign key country in the Global\_Crop\_Yield table.Same goes to other relationship between fact anddimension tables are all one-to-many relationship with the country as foreign key to the Global\_Country table.

Overview of database relationship:

**Fact Table:**

1. Global\_Country: Country (Primary Key), Region, Population, Area (sq km)

**Dimension Tables:**

1. FoodSupply\_Port\_Transportation: Country(Foreign Key), PortName, VesselsinPort, Average\_Depatures, Average\_Arrivals, Type, AreaLocal, AreaGlobal
2. Global\_Crop\_Yield: Country(Foreign Key), FoodCode, FoodItem, Year, Unit, YieldValue
3. Global\_Food\_Consumption: Country(Foreign Key), Year, Population, Beef(kg/capita), Poultry(kg/capita), Sheep(kg/capita), Maize(1000tonnes), Rice(1000tonnes), soy(1000tonnes), Wheat(1000tonnes)
4. Global\_Weather\_Condition: Country(Foreign Key), Date, Latitude, Longitude, AverageAirTemperature(celsius), MinimumAirTemperature(celsius),

MaximumAirTemperature(celsius), AverageWIndDirection(degree), AverageWindSpeed (kn/h) ,AverageSeaLevelAirPressure(km/h)

**4.0 Result and Data Analysis**

### 4.1 OLAP Operation

#### 4.1.1 Slicing

SELECT "Global\_Country"."Country",

"Global\_Crop\_Yield"."FoodItem",ROUND(AVG("Global\_Crop\_Yield"."YieldValue"),2

) as "AverageYieldValue"

FROM "public"."Global\_Country"

INNER JOIN "public"."Global\_Crop\_Yield"

ON public."Global\_Country"."Country" =public."Global\_Crop\_Yield"."Country"

GROUP BY "Global\_Country"."Country", "Global\_Crop\_Yield"."FoodItem"

HAVING ROUND(AVG("Global\_Crop\_Yield"."YieldValue"),2) >10000

ORDER BY "Global\_Country"."Country", "Global\_Crop\_Yield"."FoodItem";



### Figure 4.1: Output Slicing

The output of slicing of OLAP operation in shown in Figure 5a. This operation filters the unnecessary portions.Two tables involved in this operation that are Global\_Country and Global\_Crop\_Yield. We conduct slicing operation by selecting some attributes through inner join of two tables and conditions using <HAVING>,<GROUP

BY> and<ORDER BY> and aggregation on some attribute by

AVG("Global\_Crop\_Yield"."YieldValue") with total 241 rows of result table. Through Slicing operation, we can filter the AverageYieldValue of different FoodItem from all the countries.For example, we can observe the Algeria country has the crop of Maize, Potatoes and Rice,paddy with their respective AverageYieldValue.

#### 4.1.2 Pivot

SELECT \* FROM crosstab(

'SELECT "Country", "Year", "Rice(1000 tonnes)"

FROM public."Global\_Food\_Consumption"

ORDER BY 1,2'

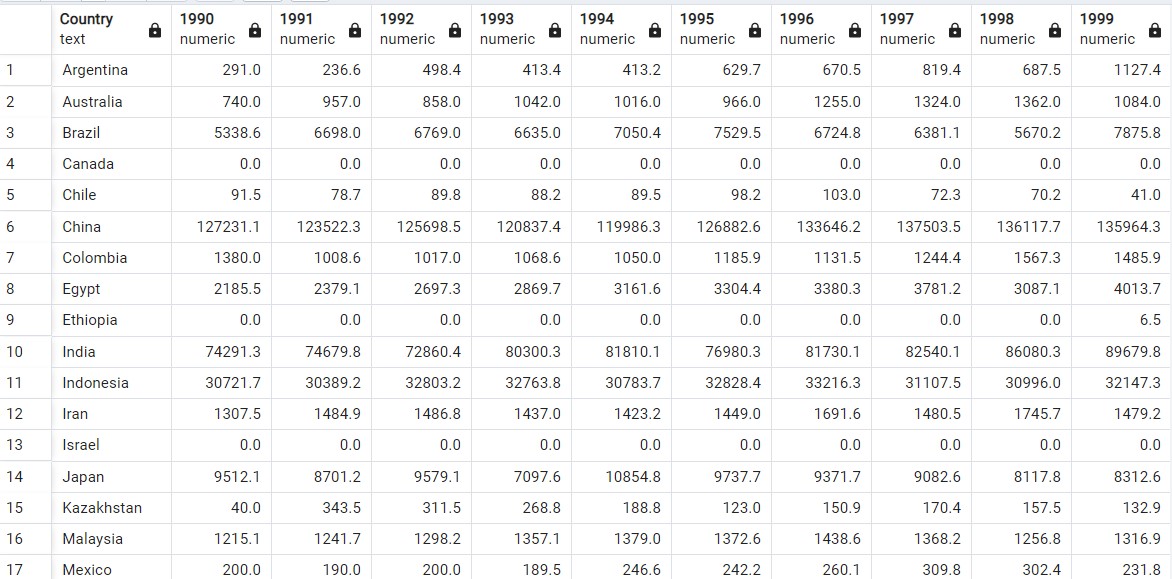
)AS Country("Country" text, "1990" numeric,"1991" numeric,"1992" numeric, "1993" numeric,

"1994" numeric,"1995" numeric,"1996" numeric,"1997" numeric,"1998" numeric,"1999" numeric,

"2000" numeric,"2001" numeric,"2002" numeric,"2003" numeric,"2004" numeric,"2005" numeric,

"2006" numeric,"2007" numeric,"2008" numeric,"2009" numeric,"2010" numeric,"2011" numeric,

"2013" numeric,"2014" numeric,"2015" numeric,"2016" numeric,"2017" numeric,"2018" numeric);



### Figure 4.2: Output Pivot

Based on Figure 4b, the output pivot indicates the purpose is to pivot the information from the "Global\_Food\_Consumption" table, specifically focusing on rice consumption data with a unit of 1000 tonnes. The resulting output will present the countries as rows and each year's rice consumption as separate columns with a total 33 rows after pivoting. By organizing the data in this manner, it becomes easier to compare and analyze the rice consumption trends across different countries and years. The query utilizes the crosstab function in PostgreSQL, which is a useful tool for transforming data from a rowbased structure into a column-based format

#### 4.1.3 Roll Up

SELECT "Country", "Type", sum("VesselsinPort")

FROM public."FoodSupply\_Port\_Transportation"

GROUP BY ROLLUP (1, 2)

ORDER BY 1, 2;

SELECT CASE WHEN "Country" IS NULL

THEN 'TOTAL' ELSE "Country" END,

CASE WHEN "Type" IS NULL

THEN 'TOTAL' ELSE "Type" END,

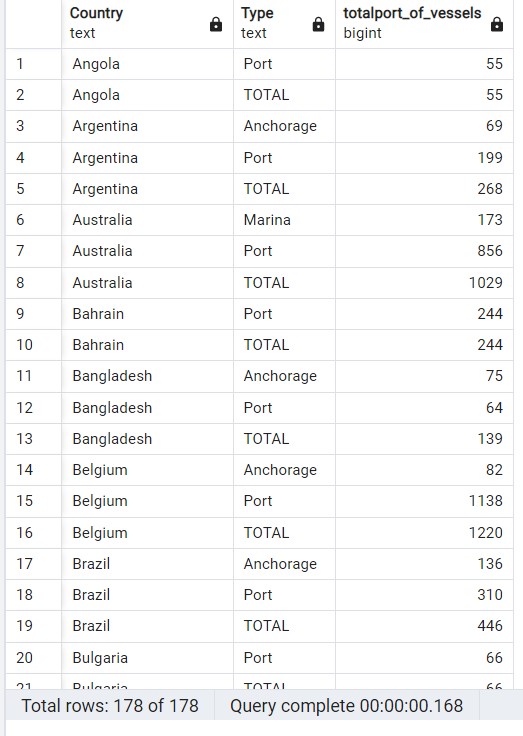
TotalPort\_of\_Vessels

FROM(SELECT "Country", "Type", sum("VesselsinPort") as TotalPort\_of\_Vessels

FROM public."FoodSupply\_Port\_Transportation"

GROUP BY ROLLUP (1, 2)

ORDER BY 1, 2) AS x;



### Figure 4.3: Output Roll Up

The output for roll up operation is shown in Figure 4c. This operation is zooming out on the data cube . Roll -up is used to provide abstract level details to user . It performs further aggregation of data by reduction in dimension or by stepping up a concept hierarchy for adimension.One table (FoodSupply\_Port\_Transportation) only is involved in this operation. Abstract level like the overview of TotalPort\_of\_Vessels in port and anchorage of each country is displayed in table by using sum("VesselsinPort") with total 178 rows in result table.

#### 4.1.4 Dicing

SELECT "Global\_Country"."Country", "Global\_Country"."Population",

ROUND(AVG("Global\_Weather\_Condition"."AverageWindSpeed(km/h)"),2) AS

Average\_Wind\_Speed

FROM public."Global\_Country"

JOIN public."Global\_Weather\_Condition" ON public."Global\_Country"."Country" = public."Global\_Weather\_Condition"."Country"

GROUP BY "Global\_Country"."Country", "Global\_Country"."Population"

HAVING ROUND(AVG("Global\_Weather\_Condition"."AverageWindSpeed(km/h)"),2)

> 5 OR "Global\_Country"."Population">10000

ORDER BY "Global\_Country"."Country" ASC;



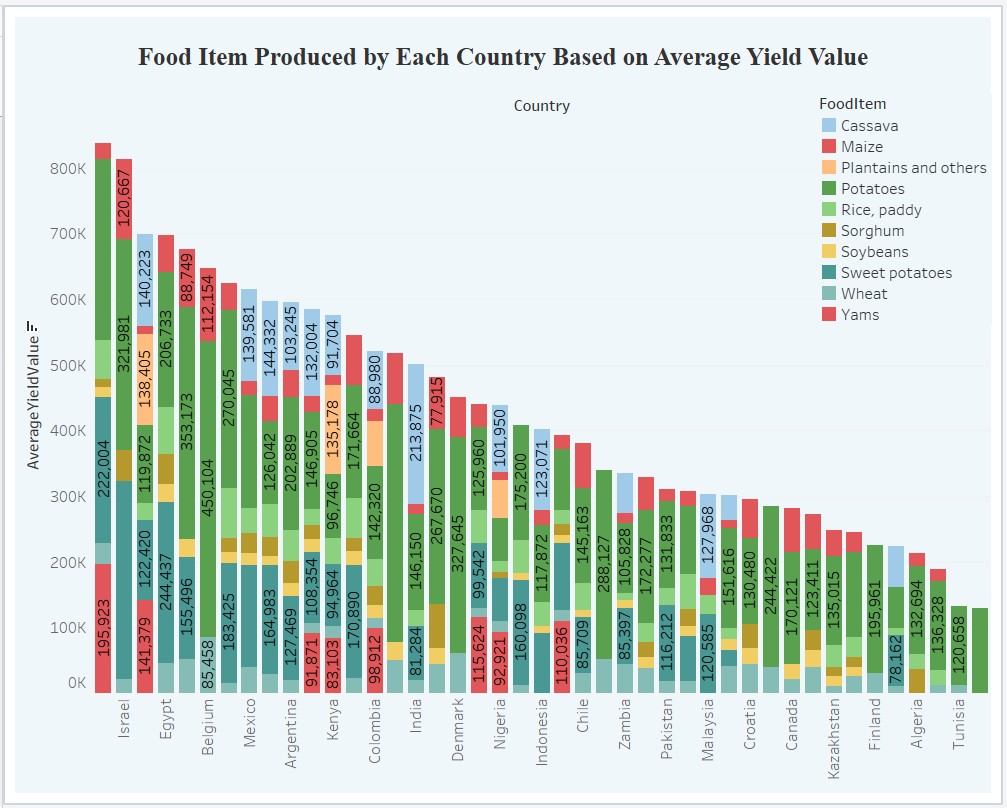
Figure 4.4: Output Dicing

The output of dicing of OLAP operation is shown in Figure 4d. This operation emphasizes two or more dimensions from a cube given and suggests a new sub-cube.Two tables involved in this operation are Global\_Country and Global\_Weather\_Condition. We conduct dicing operation by selecting some attributes through inner join of two tables and conditions using <HAVING>,<GROUP BY> and<ORDER BY> and aggregation on some attribute by AVG("Global\_Weather\_Condition"."AverageWindSpeed(km/h)") with total 31 rows in result table.The main difference compared to slicing operation is that our condition apply logical operator like <AND> and <OR> to filter more detail information from different dimensions.Through dicing operation, we can filter the

Average\_Wind\_Speed of different countries that is greater than 5 and population of all countries that is greater than 10000 with requirement to our objectives.

#### 4.2 Visualization

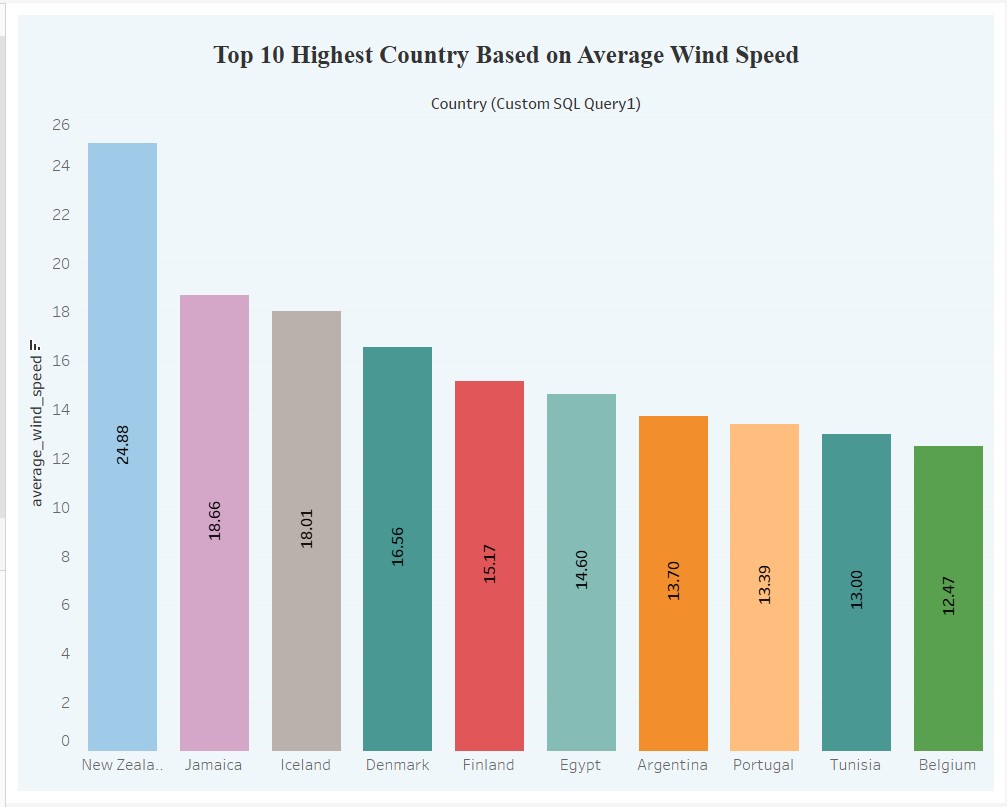
##### 4.2.1 Food Item Produced by Each Country Based on Average Yield Value



## Figure 4.2.1: Food Item Produced by Each Country Based on Average Yield Value

The objective of improving food supply chain operations based on consumption patterns can be achieved by examining the rankings of countries in terms of average yield value for food items and the specific production quantities of different food items. Japan, with the highest average yield value above 800k, and Israel, ranking second around 800k, exemplify efficient food supply chain operations. Jamaica follows closely with an average yield value of 700k. Analyzing their strategies can inform improvements in other regions. Additionally, Belgium leads in potato production, generating 450,104, while Iceland focuses solely on potatoes but has the lowest average yield value. By studying these patterns and relationships, targeted measures can be implemented to optimize food supply chains and enhance overall efficiency.

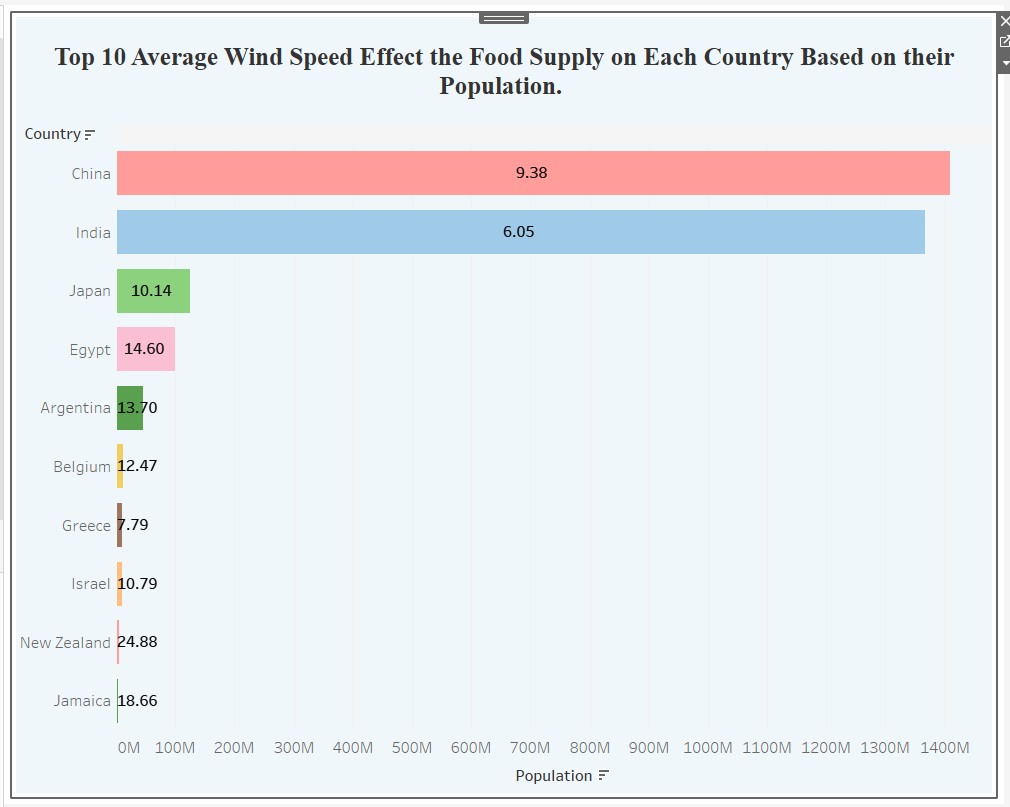
### 4.2.2 Top 10 Highest Country Based on Average Wind Speed



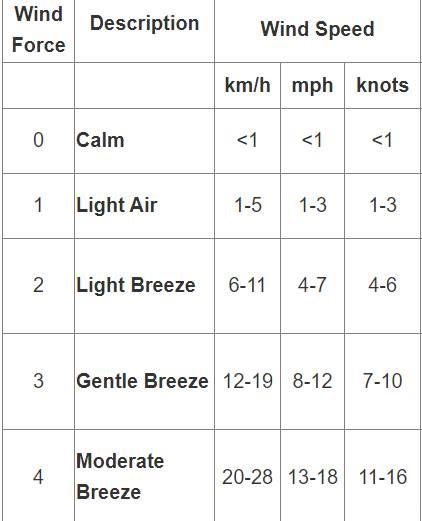
## Figure 4.2.2: Top 10 Highest Country Based on Average Wind Speed

Based on Figure 4.2.2, vertical bar shows the visualization of top 10 highest country based on average wind speed. New Zealand has the highest average wind speed which is 24.88 km/h may experience more significant challenges in transporting crops due to potentially stronger winds. This information can help policymakers, transportation logistics professionals, and stakeholders in the agricultural industry to understand the specific risks and plan appropriate measures for crop transportation in windy conditions. Followed by Jamaica at 18.66 km/h, Iceland at 18.01 km/h, Denmark at 16.56 km/h, Finland at 15.17 km/h, Egypt at 14.60 km/h, Argentina at 13.70 km/h, Portugal at 13.39, Tunisia at 13.00 km/h and lastly in Belgium at 12.47 km/h may also face transportation challenges due to wind-related factors. By considering weather data and average wind speeds in these regions, stakeholders can take proactive steps to mitigate risks, such as reinforcing transportation infrastructure, optimizing logistics routes, or adjusting transportation schedules to minimize the impact of adverse weather conditions.

**4.2.3 Top 10 Average Wind Speed Effect The Food Supply on Each Country Based On Their Population.**



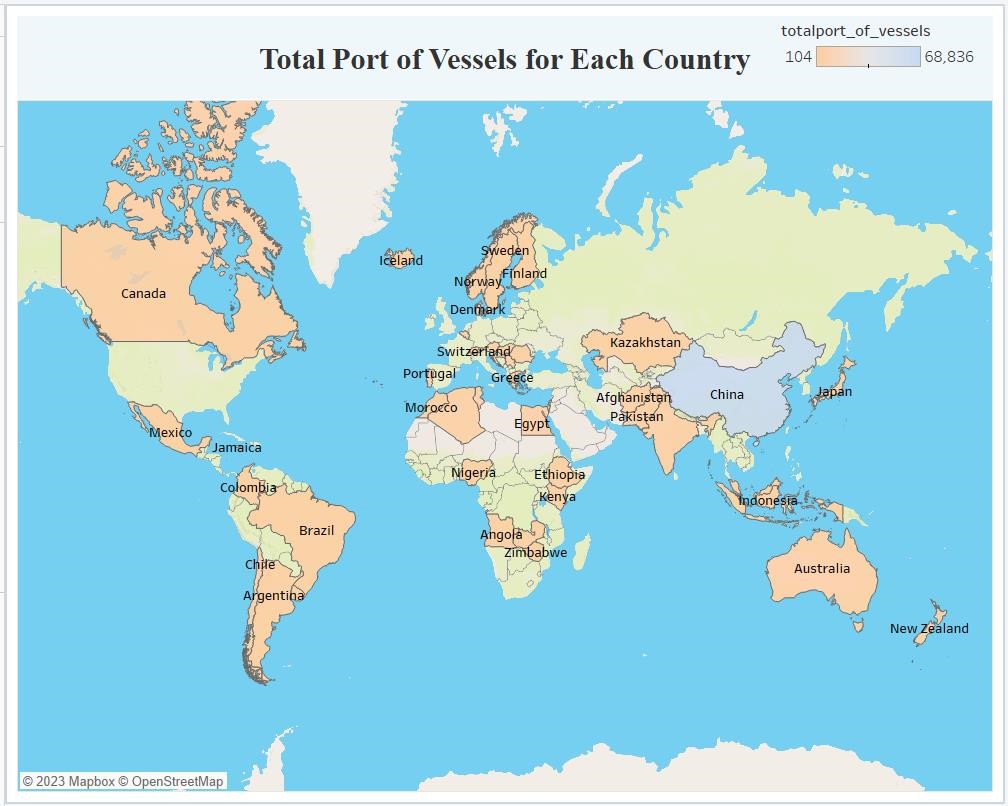
**Figure 4.2.3: Top 10 Average Wind Speed Effect The Food Supply on Each Country Based On Their Population.**



**Table 4.2.3: Safe wind Force Index.**

The objective of this visualization is to utilize weather data, specifically average wind speeds, to ensure a reliable food supply. China has the highest population, exceeding 1.4 billion, followed by India with 1.3 billion people, and Japan with over 100 million. As population size increases, so does the demand for food, necessitating a larger food supply. However, the transportation of food items by sea is influenced by average wind speeds, which can affect the quality and condition of the goods. Fortunately, all three countries are strategically located in regions with safe average wind speeds for exporting and importing food items. These safe average wind speeds can be verified by referring to Table 4.2.3. By considering this vital weather data, stakeholders can make well-informed decisions to ensure a consistent and secure food supply for these densely populated nations.

### 4.2.4 Total Port of Vessels for Each Country



## Figure 4.2.4: Total Port of Vessels for Each Country

Based on the figure 4.2.3. China has the highest total port of vessels for each country which is 68836. Followed by Belgium, Indonesia, and Australia at 2440, 2084 and 2058 respectively. The lowest total port of vessels goes to Chile with only 142 number of ports. This represents that China is the highest total port of vessels since China has the highest population which is 1,409,517,397 billion as shown in the figure 4.2.3 above. The significant presence of numerous vessels in Chinese ports highlights the necessity for efficient food distribution systems to meet the consumption patterns and requirements of its large population. By prioritizing enhancements in food supply chain operations aligned with consumption patterns and optimizing the efficiency of food distribution systems, we can ensure a dependable and efficient flow of food, effectively meeting the demands of China's population while also contributing to global food security.

### 5.0 Conclusion

In conclusion, the 5 different table from different data sources were loaded into a data warehouse and lastly loaded into a database. The ETL process is carried out so that cleaned datasets are obtained. The transforming process was done by using Python to clean the data then loading into postgresql to perform OLAP operation. Data visualization is practiced according to the result of OLAP operation.

Summary of data visualization result of OLAP operation is showing that analyzing consumption patterns, yield values, production quantities, and weather data provides valuable insights to enhance food supply chain operations. Countries like Japan, Israel, and Jamaica exhibit efficient supply chains, while Belgium excels in potato production. Considering average wind speeds helps identify transportation challenges, and proactive measures can be taken. China's large population necessitates efficient distribution systems for food security. By leveraging this information, stakeholders can optimize supply chains and meet the demands of densely populated regions.

Each group member encountered various challenges during the project. One major difficulty was finding suitable datasets that covered all the required questions and aligned with our system. We found it difficult to find the ideal datasets to cover those question especially when we tried to be looking for a dataset that consist of all the table that match our system. Another challenge was sharing the transformed data within the group's database. Each member had to update the database regularly, which consumed time and required careful attention to avoid any errors and catch up the progress. Furthermore, relating the project objectives to OLAP operations posed another challenge. We needed a comprehensive understanding of the database to determine if OLAP operations would provide efficient outputs to solve our project's problem. Lastly, we faced difficulties in establishing a connection between the database and visualization tools. We had to carefully evaluate whether PowerBI or Tableau would be the most suitable tool for visualization, considering that different group members were more proficient in different tools.

### 6.0 References

*Goal 2: Zero Hunger - The Global Goals*. (n.d.). The Global Goals. <https://globalgoals.org/goals/2-zero-hunger/>

*How to Connect To PostgreSQL Database Server Using Python*. (n.d.). How to Connect to PostgreSQL Database Server Using Python - CommandPrompt Inc. https:///education/how-toconnect-to-postgresql-database-server-using-python/

*Pandas Merging 101*. (2018, December 6). Stack Overflow.

<https://stackoverflow.com/questions/53645882/pandas-merging-101>

*Star Schema in Data Warehouse modeling - GeeksforGeeks*. (2018, October 12). GeeksforGeeks.

<https://www.geeksforgeeks.org/star-schema-in-data-warehouse-modeling/>

*PostgreSQL Connector - Tableau*. (n.d.). PostgreSQL Connector - Tableau - Tableau. <https://help.tableau.com/current/pro/desktop/en-us/examples_postgresql.htm>

*Tableau and PostgreSQL: How To Make Them Work Together*. (n.d.). Tableau and PostgreSQL: How to Make Them Work Together. [https://blog.panoply.io/tableau-and-postgresql-two-ways-ofworking-together](https://blog.panoply.io/tableau-and-postgresql-two-ways-of-working-together)

*Pandas DataFrames - Reading from and writing to PostgreSQL table | Pythontic.com*. (n.d.). Pandas DataFrames - Reading From and Writing to PostgreSQL Table | Pythontic.com. <https://pythontic.com/pandas/serialization/postgresql>

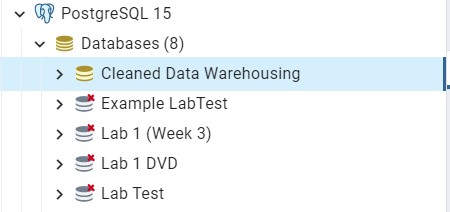
*Sustainable Development Goal 2 - Wikipedia*. (2022, June 15). Sustainable Development Goal 2 -

Wikipedia. <https://en.wikipedia.org/wiki/Sustainable_Development_Goal_2>

# 7.0 APPENDICES

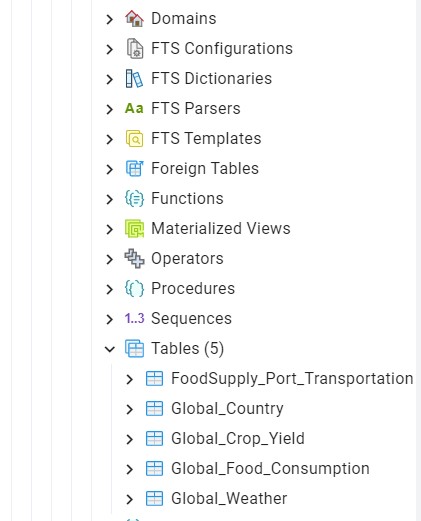
## Steps to load data into database (PostgreSQL)

1. Created database ‘Clean Data Warehousing’ PostgreSQL



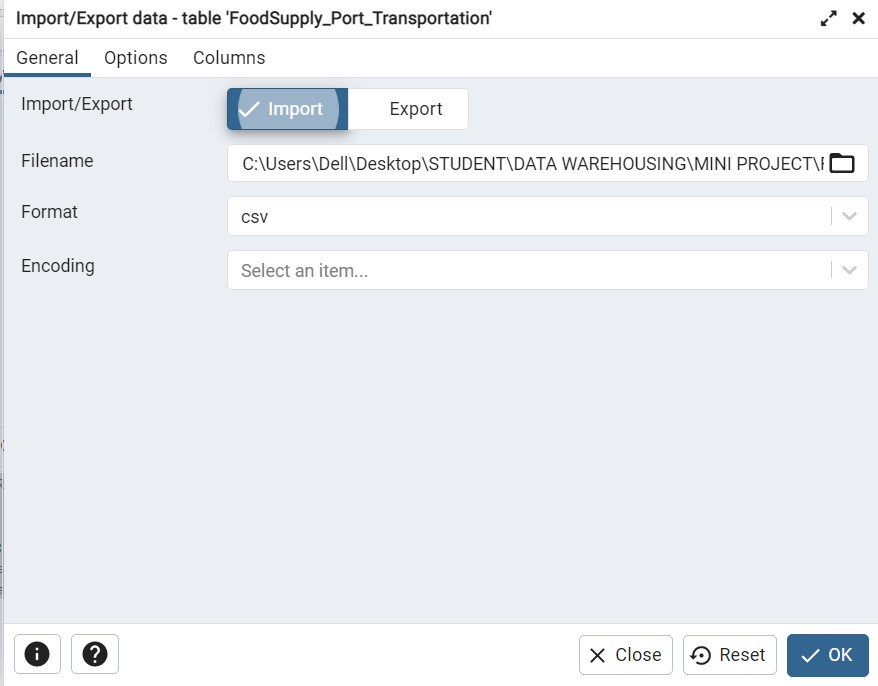
1. Created all of the five table ‘Global\_Country’, ‘Global\_Crop\_Yield’,

‘Global\_Food\_Comsumption’, ‘Global\_Weather’ and ‘FoodSupply\_port\_Transportation’

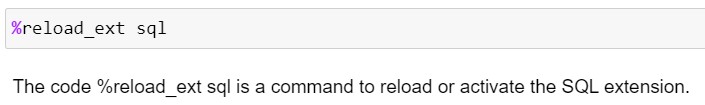
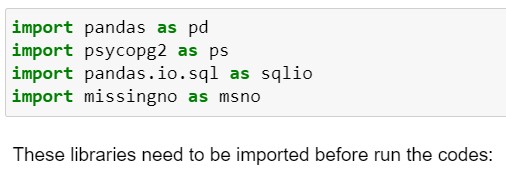
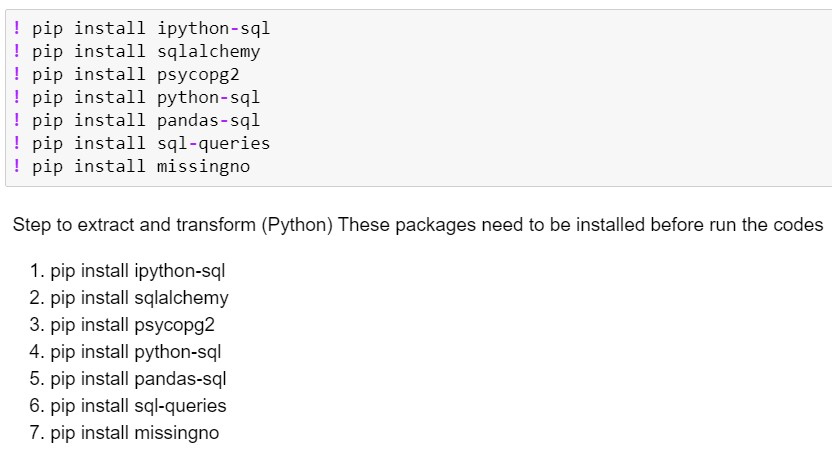


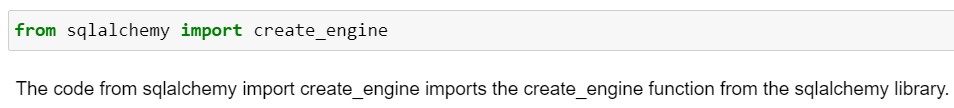
1. Import dataset ‘Global\_Country’, ‘Global\_Crop\_Yield’, ‘Global\_Food\_Comsumption’, ‘Global\_Weather’ and ‘FoodSupply\_port\_Transportation’ into table ‘Global\_Country’, ‘Global\_Crop\_Yield’, ‘Global\_Food\_Comsumption’, ‘Global\_Weather’ and

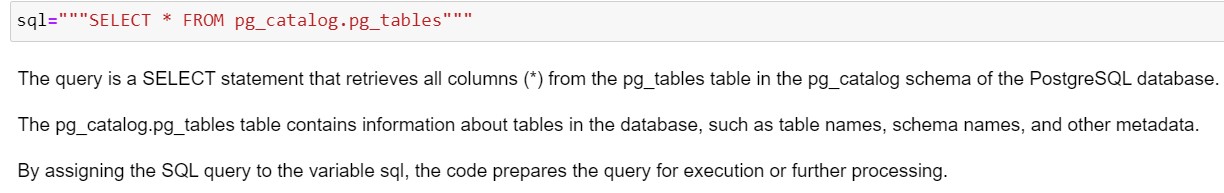
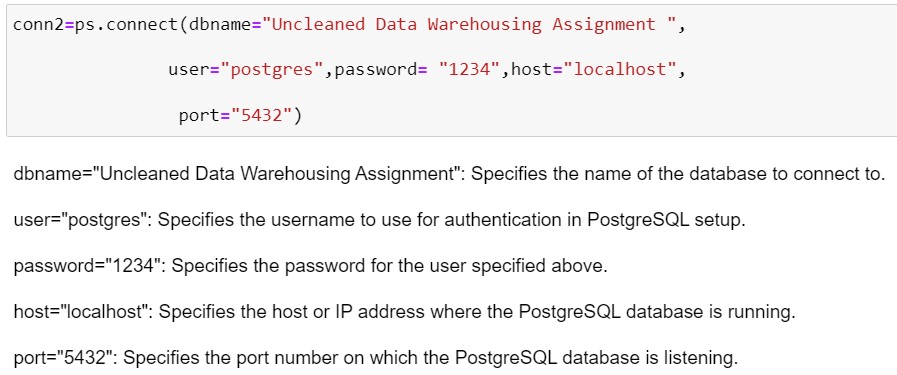
‘FoodSupply\_port\_Transportation’



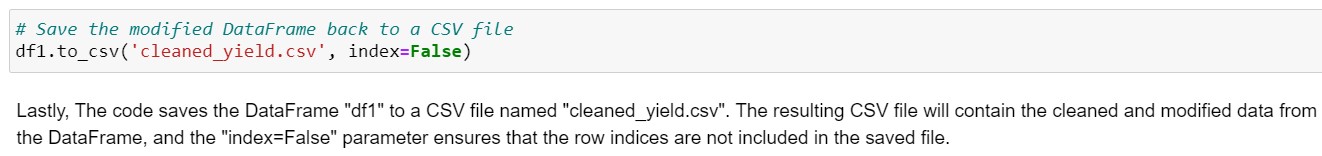
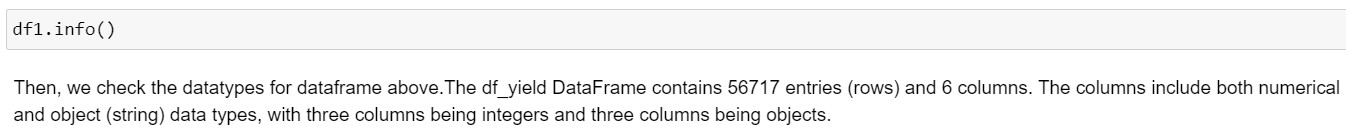
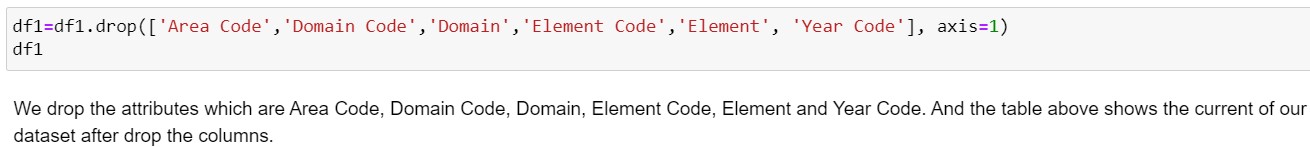
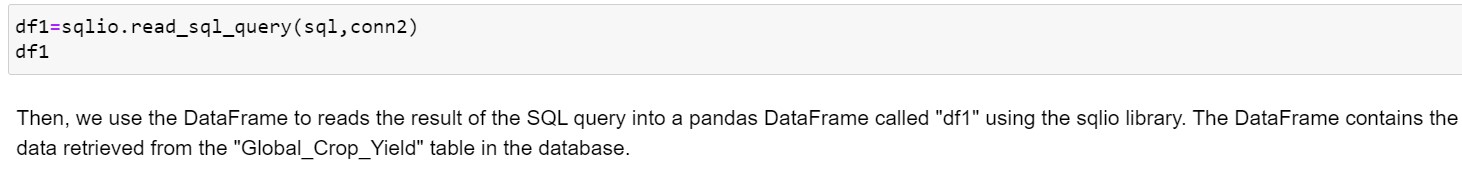
## Step to extract and transform



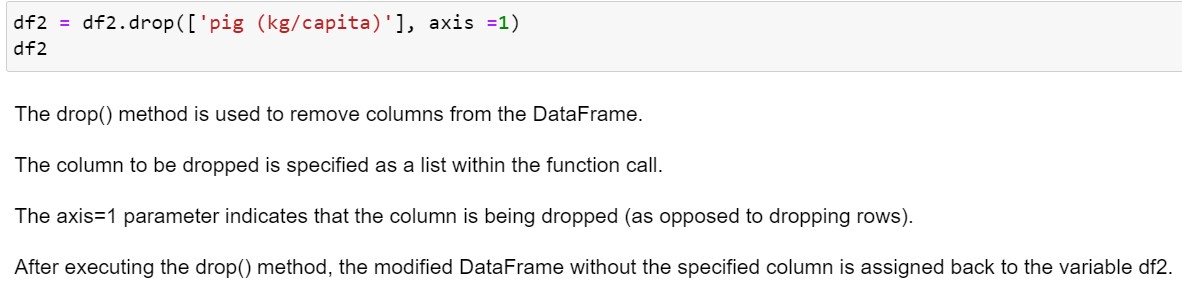
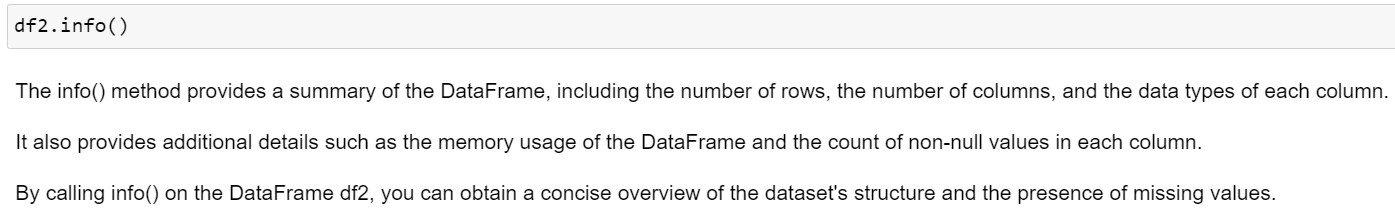
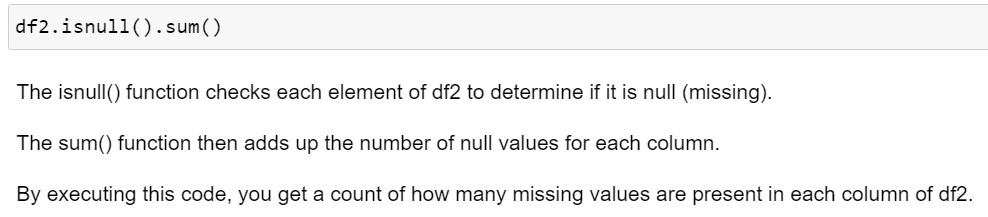
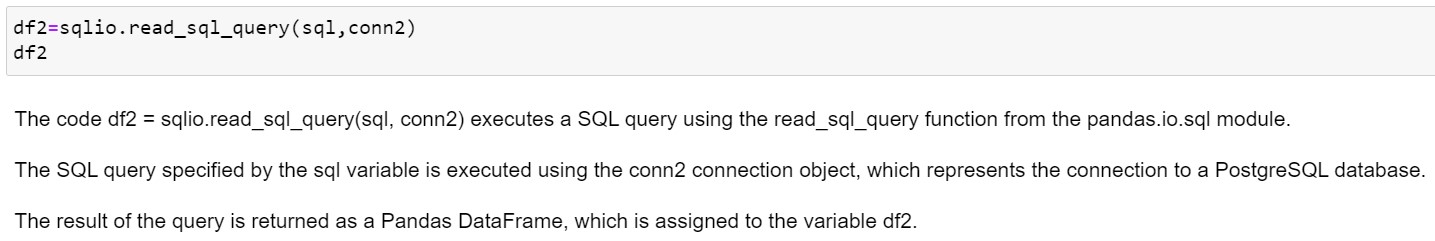
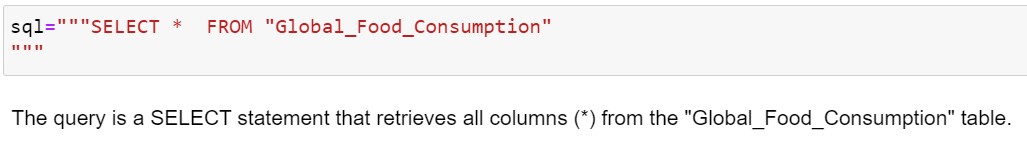


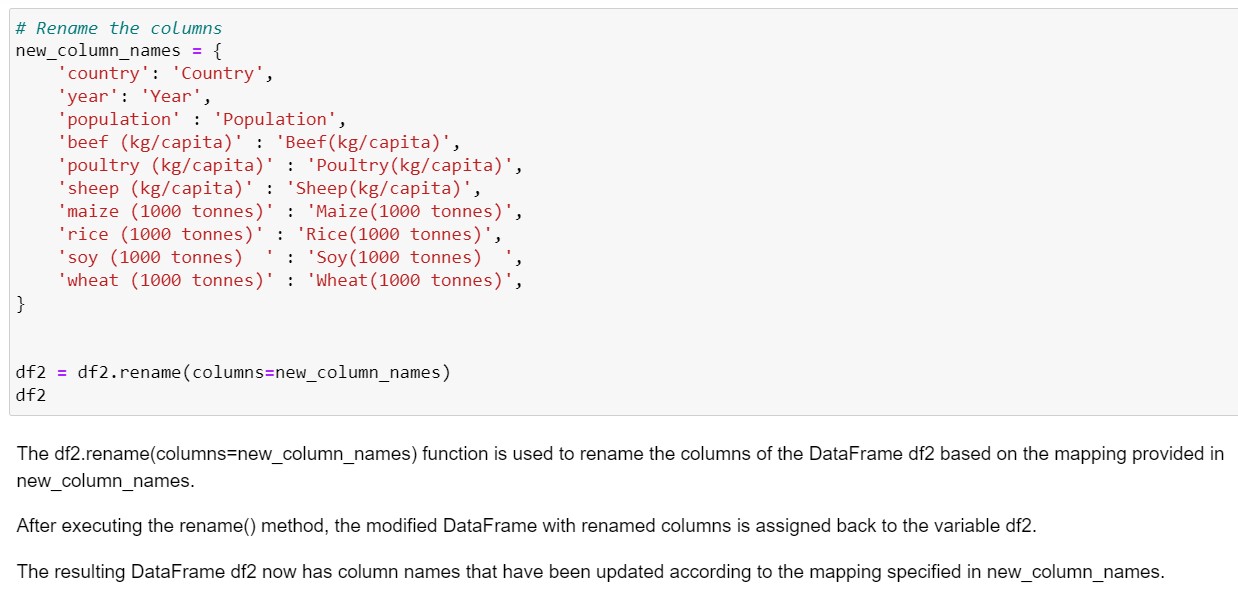


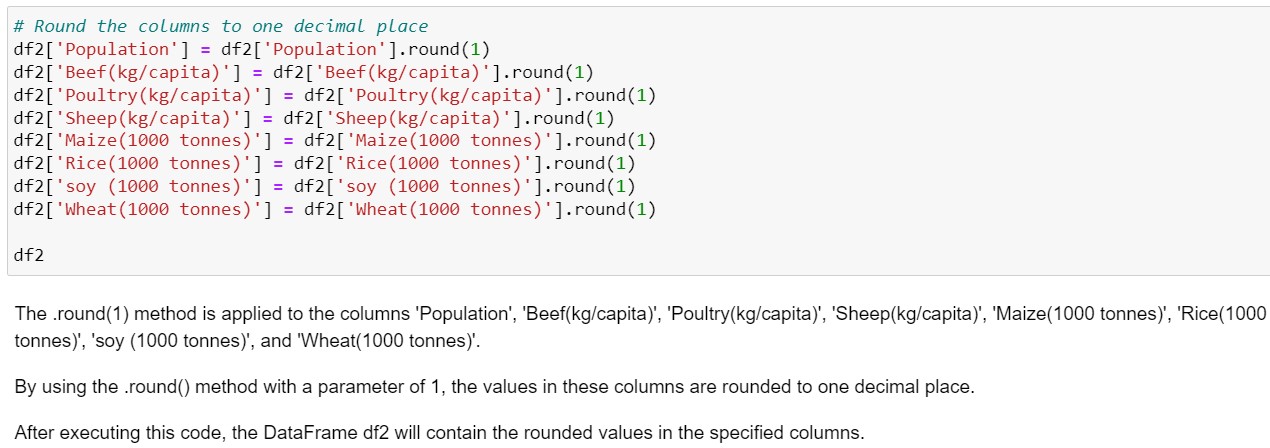
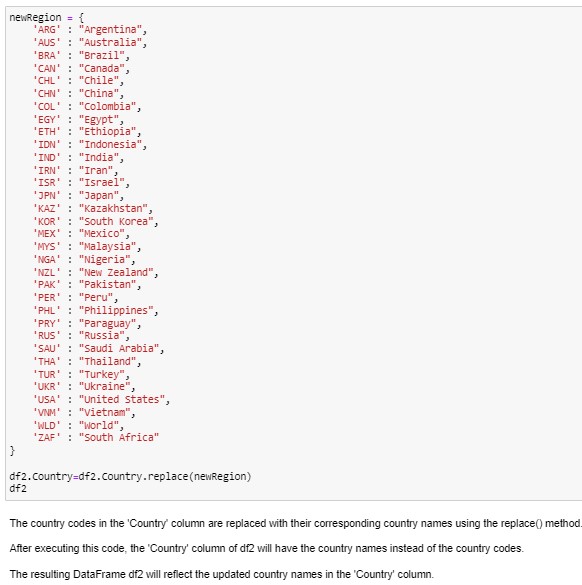
### Global\_Crop\_Yield table cleaning

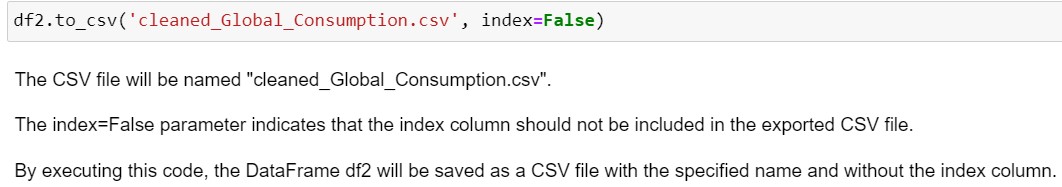


### Global\_Food Consumption table cleaning

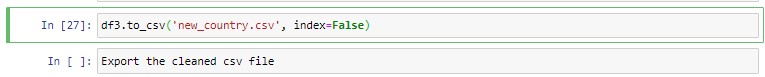
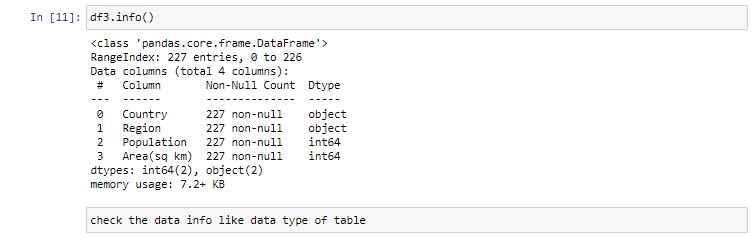
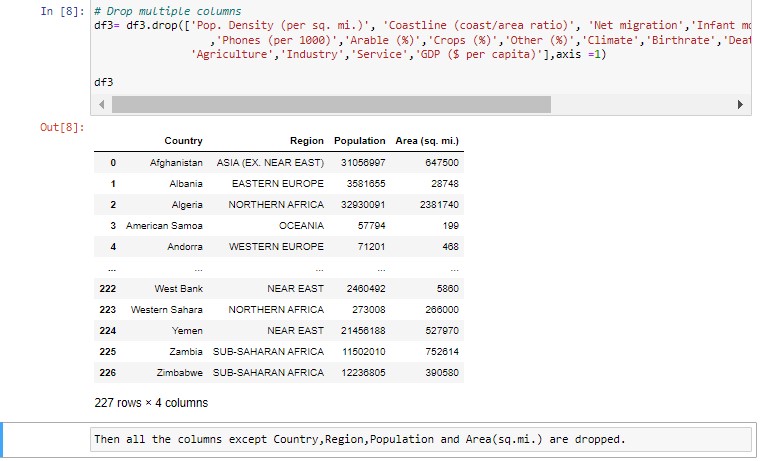
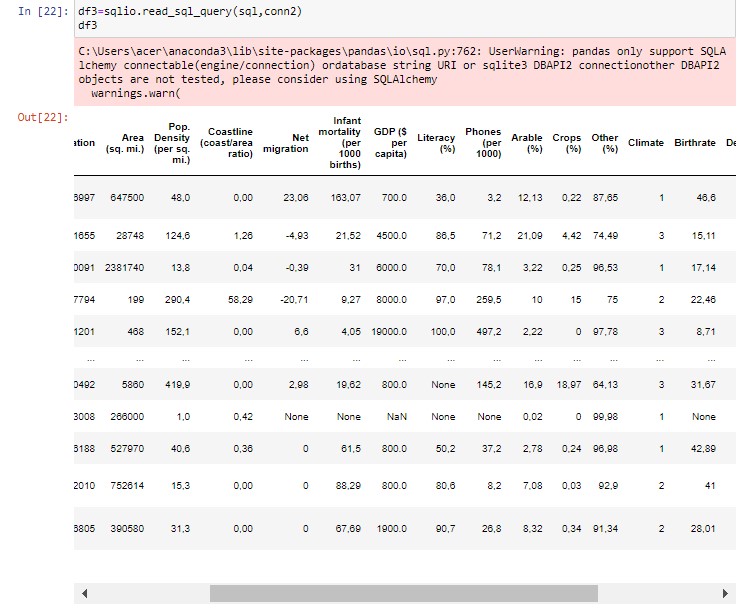
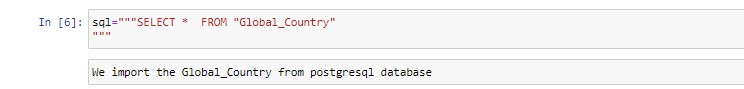




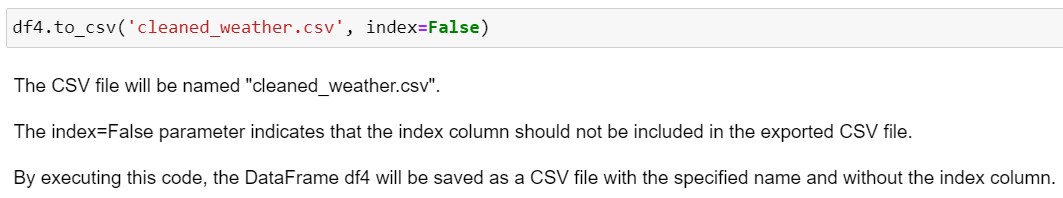
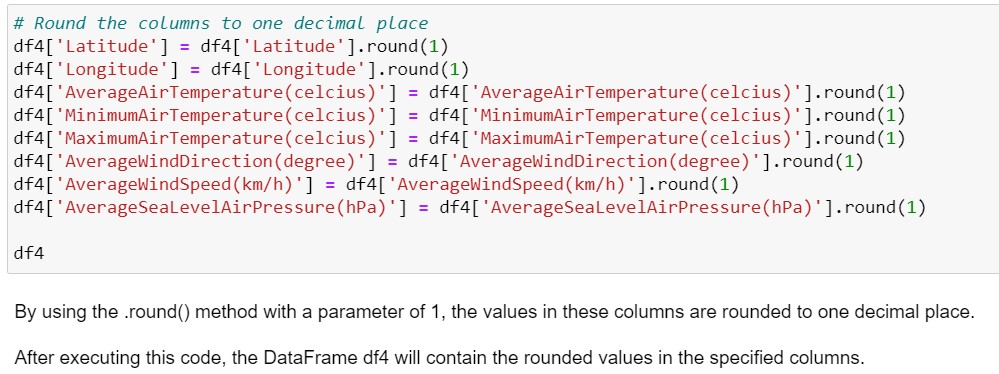
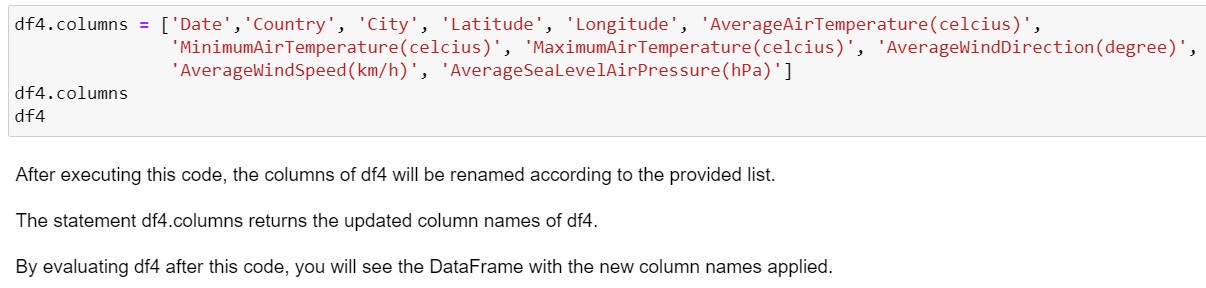
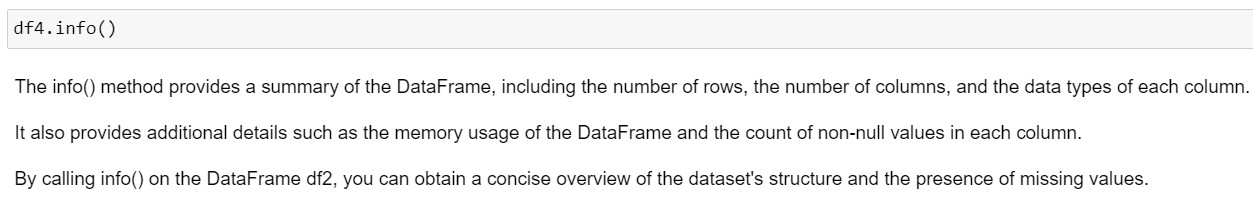
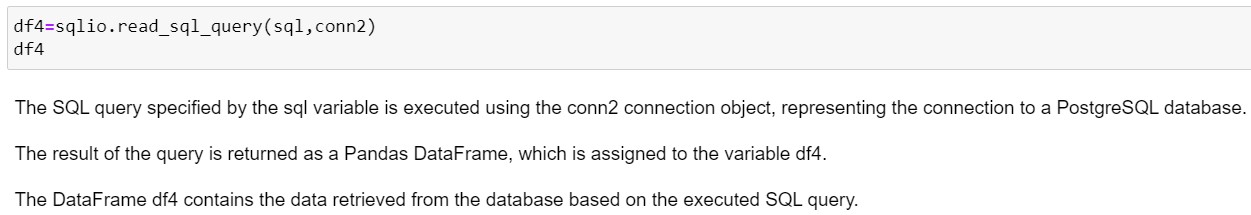
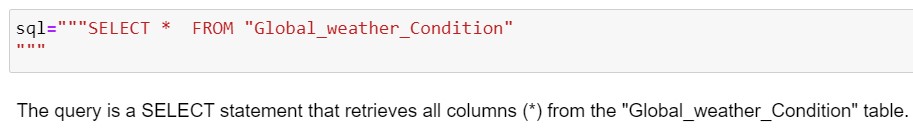




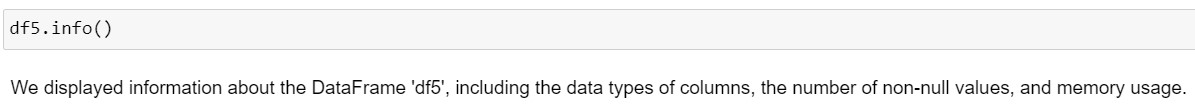
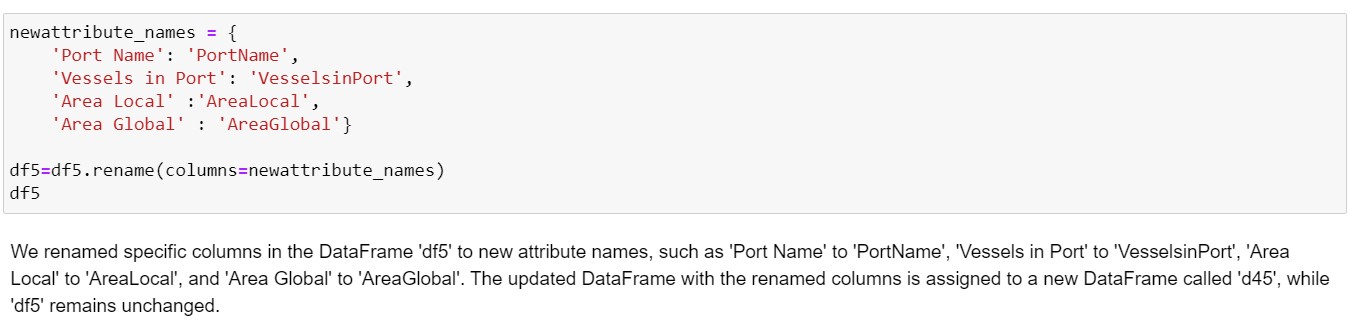
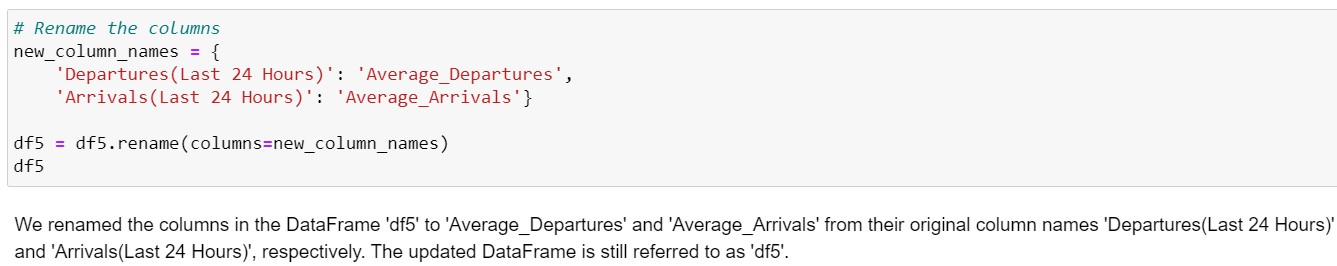
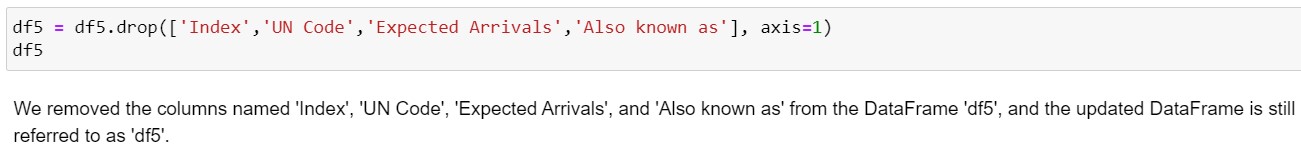
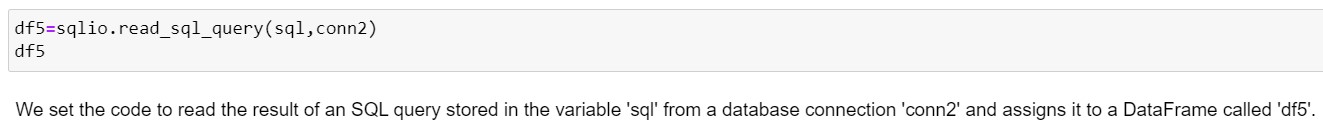
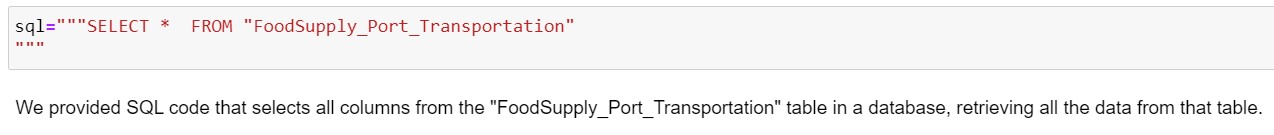
### Global\_COuntry table cleaning



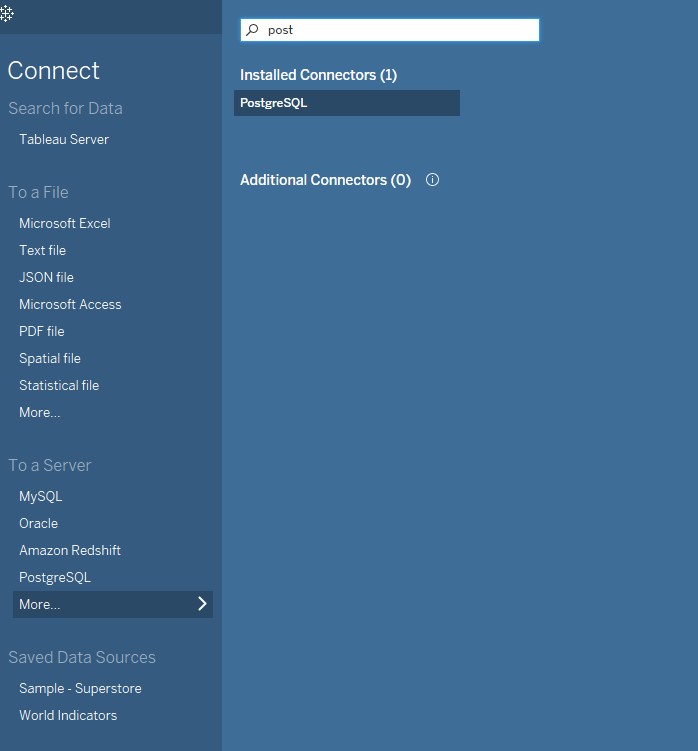
### Global\_Weather\_Condition table cleaning



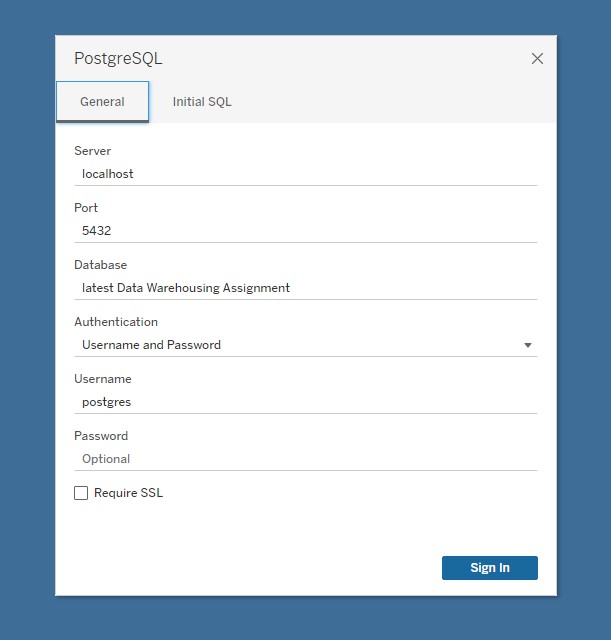
### FoodSupply\_Port\_Transportation table cleaning



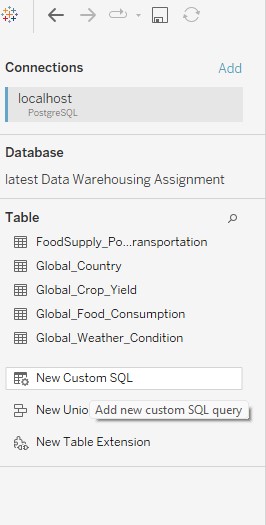
## Step to Load OLAP Data into Visualisation Tool (Tableau)



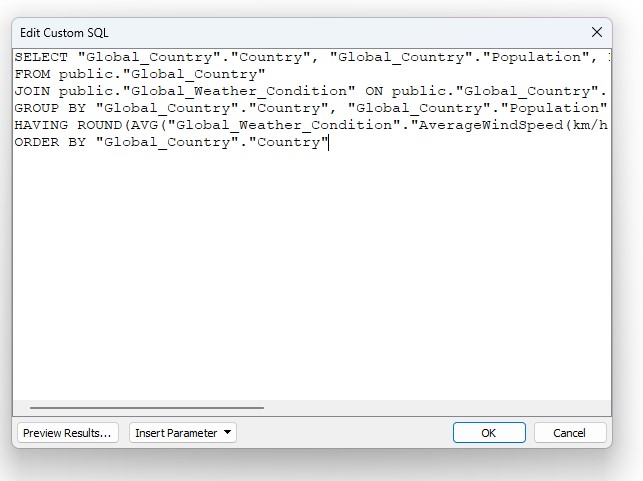
Launch Tableau and navigate to the Connect option. From there, choose PostgreSQL as the desired connection.



To establish a connection in Tableau, provide the server name and port where the database is hosted. Additionally, input the name of the specific database you wish to access. Next, choose your preferred method of signing in to the server, whether it's through Integrated Authentication or by providing a username and password. If the server requires a password and you're not operating in a Kerberos environment, you'll need to enter the appropriate credentials. Finally, click on the Sign In option to initiate the connection process.

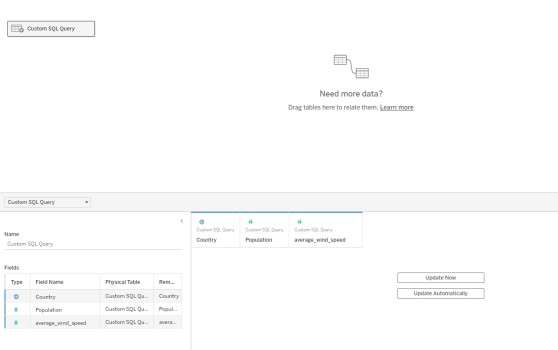


After this we can create the table directly based on sql coding of OLAP operation.Double click the New Custom SQL at the Table part of data source after connected to Tableau.

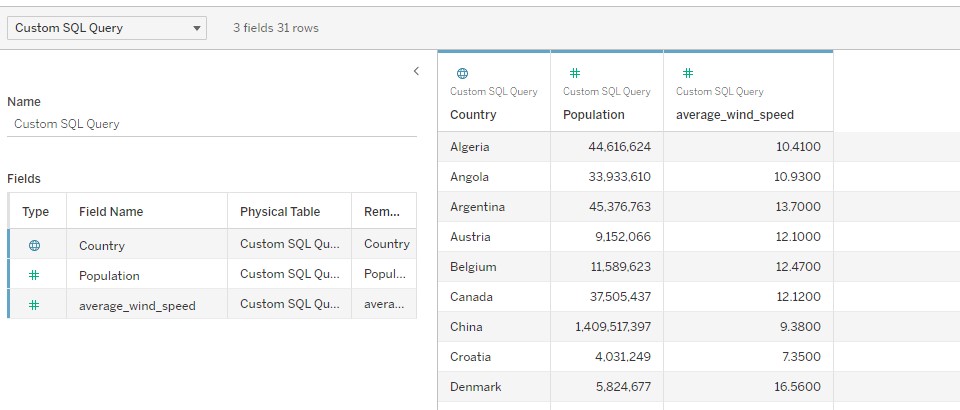


Then we are asked to copy and paste the SQL coding from postgresql and then click OK button.

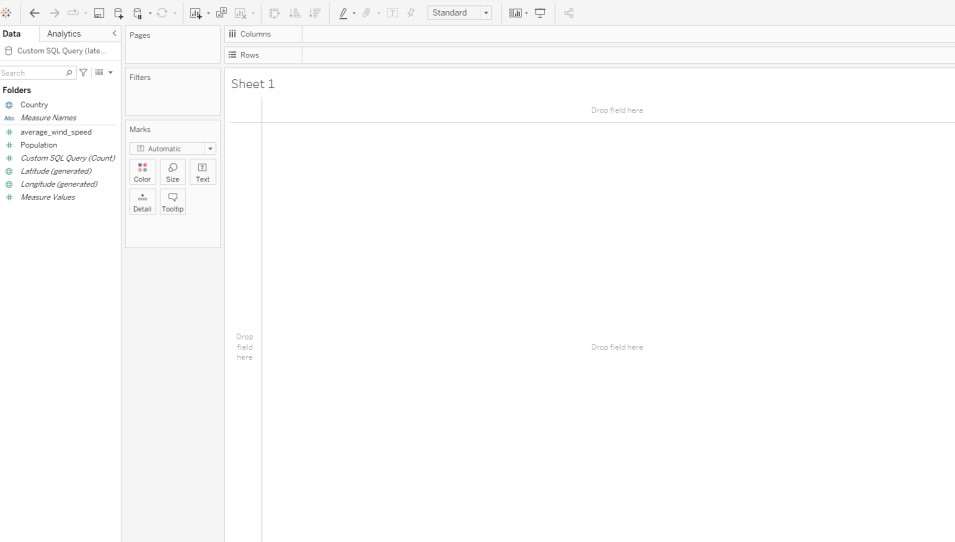
In this case, we copy the dicing operation from postgresql and paste in tableau custom sql.



Then one custom SQL query is created and we can click the button update now to check the result of SQL coding.



So we can see the table now, we can double check the result of table with the postgresql.The output should be exactly the same between them like number of rows and attribute.This means that we are successfully load the OLAP data from postgresql to tableau.



We can now start to visualise the graph based on the attribute of table result created just now.The process same goes to other OLAP operation as well.