

**A**

**MINI PROJECT REPORT ON**

**”Global Trends and Patterns in Sugar Consumption”**

**FOR**

Term Work Examination

***Bachelors of Computer Application in Artificial Intelligence and Machine Learning (BCA - AIML)***

**Year 2024-2025**

[**Ajeenkya DY Patil University, Pune**](http://www.nmu.ac.in/)

**-Submitted By-**

Mr. Shubham Sanjay Kendre

**Under the guidance of**

Prof. Vivek More

**Ajeenkya DY Patil University**

D Y Patil Knowledge City,  
Charholi Bk. Via Lohegaon,  
Pune - 412105  
Maharashtra (India)



Date: 16/04/2025

**CERTIFICATE**

This is to certified that\_Shubham Sanjay Kendre\_\_

A student’s of **BCA(AIML) Sem-4** URN No 2023-B-26112005 has Successfully Completed the Dashboard Report On

**“Global Trends and Patterns in Sugar Consumption”**

As per the requirement of

**Ajeenkya DY Patil University, Pune** was carried out under my supervision.

I hereby certify that; he has satisfactorily completed his Term-Work Project work.

Place: - Pune

**Examiner**

|  |
| --- |
| Index |
| Introduction and Objective |
| Methodology and Approach |
| Implementation and Code |
| Result and Visualization |
| Conclusion and Future Scope |

**Introduction**

Sugar consumption has become a subject of growing interest among researchers, policymakers, and health organizations due to its strong links to public health concerns, particularly obesity, diabetes, and cardiovascular diseases. This dataset provides a comprehensive overview of sugar consumption patterns across various countries and years, enabling an in-depth analysis of trends at both national and global levels. It includes data on total sugar consumption in tons, population sizes, and per capita sugar consumption, measured in kilograms per person. Such data is instrumental in identifying consumption trends, drawing correlations with socio-economic indicators, and assessing the effectiveness of health policies over time.

Understanding how sugar consumption evolves across different regions and periods is essential for governments and health institutions aiming to formulate effective dietary guidelines and interventions. Additionally, the dataset may help uncover underlying factors contributing to high or low sugar intake, such as economic development, cultural preferences, or policy changes. With the increasing prevalence of non-communicable diseases linked to high sugar intake, evaluating this data can also help support international health initiatives and promote awareness around healthier dietary choices.

The insights drawn from this dataset can also serve the food industry, aiding in the development of lower-sugar alternatives and helping businesses align with shifting consumer preferences. In sum, this dataset serves as a vital tool in the exploration of global sugar consumption dynamics and their multifaceted implications.

**Objectives:**

1. To analyze trends in sugar consumption over time across countries.
2. To compare per capita sugar consumption across different regions.
3. To identify correlations between population and total sugar intake.
4. To explore the potential impact of socio-economic and policy factors on sugar consumption.
5. To provide actionable insights for public health strategies and awareness campaigns.

**Methodology and Approach:**

The methodology adopted for analyzing the sugar consumption dataset involves a combination of **data preprocessing, exploratory data analysis (EDA), statistical techniques**, and **visualization** to draw meaningful insights. The following steps outline the approach in detail:

**1. Data Collection and Import**

The dataset was imported in CSV format, containing records on sugar consumption, population, and per capita values across different countries and years. The data serves as a secondary source compiled for analytical purposes.

**2. Data Cleaning and Preprocessing**

* Checked for **missing values**, duplicates, and inconsistencies.
* Standardized **country names** and formatted **year** as a datetime or numeric variable.
* Converted sugar consumption units to ensure consistency across entries, if needed.

**3. Exploratory Data Analysis (EDA)**

* Conducted **descriptive statistics** to summarize key attributes.
* Plotted **time series graphs** to observe trends in sugar consumption over the years.
* Used **bar charts and heatmaps** to compare per capita consumption among countries.
* Calculated **growth rates** and **averages** for individual countries and regions.

**4. Comparative Analysis**

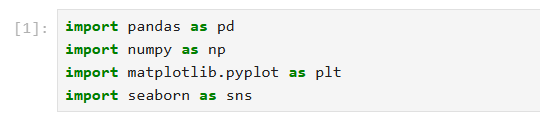
* Compared sugar consumption with population data to assess per capita trends.
* Grouped data by region or income level (if metadata is available or added) for broader comparisons.

**5. Insights and Reporting**

* Summarized key findings using graphs and tables.
* Highlighted actionable insights for health policymakers and stakeholders.

**Implementation and Code**

1.Importing Libraries

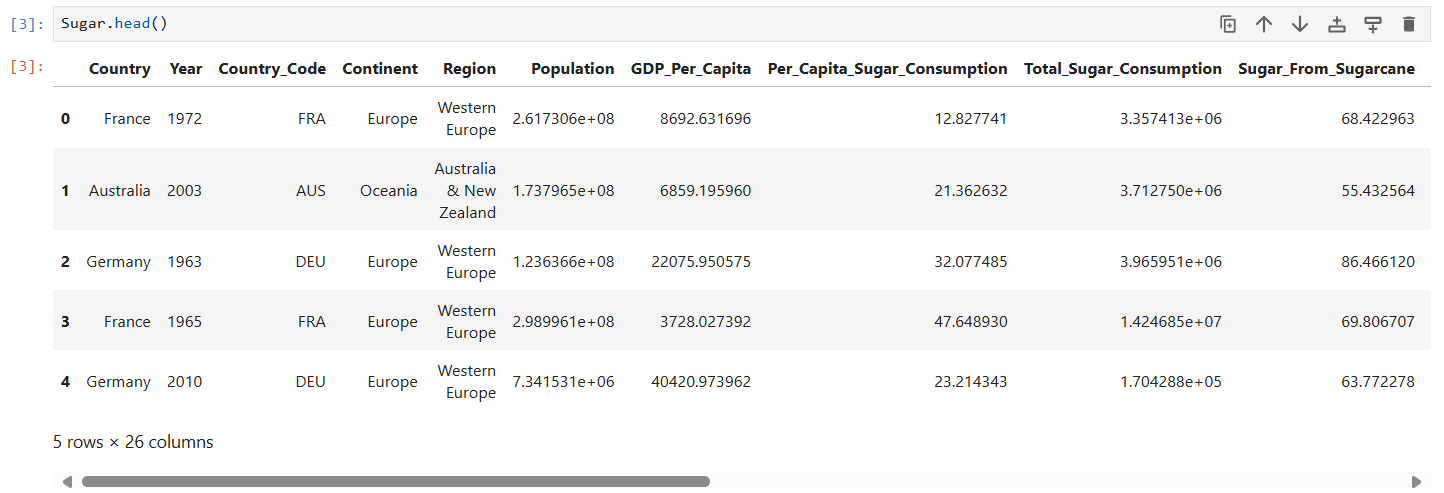


2.Loading the Dataset



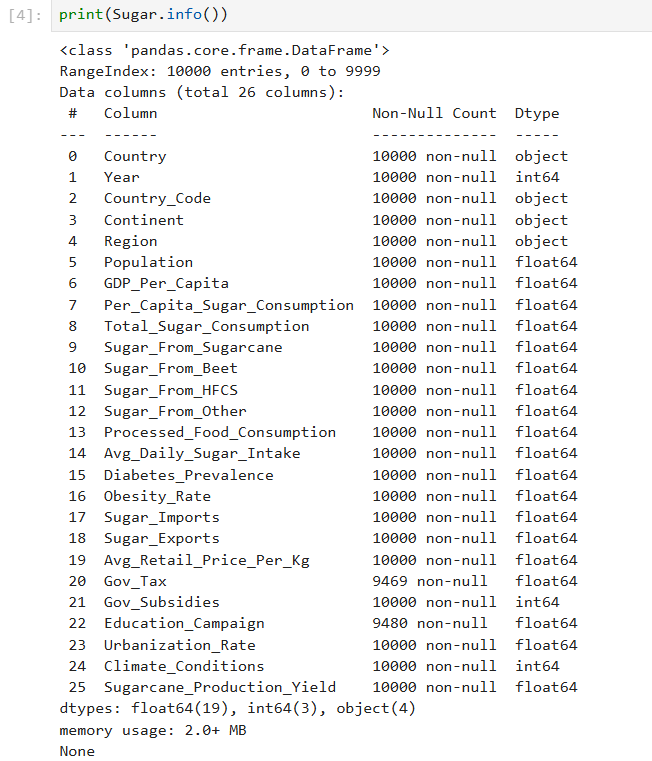
**Theory:** The pandas function read\_csv() is used to load data from a CSV (Comma Separated Values) file into a DataFrame. In this case, the dataset likely contains information about sugar consumption across different countries, years, and population details.

3. Initial Data Preview



**Theory:** The head() method displays the first five rows of the DataFrame, giving a quick preview of the data structure. This helps identify any obvious issues, such as missing values or inconsistencies, before proceeding with further cleaning.

4.Checking Basic Info



* **Theory:** info() provides metadata about the DataFrame, such as:
  + The number of rows and columns
  + The data types of each column (e.g., numeric, object, datetime)
  + The number of non-null values in each column, which helps identify if there are missing values.

This step allows for a better understanding of the data types and the presence of any missing or malformed data.

5. Dropping Rows with Missing Values



**Theory:** dropna() removes any rows with NaN (Not a Number) values, which could be the result of incomplete data entries. By using inplace=True, the DataFrame is modified directly, rather than creating a new copy. This step ensures that the dataset doesn't have any missing entries that might affect subsequent analysis.

1. Converting **the 'Year' Column to Numeric**

****

Theory: The pd.to\_numeric() function converts the values in the 'Year' column to a numeric format. The errors='coerce' argument forces any non-numeric entries to be converted to NaN. This ensures that the 'Year' column consists of numeric values only, avoiding issues in analysis.

1. Cleaning 'Country' Names



**Theory:** str.strip() removes any leading or trailing whitespace characters from the 'Country' column. This is important because extra spaces can cause issues when comparing country names or performing groupings.

1. Removing Duplicate Rows



**Theory:** The drop\_duplicates() function eliminates rows that are identical across all columns. Duplicates can distort analysis results and lead to incorrect insights, so it's essential to remove them.

1. Dropping Rows with NaN Values After Conversion



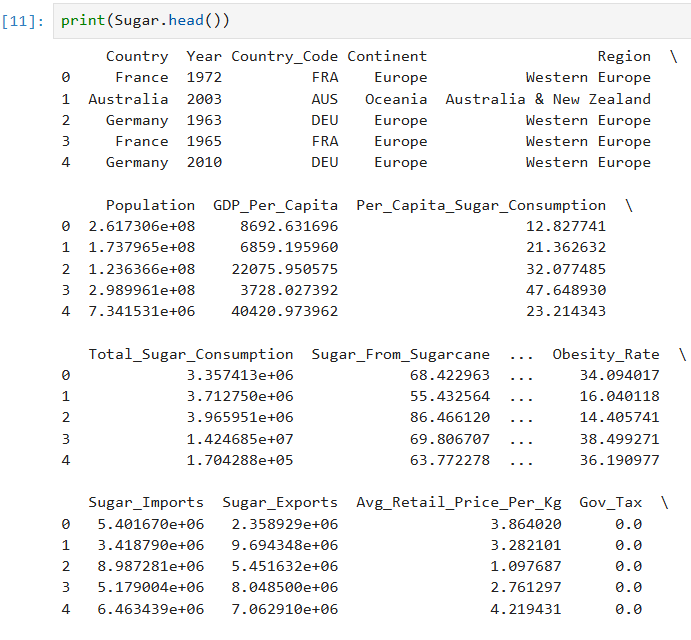
**Theory:** After converting columns to numeric types, some entries might have been turned into NaN (e.g., if the conversion failed for certain rows). This step removes those rows to ensure that no missing values remain in the dataset.

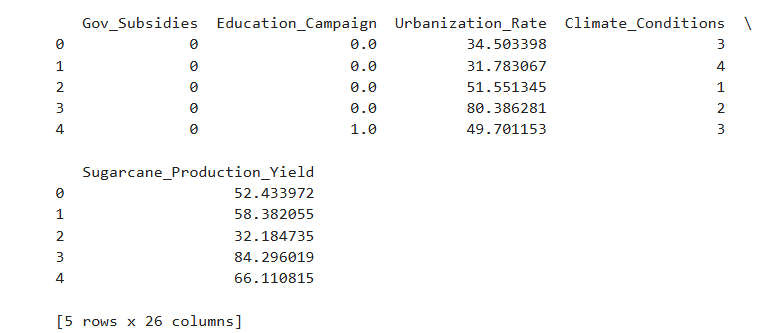
10.Resetting the Index



**Theory:** After dropping rows and making changes to the dataset, it's common to reset the DataFrame index, which might have gaps after rows were removed. The drop=True argument ensures that the old index is discarded and the DataFrame gets a clean, consecutive index.

11.Final Data Preview

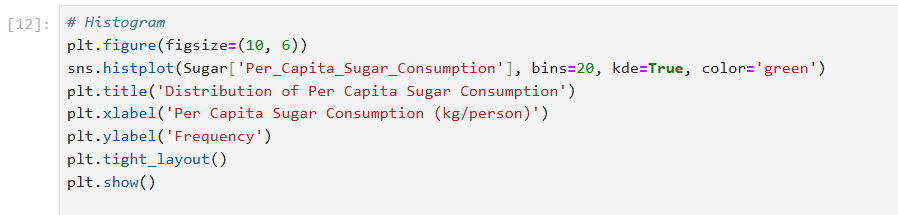


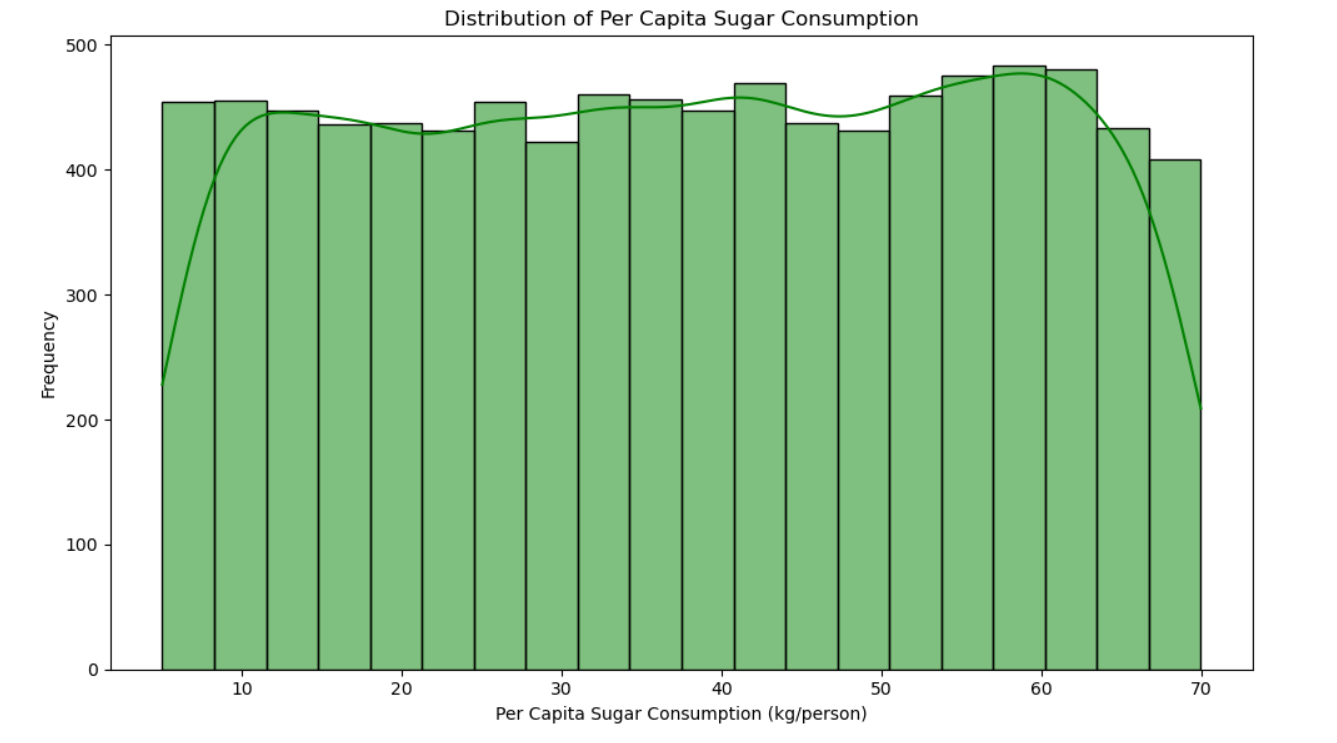


**Theory:** A final preview of the cleaned dataset ensures that all the cleaning steps were successful, and it helps verify that the data is now in a suitable form for analysis.

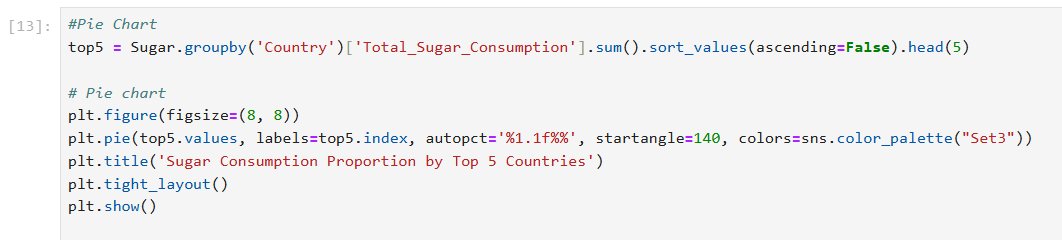
**Results and Visualization**

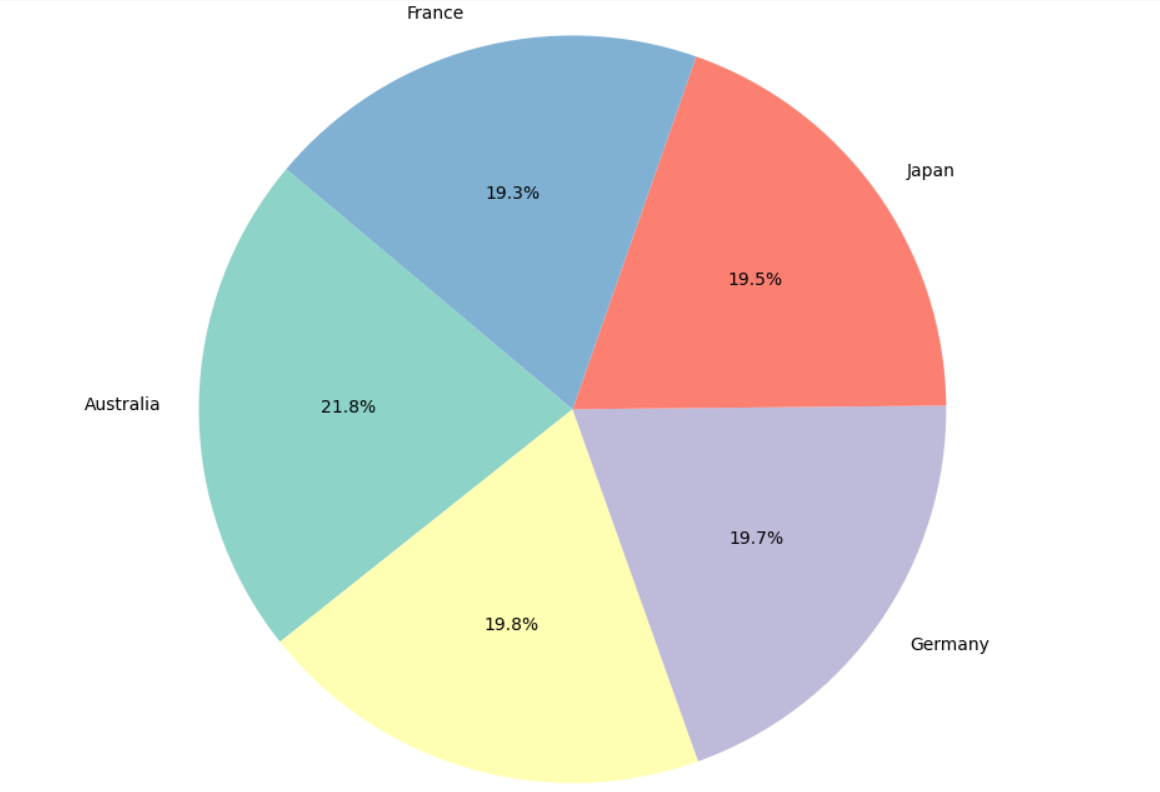
* 1. Histogram — Per Capita Sugar Consumption Distribution





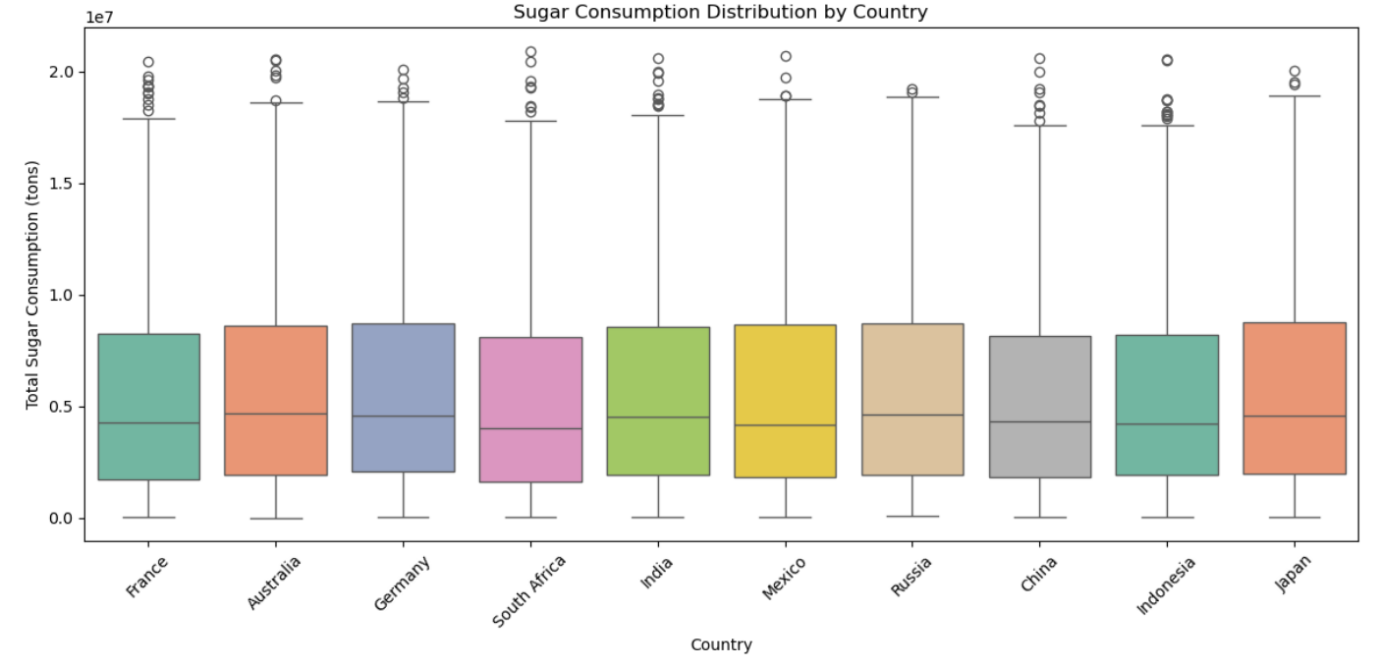
* 1. Pie Chart — Sugar Consumption by Top 5 Countries



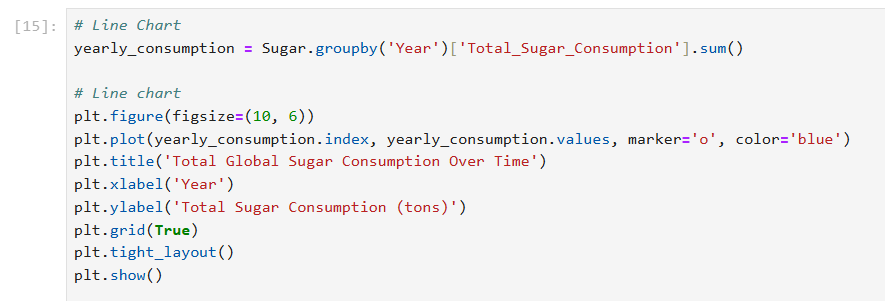


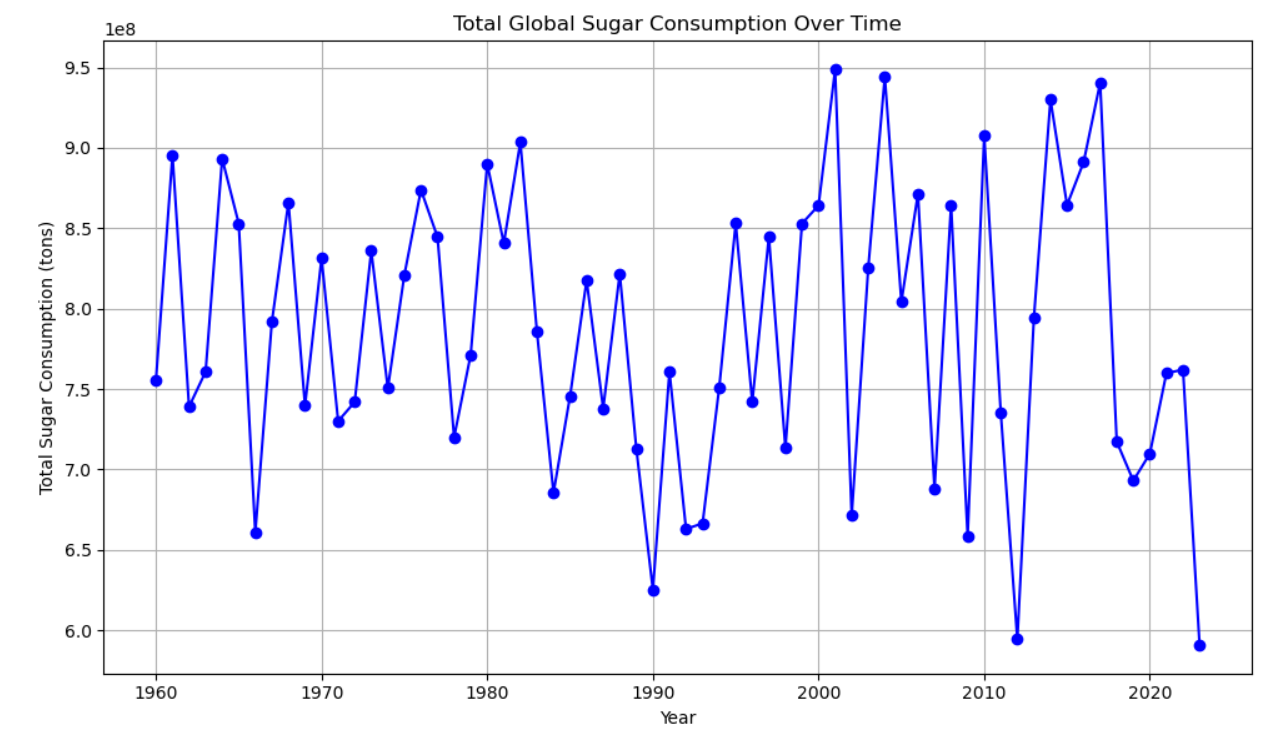
3.Box Plot — Distribution by Country (Top 10 Represented Countries)



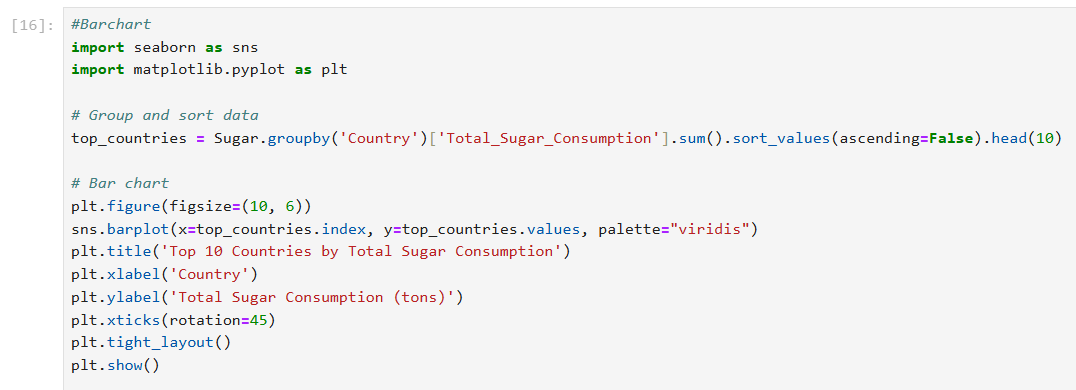


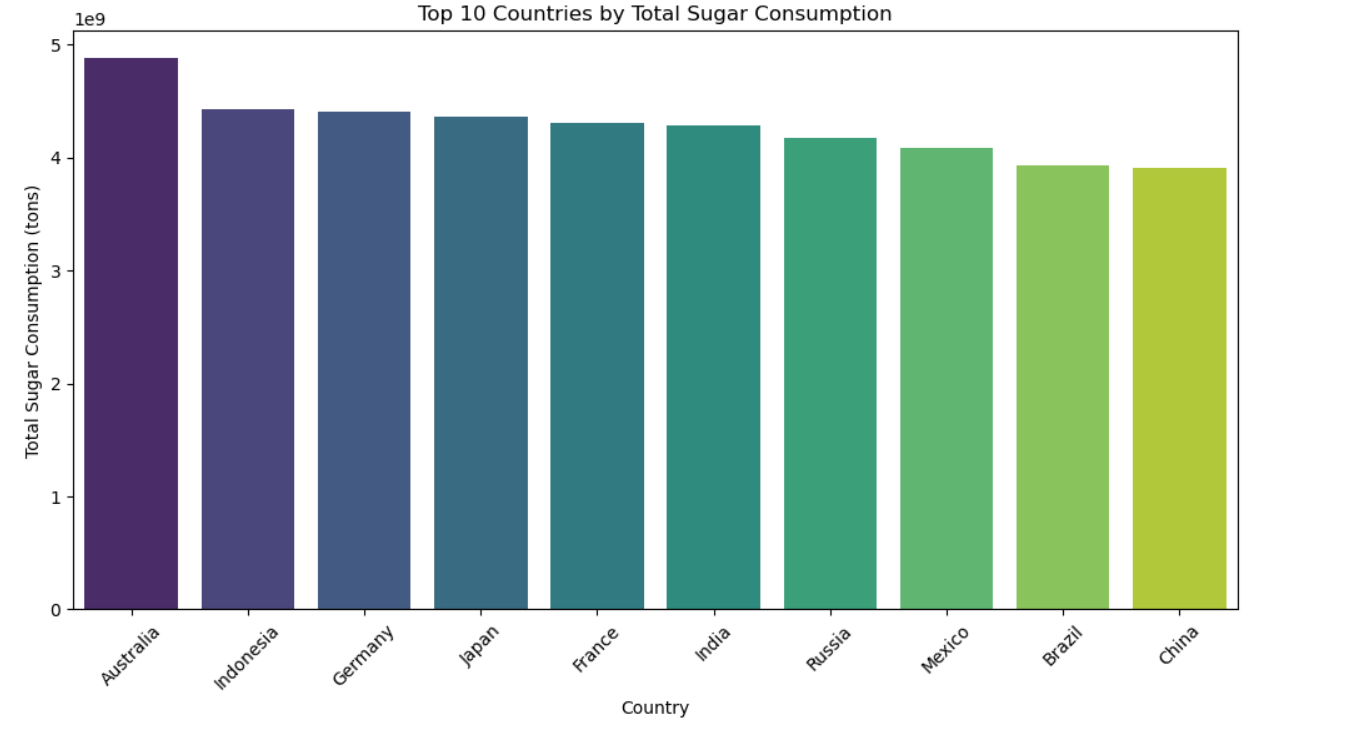
4. Line Chart — Global Sugar Consumption Over Time





5. Bar Chart — Top 10 Countries by Total Sugar Consumption





**Conclusion**

The exploratory data analysis (EDA) conducted on the sugar consumption dataset reveals several insightful trends and distributions. The bar chart indicates that a few countries dominate global sugar consumption, which is crucial for understanding which nations are the highest contributors. The line chart shows the trend over time, allowing us to observe whether global sugar consumption is increasing or decreasing. The histogram highlights the variation in per capita sugar consumption, showing where most countries lie in terms of individual consumption. The pie chart visualizes the distribution of sugar consumption across the top 5 countries, emphasizing the disparities in consumption between nations. Finally, the boxplot provides a clear picture of the spread and variability in sugar consumption by country, including identifying countries with extreme outliers or highly variable consumption patterns.

Overall, the analysis shows a skewed distribution of sugar consumption across countries, with some nations consuming far more sugar per capita than others, while the global trend of sugar consumption appears to fluctuate over time. The presence of outliers and variations across countries suggests that policy interventions, health awareness programs, and international regulations could play a role in addressing these disparities.

**Future Scope**

There are several opportunities to extend this analysis and further explore the data. First, a correlation analysis could be conducted to understand the relationship between sugar consumption and other factors, such as GDP, health statistics, or government policies. Further, incorporating data from multiple years (if available) could help analyze the impact of specific events or policy changes on sugar consumption. The dataset could also be enriched by integrating data on health outcomes, such as obesity rates or diabetes prevalence, to explore the health impact of high sugar consumption. Additionally, the analysis could be expanded to cover more countries or regions for a more comprehensive global comparison, and advanced visualizations like heatmaps could be used to explore patterns across geographic locations. Finally, predictive modeling could be introduced to forecast future sugar consumption trends based on historical data, helping policymakers make informed decisions about public health interventions and regulations.

These avenues would provide a deeper understanding of the factors driving sugar consumption and offer potential strategies for reducing health risks associated with high sugar intake.