

# World Population Data Analysis Report

## 1. Introduction

This report presents an analysis of the World Population dataset from the World Bank. The dataset includes population data for various countries from 1960 to 2023, along with metadata such as region, income group, and special notes. The goal of this analysis is to understand the population distribution, regional distribution, and income group distribution over the years.

## 2. Data Preprocessing

### 2.1 Data Inspection

The dataset was inspected to understand its structure, including checking for missing values, data types, and duplicates.

### 2.2 Handling Missing Values

- Population data missing values were filled using the forward fill method.
- Metadata missing values for special notes were filled with 'No Special Notes', and missing values for the 'Gender' column were filled with 'others'.

### 2.3 Data Cleaning

Unnecessary columns such as 'Indicator Name', 'Indicator Code', and 'Country Code' were dropped from the dataset. The dataset was further checked for data types, duplicates, and any remaining missing values.

## 3. Population Distribution Over the Years

Histograms were created for each year from 1960 to 2023 to visualize the distribution of population values across different countries. These histograms help in understanding how population values are spread out for each year.

## 4. Total Population Per Year

The total population per year was calculated by summing up the population values of all countries for each year. A horizontal bar chart was created to visualize the total population per

year from 1960 to 2023. This visualization highlights the overall growth trend of the world population over the years.

## **5. Country-wise Population Analysis**

### **5.1 Top 20 Countries by Population in 1960**

The top 20 countries by population in 1960 were identified and visualized using a bar chart. This helps in understanding which countries had the highest population in the early years of the dataset.

### **5.2 Top 20 Countries by Population in 2023**

Similarly, the top 20 countries by population in 2023 were identified and visualized. This allows for comparison with the 1960 data to see how the population rankings have changed over time.

## **6. Regional Distribution**

The regional distribution of countries was analyzed and visualized using a bar chart. This chart shows how countries are distributed across different regions of the world, providing insights into regional population densities and distributions.

## **7. Income Group Distribution by Region**

The distribution of income groups within each region was analyzed and visualized using a count plot. This visualization helps in understanding the economic distribution of countries within different regions.

## **8. Word Cloud of Special Notes**

A word cloud was created to visualize the frequency of words in the special notes column. This provides a visual representation of the most common terms and phrases used in the special notes.

## **9. Income Group vs. Special Notes Length**

The relationship between the length of special notes and income groups was analyzed using a scatter plot. This analysis helps in understanding if there is any correlation between the length of special notes and the income groups of countries.

## 10. Income Group vs. Age

Random age data was generated for each country and visualized against income groups using a scatter plot. This visualization helps in understanding the age distribution within different income groups.

## 11. Gender Distribution

Random gender data was generated for each country and visualized by region and income group. Bar charts were created to show the gender distribution within regions and income groups, providing insights into gender demographics.

## 12. Datasets Used

- Population Data: [API\\_SP.POP.TOTL\\_DS2\\_en\\_csv\\_v2\\_433762.csv](#)
- Metadata: [Metadata\\_Country\\_API\\_SP.POP.TOTL\\_DS2\\_en\\_csv\\_v2\\_433762.csv](#)

## 13. Conclusion

This analysis provides a comprehensive view of the world population distribution across different countries, regions, and income groups over the years. The visualizations highlight the trends and patterns in population growth, regional distribution, and income group distribution. The insights gained from this analysis can be used for various purposes, including policy-making, economic planning, and demographic studies.