REPORT

Credora Internship **– Data Science**

Task 01

[Population Data Visualization]

Submitted by: DNYANESH SHINDE

DATE: 19-5-2025

Introduction

This project is a part of my Data Science Internship at Credora. The goal was to understand and visualize the distribution and growth of global population using real-world data from the World Bank. The analysis involved data preprocessing, handling missing values, merging metadata, and creating informative visualizations using Python.

Dataset Description

The analysis was performed using three datasets:

1. API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv – Contains total population of each country for each year from 1960 onwards.
2. Metadata_Country_API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv – Contains country metadata such as Region, Income Group, etc.
3. Metadata_Indicator_API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv – Describes the indicator used in the dataset.
4.

Tools & Libraries Used

- Python (Google Colab / Jupyter Notebook)
- pandas
- matplotlib
- seaborn
- plotly.express
-

Data Cleaning & Preprocessing

- Loaded all three datasets
- Removed unnecessary columns like 'Indicator Name' and 'Indicator Code'
- Converted the dataset from wide format (years as columns) to long format
- Converted year column to integer
- Merged with country metadata using Country Code
- Removed rows with missing population values

Visualizations

1. ***Bar Chart – Population by Region (Latest Year)***

This chart shows total population grouped by region for the most recent year available. Asia has the largest population share followed by Africa.

---

2. ***Histogram – Population Distribution Across Countries***

This histogram shows that the majority of countries have a smaller population, while a few countries have extremely large populations.

---

3. ***Line Chart – Population Growth Over Time***

The line chart visualizes population growth trends of selected countries like India, China, the USA, Brazil, and Nigeria. All show significant growth trends, especially in India and Nigeria.

---

4. ***Choropleth Map – World Population by Country***

This interactive map shows population distribution by country. Countries with higher populations are highlighted with more intense color shades. It provides a global view of the population landscape.

Key Insights

Asia remains the most populous region globally.

- CN China and IN India continue to dominate global population numbers.
- Population distribution is heavily skewed – most countries have a low population.
- The choropleth map provides excellent geographical context to population data.

Conclusion

The population visualization task offered a deep understanding of how population data can be analyzed and presented meaningfully. With Python and visualization libraries, we translated large datasets into easy-to-understand charts and maps. This helped identify global and regional population trends.

🔗 Reference

- **World Bank Population Dataset**