# REPORT

Credora Internship- Data Science

[Task-01]

[Population Data Visualization]

Submitted by :Alamuru Venkata Harshitha

Date:20/05/2025

# Introduction :-

This project is part of my **Data Science Internship at Credora**. The objective was to analyze and visualize global population trends using publicly available data from the **World Bank**. Population data is a crucial indicator for policymaking, urban planning, and global development tracking. Through this project, I aimed to gain practical experience in:

- Extracting and cleaning real-world datasets
- Working with metadata to enhance analysis
- Visualizing trends and distributions in population
- Drawing insights from historical and regional demographic changes

## ## 💼 Dataset Description :-

The analysis uses three main datasets:

1. `API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv`
   → Contains total population data per country from 1960 onwards.

2. `Metadata_Country_API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv`
   → Metadata for countries including Region and Income Group.

3. `Metadata_Indicator_API_SP.POP.TOTL_DS2_en_csv_v2_85220.csv`

## Tools and Libraries Used:-

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

## 🔧 Data Cleaning & Pre-processing:-

- Loaded all three datasets
- Removed unnecessary columns like `'Indicator Name'`, `'Indicator Code'`
- Transformed from wide to long format
- Merged datasets using `Country Code`
- Removed missing values

## Visualizations Created:-

1. **Bar Chart – Population by Region (Latest Year)**
   > Highlights that Asia has the highest population share.

2. **Histogram – Population Distribution Across Countries**
   > Most countries have smaller populations; a few are very large.

3. **Line Chart – Population Growth Over Time**
   > Tracks growth in countries like India, China, USA, Brazil, Nigeria.

4. **Choropleth Map – Global Population**
   > Visual interactive map colored by country population levels.

--

## Key Insights:-

- Asia is the most populous region globally.
- China CN and India IN lead global population figures.
- Population is unevenly distributed: many low-population countries vs. few high-population ones.

## ✓ Project Outcome:-

This task helped me learn how to:
- Extract and clean real-world datasets
- Merge and process structured metadata
- Create visualizations that make data insights clear and accessible
- Understand global demographic trends using visual tools

## 📎 References:-

- 📊 Dataset: [World Bank Population](https://data.worldbank.org/indicator/SP.POP.TOTL)
- 📄 Report: `population_data_visualization_report.pdf` (see in this repository)