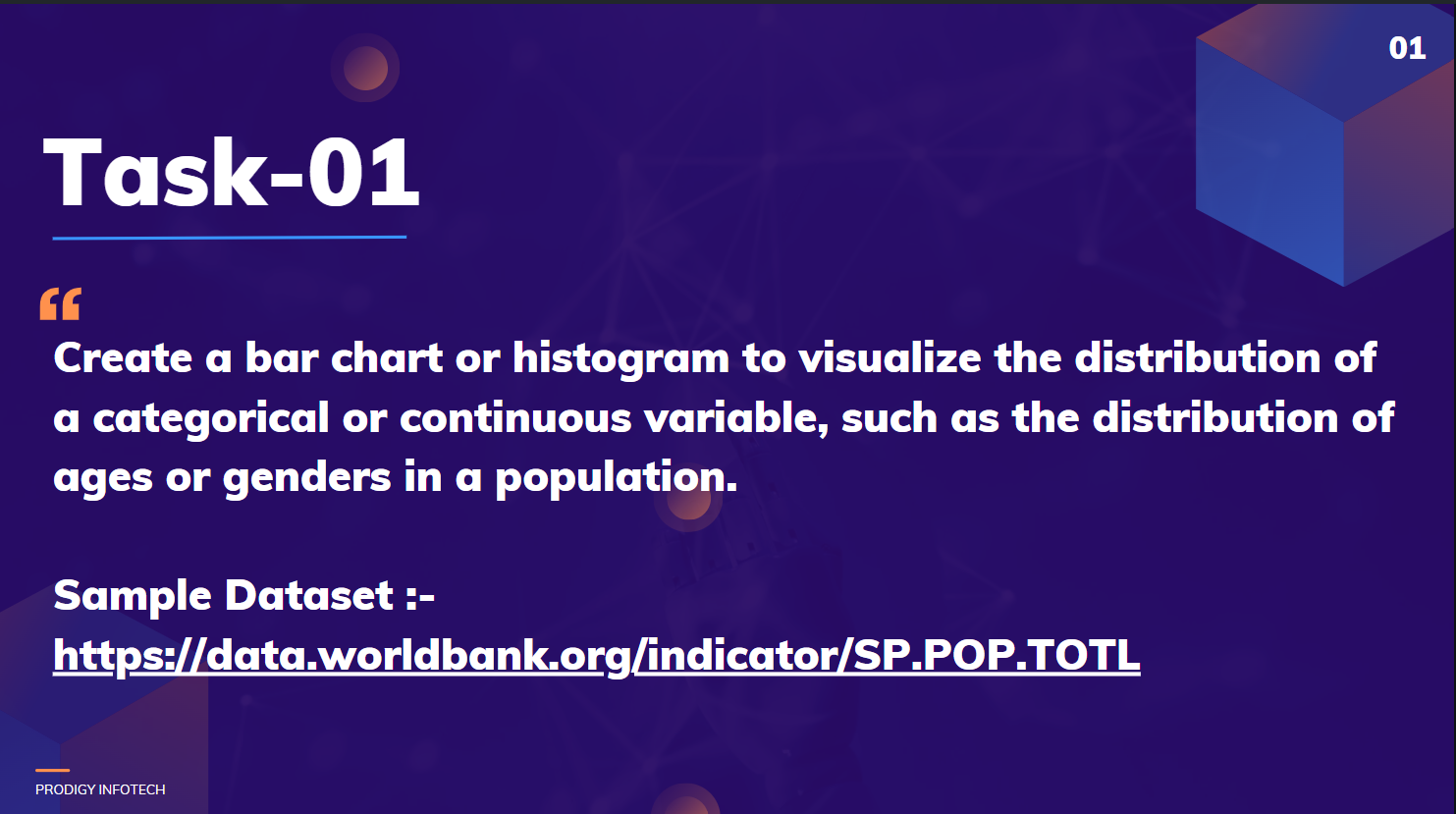
**Prodigy Infotech Task 1**

[](https://github.com/Ksheeraja23/Prodigy_infotech_task1/blob/main/ds1.png)

Welcome to my submission for Task 1 of the Data Science Internship at Prodigy Infotech. In this task, I have performed Exploratory Data Analysis (EDA) on a dataset provided, focusing on creating a visualization to represent the distribution of a categorical or continuous variable.

**Dataset**

The dataset used for this task is [world\_population\_dataset](https://github.com/kindo-tk/PRODIGY_DS_01/blob/main/worldpopulationdata.csv). This dataset contains records of population from the year 2001 to 2022.

**Tools and Libraries used**

* Jupyter notebook
* Pandas
* Numpy
* Matplotlip & Seaborn for visualization

**Exploratory Data Analysis (EDA):**

1. Data Loading: The dataset is loaded into a Pandas DataFrame for analysis.
2. Data Exploration: The unique countries in the dataset are identified, and the total number of unique countries is printed.
3. Top Ten Countries: The top ten countries by total population for the years 2001, 2012, and 2022 are extracted and visualized using bar charts.
4. Bottom Ten Countries: Similarly, the bottom ten countries by total population for the same years are extracted and visualized.
5. Gender Population Analysis: The project also analyzes the male and female populations separately, identifying the top ten countries for each gender across the specified years.

**Visualizations:**

Bar Charts: The project generates bar charts to visually represent the top and bottom ten countries by population for the years 2001, 2012, and 2022. These visualizations help in understanding population distribution and trends over time. Histogram: To represent the data of the Top and Bottom 10 countries with both male and female population, comparing the data over the span of 10 years each between 2022, 2012, and 2001.

**Conclusion**

This analysis provides insights into global population trends, highlighting the countries with the largest and smallest populations over the years. It also focuses on the gender distribution or age distrubution within the population for which, I've personally decided to choose the gender distribution.