Link to github <https://github.com/AadyaShakti01/045001_Aadya/commits?author=AadyaShakti01>

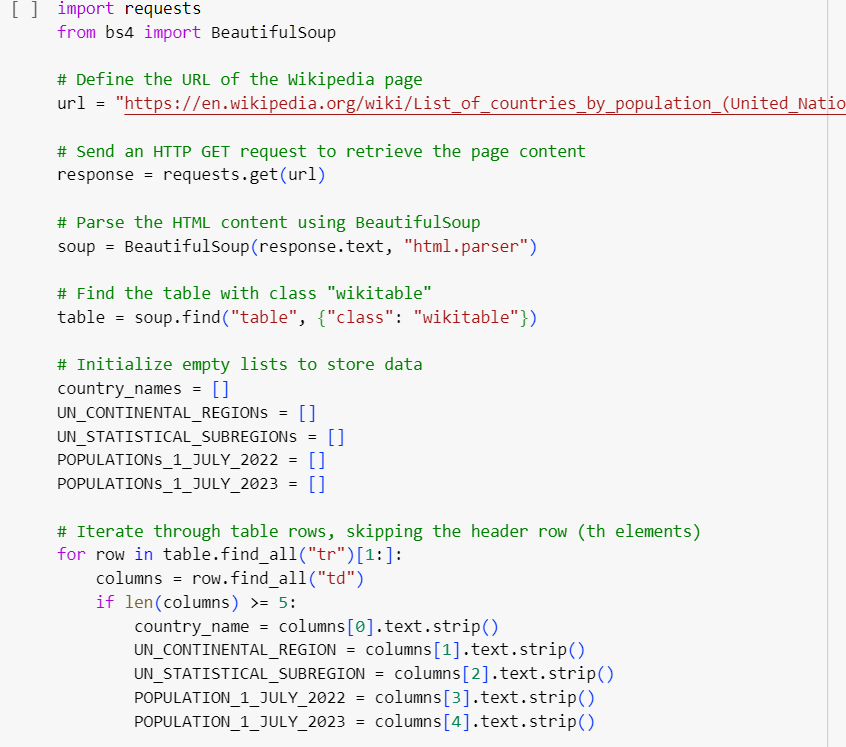
PROJECT OBJECTIVES : To understand the trends of population from 2022 to 2023. To enhance our knowledge on statistics and understand the relationship between population in 2022 and 2023. Also, understand the reasons for growth or decline of population among different countries whose data is available to us.

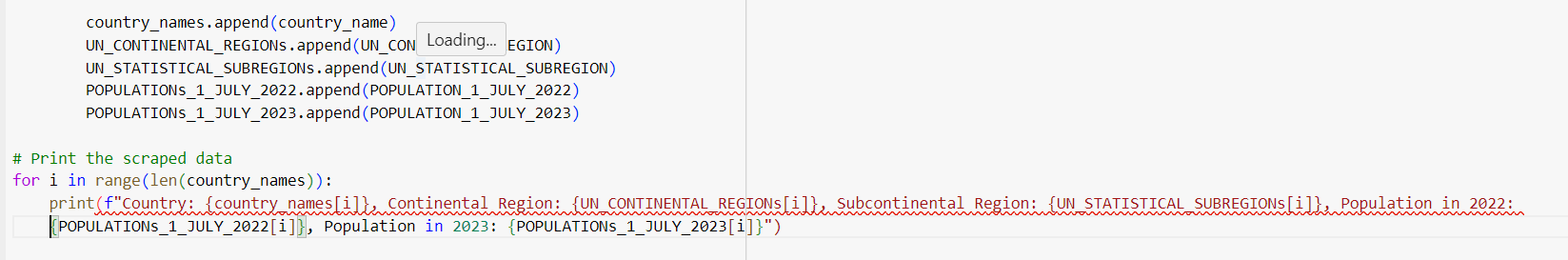
Here I have described a few codes which were used while coding and rest can be referred to in google colab file.

The data being scrapped over here is from Wikipedia. It displays list of countries by population (approx. 8 billion). It has been released by UN (United nations).

STEP 1:

In this code we need to import requests and BeautifulSoup which helps us to scrap data from any website. Here we have given the url of the Wikipedia page we need to scrap. We scrap the data of a table given on the page. We do this with help of html codes. The html.parser is used to derive data of the table within html tags.





Now , we wish to initialize the columns to enter the required data in respective columns. Then we print the scrapped data.

We get the OUTPUT in this form:

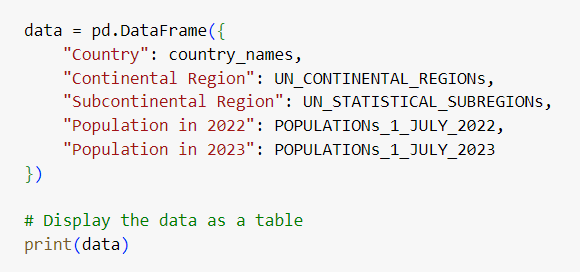


There are around 235 such rows.

STEP 2:

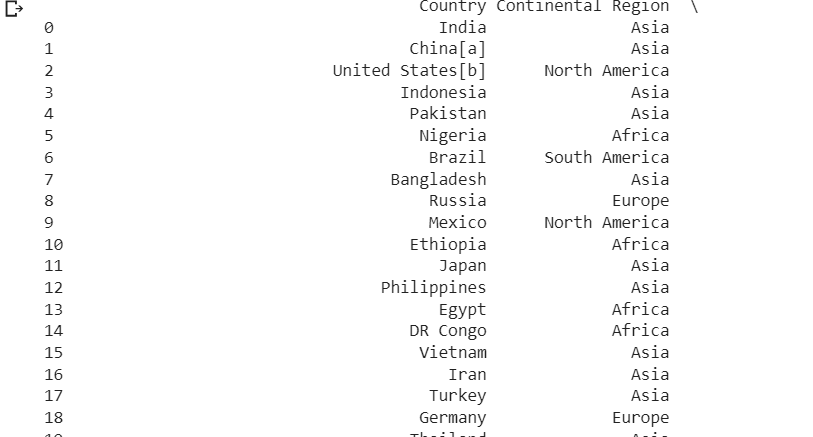
Now we convert this data into a dataframe and display the output in table form.

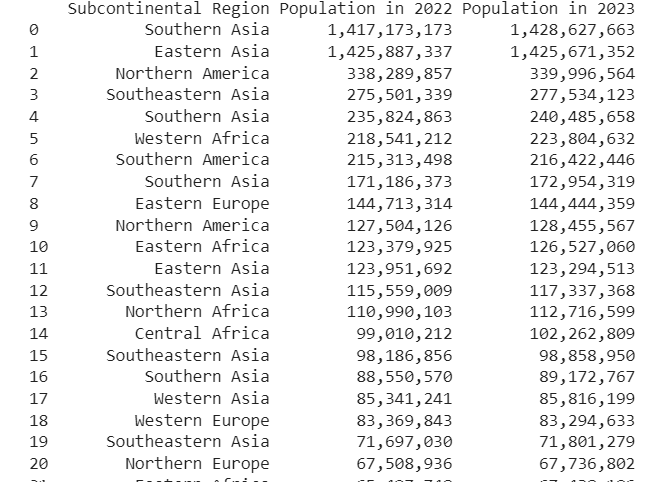
Following is the code used:



OUTPUT

We get a well defined table

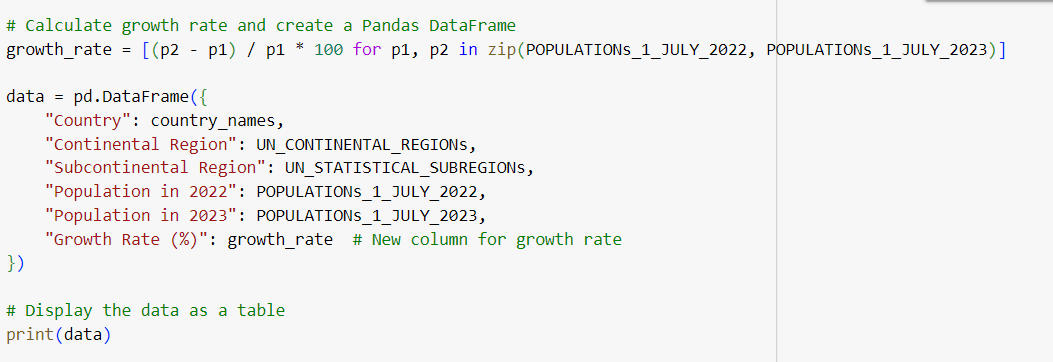




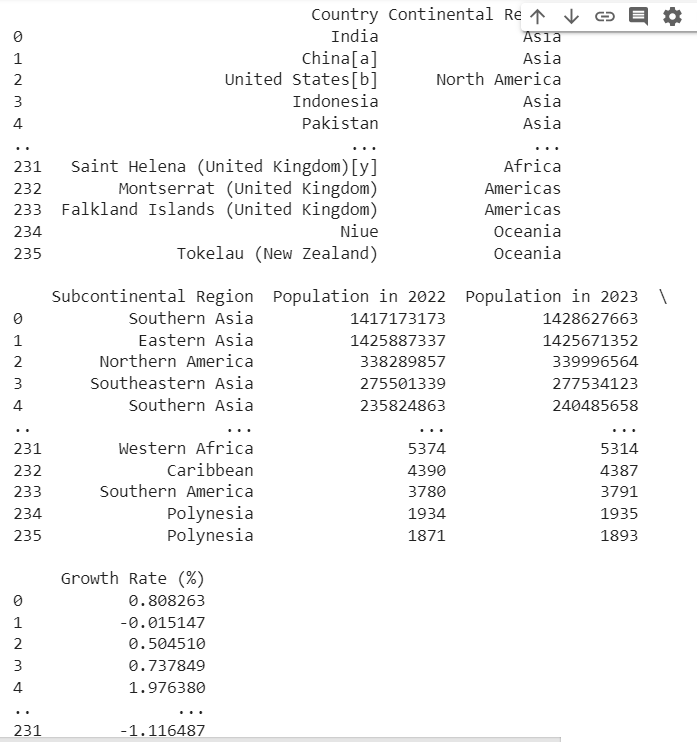
And 235 such rows follow……………………..

STEP 3:

Now we wish to find growth rate of population from 2022 to 2023 on our own and display the same in new column…



OUTPUT:

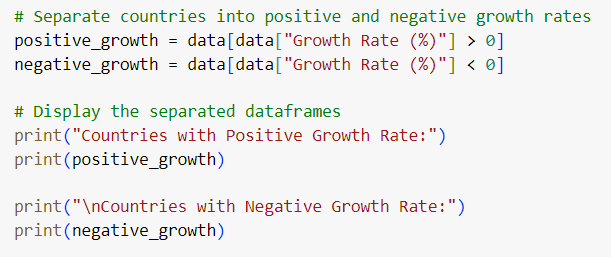


We get an additional sixth column with row heading as growth rate.

STEP 4 :

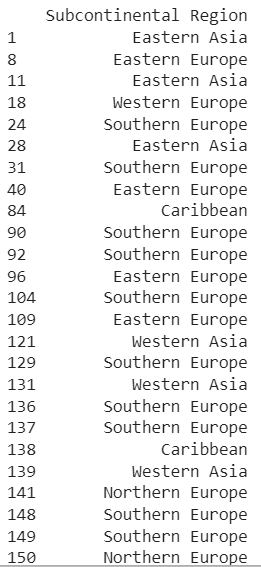
Now we wish to find out how many countries have positive or negative growth rate.

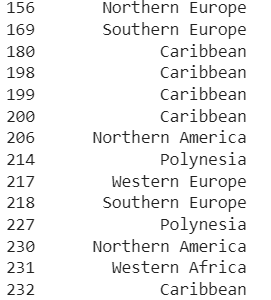
We use the following code for the same.



OUTPUT :

We have around 39 countries which have negative growth rate.





For the rest of the analysis, we have used different codes which can be viewed in colab file attached……

Unfortunately, we Could not build wheels for pystan and thus was not able to forecast future trends.

The managerial report and insights are shared below.

**REPORT**

* We have scrapped the data from Wikipedia page which provides the table having the data showing the population of different countries in the world.
* Here it is tried best to calculate the growth rate of different countries from the population in 2022 and population in 2023.
* From the data received we can tell that there are around 39 countries from 235 which have a negative growth rate.

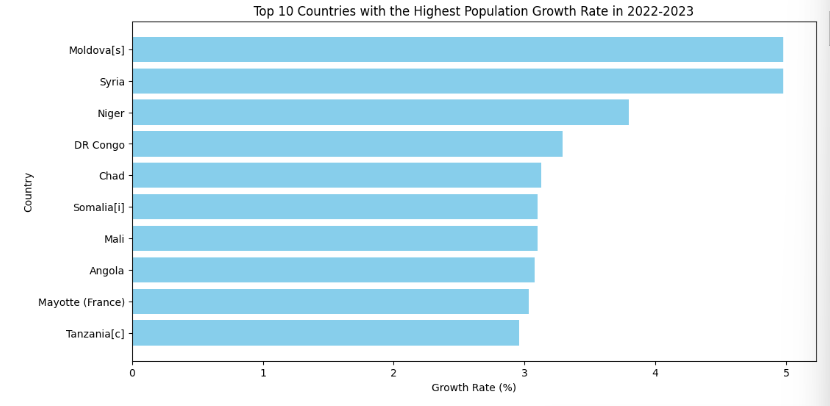
1. Negative growth rate is indicative of a declining population. This means that the number of births in a given population is lower than number of deaths and other factors might include immigration.

Other factors might be:

* Low Birth Rates
* High death Rates
* Net OutMigration
* Aging Population

1. Positive growth rate is indicative of an increasing population. This means that number of births in a given population is greater than the number of deaths and it may also be due to immigration. Other factors might include :

* High Birth rates
* Low Death rates
* Net Inmigration
* Youthful Population
* Also, here we have found the top 10 countries with highest population growth rate in 2022 – 2023.



The highest population growth rates in some countries indicate several demographic and socioeconomic factors and can have a range of implications. Here are some of the key implications and factors associated with high population growth rates:

1. High Fertility Rates

2. Young Population

3. Economic Challenges:

4. Pressure on Resources

5. Healthcare and Education Challenges

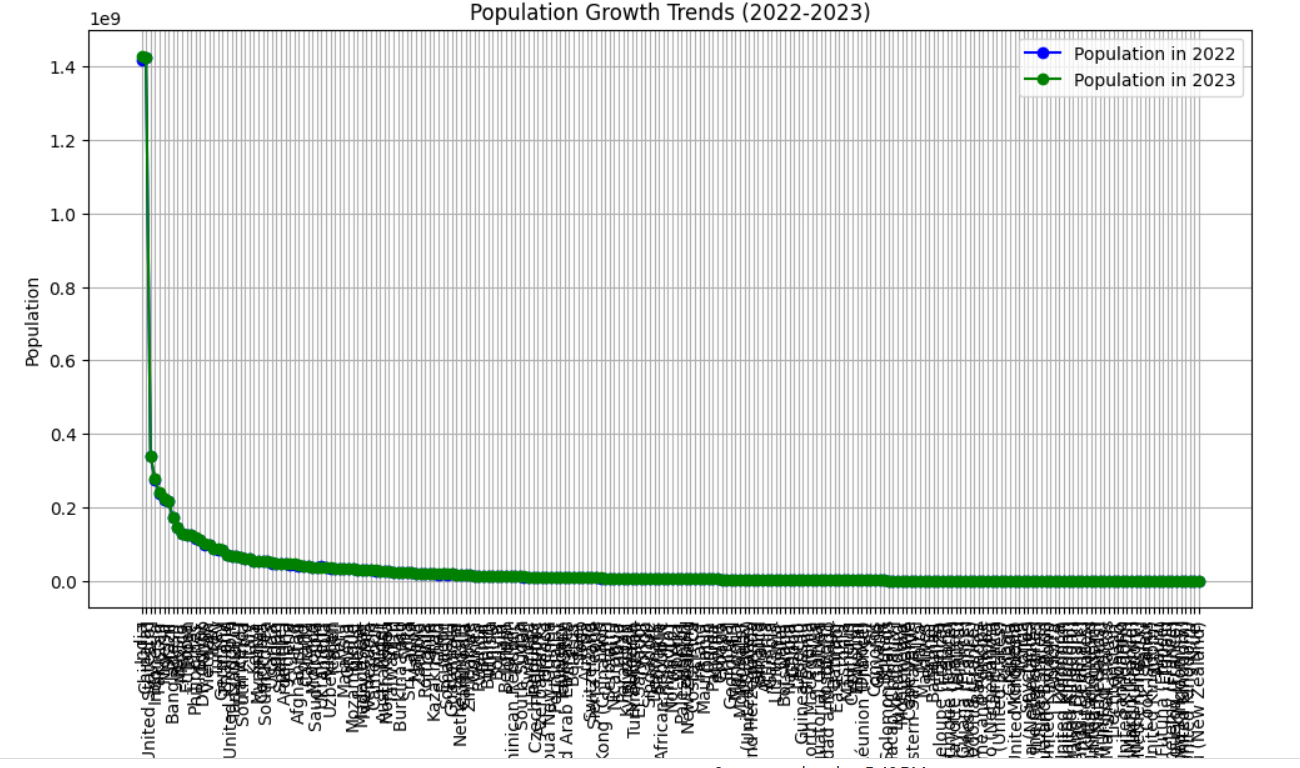
6. Urbanization

7. Gender Equality

8.Government Policies

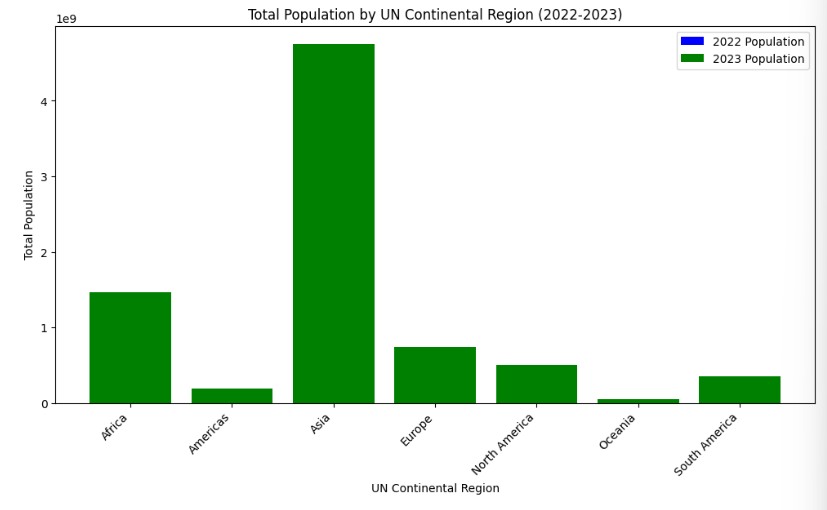
It's important to note that the implications of high population growth rates can vary widely depending on the specific circumstances of each country. Some countries may effectively manage the challenges associated with high population growth rates and leverage their youthful populations for economic development, while others may struggle to provide essential services and opportunities for their citizens. Sustainable development strategies often aim to address these challenges by promoting access to education, healthcare, family planning, and economic opportunities while considering the unique context of each country.

* We have tried to get the population growth trends in 2022 – 2023. These growth trends indicate that



Since the data here is inclusive of 235 countries it is a bit difficult to checkout for each country. But it does show that growth rate is more or less the same for about 80% of the countries presented here.

* Now we have calculated the total population in each continent in the year 2023.



This bar graph shows that maximum population of the world resides in Asian continent. The least population is found in Oceania.

The distribution of the world's population across different continents is influenced by a complex interplay of historical, geographical, cultural, and economic factors.

1. Geographical Size: Asia is the largest continent on Earth, covering a vast land area. Its landmass provides ample space for a large population to inhabit. Oceania, on the other hand, is much smaller in comparison and consists of numerous islands, many of which are remote and sparsely populated.

2. Historical Settlement Patterns: Historically, Asia has been home to some of the world's oldest civilizations, including those in the fertile river valleys of Mesopotamia, the Indus, the Ganges, and the Yellow River. These areas have supported dense populations for thousands of years. In contrast, Oceania was one of the last regions to be settled by humans, and its remote islands limited population growth.

3. Agriculture and Food Production: Asia has a long history of agriculture, with a wide range of crops and domesticated animals that can support large populations. The development of agriculture allowed for food surplus, which in turn supported population growth. In Oceania, agriculture was more limited due to the fragmented island geography and limited domesticated plants and animals.

4. Urbanization and Economic Opportunities: Asia is home to several of the world's most populous and rapidly growing countries, such as China and India. These countries have experienced significant urbanization and economic development, which has drawn people from rural areas to cities in search of better economic opportunities. Oceania's smaller economies and isolated geography have limited urbanization to a smaller scale.

5.Cultural and Historical Factors: Cultural and historical factors, including migration patterns and family structures, have influenced population growth and distribution in both regions. In Asia, large family sizes have historically been common especially people with old ages, contributing to population growth. In Oceania, the population is more dispersed across numerous island nations and territories.

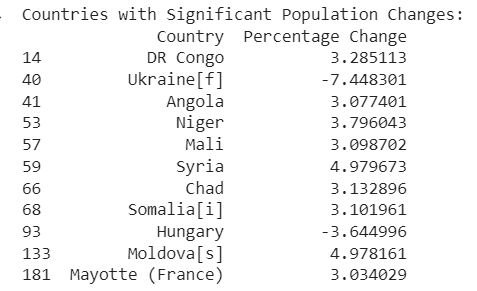
6. Natural Resources and Infrastructure: Asia's vast and diverse landscapes provide a wide range of natural resources, supporting economic activities and infrastructure development. Oceania's smaller landmass and remote islands have limited access to some resources, affecting economic development and population growth.

7. Climate and Environment: Climate and environmental factors can also influence population distribution. Some parts of Oceania have less hospitable climates or face environmental challenges such as cyclones and rising sea levels, which can impact population growth and stability.

It's important to note that within each continent, there is considerable diversity in population density and distribution. Some regions within Asia are densely populated, while others, such as deserts and mountainous areas, have sparse populations. Likewise, Oceania includes densely populated areas like Australia and New Zealand, in addition to many sparsely populated islands.

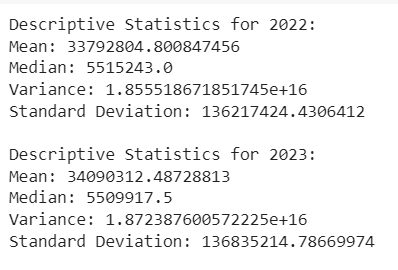
* Here we have tried to find out countries having growth rate more than 3%. This indicates the countries having significant changes in population of each country from 2022 to 2023.

It indicates rapid population growth.



* DESCRIPTIVE STATISTICS

The descriptive statistics provided for the data in 2022 and 2023 give valuable insights into the distribution and characteristics of the population data. Here's what each statistic means:



1. Mean (Average):

- The mean represents the average population size for all countries in the dataset for a given year.

- In 2022, the mean population size is approximately 33,792,805, while in 2023, it's approximately 34,090,312.

- Easily affected by outliers.

2.Median (Middle Value):

- The median is the middle value in the dataset when the data is sorted in ascending order.

- In 2022, the median population size is 5,515,243, and in 2023, it's 5,509,917.

- The median is less sensitive to extreme outliers compared to the mean and gives a measure of the "middle" of the data distribution.

3. Variance:

- Variance measures how much the data points deviate from the mean.

- In both 2022 and 2023, the variance is quite large, indicating a wide spread of population sizes among countries.

- A high variance suggests that the data points are dispersed across a wide range of values.

4. Standard Deviation:

- The standard deviation is a measure of the dispersion or spread of the data.

- In both 2022 and 2023, the standard deviation is quite large, similar to the variance.

- A large standard deviation indicates that the data points are spread out from the mean.

Key Insights:

- The mean population size increased slightly from 2022 to 2023, suggesting overall population growth.

- The median population size remained relatively stable between the two years.

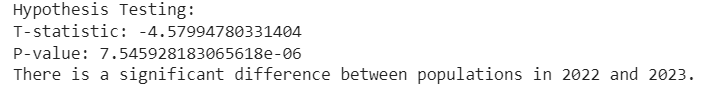
- The large variance and standard deviation indicate significant variability in population sizes among countries. Some countries have very large populations, while others have small populations.

- The data distribution appears to be right-skewed, with a few countries having exceptionally large populations, contributing to the high variance and standard deviation.

These descriptive statistics provide a summary of the data's central tendency, variability, and distribution characteristics. Further analysis and visualizations can help you explore and understand the population data more deeply.

* HYPOTHESIS TESTING

Now we wish to test a hypothesis to check if there is any significant change in the populations from 2022 to 2023.



* CORRELATION

A correlation coefficient of approximately 0.9999835249015268 indicates an extremely strong positive linear correlation between two variables. In the context of the data, this means that there is a very strong relationship between the population sizes of countries in the year 2022 and their population sizes in the year 2023.



Here's what it signifies:

1. Strength of Correlation: The correlation coefficient (r) ranges from -1 to 1, where:

- r = 1 indicates a perfect positive linear relationship.

- r = -1 indicates a perfect negative linear relationship.

- r = 0 indicates no linear relationship.

- In this case, a correlation coefficient close to 1 (0.9999835249015268) suggests a nearly perfect positive linear relationship.

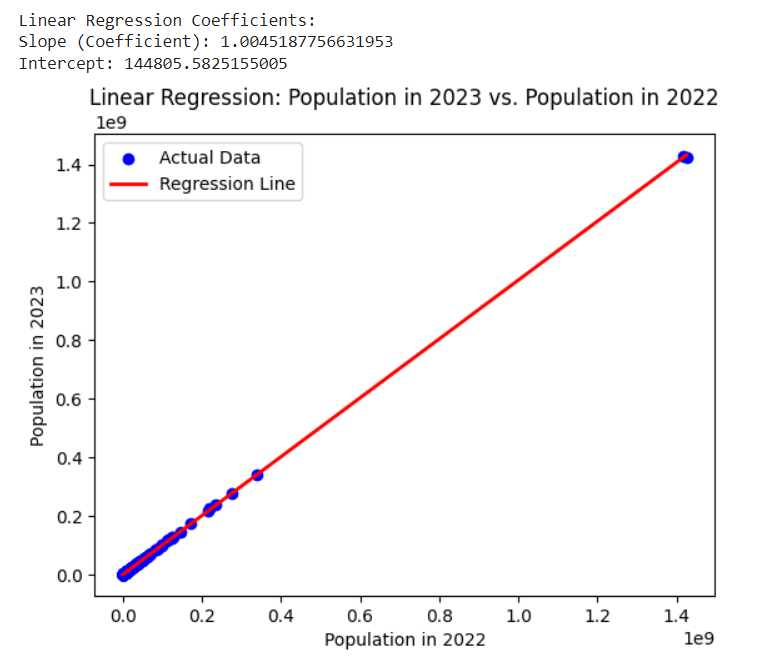
2. Positive Correlation: Since the correlation coefficient is positive, it implies that as the population size in 2022 increases for one country, the population size in 2023 tends to increase as well. In other words, countries with larger populations in 2022 are likely to have larger populations in 2023.

3. Strength of Predictability: The high correlation coefficient indicates that population sizes in 2022 are almost perfectly predictive of population sizes in 2023 for the countries in your dataset. In practical terms, if you know the population size of a country in 2022, you can make a very accurate prediction of its population size in 2023 based on this strong positive relationship.

In summary, a correlation coefficient of approximately 0.9999835249015268 indicates a nearly perfect positive linear relationship between the population sizes of countries in 2022 and 2023. This means that changes in population sizes from one year to the next are highly predictable based on the current population sizes.

* REGRESSION

The linear regression coefficients you've obtained indicate the relationship between the independent variable (population in 2022) and the dependent variable (population in 2023) in the dataset.



1. Slope (Coefficient): The slope represents the change in the dependent variable (population in 2023) for a one-unit change in the independent variable (population in 2022). The slope is approx.1.0045. This means that, on average, for every additional unit increase in population in 2022, the population in 2023 is expected to increase by approximately 1.0045 units.

In summary, the linear regression coefficients suggest that there is a positive relationship between the population sizes in 2022 and 2023. An increase in population in 2022 is associated with a slightly larger increase in population in 2023, as indicated by the slope coefficient of approximately 1.0045.