**PYTHON PROJECT**

**REPORT**

**INFORMATION ABOUT THE POPULATION DATA**

Load libraries

Load population data into the pandas dataframe, clean, explore, and transform the data.

Population data contains 2343 rows(entries) or observations and 6 columns, (2343, 6).

The datatypes are: id, country\_id, and year are integers (int64)

Population is float (float 64), and country name and continent are (object).

The descriptive statistics of population data

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **count** | **mean** | **std** | **min** | **25%** | **50%** | **75%** | **max** |
| id | 2343 | 1172.000000 | 676.510163 | 1.0 | 586.5 | 1172.0 | 1757.5 | 2343.0 |
| population | 2335.0 | 30.041542 | 120.340090 | 0.0 | 0.0 | 5.0 | 20.0 | 1330.0 |
| year | 2343.0 | 2005.000000 | 3.162953 | 2000.0 | 2002.0 | 2005.0 | 2008.0 | 2010.0 |
| country\_id | 2343.0 | 107.840376 | 61.733298 | 1.0 | 55.0 | 108.0 | 161.0 | 214.0 |

Checking for missing values in population data, the result indicates that there are 8 missing values in population column. The id with the missing population data are (667, 668, 669, 670, 671, 2289, 2290, 2291).

Checking the number of continents in population data, there are 6 continents in population data namely; (North America, South America, Europe, Asia, Africa, and Oceania).

Checking the number of countries in population data, there are 213 countries in population data using value counts to get the frequency of the unique countries. Each country occurred 11 times in the country name column.

QUESTIONS

* How many countries had no population data (0) for 2000?

154 countries had no population data (0) for 2000

* What was the total population of Africa in 2010?

The total population of Africa in 2010 is 56

* What is the average population of countries in South America in 2000?

24.5

* In 2007, which countries had a population greater than 1000?

China and India

* How much has the population of Europe grown from 2000 to 2010?

The population growth in Europe from 2000 t0 2010 is (-8).

There was decrease in population in Europe from 2000 t0 2010, during this decade there were rise and fall of population.

(2000 = 714

2001 = 713

2002 = 714

2003 = 715

2004 = 716

2005 = 715

2006 = 708

2007 = 708

2008 = 706

2009 = 708

2010 = 706)

The decline in population in Europe from 2000 to 2010 may be attributed to several factors, including demographic, economic, environmental, and societal causes.

1. **Low Birth Rates**: A decline in fertility rates, influenced by delayed marriages and childbirth, as well as the increased use of contraception.
2. **High Death Rates**: Increased mortality leading to a higher number of deaths.
3. **Emigration**: A rise in the number of people moving out of the country to seek economic opportunities elsewhere.
4. **Environmental Factors**: Climate change impacts or natural disasters forcing migration and contributing to population decline.

However, there was increase in the overall population across the year.

(2000 = 6004.0

2001 = 6078.0

2002 = 6150.0

2003 = 6228.0

2004 = 6303.0

2005 = 6380.0

2006 = 6448.0

2007 = 6521.0

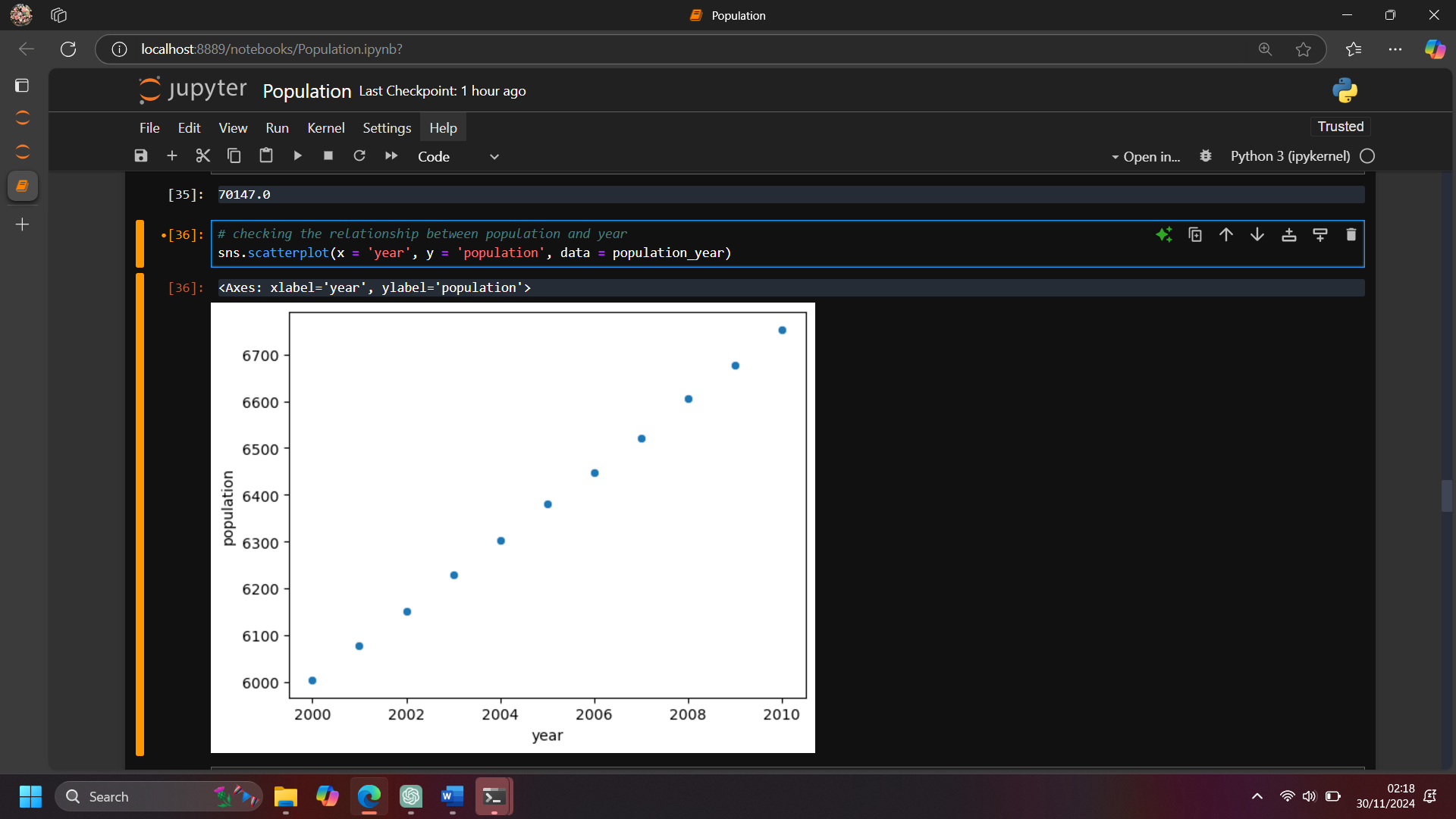
2008 = 6606.0

2009 = 6676.0

2010 = 6753.0)

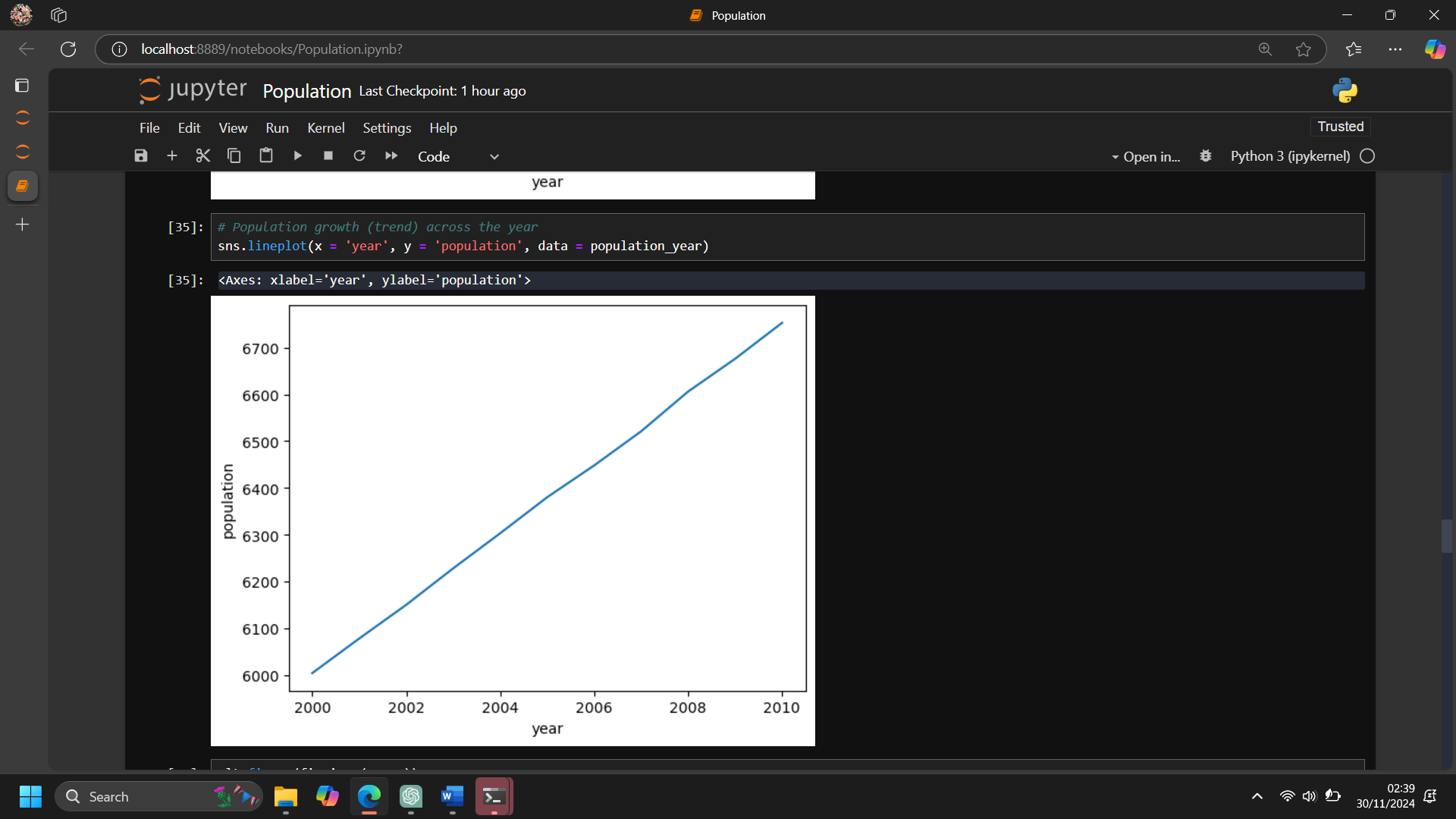
This brings the overall summation of population to (70147.0)

Checking the relationship between population and year



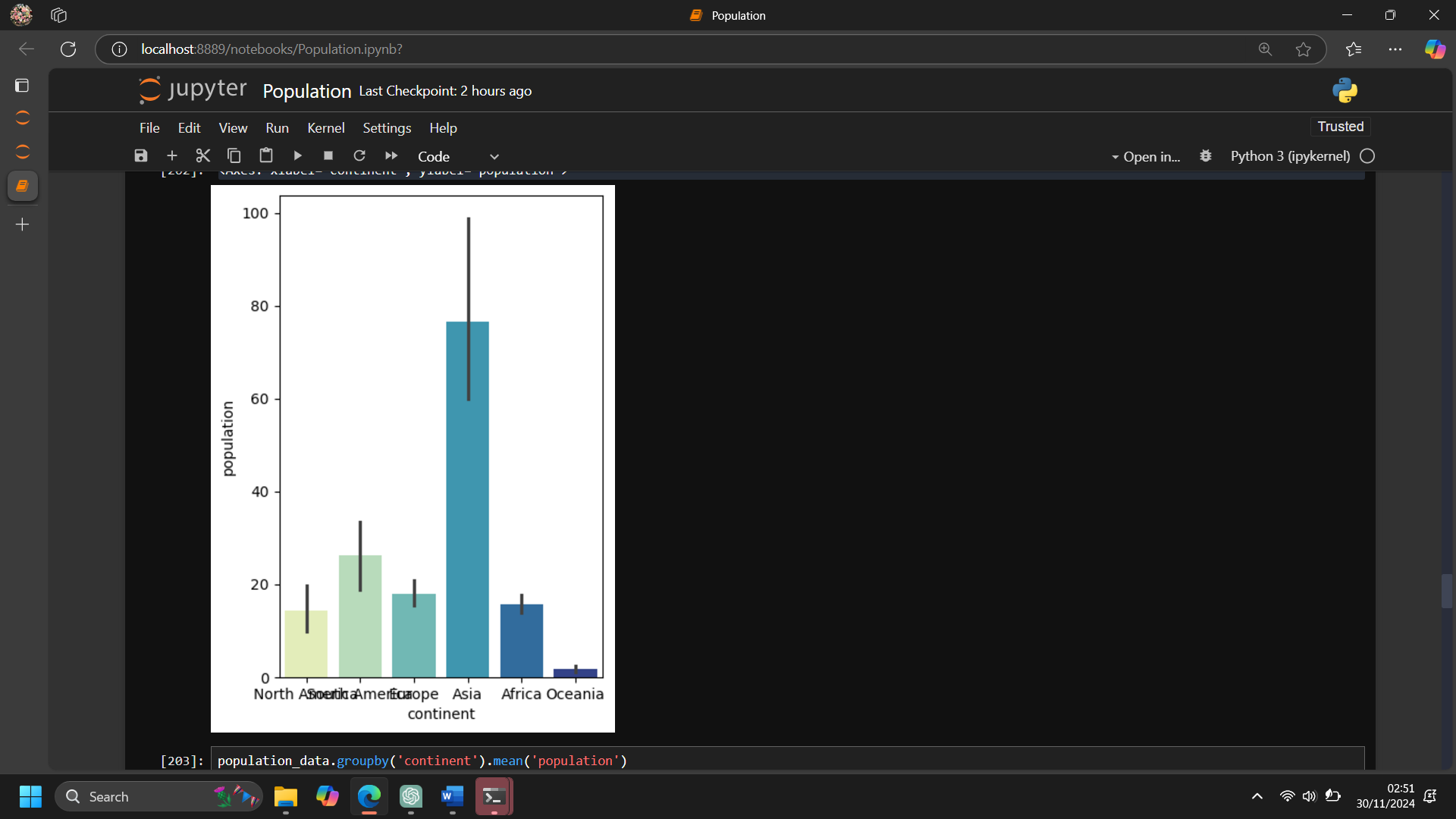
**Figure 1.0 scatterplot**

Figure 1.0 indicates a strong positive linear relationship between population and year. As the year increases, the population increases as well. This is a linear graph, and it demonstrates a consistent increase in population over the years.



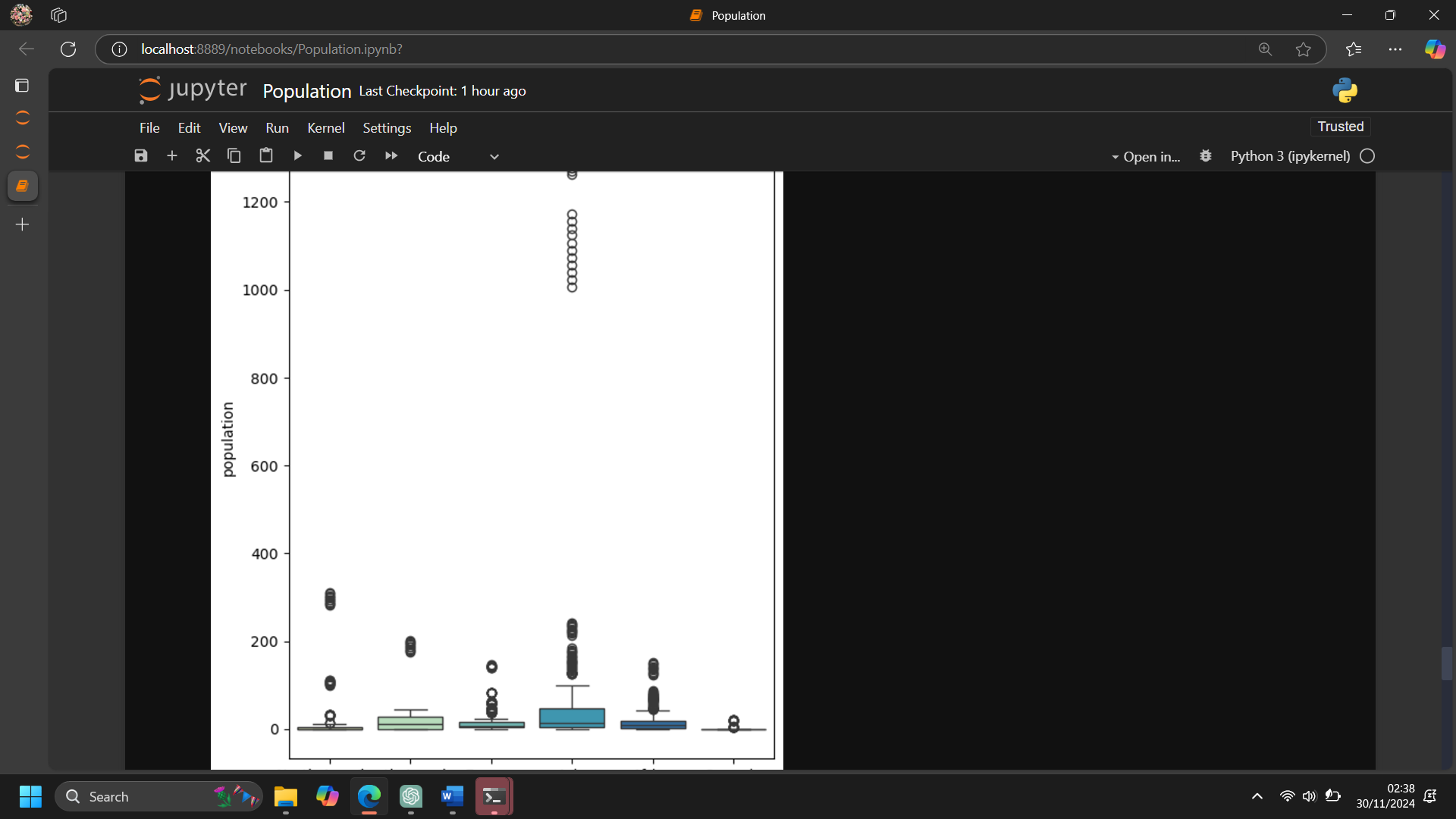
**Figure 2.0 line plot**

Figure 2.0 shows a growing trend in population across the year. The graph corroborates the increase in population from (2000 to 2010). The graph clearly demonstrates a steady increase in population, particularly from (2000 to 2010), highlighting significant demographic growth during this period. This trend underscores the continued expansion of the population over time.



**Figure 3.0 bar plot**

Figure 3.0 presents a bar plot displaying the population data for each continent. Asia stands out with a population of (76.7), which is exceptionally high and nearly triple that of South America (26.1). Following Asia, Europe accounts for (18.0), Africa for (15.71), North America for (14.36), and Oceania, with the lowest population, at (1.67). This highlights the stark differences in population distribution across continents.



**Figure 4:0 boxplot**

Figure 4.0 depicts a boxplot of population data across continents. A boxplot is a useful tool for identifying outliers in the data. In this case, Asia exhibits an outlier due to its exceptionally high population, which is prominently visible in the plot. This emphasizes the significant disparity in population distribution compared to other continents.

**Conclusion:**

1. **Population Growth Trends**: The data reveals a clear upward trend in global population over the years, particularly during the period from 2000 to 2010, which signifies rapid demographic growth in certain regions.
2. **Uneven Distribution**:

* Asia has an overwhelmingly large population compared to other continents, which also contributes to a significant portion of the global population.
* South America, Africa, and North America show moderate populations, while Oceania has the smallest population.

1. **Presence of Outliers**: The boxplot analysis identified outliers, particularly in Asia, reflecting extreme population values in specific countries or regions within the continent.
2. **Global Disparities**: The stark differences in population data across continents highlight disparities in population density, resource distribution, and potential challenges related to urbanization and sustainability.

**Recommendations:**

1. **Focus on Population Management**:

* Regions with high populations, such as Asia, should prioritize policies aimed at managing population growth through improved education, healthcare, and family planning programs.
* Countries with rapidly growing populations should ensure adequate infrastructure, housing, and resource management to maintain sustainable growth.

1. **Address Regional Disparities**:

* Governments and organizations should explore ways to address uneven population distribution by promoting regional development in underpopulated areas like Oceania.

1. **Monitor Outliers**:

* Conduct further analysis on countries or regions within Asia contributing to the outlier values. This can provide insights into specific factors driving population surges and help design targeted interventions.

1. **Advice for Europe in Addressing population decline**

Governments should introduce or enhance policies that support families, such as paid parental leave, childcare subsidies, and flexible work arrangements to encourage couples to have more children.

Provide financial incentives, such as tax benefits or housing support for families with children, to reduce the financial burden associated with raising children.

Offer subsidies for education and healthcare to make family life more affordable.

Encourage gender equality in the workforce by supporting women in pursuing careers and balancing work with family life. Policies that enable women to participate fully in the workforce while raising children can positively impact birth rate