Lab 01: Project Coronavirus Analysis

Data Visualization

Group 5

2023-03-10

Contents

1 General Introduction

In this project, we collect Covid-19 data on Worldometer to analyze and visualize the properties of the data set to understand the situation of the Covid-19 epidemic and its influence in the world. The duration of the pandemic outbreak through the relationships of the attribute sets.
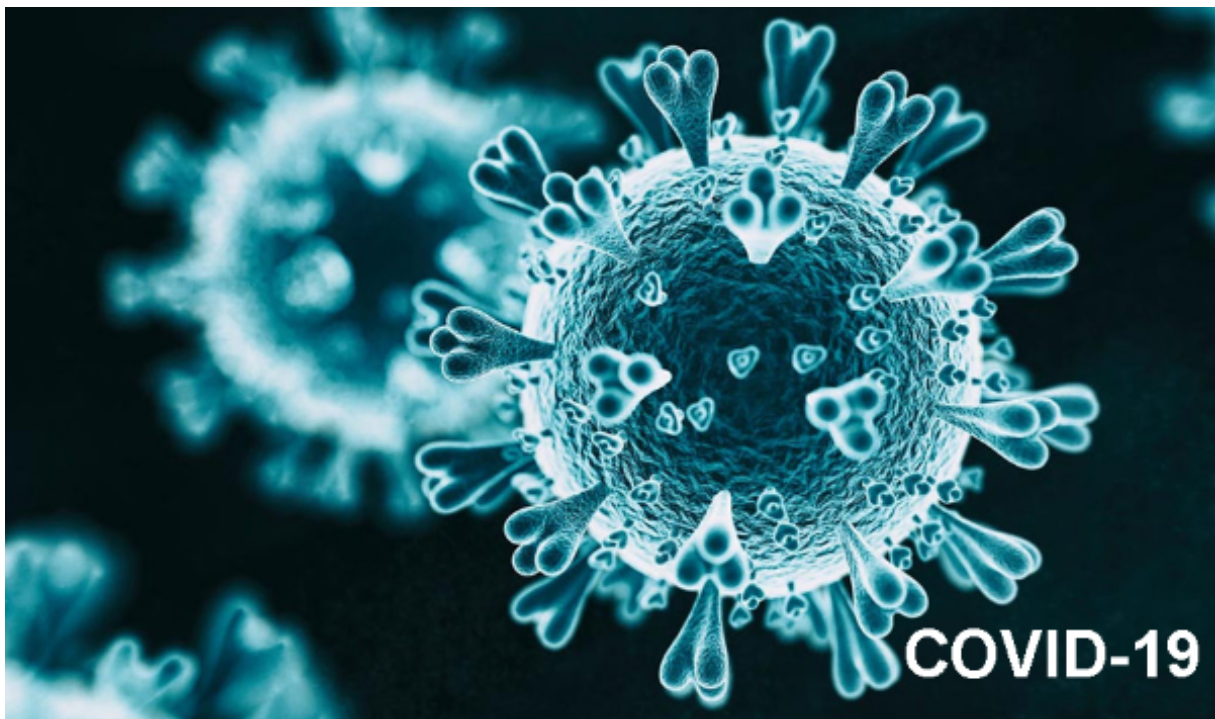
Figure 1.1: Covid - 19

2 Member Information

MSSV	Full name	Tasks	Rate
20127082	Lam Thi Xuan Thy	Data preprocessing, write report, linear regression	100%
20127500	Dao Ngoc Hoa	Data preprocessing, data analysis the relationship attributes	100%
20127166	Nguyen Huy Hoan	Data preprocessing, data analysis the relationship attributes	100%
20127546	Vo Thanh Lam	Crawl data and data preprocessing	100%
20127402	Bui Thanh Lam	Crawl data and data preprocessing	100%

3 Collecting data

See more information in Crawl.pdf below. We have Crawl.pdf and Crawl.ipynb to test and check the result below. Please choose the link to know more

Crawl.pdf

4 Data preprocessing

Step 1: Convert object to numeric datatypes

Convert object to numeric datatypes

```python
columns = ['TotalCases', 'NewCases', 'TotalDeaths',\
        'TotalRecovered', 'NewRecovered', 'ActiveCases',\
        'Serious,Critical', 'Tot\xa0Cases/1M pop', 'Deaths/1M pop', 'TotalTests',\
        'Population', '1 Caseevery X ppl',\
        '1 Deathevery X ppl', 'New Cases/1M pop', 'Active Cases/1M pop']
```

Step 2: Convert Date to datatime type

Convert `Date` to datatime type

```python
: big_df['Date'] = pd.to_datetime(big_df['Date']).dt.date
```

Step 3: Drop nan for attributes countries

Countries

```python
]: countries_df = big_df.dropna(subset=['#'])
   countries_df.head()
```
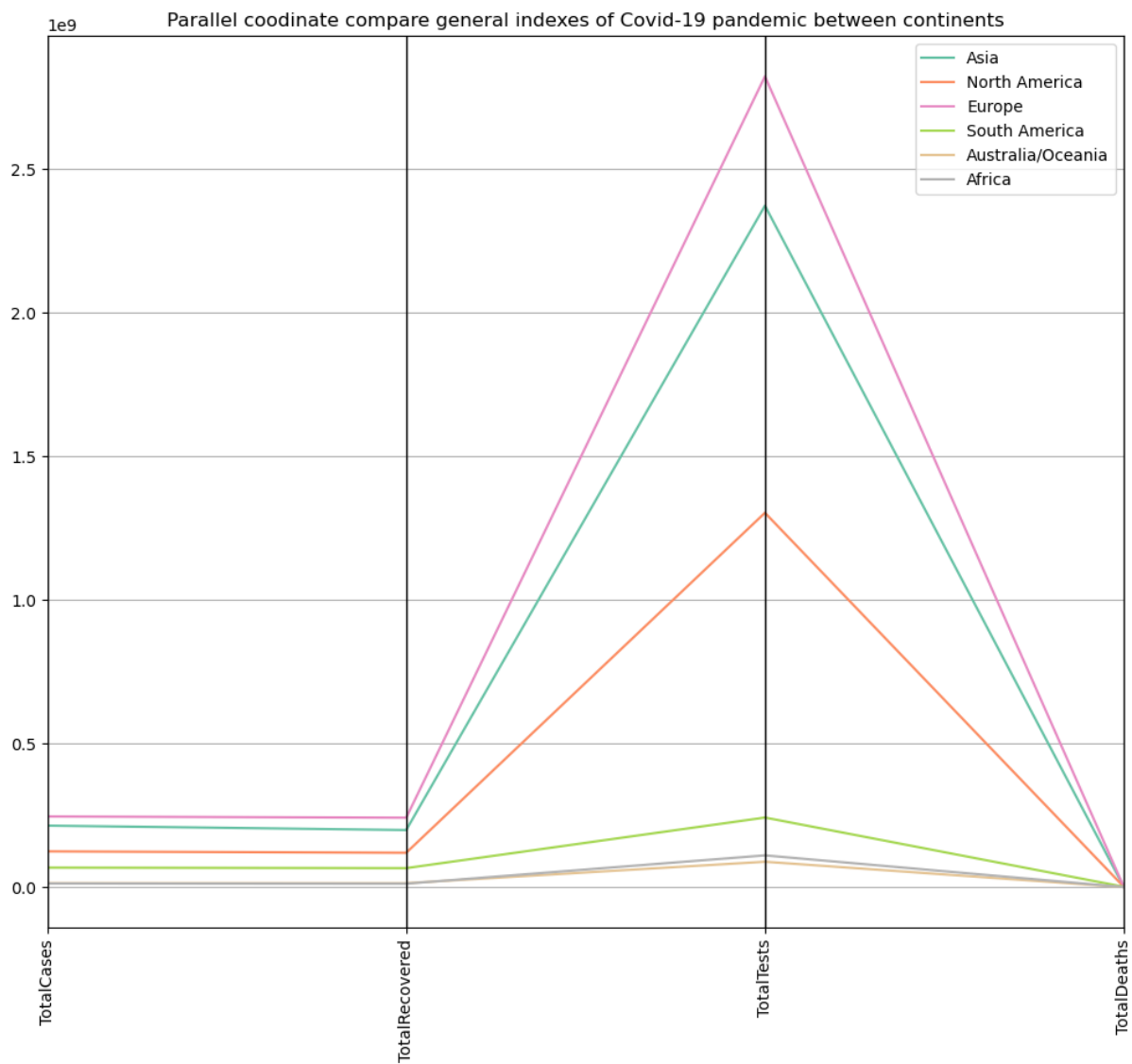```
]:
```

### Step 4: Fill nan for attributes new cases

```python
countries_df['NewCases']=countries_df['NewCases'].fillna(0)
```

5 Data relationship

5.1 General visualization about the state of continents when Covid-19 outbreak

Using parallel to compare the indexes of Covid-19 pandemic among continents
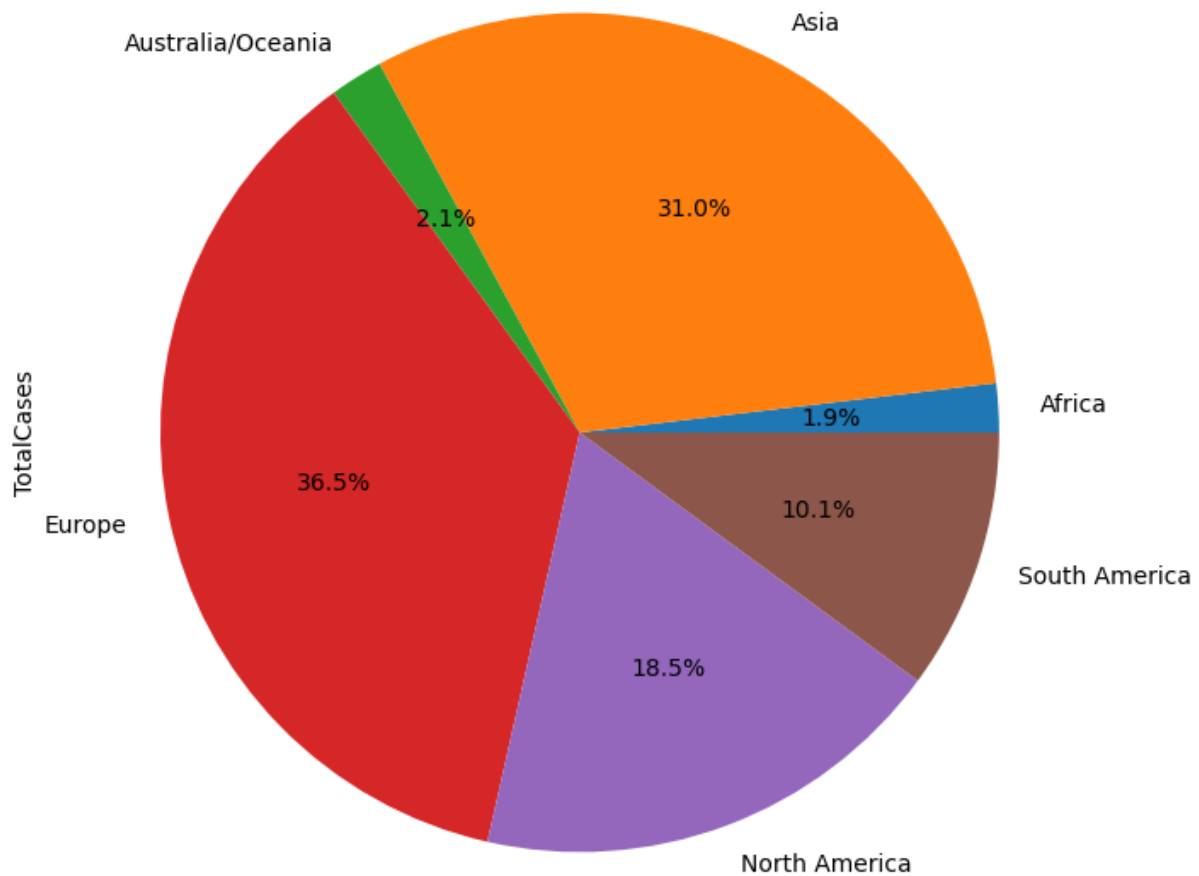
Conclusions

- It is easy to see that Europe have the highest value in all figures, whereas the second place is the Asia. It seems that Europe and Asia have experienced more serious situation than others.

- What stands out from the diagram is that the total deaths was significantly smaller than the total cases, while the number of residents who have tested was considerably larger than that of being infected.

- The total recovered was approximately equal to the total cases.

5.2 What is the relationship between Total Cases and Continents?

*How is the breakdown of total cases among continents ?*

The chart illustrate the breakdown of total cases among continents

Conclusions

- Most of Covid cases are in Europe, which made up for over a third of totalcases worldwide, while the Asia just take over 31.0%.

- Africa had the smallest proportion, just 1.9%, compared to 2.1% of Australia/Oceania.

5.3 What is the relationship between Population and Continents
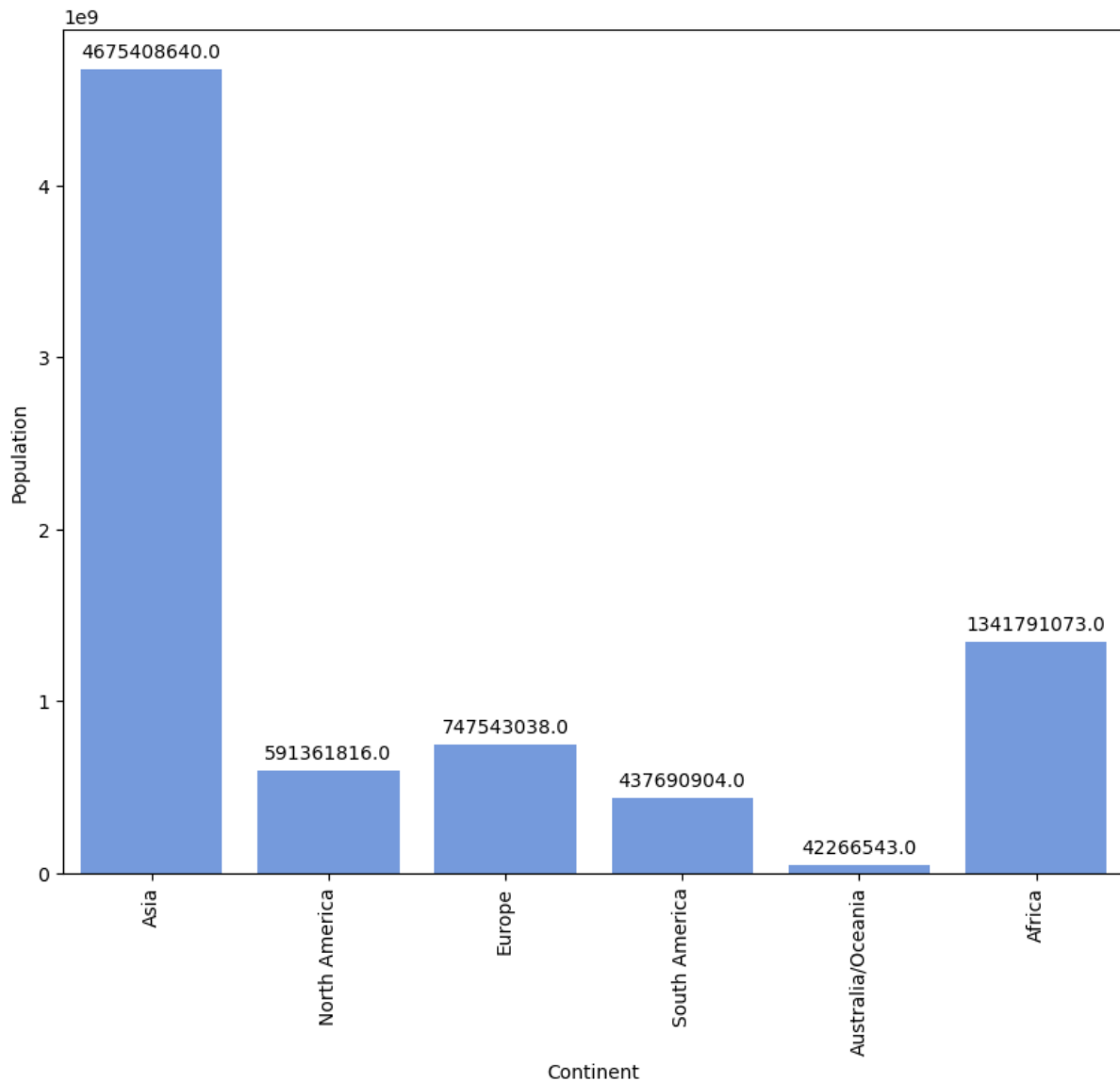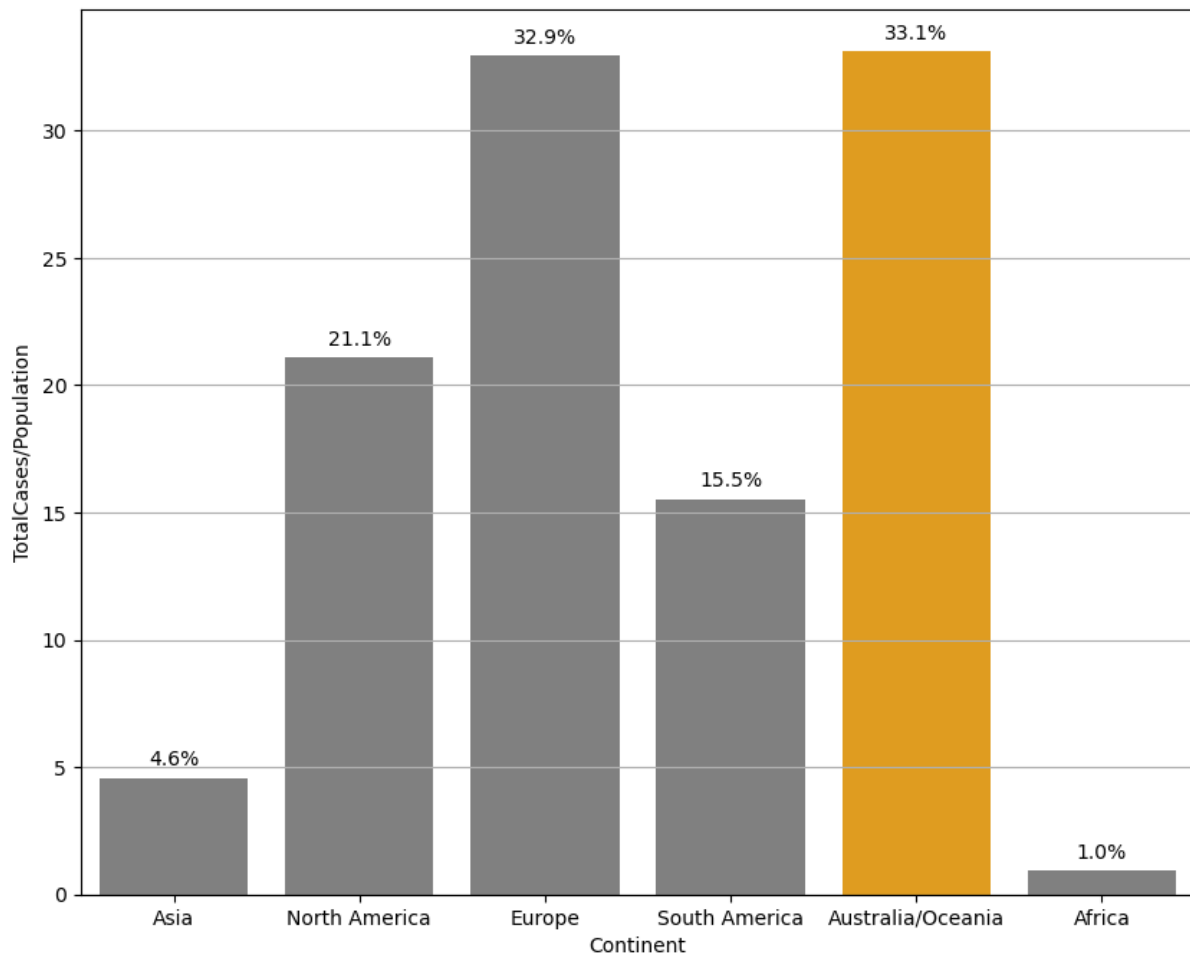

Conclusions

- Although Europe made up the largest proportion of Covid-19 cases, at 36%, all of us know that the most densely populated continent is Asia.

- This means that Europe underwent more severe pandemic than Asia. So now the question is

whether Europe was the continents having the highest covid-19 infection rate?
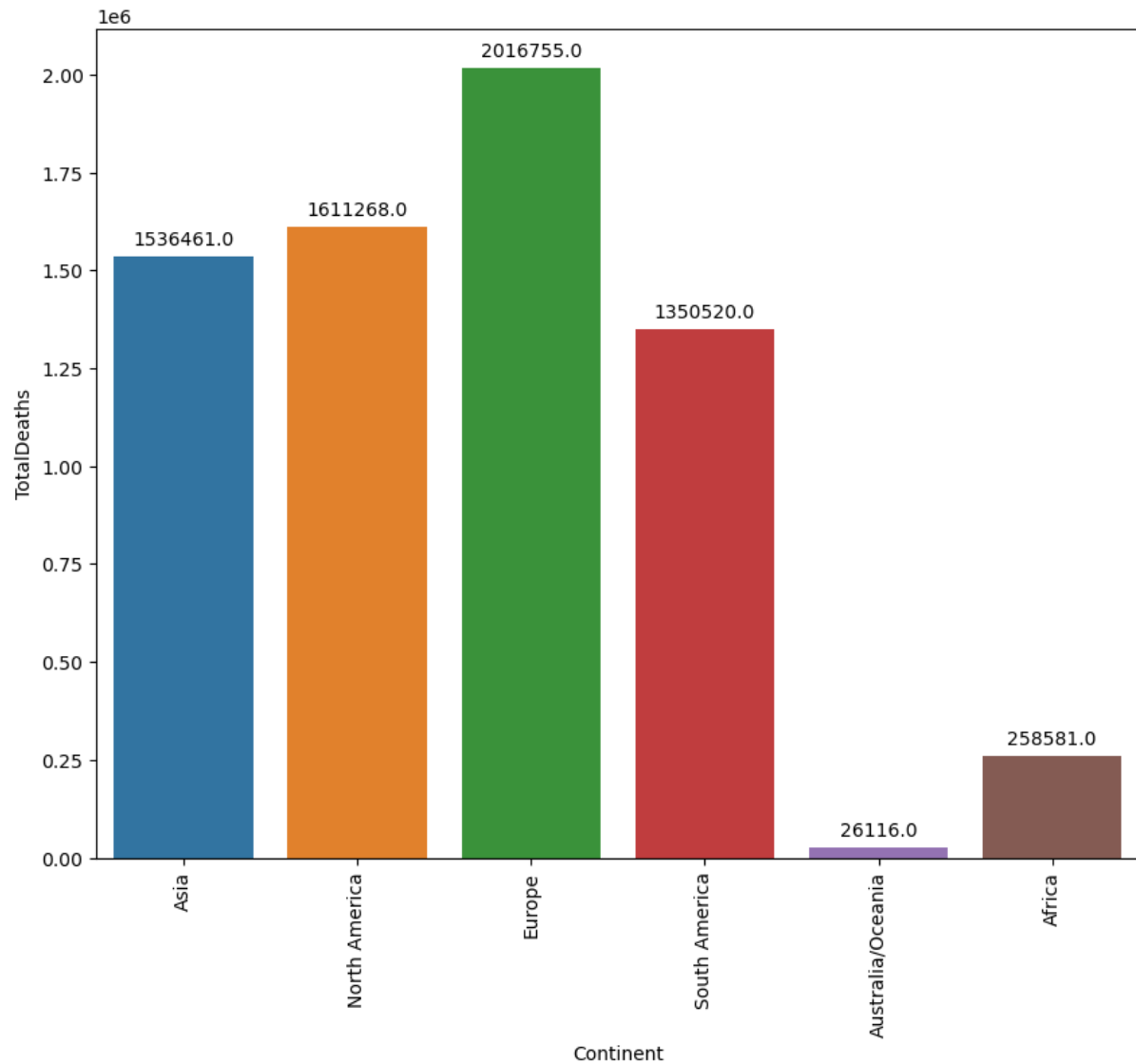
5.4 The relationship between Continent and Total Cases/Population

Conclusions

- The Australia/Oceania is the highest - 33.1%.

- The Africa is the smallest about the rate.

5.5 The relationship between Continents and Total Deaths
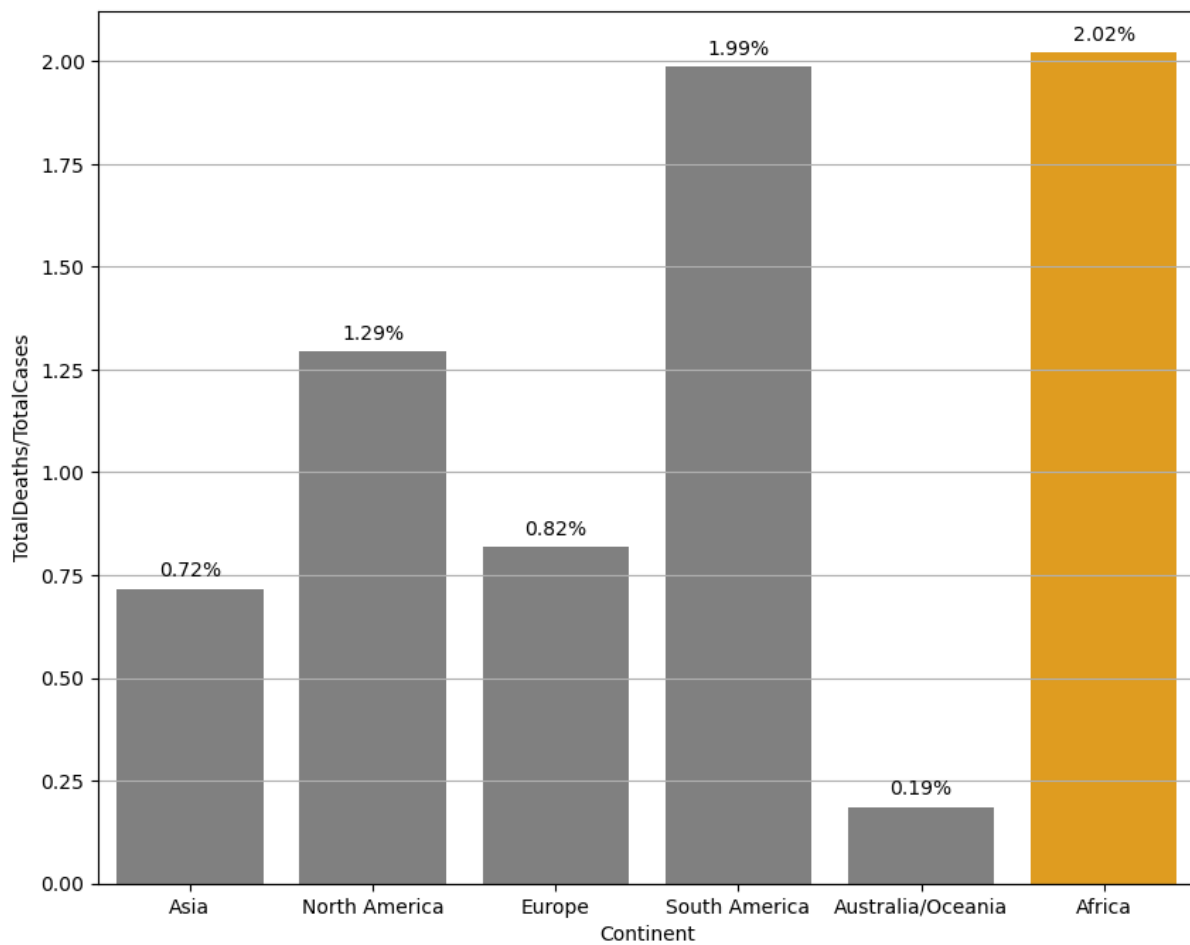
Conclusions

- It is interesting that although Europe seems having the worst situation, its deadth rate was just over a third of Africa, which was the highest deadth rate. South America was modestly smaller, 1.99% compared to 2.02%.

- Inspite of having the highest infection rate, Australia had the smallest deadth rate in our planet.

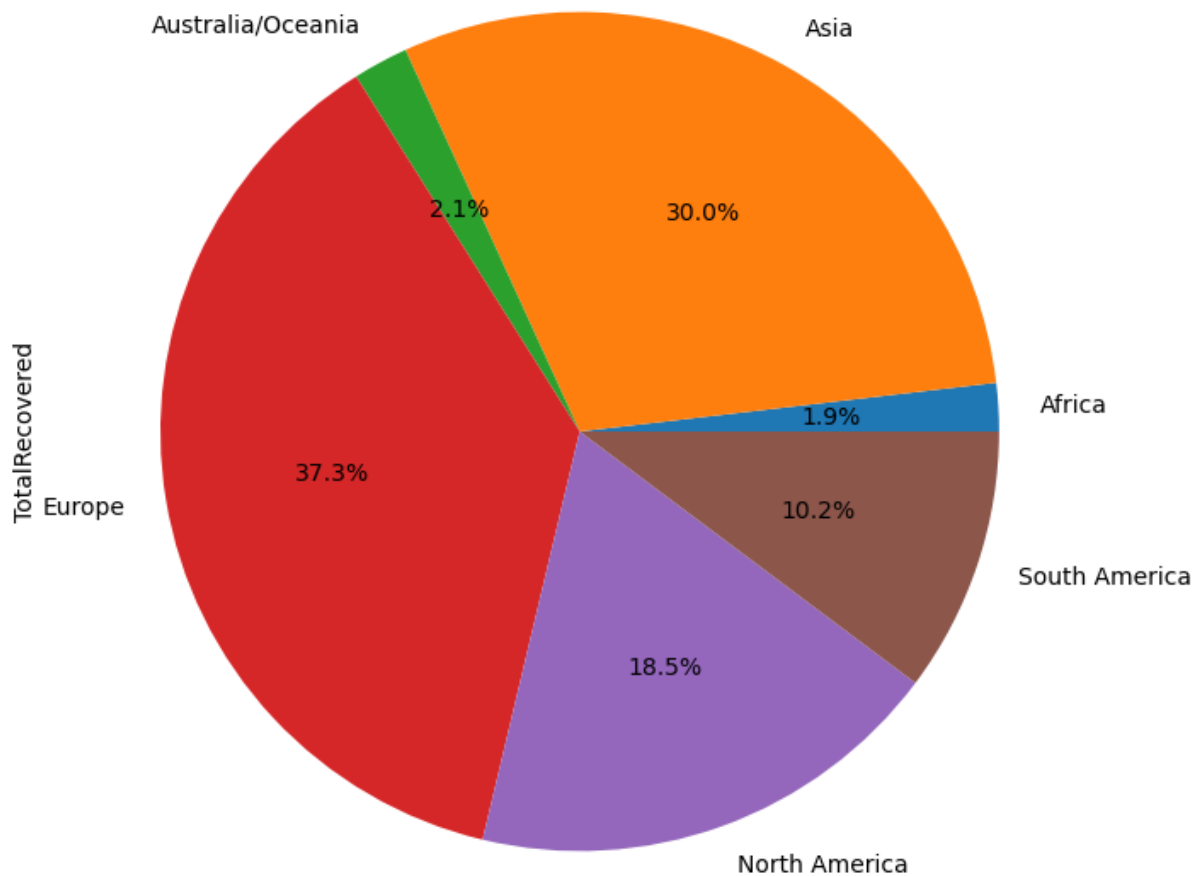## 5.6 The relationship between Total Recovered and North America

The chart illustrate the breakdown of total cases among continents

****Conclusions**

- From the pie chart, Europe was continually the continent which accounted for the largest breakdown of recovered patients, and Asia standed at the second place.

5.7 The relationship between Total Recovered/Total Cases and Continents
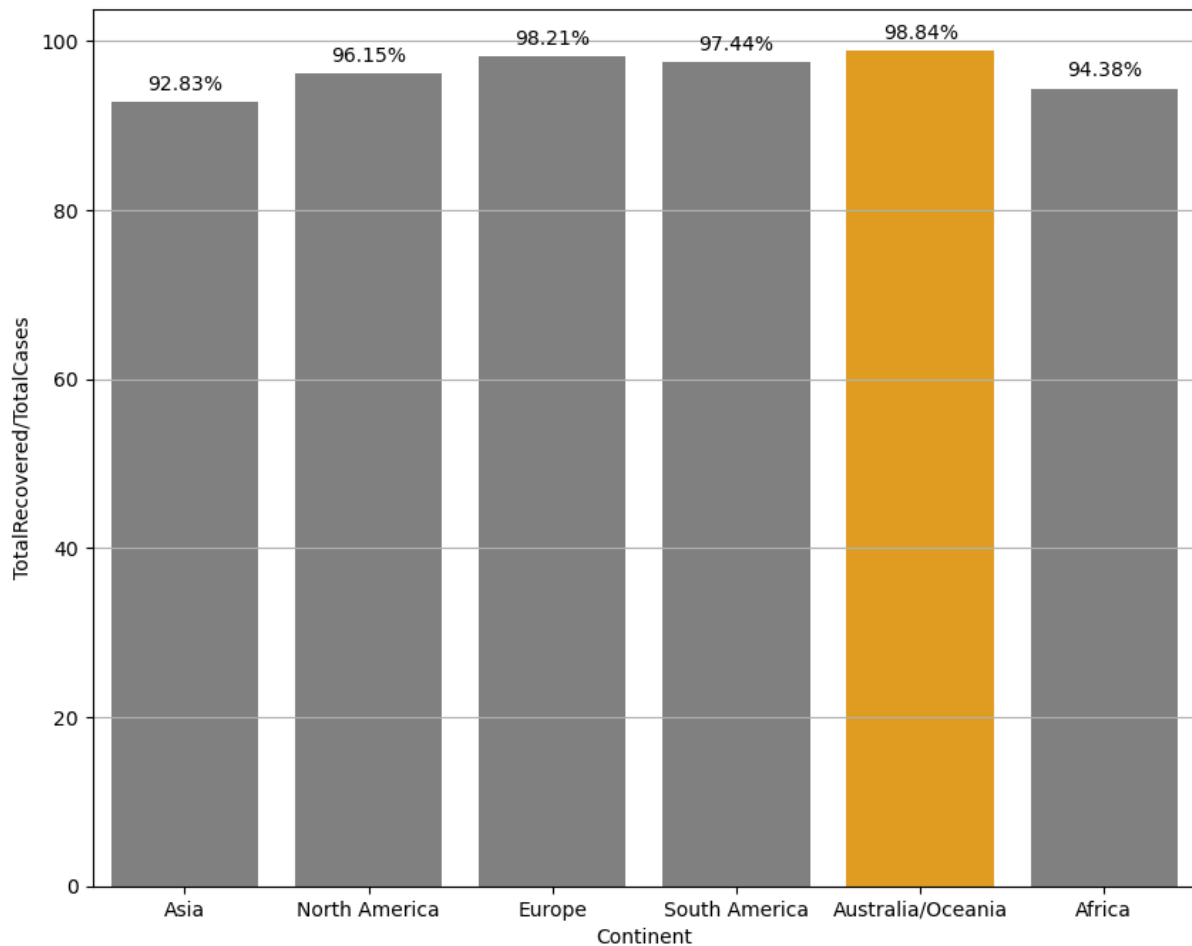
Conclusions

- Although death rate and infection rate of Asia was better than most other continents, its recovered rate was not so impressive, just 92.83%. While Australia was the best again, at 98.84%, compare to 98.21% of Europe.

5.8 The relationship between Total Cases/Total Tests and Continets
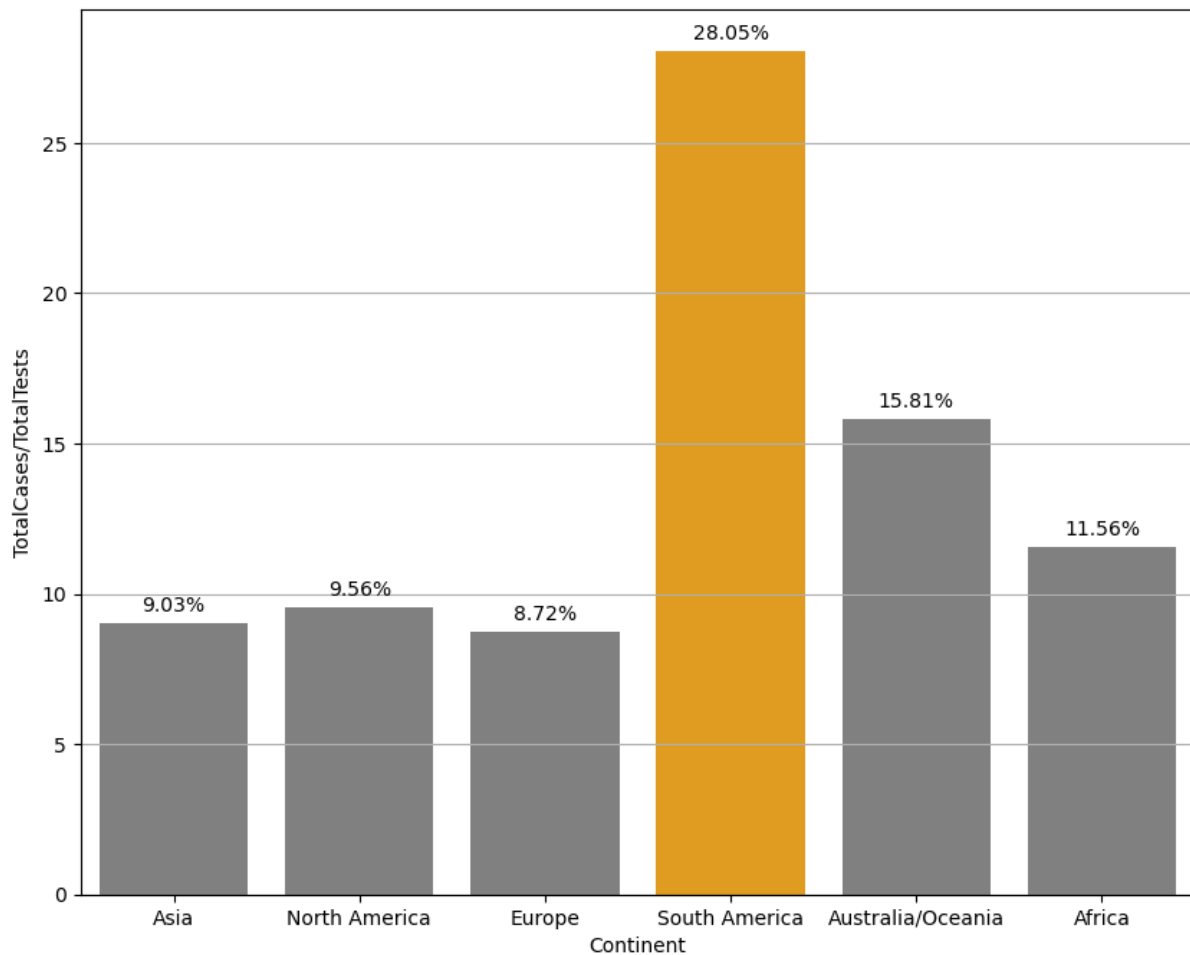
Conclusions

*It is interesting that the continent having the foremost detection rate was South America at 28.05%. Europe had the smallest propotion, at just 8.72%, which may be the reason why this continent experienced the worst Covid-19 outbreak.

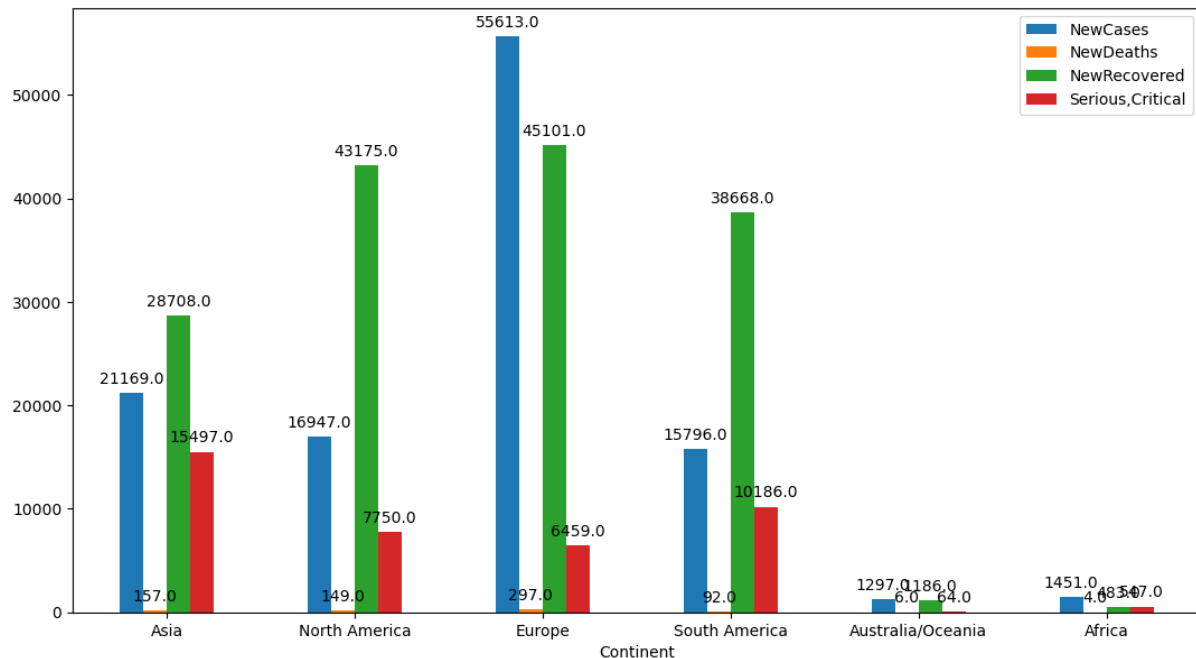## 5.9 The relationship between NewCases,NewDeaths,NewRecovered,Serious,Critica
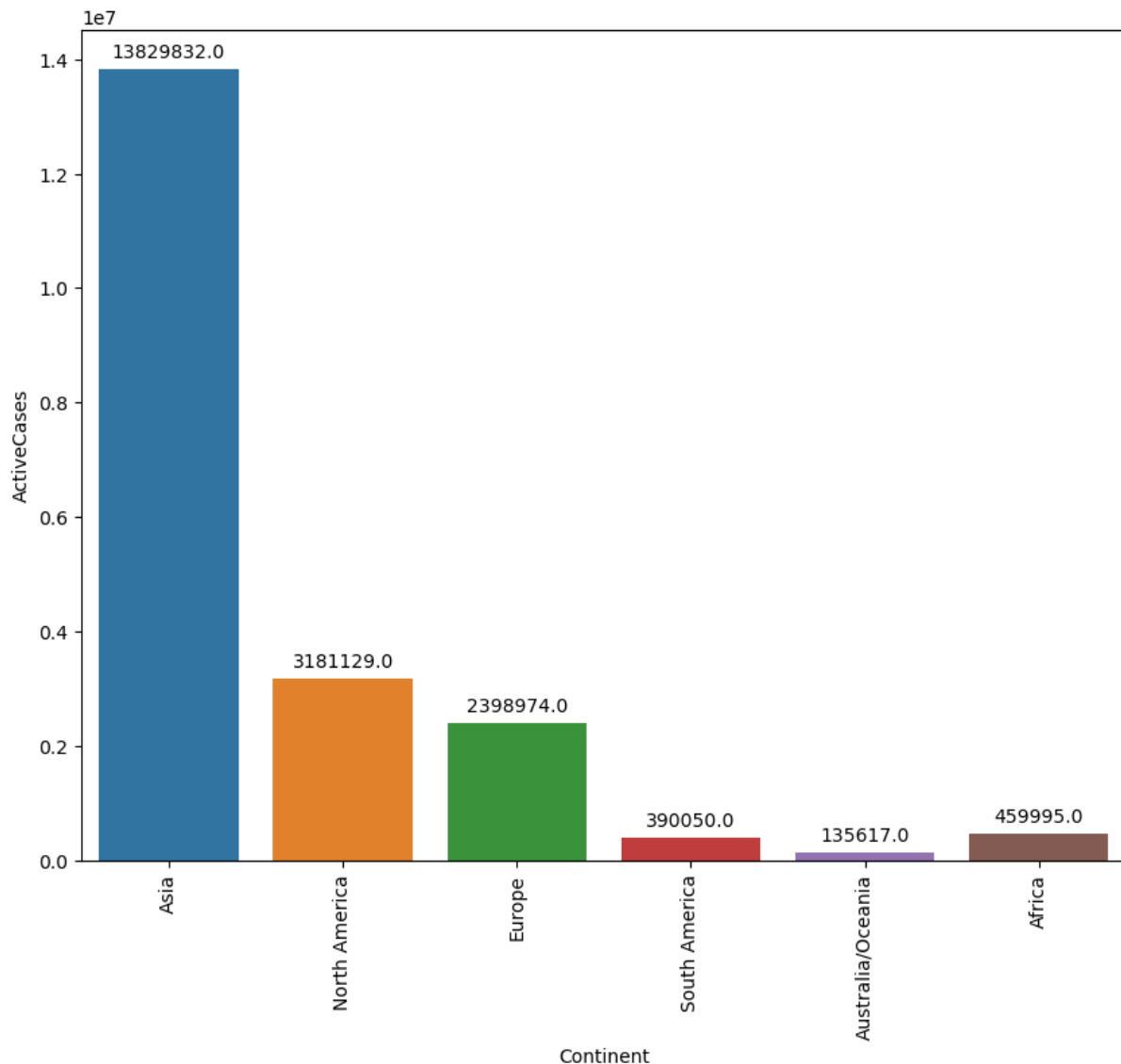

Conclusions

- The number of new infections across continents is still high, with over 50,000 new cases in **Europe** being possibly the most alarming. Africa is the continent with the second-lowest number of new infections, with fewer than 1,500 cases, followed by **Oceania**. Overall, there are still steady, ongoing epidemics with new cases in the remaining nations.

- Along with the number of new infections, there are also a lot of reinfections, with **Europe** and **Africa** having the greatest and lowest rates respectively. The other continents still have high numbers, especially Europe at the top and **North America**, which despite having an average number of new cases, still has over 16,000 cases of reinfection, demonstrating the complacency in post-Covid epidemic prevention and control still has many limitations. Continents like **Africa** and **Oceania**, which have relatively good numbers at this level, are the exceptions.

- The lowest and most steady number of severe cases continues to be in **Oceania**, followed by **Africa**. With almost 15,000 cases of critically ill patients, Asia leads the way, followed by **South America**, **North America**, and **Europe**. A greater population, a higher percentage of elderly people who may be more susceptible to the disease, and other factors may be to blame for the higher number of seriously ill cases than the continent's underdeveloped medical care system.

- The death toll on continents is also still there, but the pace has drastically dropped and stabilized when computed globally. With 297 fatalities, **Europe** has the greatest rate, while **Africa** has the

lowest.

5.10 The relationship between Continent and Activate
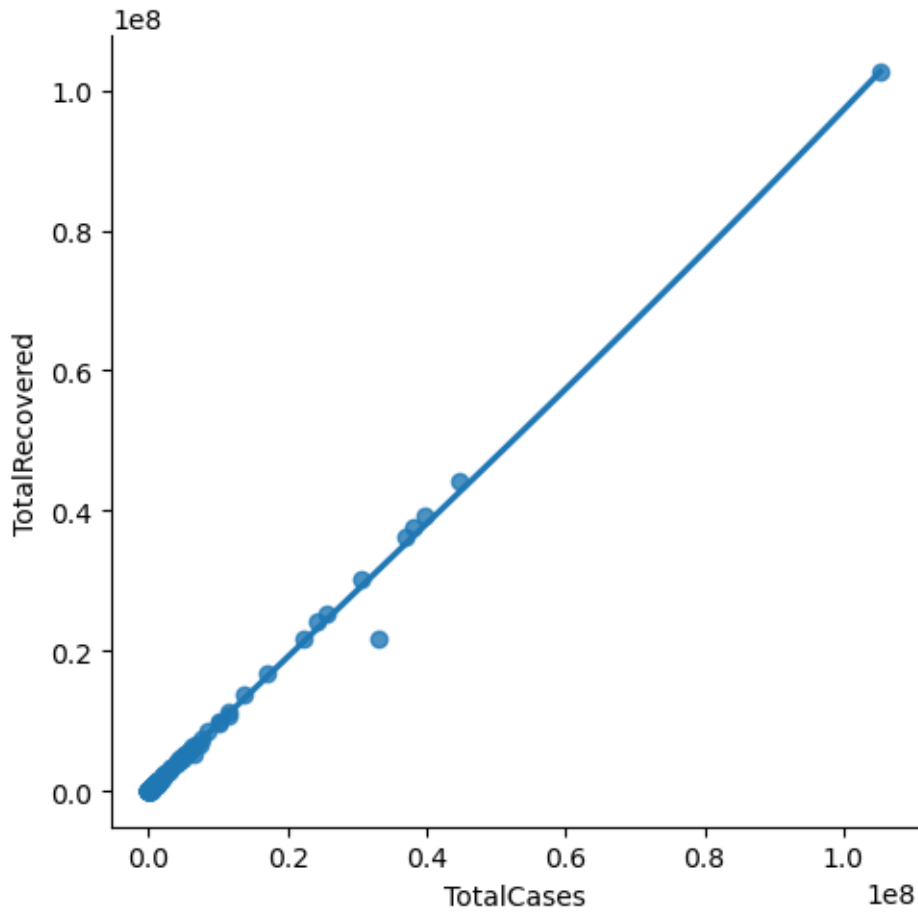

Conclusions

Comment of Active Cases Looking at the data, we can make the following observations and evaluations:

- **Asia** is currently facing the most serious situation, with the highest number of cases compared to other continents, reaching 13.8 million infections. This may be related to the fact that it has a

large population and many countries have underdeveloped healthcare systems to control the pandemic.

• **Europe** is third in terms of the number of infections, but is still one of the regions most affected by the pandemic, with many countries reporting high numbers of infections and deaths.

• **North America** is second in terms of the number of infections, with over 3 million cases. However, compared to Asia, the number of cases in North America is still much lower. Many countries and states in this region have implemented measures to reduce transmission and control the pandemic.

• **South America** reports a relatively low number of infections compared to other continents, with 390,050 cases. However, there are still many challenges in controlling the pandemic, especially in Brazil - the country facing a serious pandemic situation with the highest number of cases in the region.

• **Australia** and **Africa** report the lowest number of cases compared to other continents, with 135,617 and 459,995 cases respectively. However, it is important to note that controlling the pandemic in these regions also depends on many factors such as government policies, the average age of the population, healthcare infrastructure, and measures to prevent and control the spread of the virus.

5.11   The relationship between Totalcases and TotalRecovered below linear regress
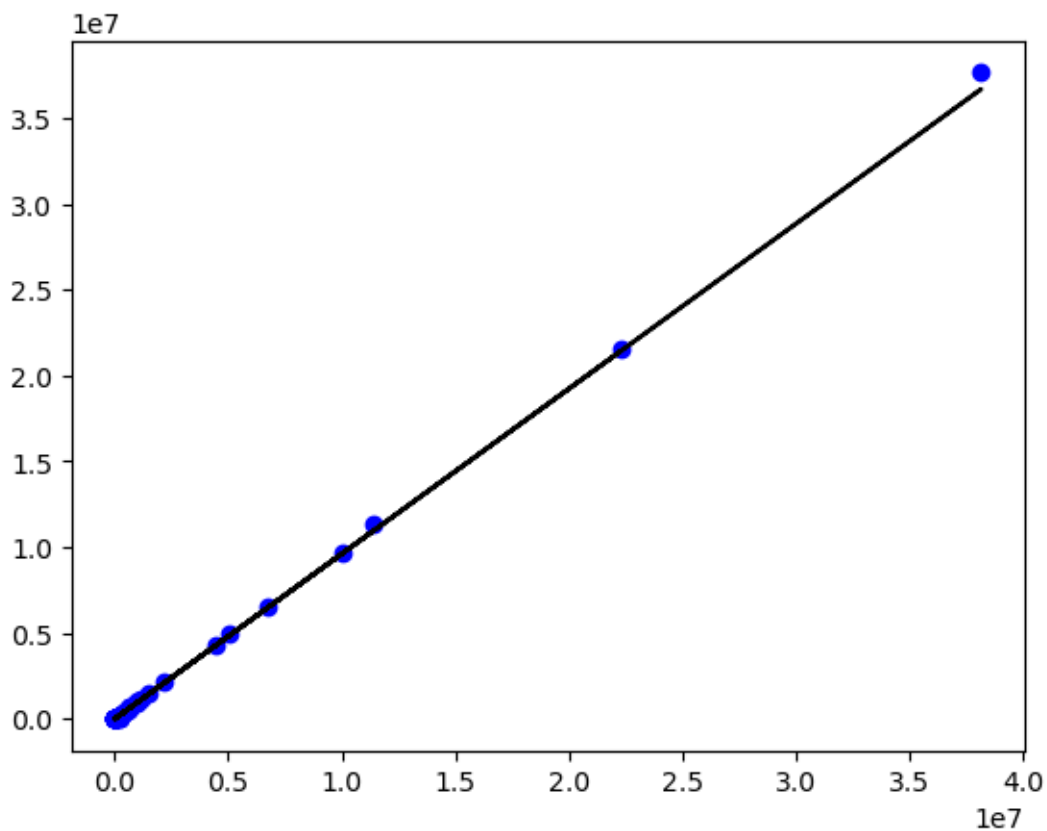
Conclusion

- Total Cases and Total Recovered has the linear regression

- In conclusion, the COVID-19 pandemic has affected each continent differently due to a range of factors. While some continents have experienced significant outbreaks with high infection and mortality rates, others have been able to contain the spread of the virus more effectively. The differences in government responses, healthcare systems, and access to vaccines have played a crucial role in shaping the impact of the pandemic on different continents. To effectively tackle the global pandemic, it is important to acknowledge and address these differences, and work together to ensure that resources and support are distributed equitably across all continents.

6 Group-self evaluation

- In this project we have a bit of trouble finding the right graphs . However we got over it.

- Machine learning tools and knowledge are lacking and need improvement

7 References

Link 1

1. Linear Regression

Link 2

2. Chart

All link about data visualization on Geeks.