**COVID-19 Tableau Analysis**

**INFO 5709 – SUMMER 2023**

**DATA VISUALIZATION AND COMMUNICATION**



**PROFESSOR:** Dr. Gahangir Hossain

**Teaching Assistant**: Yaseen Quadri Ameen Syed, Shafi Parvez Mohammed

**By**

Teja Vamshidhar Reddy Chilukala

[TejaVamshidharreddyChilukala@my.unt.edu](mailto:TejaVamshidharreddyChilukala@my.unt.edu)

**Student id:** 11641617

**Introduction:**

The covid-19 which was emerged in 2019, has created a great difficulty for our time with its impact on every country. This has affected not only people’s life, but also country’s economy. By using this analysis, people can explore and visualize various dimensions of this pandemic. Using the tableau visualizations, it will be very easy to make people understand by transforming the complex data into a visual representation that can be interpreted by wide range of people. As a citizen, it will be very crucial to get a deeper understanding on this covid-19 impact to take prior decisions and effective strategies to tackle this crisis and overcome its effects.

**Background:**

COVID-19 is a highly infectious respiratory infection that, particularly in elderly persons and those with underlying illnesses, can cause severe symptoms. When an infected person speaks, coughs, or sneezes, the virus is spread by respiratory droplets using contaminated surfaces as a surface to touch.

Lockdowns, travel bans, and social isolation have all been used in an effort to stem the virus's spread. solitude and wearing a mask. Furthermore, COVID-19 vaccinations have been created and are being distributed disseminated internationally to safeguard people and communities.

A disproportionately large number of people have died as a result of the epidemic in several nations and areas others have mostly escaped with few instances and fatalities. a number of variables, demography, healthcare systems, governmental regulations, and cultural and social customs may be a factor in these discrepancies.

**Dataset:**

This is a very huge dataset, which gives lot of detailed information of this outbreak in every continent and every country present in that continent. In this dataset there are columns like total deaths, total cases, total vaccinated, total tests done in every country with also people who died with cardiovascular disease during this pandemic.

This dataset is downloaded from Kaggle: <https://www.kaggle.com/datasets/georgesaavedra/covid19-dataset>

Following are few important columns present in the dataset:

1. **Continent:** Where Country is located
2. **Location:** Country name.
3. **Date:** The date of the recording entered.
4. **Total Cases:** Shows the total cases in that location.
5. **New Cases:** Cases newly added in that location.
6. **Total Deaths:** Total of people deaths in every location.
7. **New Deaths:** Newly deaths of people in every location.
8. **Total Recovered:** Number of people recovered after treatment.
9. **Total Vaccinations:** Total vaccinations done in every location in each continent.
10. **Population:** The total population of each country.
11. **Total Boosters:** Total number of boosters done in every country.
12. **Cardiovascular Death:** People who died with cardiovascular problem while treatment.
13. **Diabetes Prevalence:** People who got diabetes.
14. **Positive Rate:** Rate of covid positive for a number of tests completed.
15. **GDP Per Capita:** GDP of every country during the crisis.
16. **People Fully Vaccinated:** People who completed all doses of vaccination.

And also, there are other columns like people are over 65 years and 70 years. Male and female smokers, excess mortality, population index, life expectancy and reproduction rate.

**Tools Used:**

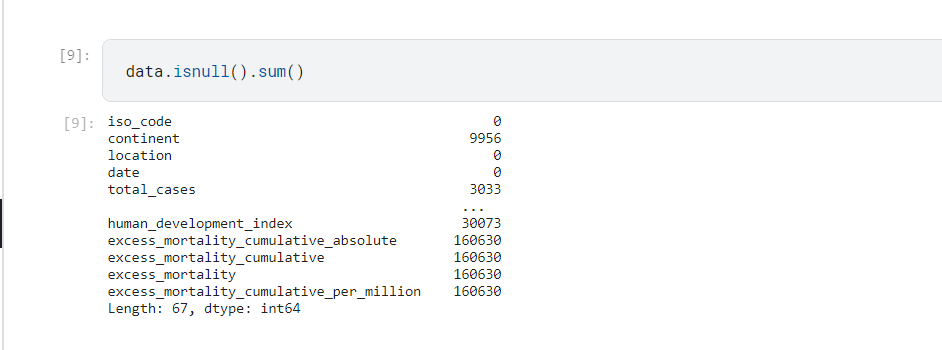
**Python**

**Tableau**

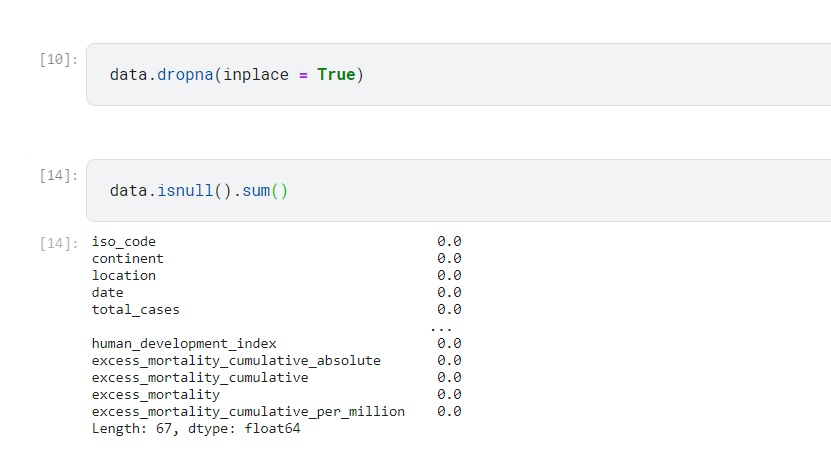
**Exploratory Data Analysis (EDA):**

Before starting the analysis, as this dataset has lot of columns it is always better to perform exploratory data analysis for getting best output. This helps to find new patterns and also the missing values in the dataset. If there are any missing values, they are removed from the dataset to maintain consistency. For doing this analysis, Python has been used.

**Checking for Null Values:**

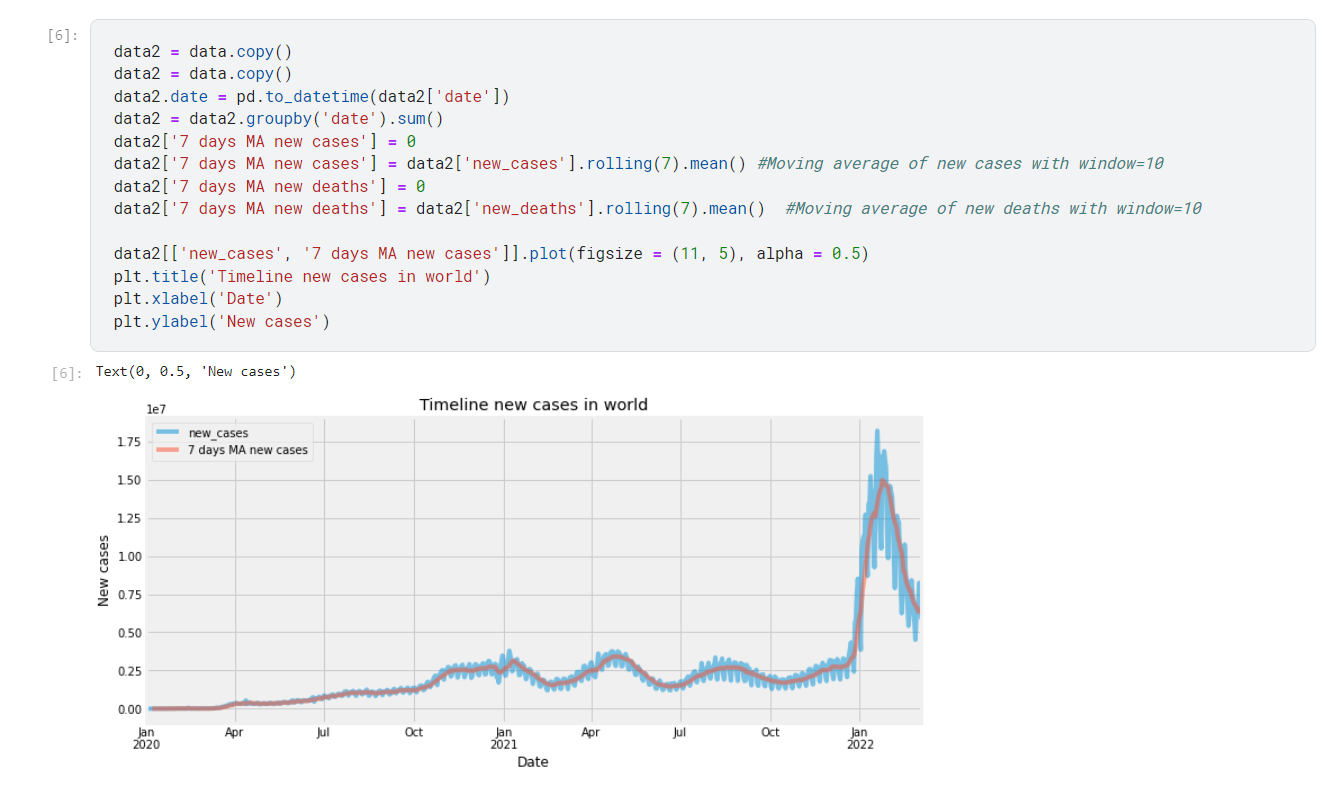
****

**Clearing the Null Values:**

****

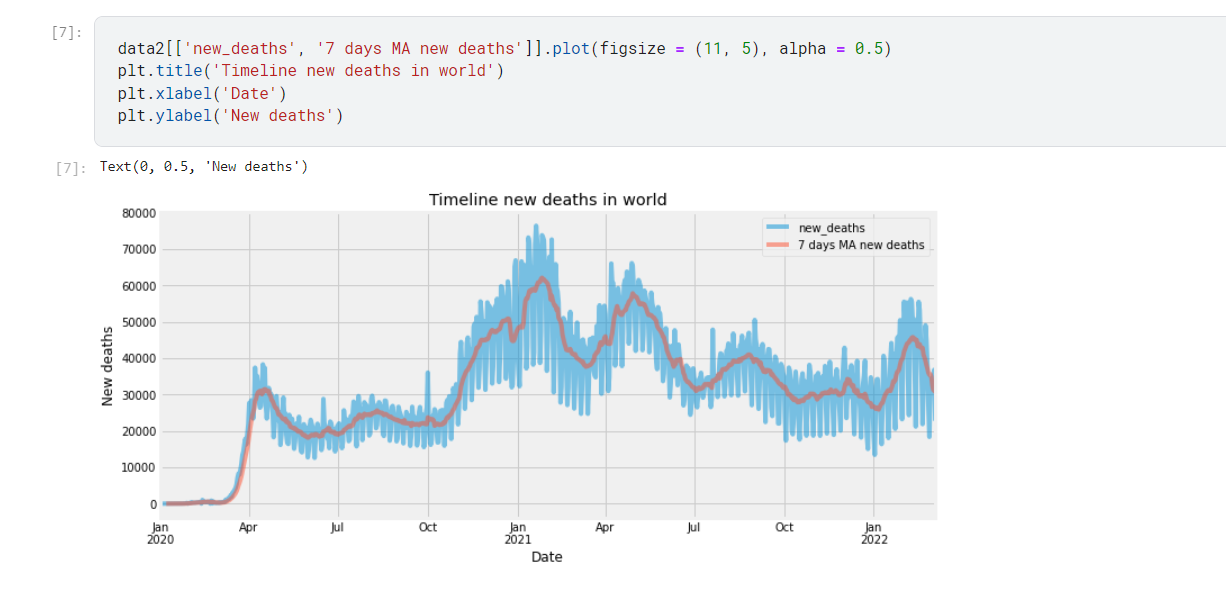
**Visualizations in Python Code:**

**Timeline series of new cases added across the world:**

****

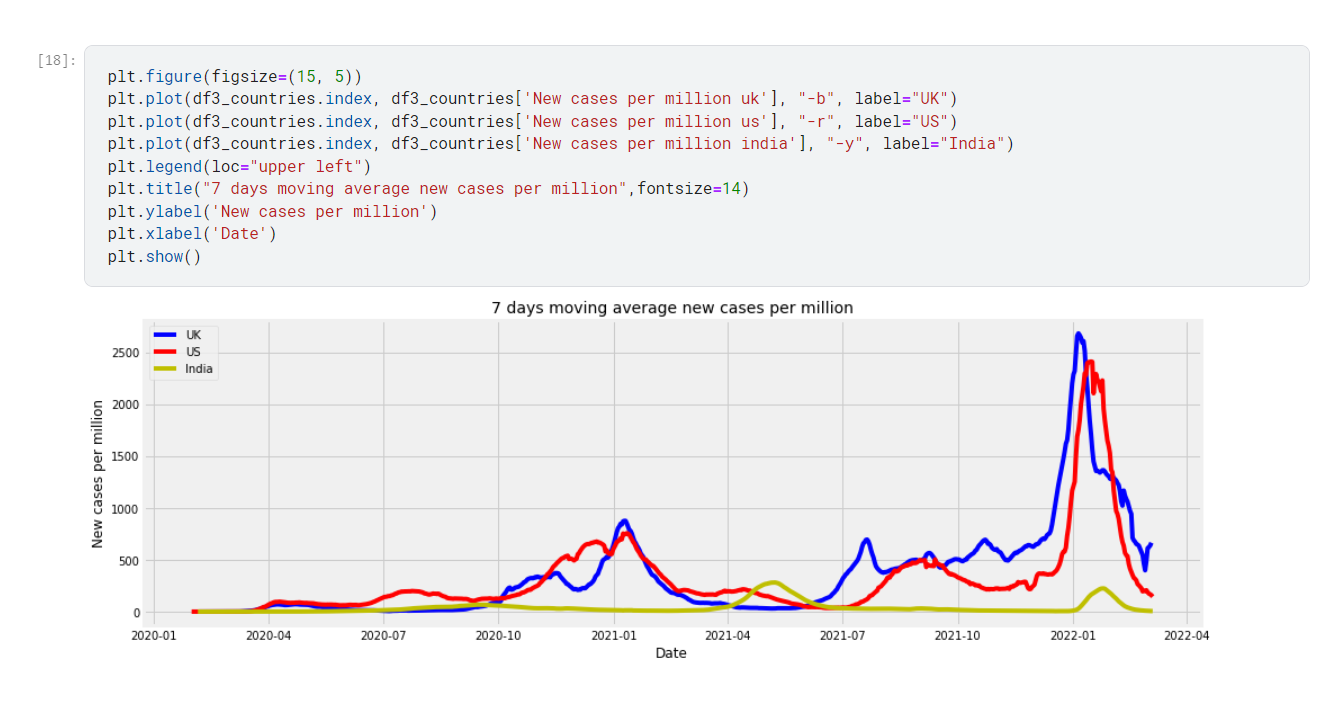
From the above time series graph, it shows there’s increase of new cases from January 2021 and in January 2022 has the highest cases added.

**Timeline series of new deaths across the world:**

****

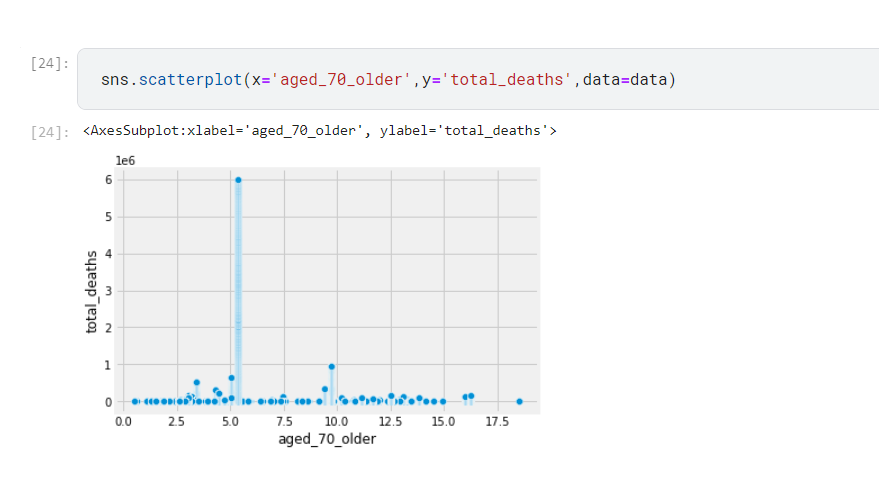
The above graph shows the timeline of new deaths in the world. It shows that from October 2020 the deaths were increased more and January 2021 has got highest deaths.

**Comparison of new cases per million people in US, UK and India:**

****

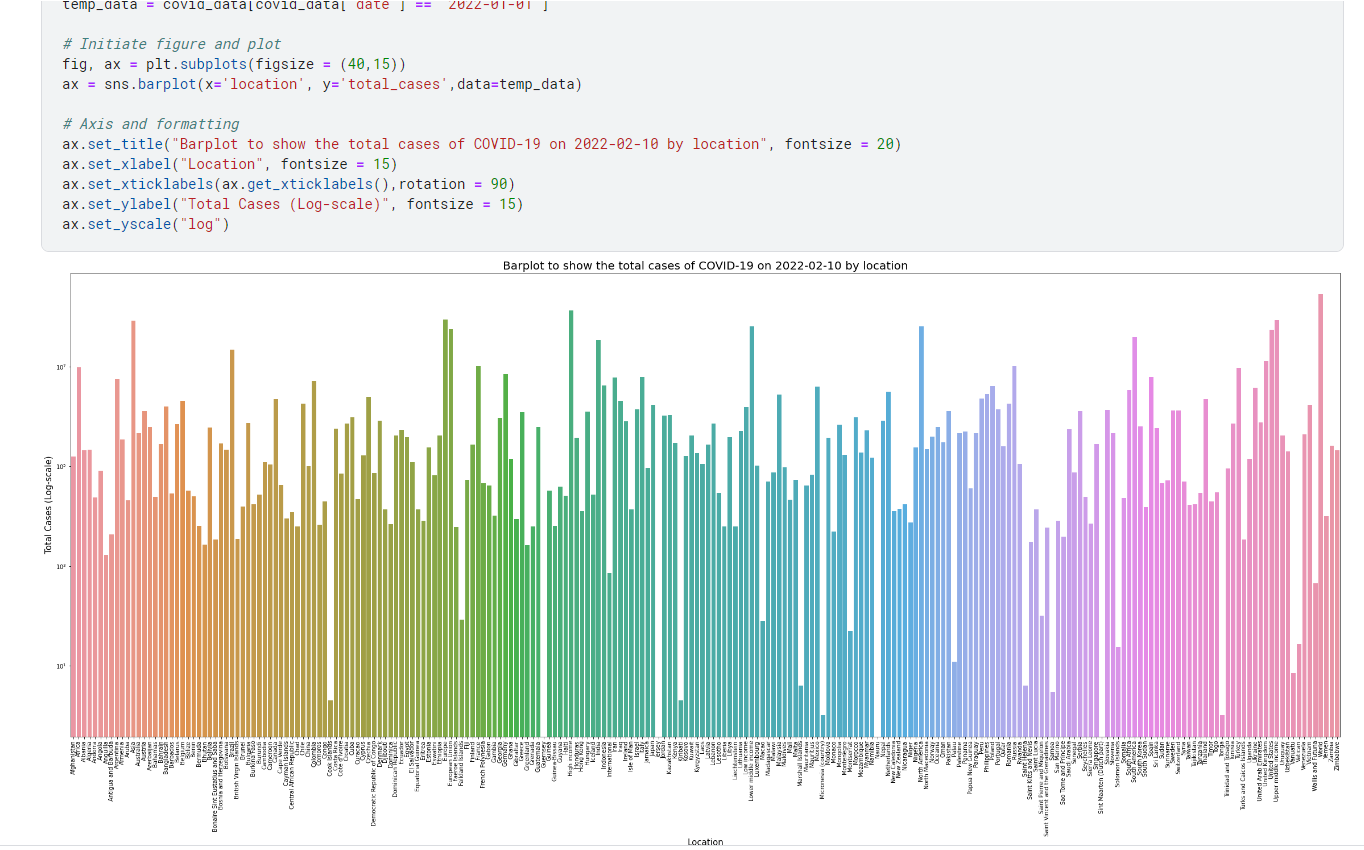
The above graph shows the new cases per million people who got affected by US, UK and India. In January 2022, UK and US have the highest new cases per million population which is over 2500 cases. Where India has got less than 500 cases for every million population.

**Total Deaths of People who are of age 70 or older:**

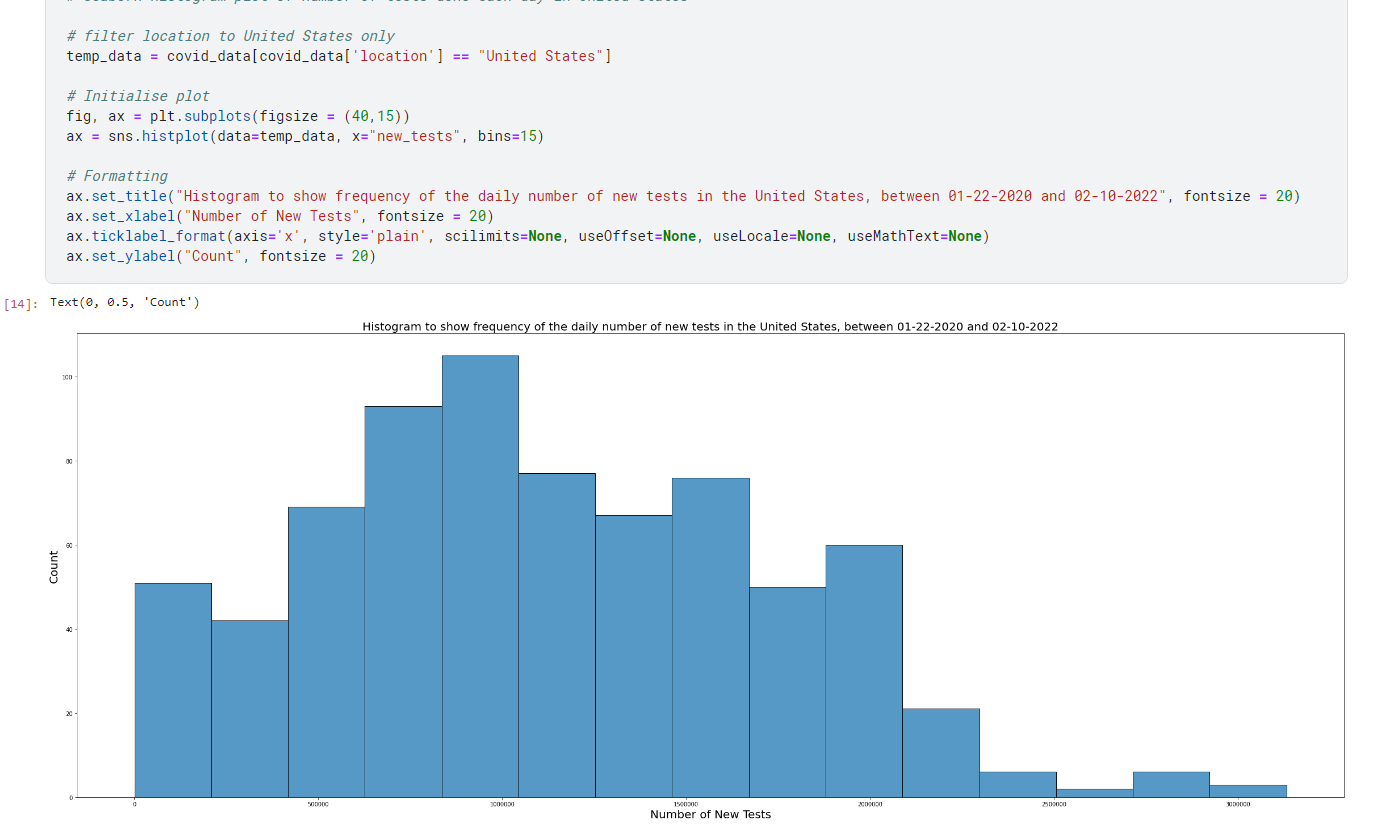
****

The above scatter plot shows the comparison of total deaths of people who are of age 70 and older.

**Total Cases of Covid-19 on January 1st 2022:**

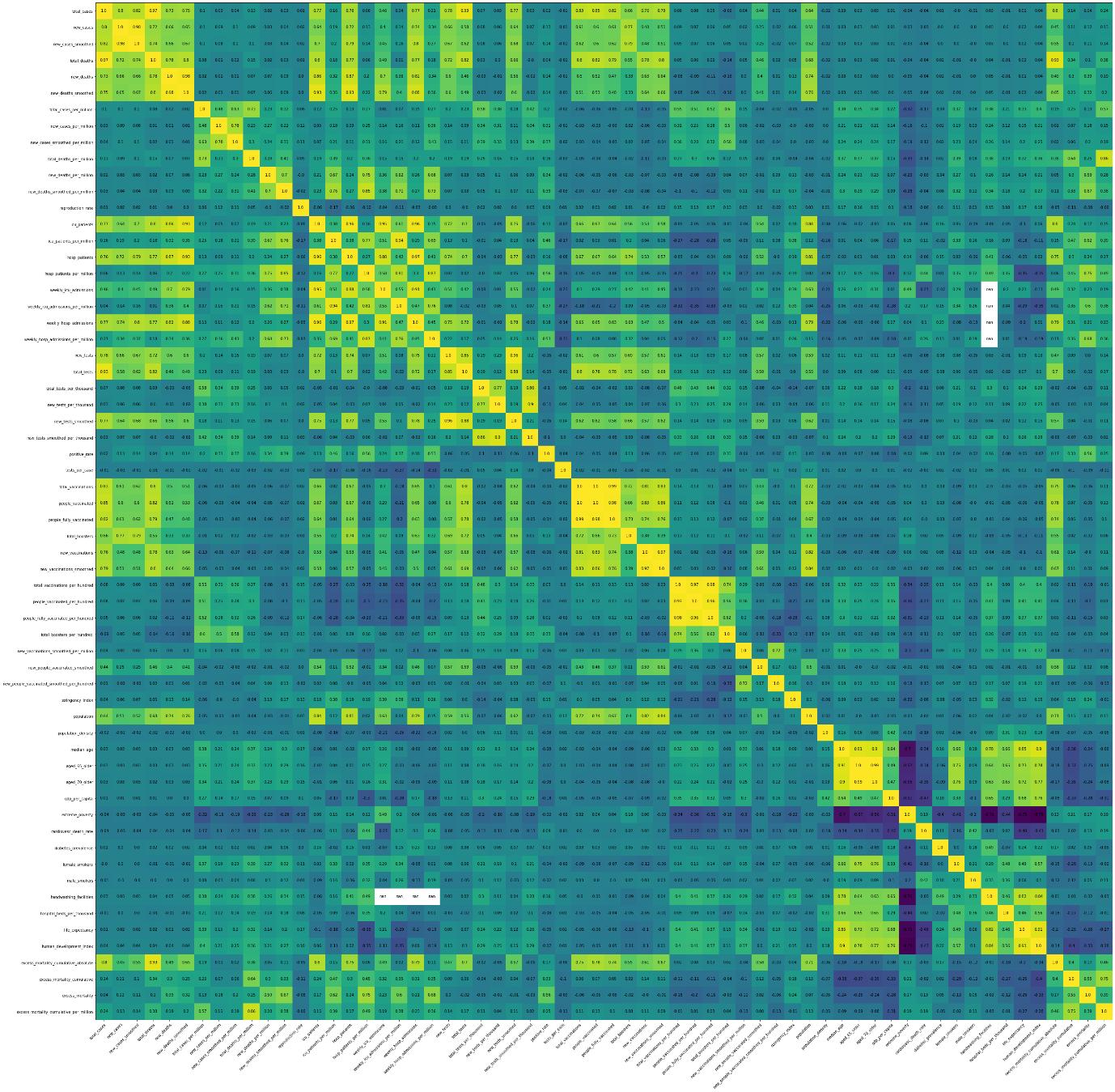
****The above bar graph shows the total number of cases based on every location. The log scale is used to refer the total cases on the y-label. And the color shades represent individual country.

**Histogram to show number of tests done in US:**



The above histogram shows the number of new tests done in the United States.

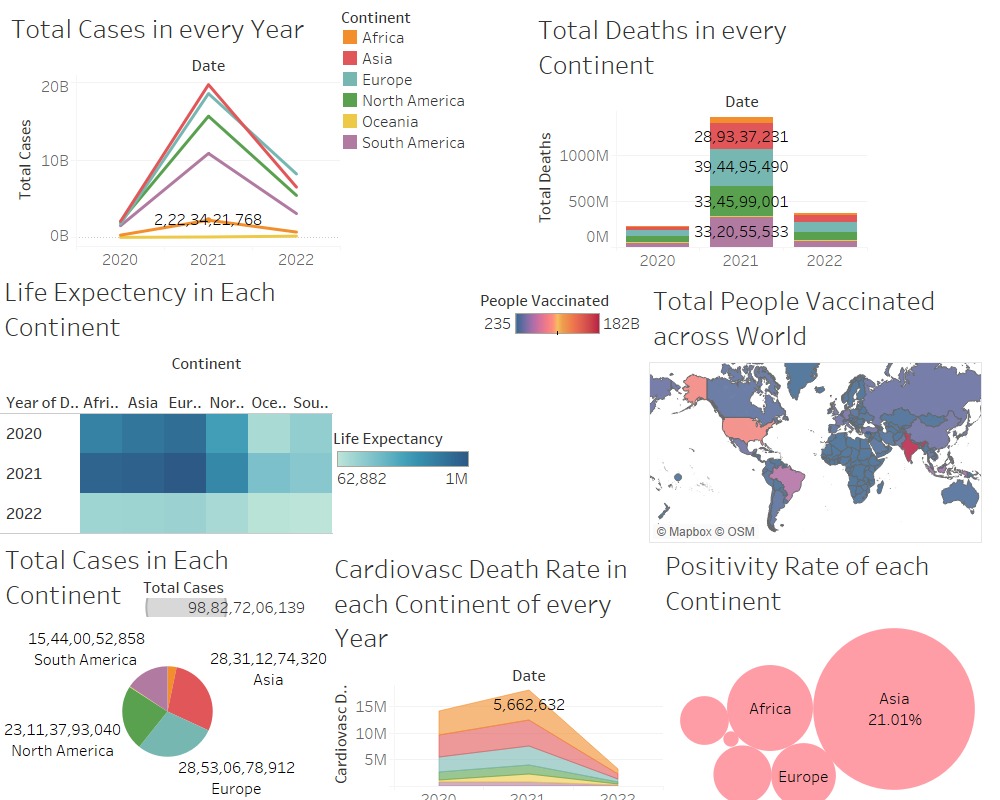
**Heatmap of every Column in dataset:**



The above heatmap shows the correlation matrix of the columns present in the dataset. By this heatmap, it shows the strong relation between individual columns and can understand the dependency of the columns within the data.

**Visualizations in Tableau:**

Below is the dashboard showing few visualizations done using the tableau software. A dashboard used to view the multiple graphs in one sheet for better understanding of the analysis.



**Hypothesis 1:**

Which country has the most people vaccinated across the world?

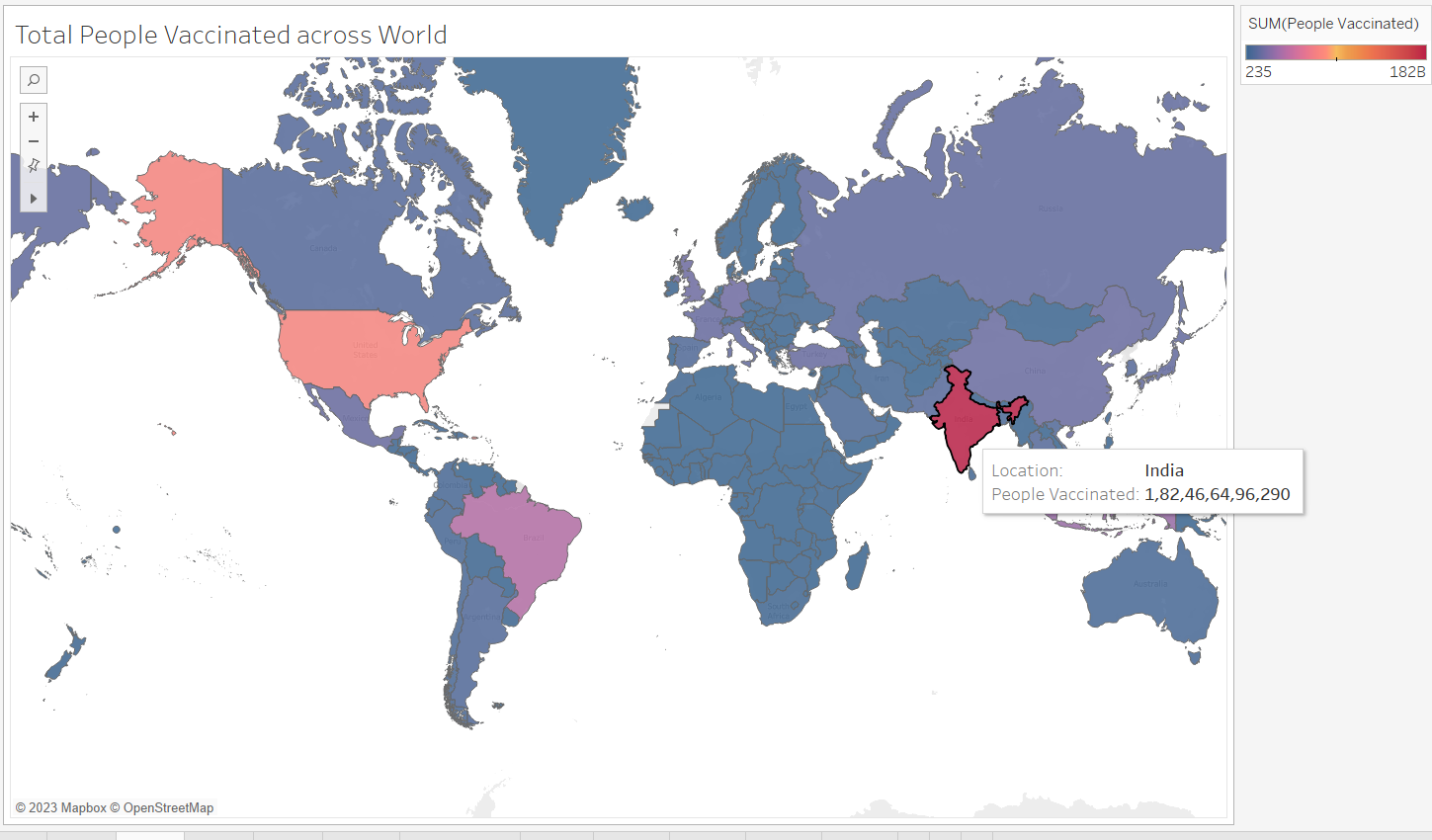
**Hypothesis 2:**

Comparison of continents in total cases, deaths, fully vaccinated, tests and boosters completed.

**Hypothesis 3:**

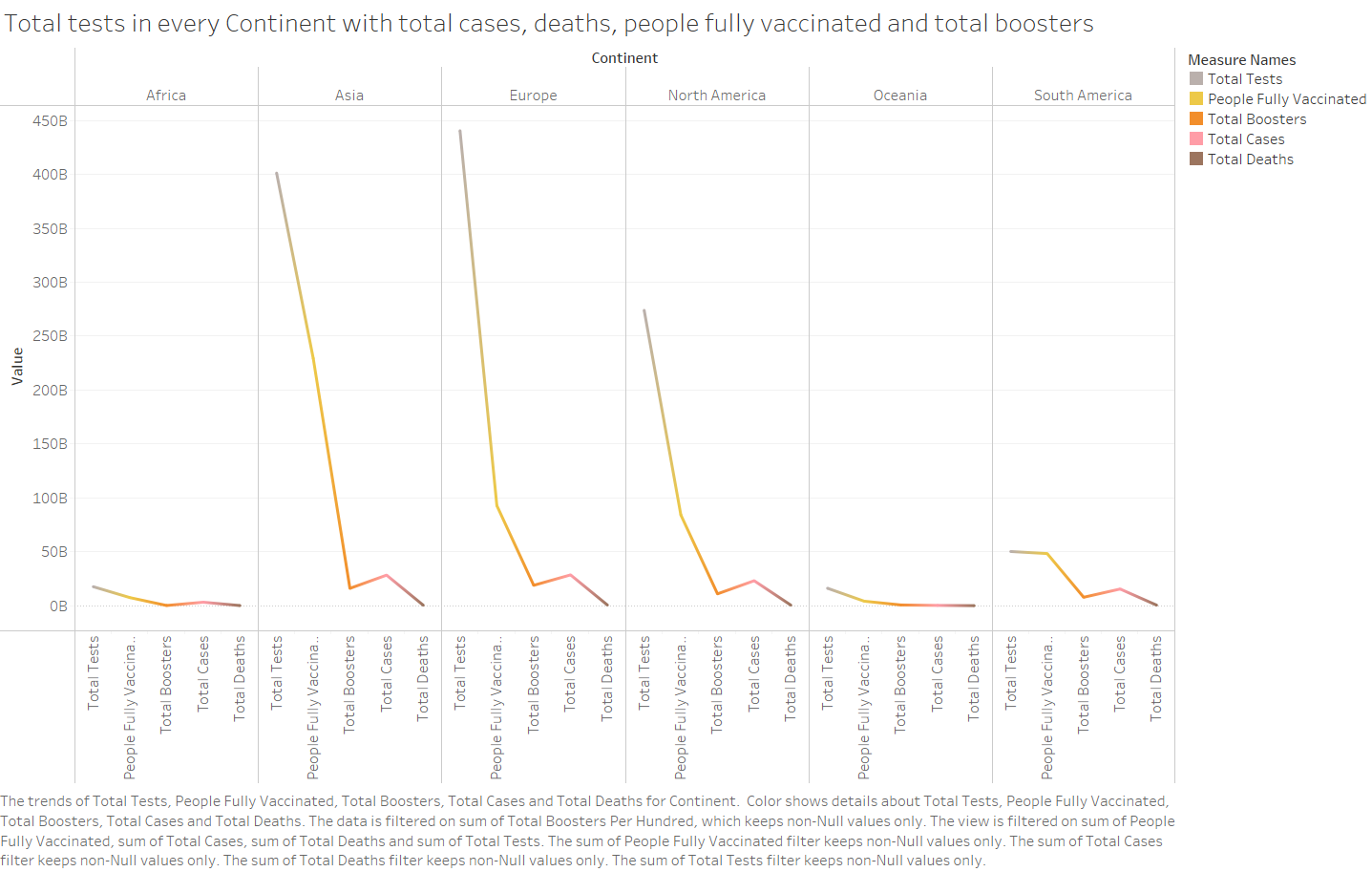
Life Expectancy of each continent during pandemic.

**Hypothesis 1:**

****

**India** has the most people vaccinated with the highest vaccinations of 1.82 billion. Which means every person living in India has got vaccinated. India has the highest vaccinations among all the other countries present in the Asia continent.

**Hypothesis 2:**



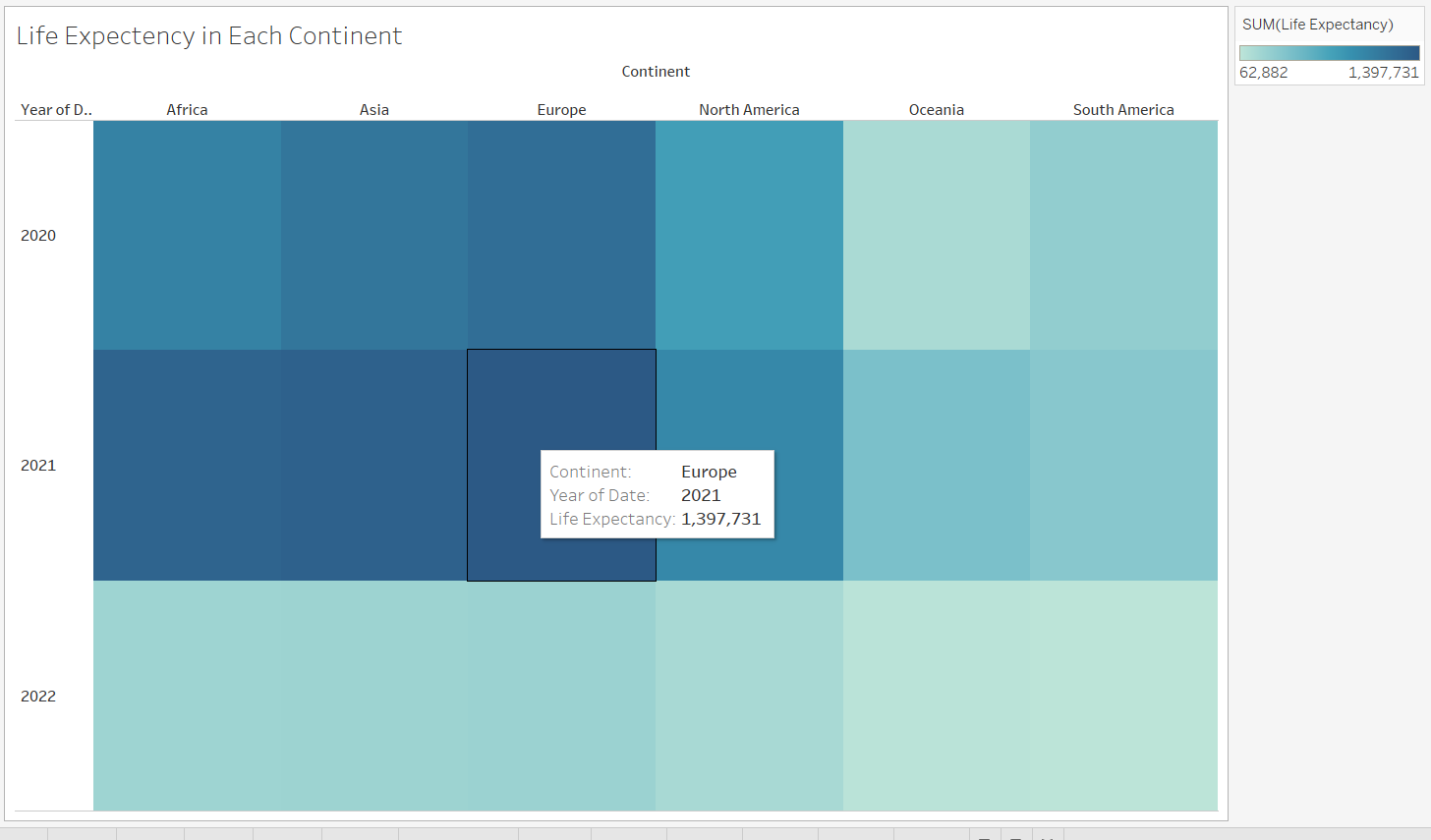
From the above line visualization, it is clear that there are clear differences among all the continents in terms of the tests conducted, people who got vaccinated, total deaths in the continent and total cases in the continent.

In the category of total tests conducted Europecontinent has the highest tests done followed by Asiacontinent. And the least tests are conducted in Oceaniacontinents**.** Whereas, for getting people vaccinated Europeis in second compared to Asiawhich is in first place.

The highest total number of cases were registered in Europe continent closely followed by Asia continent with very less difference between the two continents. In this category also, Oceania continents has the lowest cases registered. When it comes to total deaths, Europe has the greatest number of deaths compared to other continents which is next followed by South America and North America. Oceania continents has the lowest number of deaths which is close to 1.3 million deaths in those continents.

Finally, Europe has highest total number of boosters given to the public compared to other continents. And the least number of boosters can be seen in Africa continent.

**Hypothesis 3:**

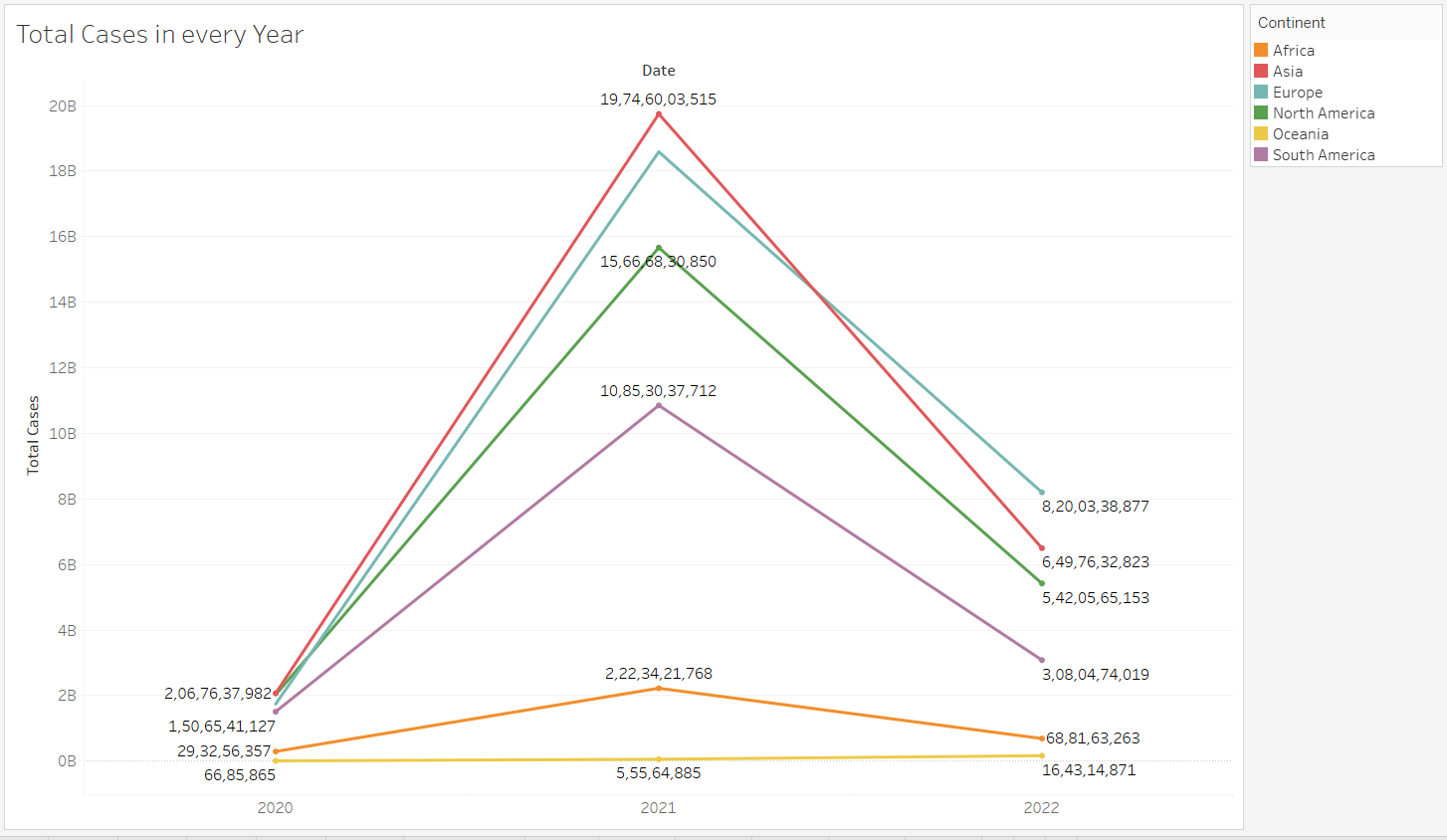


From the above visualization we can check for the continent which has the highest life expectancy compared to other continents. If the color shade is light then the life expectancy is low and brighter the color, higher the life expectancy. In the year 2020, Oceania has the lowest life expectancy of 163,713, where Europe has the highest life expectancy which is close to 1.2 million.

Here in the above graph, in the year 2021 Europe has the highest life expectancy of 1.39 million followed by Asia with 1.32 million. In the year 2021, South America has the lowest life expectancy with a total of 358,623. The lowest life expectancy can be seen in South America in the year 2022 with total of 62,882 and the highest in the year 2022 is in Europe.

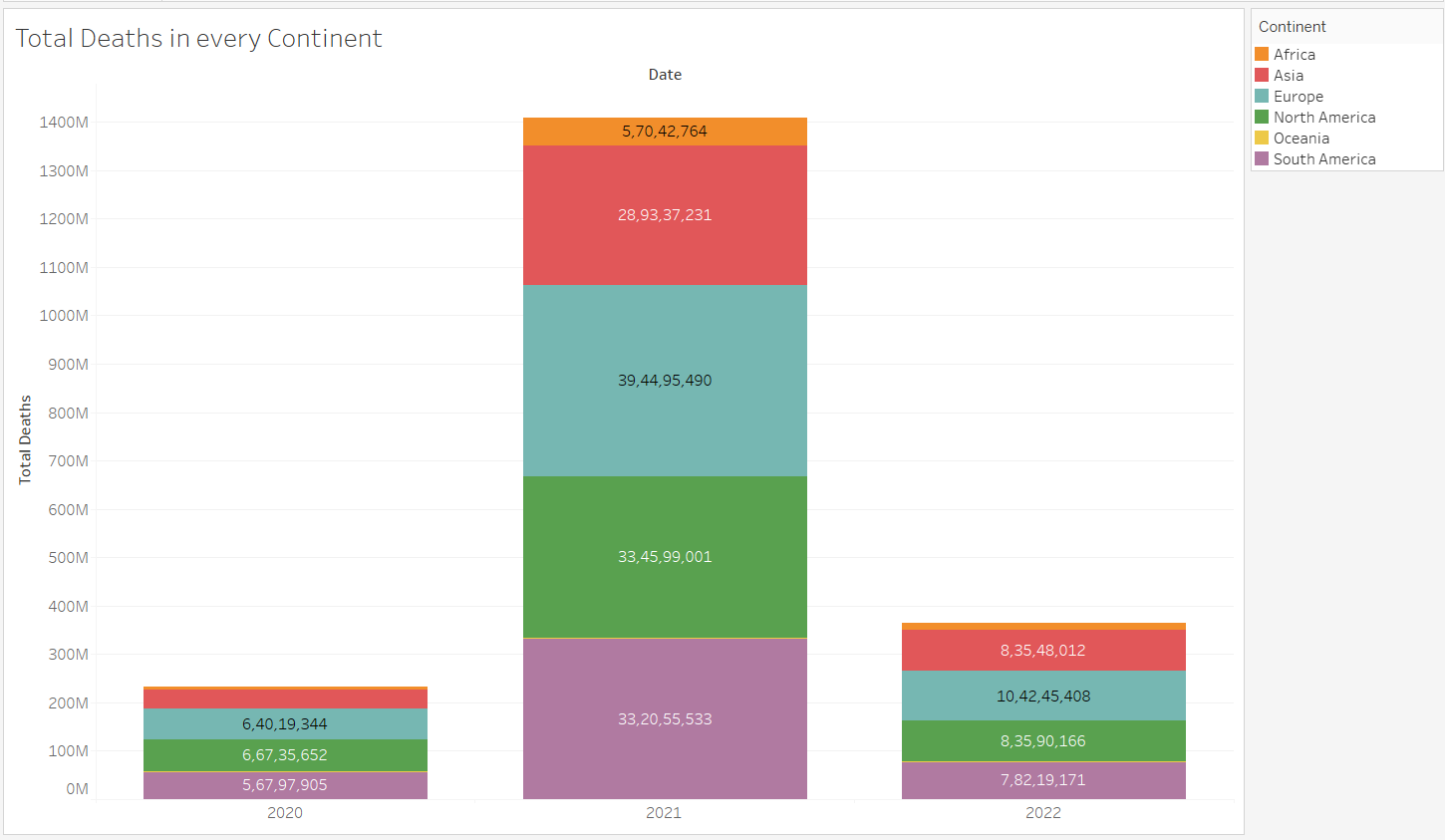
**More Visualizations:**

**Total Cases in Every Year:**

****

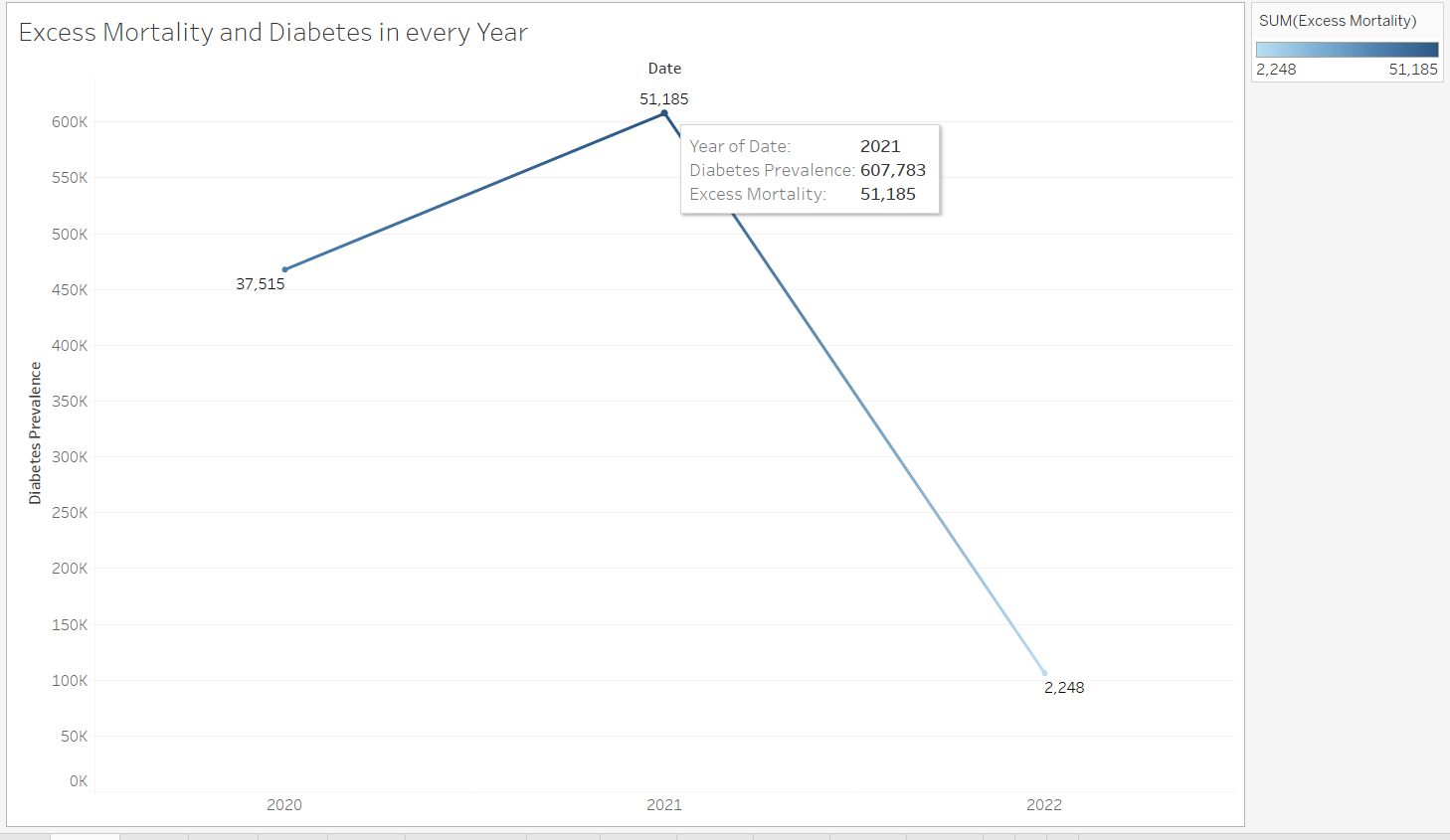
The above line graph shows the total number of cases registered in every year. In the year 2021, Asia has the highest number of cases registered, followed by Europe. The least cases are registered in Oceania continent in the year 2020.

**Total Deaths in Every Continent:**

****

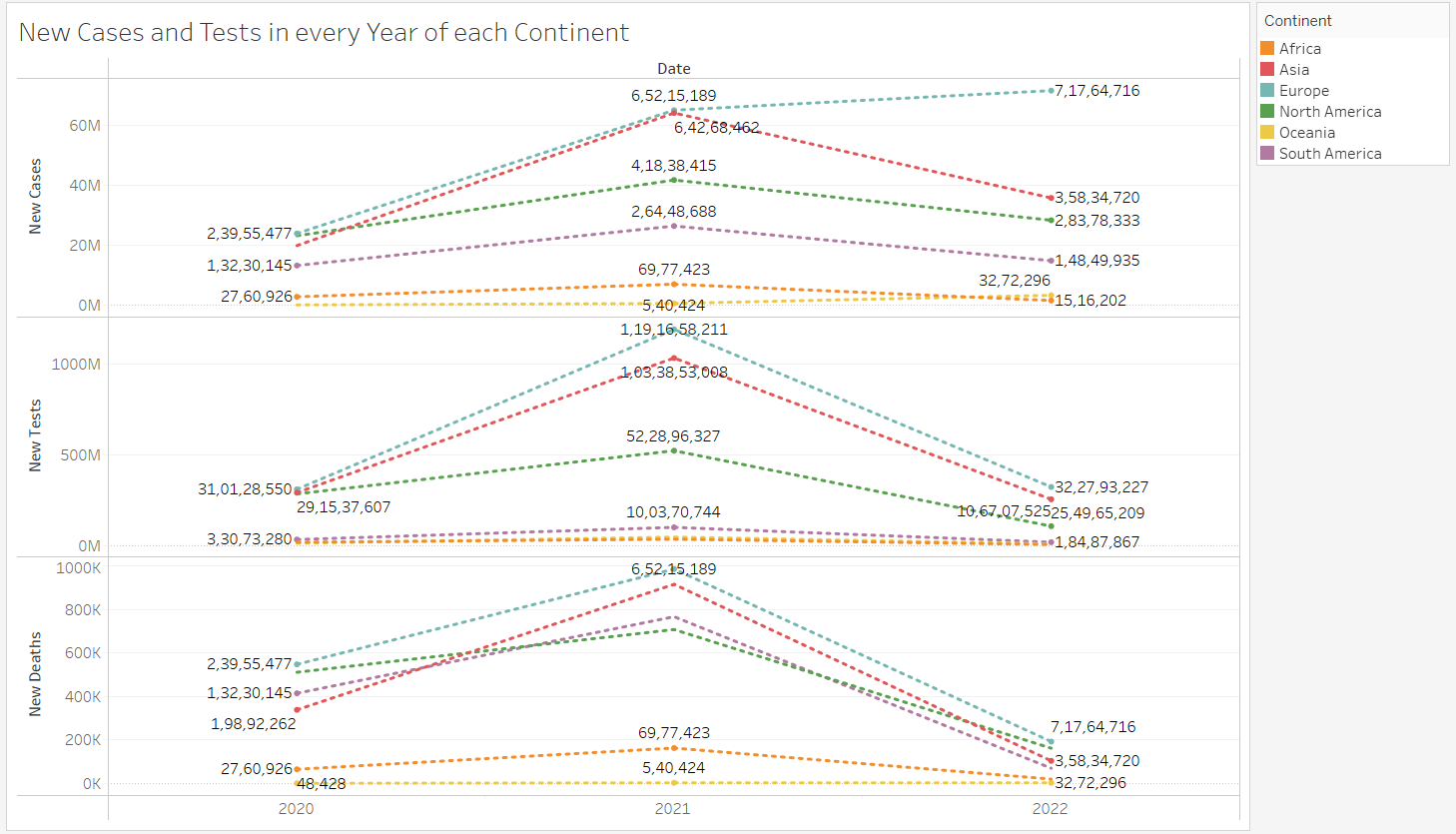
The above stacked bar shows about the total deaths in every continent during the pandemic. As from the graph, in the year 2021 Europe has the highest deaths registered and lowest deaths can be seen in the year 2020 in Africa continent

**Excess Mortality and Diabetes Prevalence in every year:**

****

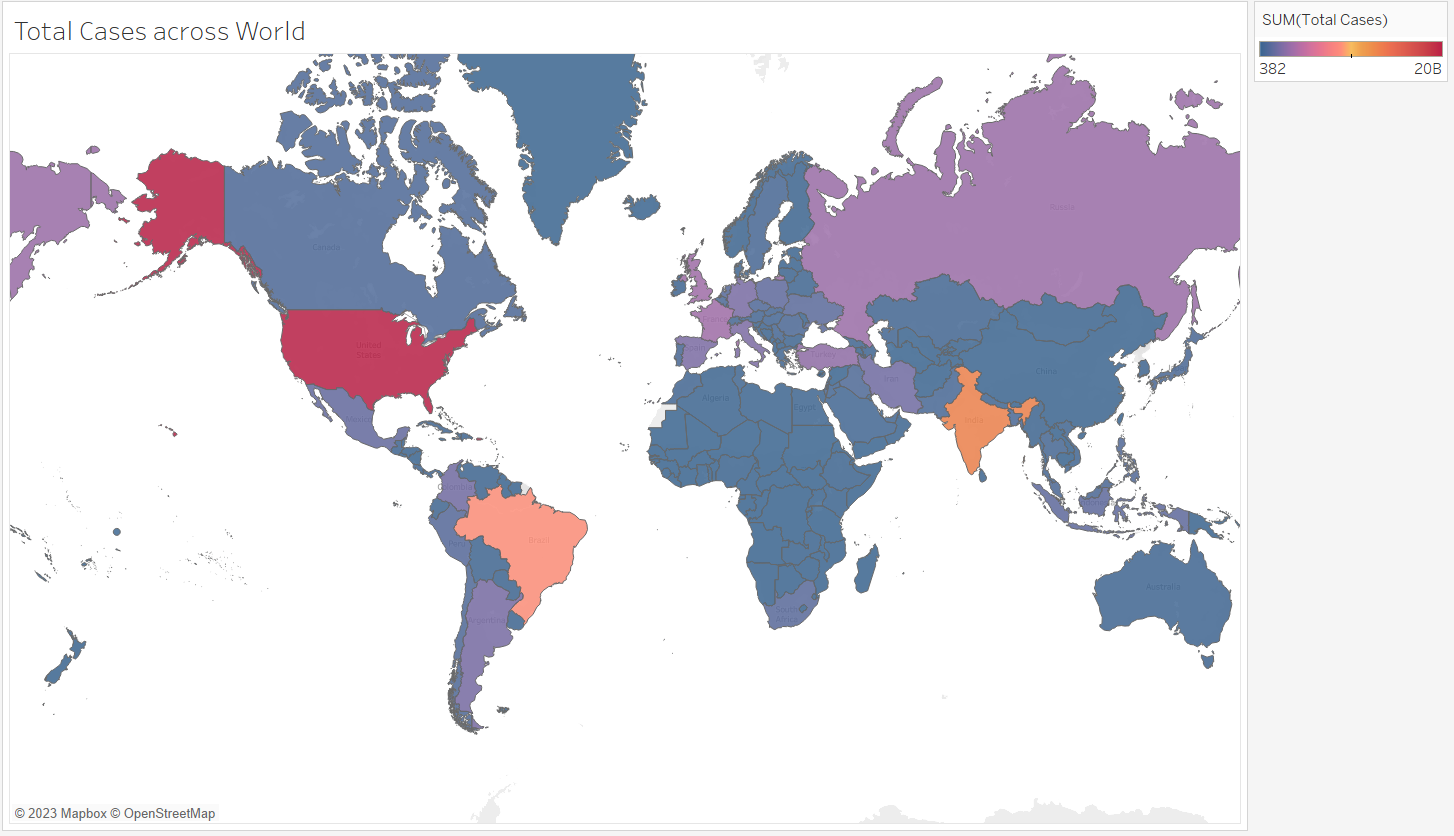
From the above visualization, in the year 2021 has the highest in both diabetes prevalence which is a total of 607,783 and excess mortality of 51,185. And the lease can be seen in the year 2022, where diabetes prevalence is about 106,319 and the mortality rate is 2,248.

**Year wise New Cases, New Tests and New Deaths in every Continent:**

****

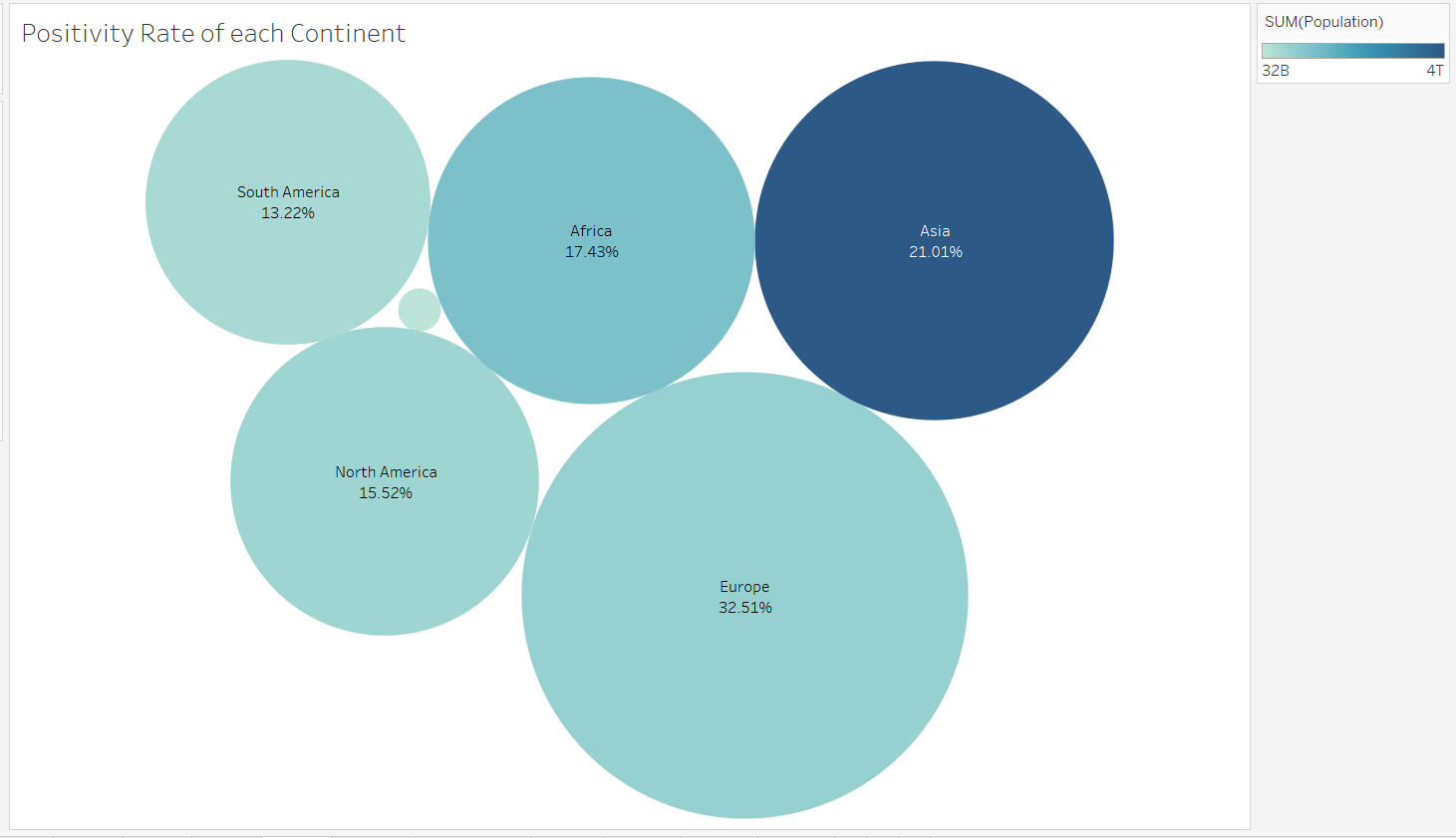
From the above sheet, which shows about the new cases, tests and deaths in every continent happened during the pandemic. Europe has the got high number of new cases in the year 2022. And in the year 2021, Europe has the greatest number of tests done when compared to other continents.

**Total Cases in Across the World:**



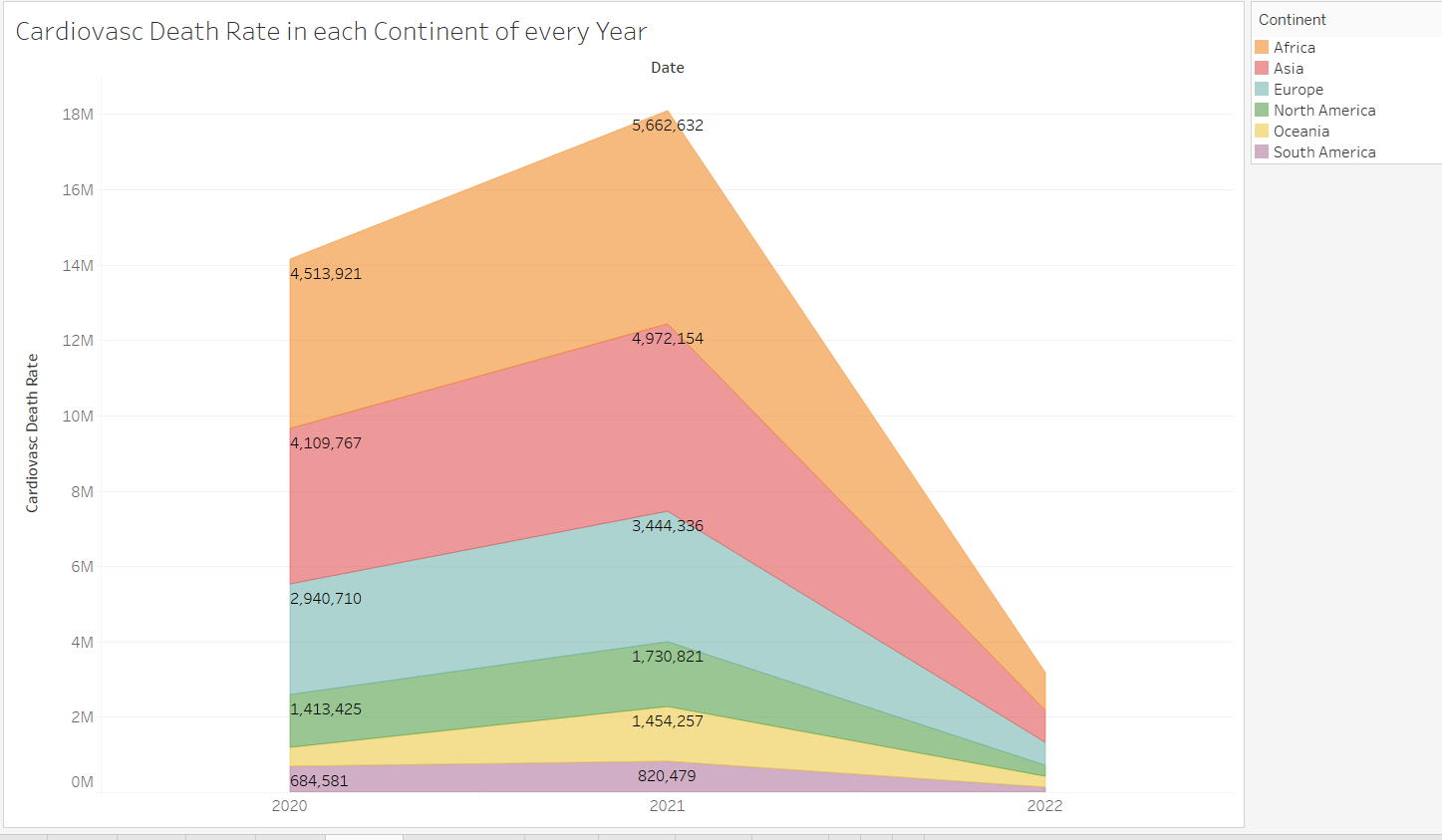
South America and North America has the highest number of cases registered.

**Positivity Rate of Each Continent:**

****

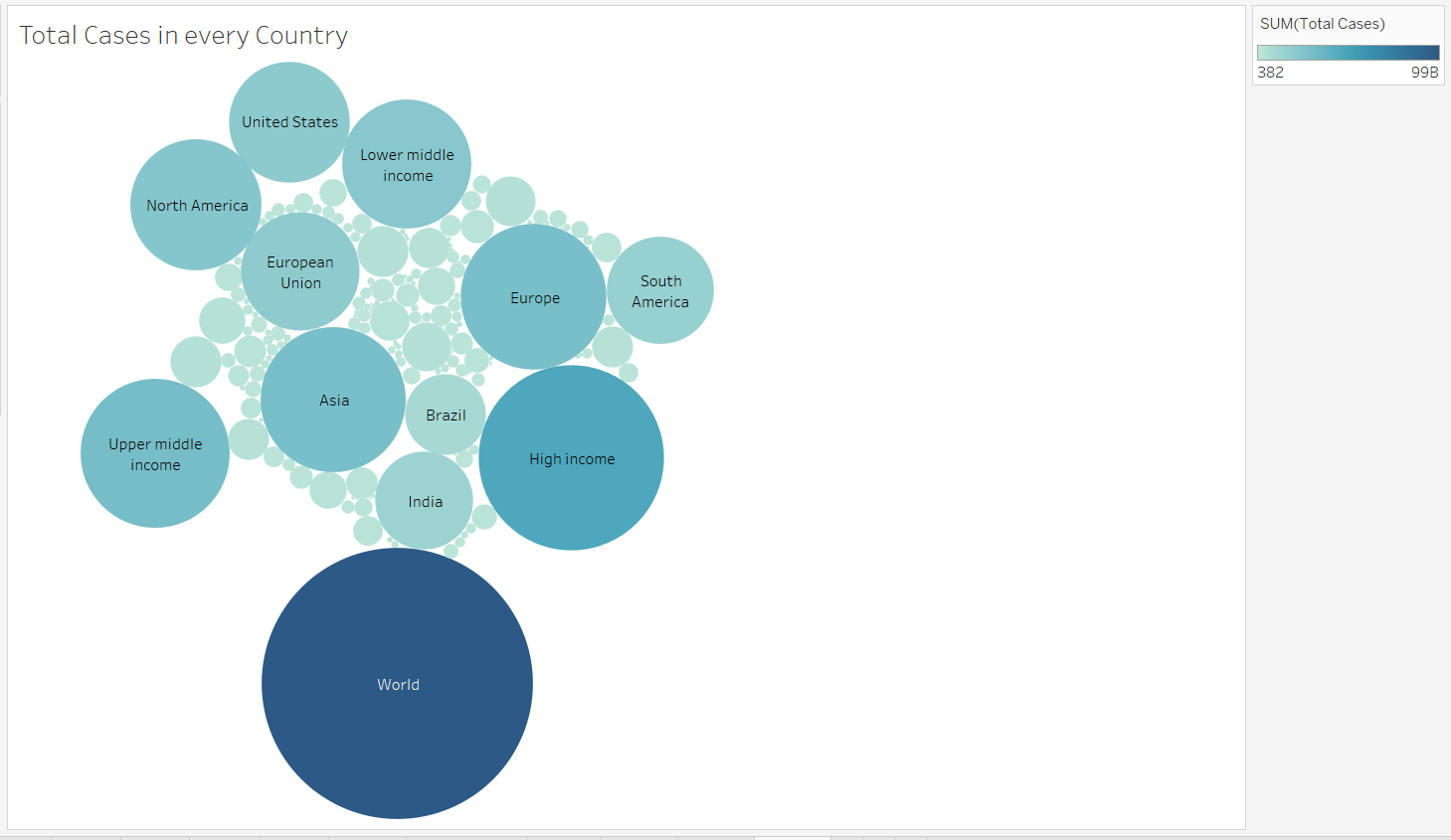
Asia has the highest population compared to other continents, but the positivity rate is highest in Europe continent with 32.51%. Whereas, Asia has a positivity rate of 21.01%. Combined percentage of South America and North America is higher than Asia which will be 28.74%.

**Cardiovascular** **Death Rate in Every Year:**

****

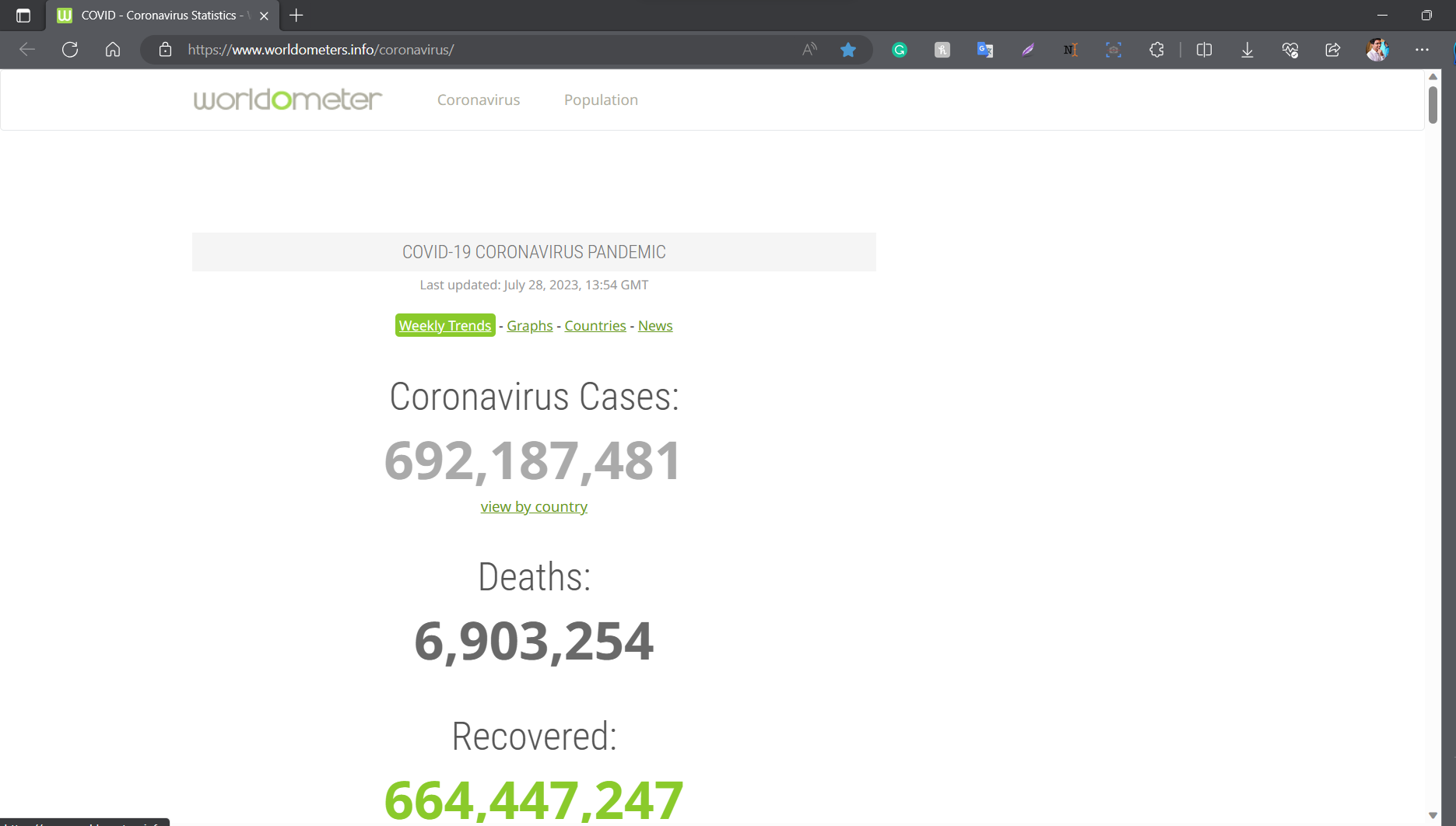
In the year 2021, Africa continent has the highest number of cardiovascular deaths which is highest in all the pandemic years. The least death rate can be seen in South America in the year 2022 with a total death rate of 143,865.

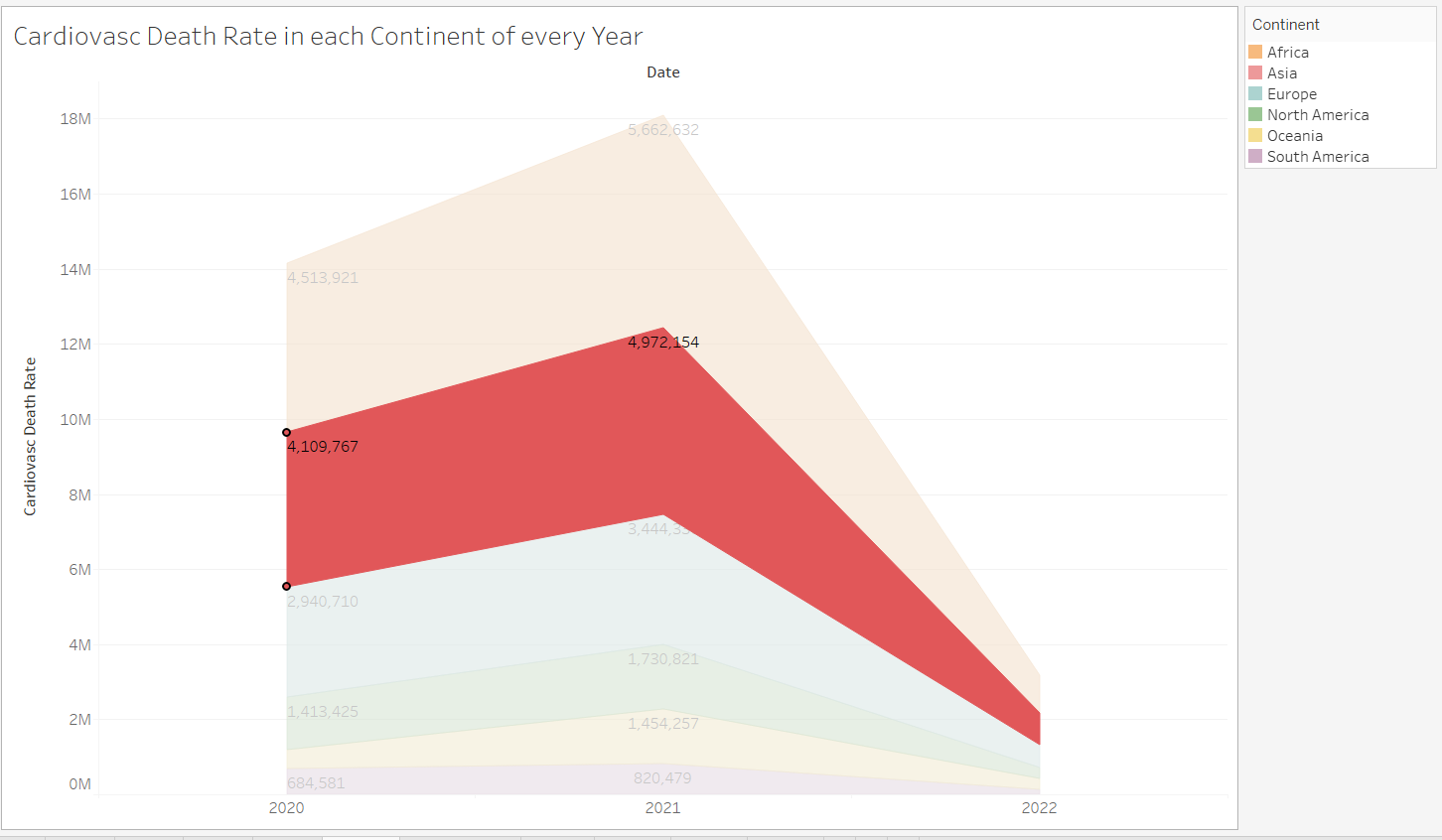
**Interactive Actions:**

****

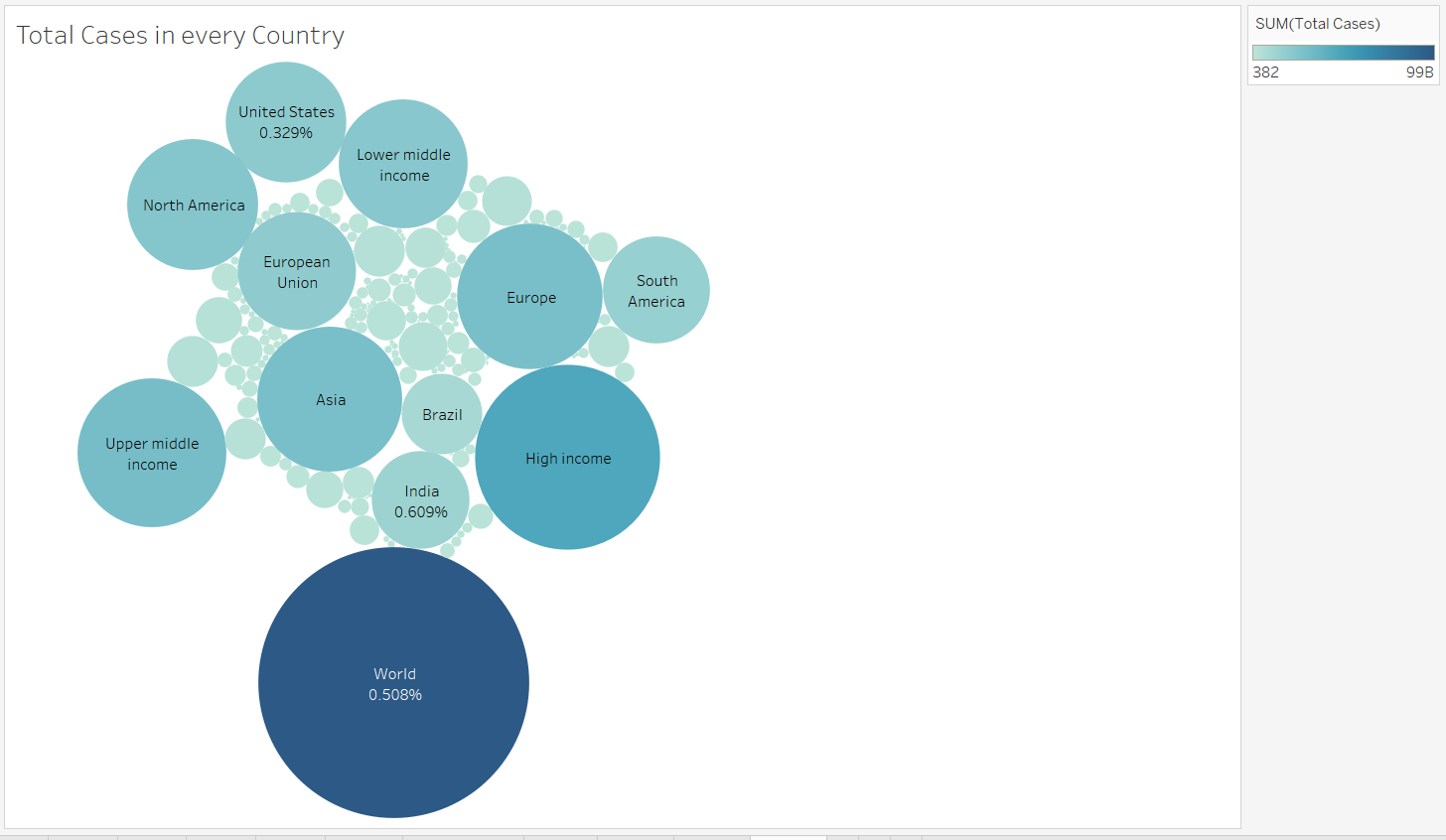
In this sheet, there’s an action implemented. When you select any of the countries it will open a website which shows the live counter of the corona cases, deaths and full tabular form data.

URL used: https://www.worldometers.info/coronavirus/





When the above sheet graph which shows about cardiovascular death rate in each continent is double clicked then it will take to the graph shown below:



This graph shows about the total cases registered in every country. Where the percentage shows the sum of total cardiovascular death rates in that country.

**Reasons for choosing the rationale of 2-dimensional Space Design:**

1. **Simplicity:** It will be very easy to work on 2-dimensional space than to work on the 3-dimentional space. As working on 2-dimensional can easily helps in creating a clean design.
2. **Compatibility:** 2-dimensional designs have most compatibility across all devices and all platforms. This provides a very consistent experience for the users.
3. **User Interaction:** The spatial layout in the 2-dimensional designs have most user interactions and usability. In 2-dimensional designs interactive elements are easy to access and easy to operate.
4. **Content Organization:** These designs have the content organized in a well-structured manner which is in logical way. This makes user to operate and navigate easier and find the information they are looking for.
5. **Visual Hierarchy:** The visual hierarchy is used to establish colors in the graphs which differentiate the main elements and secondary elements.
6. **Data Types:** It makes easy to understand the data types in a dataset which are crucial in making a visualization. Accurate representation and preventing a misinterpretation can be achieved by choosing a appropriate chart.
7. **Consistency:** Using consistent colors, labels and fonts in a visualization helps to create a visual narrative.
8. **Storytelling:** When performing a visualization on a dataset it is always best to make sure that it tells a proper story. You should make sure that the people who are viewing your visualizations can understand what you are trying to say with that graph or chart.

**Conclusion:**

Using the visualizations in the tableau, this project has given a chance to explore numerous data to gain more understanding on how much impact this pandemic has created.

From the visualizations, the maps and time series show how much the virus is progressed among different countries and how much cases increased in over a period of time. The vaccination efforts show how much the vaccination was working and available to people to decrease the effects of covid-19 which can be seen in percentage of vaccinated people and fully vaccinated people in every country. The prevalence of diabetes, an underlying health condition that may raise the risk of serious COVID-19 outcomes, was shown in the diabetes prevalence column.

The interactive features in the tableau software helps to more understandings on the data that is being used. This dynamic facility helps the people for high level of understanding which allows to do more exploration on the dataset to provide more detailed explanations.

**References:**

*Covid 19-Dataset, Kaggle.*

<https://www.kaggle.com/datasets/georgesaavedra/covid19-dataset>