

COVID 19

Introduction:

The SARS-CoV-2 virus causes Coronavirus Disease (COVID-19) which is an infectious disease.

Most of the people who are infected with covid-19 will the major symptoms based on respiratory. On other the hand, some people will be critically unwell and require assistance from doctors. The precautions to be taken are we should Stay at least 1 meter away from people, should have a worn mask, and we should wash our hands with an alcohol content sanitizer to protect ourselves from infection. As per the government's advice, we should take a vaccination.

When a covid 19 is infected, the symptoms are sneezes, Body weakness cough, and low oxygen levels Worldwide all the countries got suffered from covid 19. Till now we can experience the effects in every part of the world. Every country got affected awfully but, we can measure the effects based on the population and the data collected from health care systems.

In this project, my main thought was to work with Covid-19 data from the datasets I found on Kaggle and Covid.org

- 1. What are the overall confirmed cases in each country?**
- 2. What are the overall recovered cases in each country?**
- 3. What is India's total number of confirmed cases?**
- 4. Where can I find the top death rate?**
- 5. Where can I find the lowest death rate?**
- 6. People Vaccinated Vs People Fully Vaccinated by Location?**

Methodology:

For the above COVID-19 project, I have undertaken 2 datasets, of which one was already cleaned up, and the second dataset was changed and cleaned up by removing the unnecessary data.

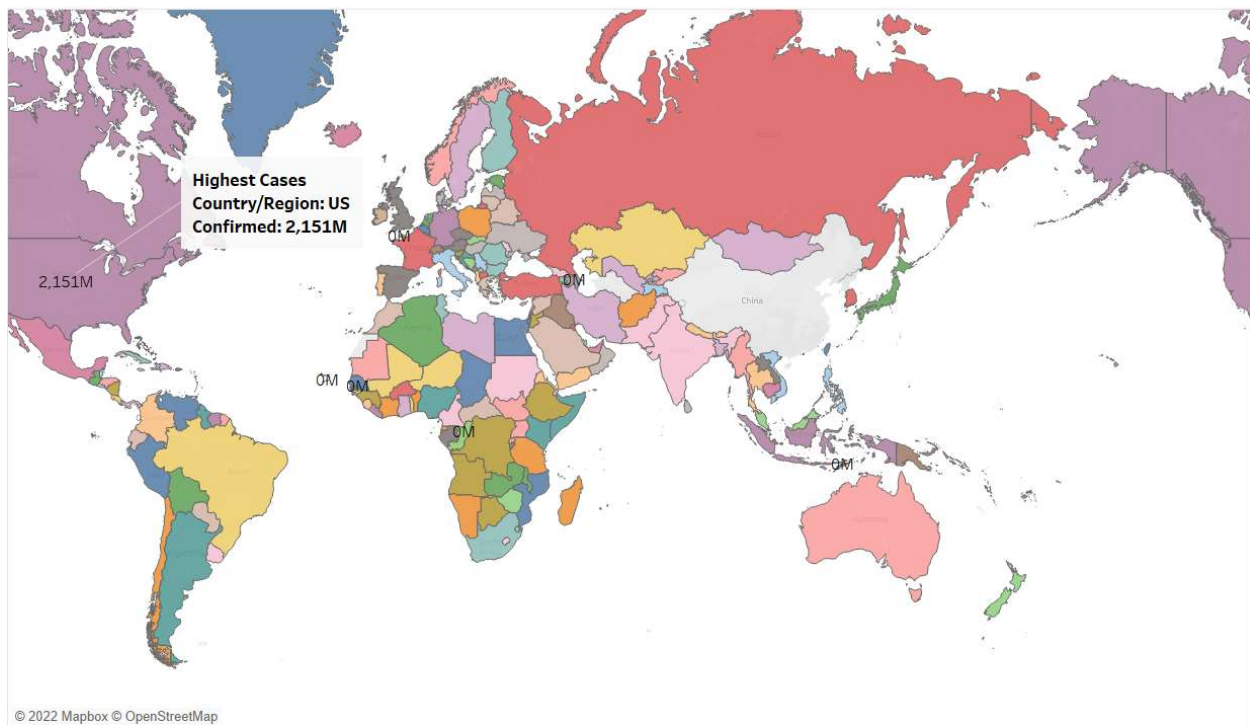
The main attributes to look for in the above datasets are province/state, country, death cases, recovered cases, and confirmed cases.

Methods and case study:

Positive covid 19 by countries

In this data-centric era, we can analyze how far the world come across these challenging days by covid 19 from various resources. In our case, we took a dataset from Kaggle. In this dataset the data is recorded based on the 'Observation Date' and 'Last Update', which is on a date (24 hours) it records the number of people got confirmed as a covid patient is recorded as 'Confirmed' and patients who successfully won the covid battle is recorded as 'Recovered'. Unfortunately, the people who lost their life is recorded as 'Deaths'. There are data available about the 'country/Region' and 'Province/State' along with the information on the number of patients.

Confirmed Cases based on Country



By using Tableau visualization software, we can do visual analysis to find an overall number of instances in each country. We can also easily differentiate each country by the number of cases by giving different color shades to them. By the color shade, the graph becomes more visually pleasing and easily understandable for anyone.

To visualize the geographical data the best visual chart is a 'Map', for that we need at least 1 geographical dimension data value and 0 or 1 measure value. Here, to find each country's overall instances, drag and drop the 'Longitude(generated)', and 'Latitude(generated)' dimension values in columns and row shelves respectively to plot the map view. Then drag and drop the 'Confirmed' measure value to the Color Marks card then it visualizes the countries with different shades of default color by aggregate value (sum) measure of Confirmed cases in the way of a darker shade to largest aggregate measure and lightest shade to smallest aggregate measure. We can change the color to make the visualization catchier and easier to understand and drag and drop the 'Country/Region' dimension value to the Details option in the Mark card. It will show the details

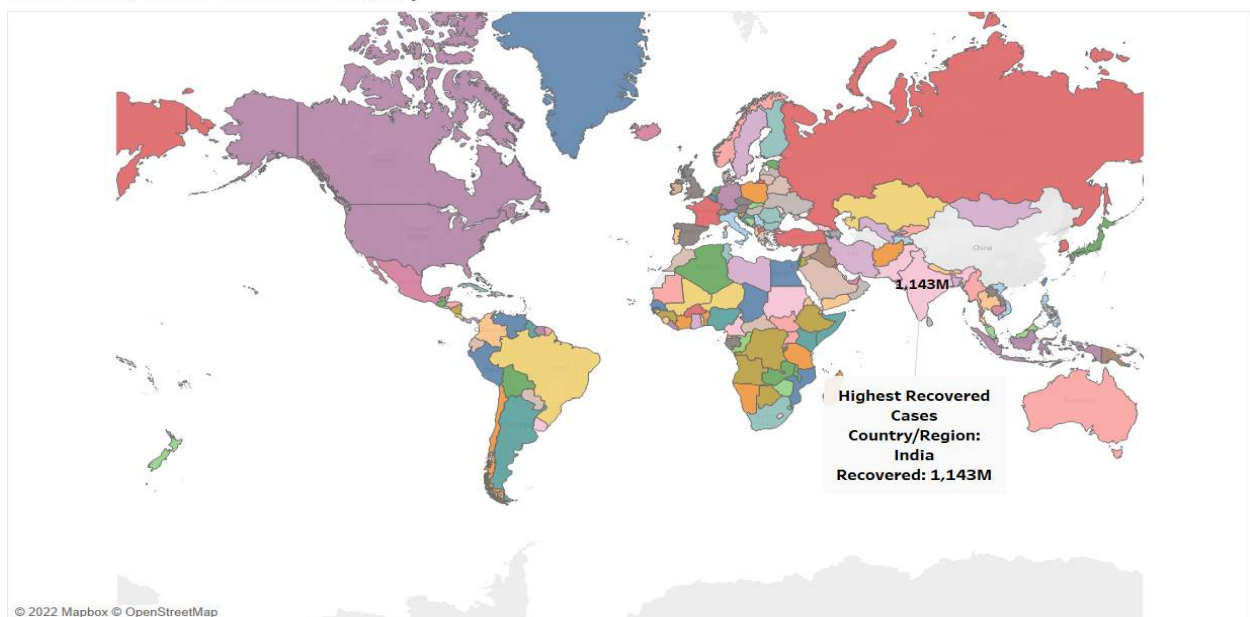
about the country name and its overall instances when we move the mouse pointer over any of the countries on the map.

As a result of this visualization, we can easily find the overall covid instances of the countries in the world. The color shaded purple indicated the most affected region and from that, we can see the US is affected with the highest number of cases 2,151M patients, and other countries are eventually shaded according to their rate color shade.

Recovery by countries:

After these tragic effects of covid, one of the main aspects to analyse the 'Recovered' data among the countries. It helps to understand the immunity of people from various regions and health care systems. As we did in the previous visualization, we can visualize the 'Recovered' number of cases data around the world by using a Map chart.

Recovered Cases based on Country



Map based on Longitude (generated) and Latitude (generated). Color shows details about Country/Region. The marks are labeled by sum of Recovered. Details are shown for Country/Region. The view is filtered on Latitude (generated) and Longitude (generated). The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only.

We follow the same steps which were followed in the earlier graph to generate a geographical map. Drop the 'Recovered' measure value to the Color Marks card then it visualizes the countries with different shades of colors by an aggregate value (sum) measure. The number of Recovered cases in the way of a lighter shade to largest aggregate measure and darkest shade to smallest aggregate measure.

As a result of this visualization, we can easily find the overall number of cases recovered from the countries available on the map. Also, we can find that 'India' with the lightest color shade has the most recovered measure of 1143M. This high number of recoveries might be due to effective medical system measures, their generic immunity, and mainly might be due to their unique food habits.

Positive Covid 19:

Based on the earlier visualizations, India has a positive prominent impact, so we need to analyze India's covid data. This analysis would be helpful for some good decision-making or any statistical study for data analysts in health care and governing bodies to take measures against the covid in other parts of the world.

In the beginning, we could find India's total number of confirmed cases. For this visualization, we use the symbol map chart to visualize India's total number of confirmed cases. For this symbol map chart, we need 1 geographical dimension and 0 to 2 measures to make visualization.

Confirmed Cases from india



Map based on Longitude (generated) and Latitude (generated). Color shows details about Country/Region. The marks are labeled by Country/Region and sum of Confirmed. Details are shown for Country/Region. The view is filtered on Latitude (generated), Longitude (generated) and Country/Region. The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only. The Country/Region filter keeps India.

Drop the aggregate(sum) confirmed measure value to the text option in the marks card and filter the Country/Region dimension by India. This modification focuses on the visualization of India on the map.

Our scope is to find India's total number of confirmed cases. Now, this can be achieved by using the 'Filters' in the Tableau workspace. After selecting India in the filter, we just need to add the aggregate of confirmed cases to the color to focus on the number of confirmed cases in the Indian country. From the details, we could see that India's total number of Confirmed cases is 1283M.

Recovery:

From the earlier visualizations, we know that India ranks in the number 1 position for the recovery cases in the world. Now we going to explore the total number of cases that are recovered in India. For this visualization, we make use of the symbol map chart to visualize India's total number of Recovered cases.

Like the above steps to get the confirmed cases in India we just drop the Country/Region dimension value in the Color and Text option in the mark card. Drag and drop the aggregate of recovered measure value to the text option in the marks card and filter the Country/Region dimension by India. Now we get the visualization of the number of recovered cases from India.

Recovered Cases from india



Map based on Longitude (generated) and Latitude (generated). Color shows details about Country/Region. The marks are labeled by Country/Region and sum of Recovered. Details are shown for Country/Region. The view is filtered on Latitude (generated), Longitude (generated) and Country/Region. The Latitude (generated) filter keeps non-Null values only. The Longitude (generated) filter keeps non-Null values only. The Country/Region filter keeps India.

As the result of this visualization, we see that India's total number of Recovered cases is about 1143M. This high number of recovery cases might be due to various aspects of the country. One main reason might be the country's population, as the population of the country is high the number of confirmed cases and the recovery cases are at a higher level. Another important aspect of the larger number of recoveries is the medical infrastructure and awareness among the people for medical vaccination for any disease.

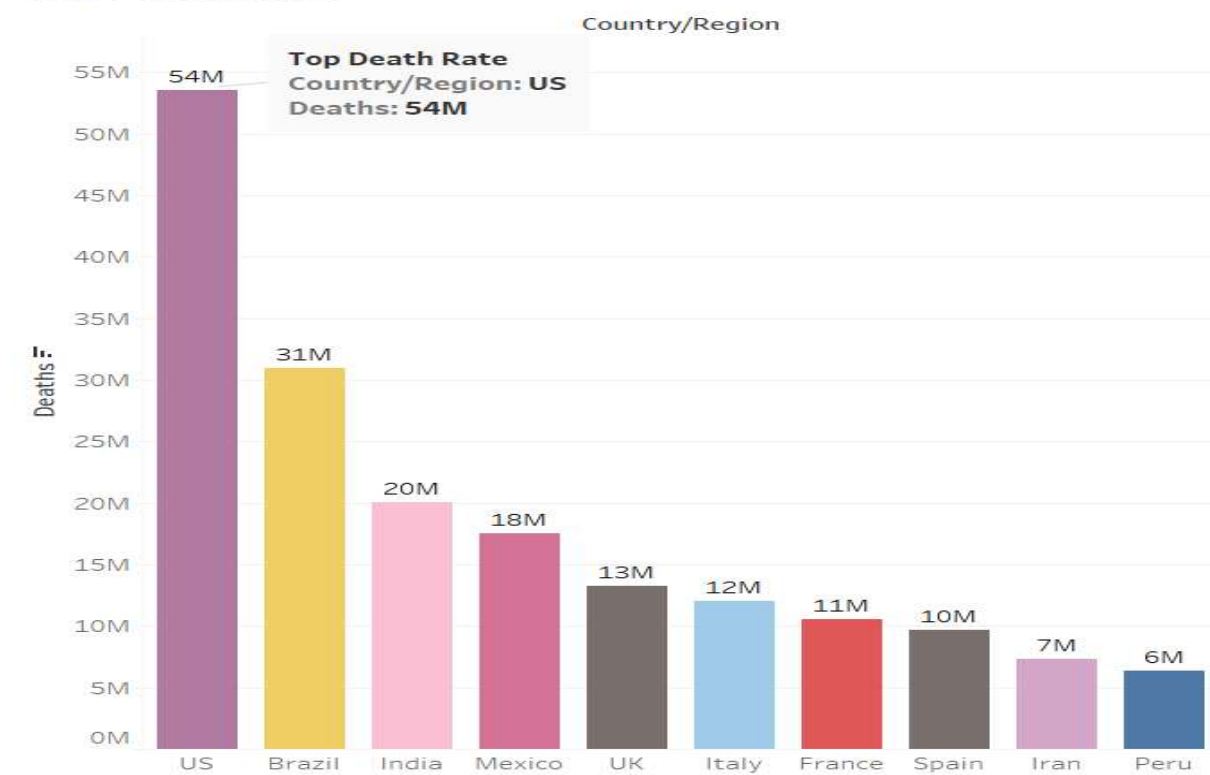
Deaths:

Covid has its impact all over the world and covid has an even darker side where it might also lead to the death of an individual. The virus is so harmful that it affects the immunity of some people and completely makes them weak and eventually kills them. This mainly affects the lungs of the people and once the breathing is destroyed automatically all the functions in the body become weak as the blood flow becomes weak.

Let's now explore which region has the greatest number of unfortunate deaths due to this Covid virus. To make this visualization we have just created a bar graph with countries in columns and the number of deaths in the rows.

The graph gives us clear information about the deaths that happened due to covid and unfortunately, the United States of America is the country where there is the greatest number of deaths recorded. Over 54M people have lost their lives in the US due to this harmful virus. The second country where there are a high number of deaths in Brazil with 31M deaths.

Top Death Rate



Sum of Deaths for each Country/Region. Color shows details about Country/Region. The marks are labeled by sum of Deaths. The view is filtered on Country/Region, which keeps 10 of 226 members.

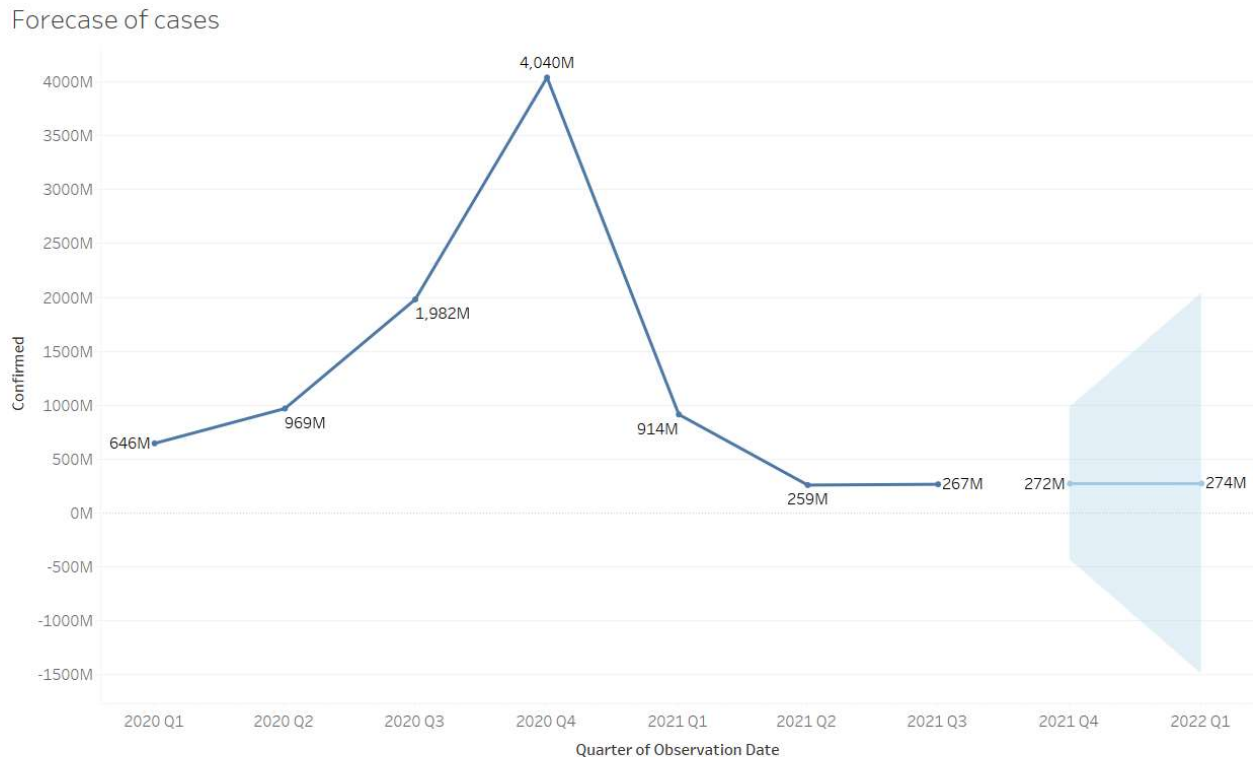
Analysis:

Time series covid forecast

Covid 19 has been so harmful for the past years and in this part, we do some analysis to forecast the number of covid 19 instances in the future. We can forecast the covid 19 Confirmed cases by

using the observed date dimension variable. But the format of the confirmed date dimension is not applicable for use in forecasting. So, we modify the dimension value to a measure by dropping the observed year dimension value in the columns shelf and changing its values to quarters by right click on the variable and clicking the measure option, once the option opens up we select the quarters. It shows how the Covid data spread across the year for each quarter. Use the sum aggregation of confirmed cases to rows. Usually, a linear line chart is the best way to visualize a time series information and we make use of the same to examine the distribution of the covid cases.

To perform analysis and predict the future cases and forecast the covid 19 cases in the future, we could make the forecast model. For that just click on the analytics tab in the toolbar and choose the model and in the model, the menu selects forecast.



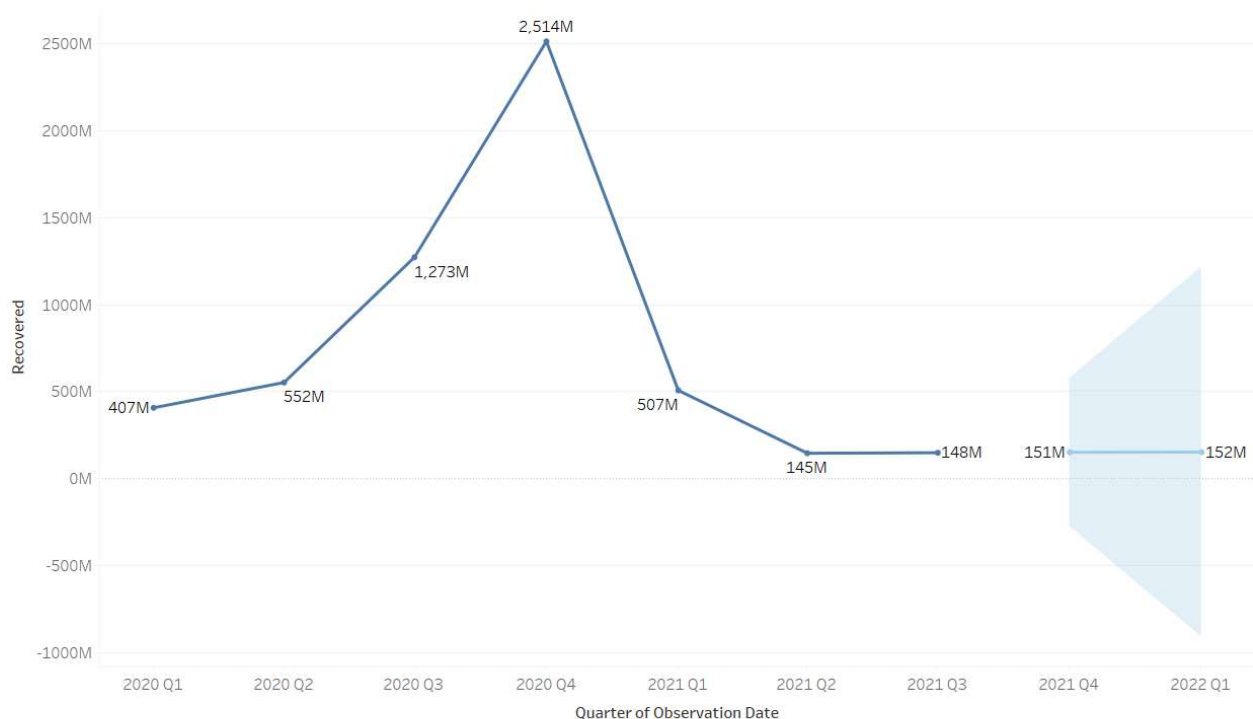
The trend of sum of Confirmed (actual & forecast) for Observation Date Quarter. Color shows details about Forecast indicator. The marks are labeled by sum of Confirmed (actual & forecast) .

This forecast analysis shows the covid 19 confirmed cases for the years 2021, and 2022 quarter by quarter. The actual data and the estimated data are shown in a different colors in the forecast chart. We can make the visualization more appealing change the color option of actual data and estimating data indicators. From the graph and the analysis, we could see that there is an estimated slight increase in the Q4 of 2021 and again followed by a slight increase in the cases in the quarter

1 of 2022. This analysis says that we must be even more careful with this virus contamination and spread still more awareness among people and reduce the spread. Mainly based on this forecast we should also keep the medical units ready so that it would be very helpful in serving the people when they get a confirmation of the virus.

We perform the same similar forecast for recovery. For this plot, we just replace the confirmed cases with the recovery in the aggregate sum measure. Once again, we use a line chart to visualize the distribution of covid data as it is a very effective way of representing time-series information.

Forecast of Recovered cases



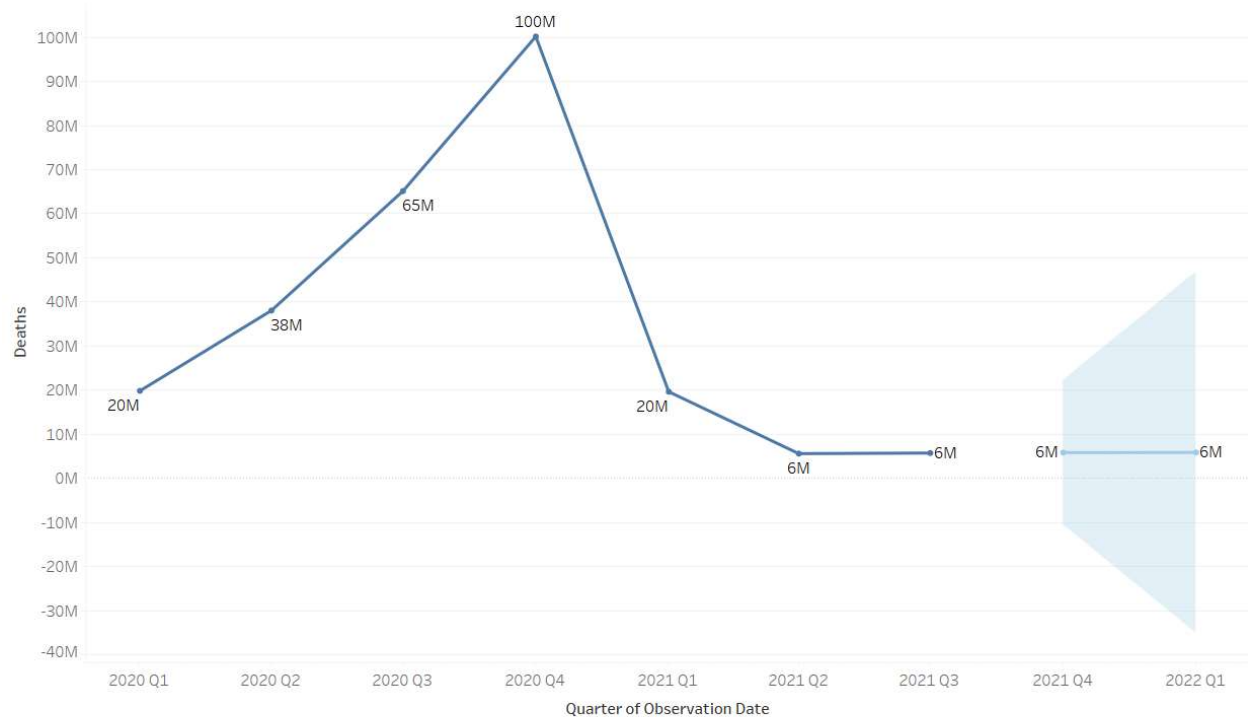
The trend of sum of Recovered (actual & forecast) for Observation Date Quarter. Color shows details about Forecast indicator. The marks are labeled by sum of Recovered (actual & forecast) .

From the analysis plot, we could see as the confirmed cases which we saw earlier increased slightly the recovered cases has also increased slightly for 2021 Q4 and 2022 Q1. In the final quarter of 2021, the forecast is about 151M people will recover from this disease, and at the beginning of 2022, there will be about 152M people recovering from the infection.

Similarly, we have also forecasted the total death cases from the data from the visualization below we could see how the forecast trend is moving.

The graph is plotted in a similar way but by just changing the recovered parameter to the death parameter from the earlier plot.

Forecast of Deaths

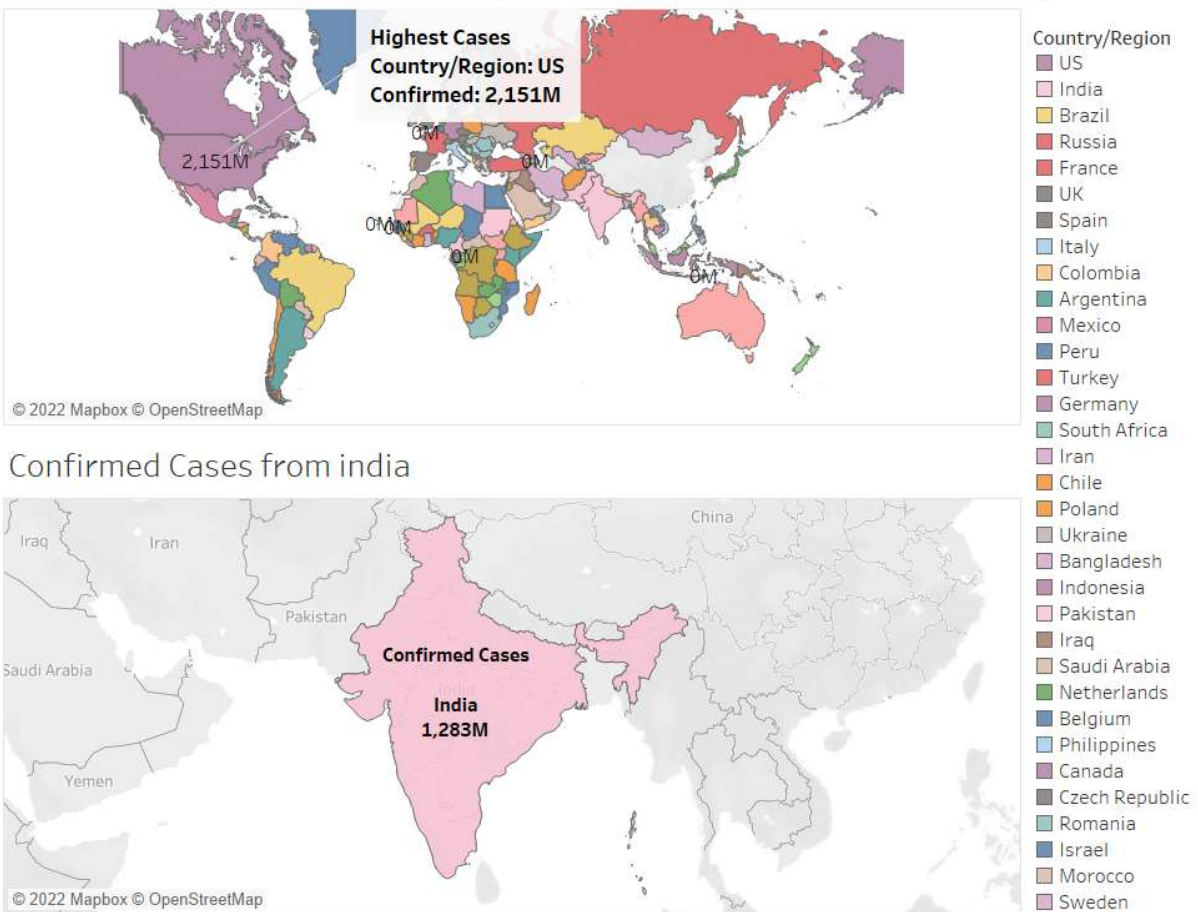


The trend of sum of Deaths (actual & forecast) for Observation Date Quarter. Color shows details about Forecast indicator. The marks are labeled by sum of Deaths (actual & forecast).

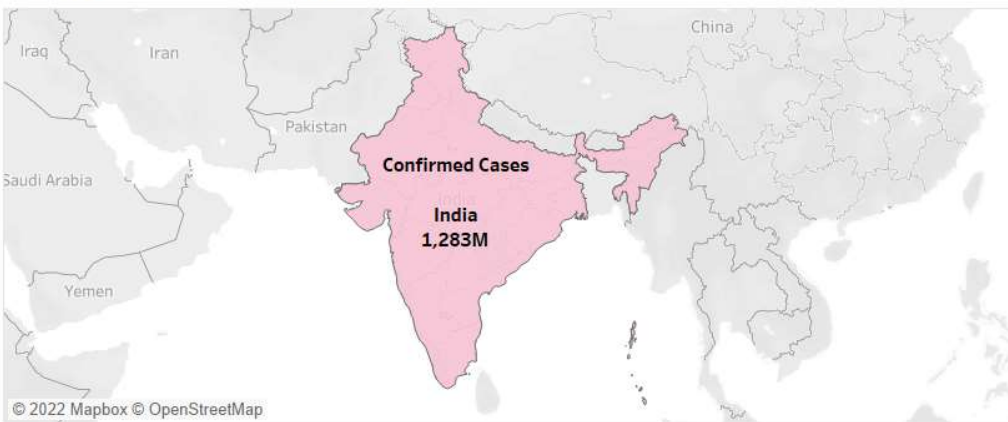
From the analytics graph, we could see that there will be steady maintenance in the amount of several deaths. Even though the number of recovery increases from the earlier graphs the number of deaths happening due to the covid infection remains the same. This is a very strong indicator that there is more danger in the future concerning the lives of the individuals who get infected with this virus.

Dashboard:

Confirmed Cases based on Country

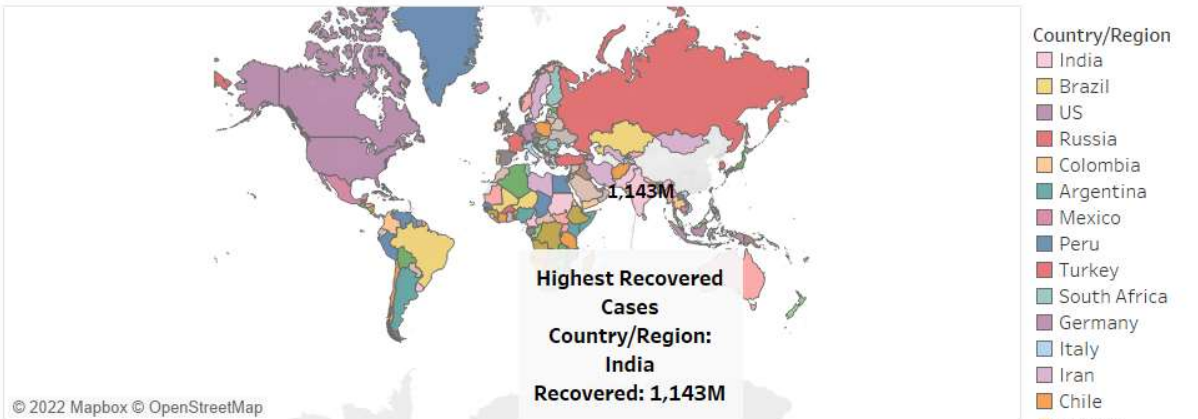


Confirmed Cases from India

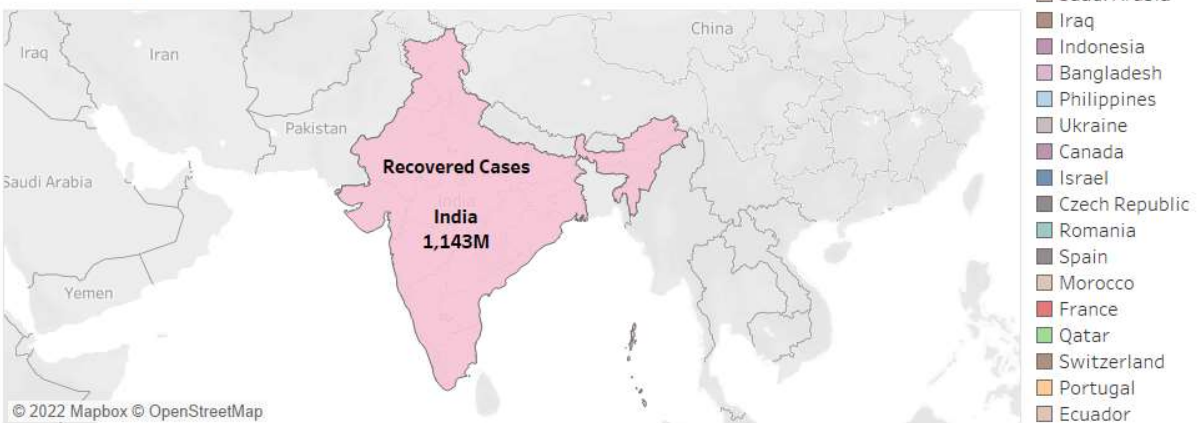


The above dashboard indicates multiple valuable information. The first graph indicates the number of confirmed cases all over the world. The geographical map is used for both the graphs in this dashboard. Each country was given a unique colour to identify them uniquely. The next graph is the geographical location of India and the number of confirmed cases in India is about 1283 million confirmed cases.

Recovered Cases based on Country

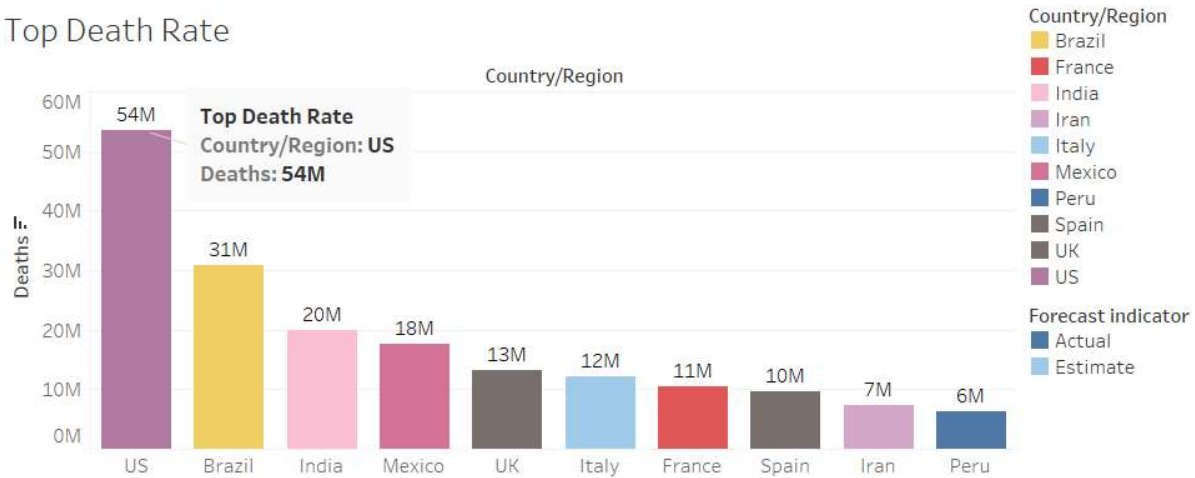


Recovered Cases from india

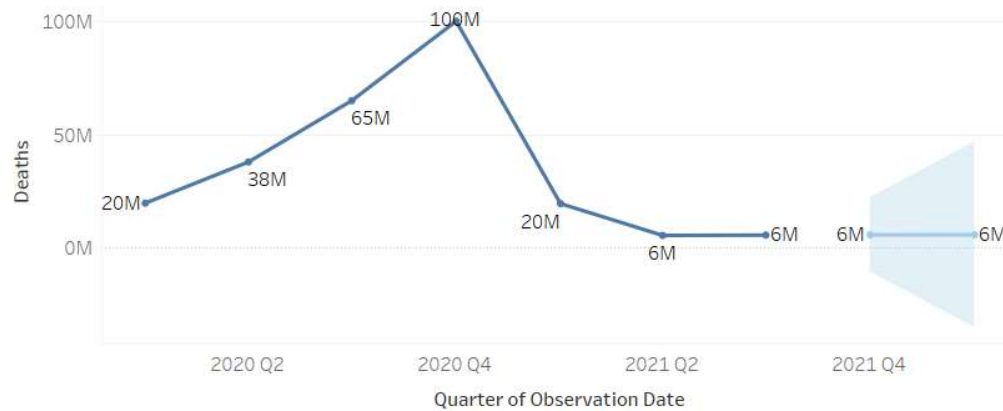


This dashboard is the same which is just an inverse of the confirmed cases. This is the number of recovered cases all over the world and the second plot indicates the number of recovered cases in India.

Top Death Rate



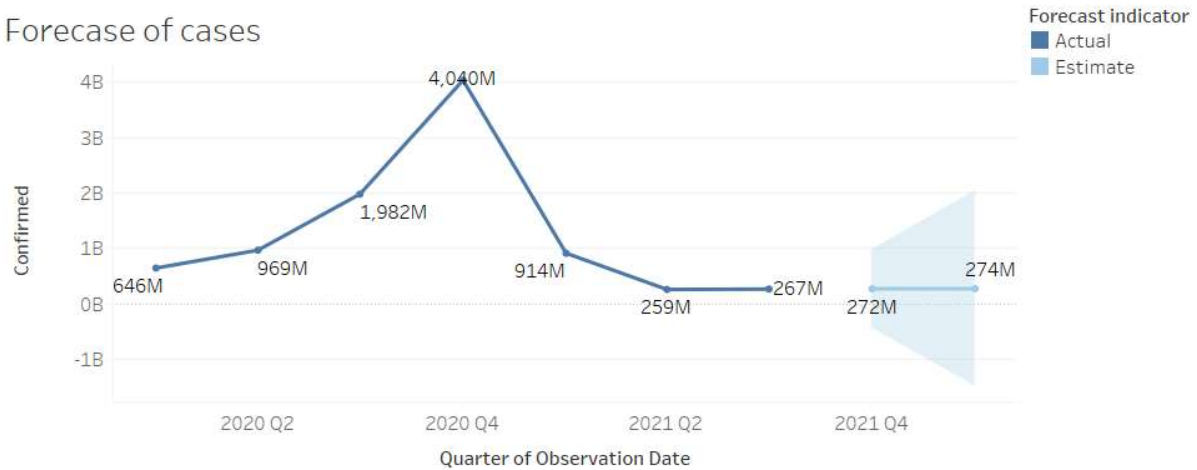
Forecast of Deaths



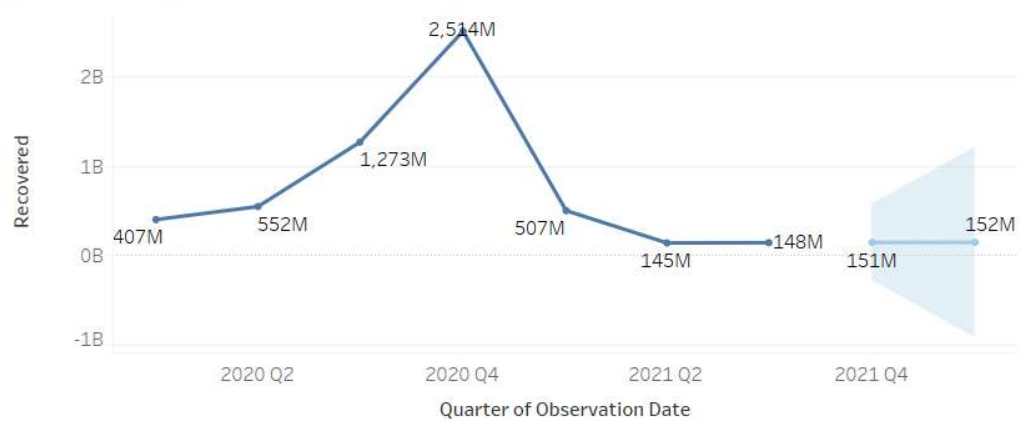
The above dashboard indicates the death rate across the world. The US has the most death rate in the world where overall 54 million deaths were recorded in that country.

The forecast of the death rate is also indicated using the line graph. The line graph data is available till 2021 second quarter and Q4 and next year's Q1 are forecasted using the line plot and the forecast shows the average of 6M deaths overall across the world.

Forecast of cases



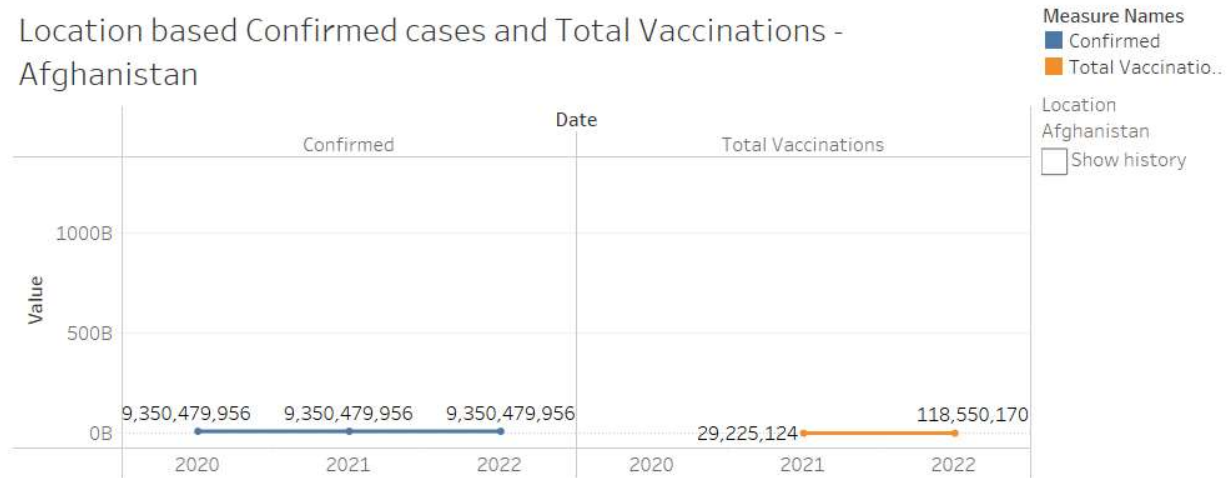
Forecast of Recovered cases



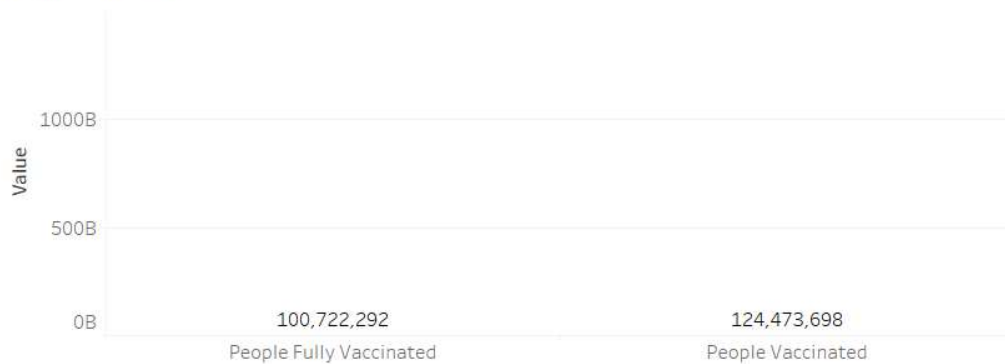
The above dashboard indicates the number of confirmed cases the peak indicates the four million data in 2020 Q4 and the forecast for the next quarter was 272 million and 274 million.

The recovered case forecast was also forecasted and in 2021 Q4 it has been forecasted a 151 million to 151 million the next quarter.

Location based Confirmed cases and Total Vaccinations - Afghanistan



People Vaccinated Vs People Fully Vaccinated by Location - Afghanistan



The location-based confirmed cases across the years are displayed in the above dashboard. It is an animation plot where when it is played the graph indicates the proper levels. The below graph indicates the people vaccinated vs fully vaccinated by location, the locations are also modified based on the filter and animation and it changes every location one by one.

Conclusion:

On exploring the data, we got very valuable insights about the covid infection rate and the geographical area where the infection rate is higher. We also explored the number of people who recovered successfully from the virus at their place of infection, and we also found a valuable insight into why India has the greatest number of recovered cases. Also, we checked the death rate due to the infection and we performed an important forecast analysis of the covid infection for the upcoming quarters. From the forecast, we got valuable information as there will be a slight increase in the number of case infections and there is also a slight increase in the number of recovers. On the other hand, the forecast for the number of deaths did not decrease and it maintained a steady 6M mark. So, the medical practitioners and the medical governing bodies have to consider this upcoming increase in mind and take necessary precautionary measures to control the spread all over the world.

References:

covid.cdc.gov/covid-data-tracker/#vaccine-delivery-coverage

<https://www.kaggle.com/datasets/imdevskp/corona-virus-report>