**Abstract:**

The analysis of COVID-19 data plays a crucial role in understanding the impact of the pandemic and formulating effective strategies for response and

mitigation. In this study, we employed Power BI, a robust business intelligence tool, to analyze and visualize COVID-19 insights. Our dataset consisted of columns such as total cases, total deaths, total recovered cases, country-wise total tests conducted, and active cases.

Prior to utilizing Power BI, we conducted exploratory data analysis (EDA) and data cleaning using Jupyter Notebook. This process allowed us to gain insights into the dataset's structure, address missing or inconsistent values, and

prepare the data for further analysis. By ensuring data quality and integrity, we aimed to derive accurate and reliable insights using Power BI.

With the cleaned dataset, we leveraged the advanced capabilities of Power BI to create interactive dashboards, reports, and visualizations. Through various charts, graphs, and maps, we visualized the progression of COVID-19 metrics over time, including total cases, total deaths, and total recovered cases.

Furthermore, we conducted country-wise comparisons, highlighting variations in testing efforts, active cases, and other relevant indicators.

The insights obtained from this analysis provide valuable information for tracking the global impact of COVID-19, identifying patterns and trends, and supporting evidence-based decision-making. By presenting the data in an

intuitive and interactive manner, Power BI facilitates effective communication of the key findings to stakeholders and the general public.

This analysis contributes to the broader understanding of the COVID-19 pandemic by shedding light on its dynamics and providing actionable insights for public health officials, policymakers, and researchers. The combination of

Power BI's powerful visualization capabilities and the meticulous data cleaning and EDA processes enhances the accuracy and reliability of the obtained insights, making them valuable for informing strategies aimed at mitigating the impact of the virus.

**CONTENTS**

1. **Introduction**
2. **Scope and Purpose of the Project**
3. **Methodology Adopted**
4. **The Dataset**
5. **Data visualization process flow**
6. **The CODE**
7. **Conclusion**

# INTRODUCTION

The COVID-19 pandemic has had a profound impact on global health and economies, making it crucial to gain insights into the data surrounding the virus. In this analysis, we utilize Power BI, a powerful business intelligence

tool, to explore and visualize COVID-19 data. Our dataset comprises various columns, including total cases, total deaths, total recovered cases, countrywise data on total tests conducted, and active cases.

Before diving into Power BI, we performed preliminary exploratory data analysis (EDA) and data cleaning using Jupyter Notebook. This allowed us to

understand the structure and quality of the dataset, identify any missing or inconsistent values, and transform the data into a suitable format for analysis. By addressing these data-related challenges, we ensure the reliability and accuracy of the insights we derive from Power BI.

With the cleaned and processed dataset at our disposal, we proceed to leverage the capabilities of Power BI. This tool enables us to create

interactive and visually appealing dashboards, reports, and visualizations, enabling a deeper understanding of the COVID-19 situation across different countries and regions.

By utilizing Power BI's rich features, we can generate informative charts, graphs, and maps that illustrate the progression of total cases, total deaths, and total recovered cases over time. Additionally, we can delve into countrywise comparisons, highlighting variations in testing efforts, active cases, and other relevant metrics.

The insights gained from this analysis will aid in tracking the global impact of

COVID-19, identifying trends, and making data-driven decisions. By visualizing the data in an intuitive and interactive manner, we can effectively communicate the key findings and contribute to a broader understanding of the pandemic's dynamics.

# SCOPE AND PURPOSE OF THE PROJECT

The scope of analyzing COVID insights is to understand and extract valuable information from the vast amount of data generated during the COVID-19 pandemic. This analysis encompasses various aspects such as

epidemiology, public health interventions, healthcare system response,

economic impact, and social dynamics. By examining COVID insights, researchers and policymakers can identify patterns, trends, and correlations to inform decision-making, develop effective strategies, and mitigate the impact of the pandemic.

The purpose of analyzing COVID insights is to inform decision-making, policy formulation, and public health strategies. By analyzing the data and insights, experts can identify patterns, risk factors, and effective

interventions to control the spread of the virus, mitigate its impact, and

develop strategies for future preparedness. The analysis of COVID insights aims to provide evidence-based recommendations for managing the

pandemic, protecting public health, and facilitating a swift and effective response to similar outbreaks in the future.

# METHODS OF DATA ANALYSIS

The dataset was retrieved from different sources which were further used for analysis and visualization methods. The methods which were used for analysis were able to track the spreading up of COVID-19

throughout the World. The data included the number of confirmed cases, recovered rate, and death rate in different countries.

## Cleaning the dataset

The very first step in our project is to get a reliable and authentic dataset for the prediction and analysis. Our search for dataset ended on which is kaggle which has provided dataset for free use and is authentic. Then next thing we did was to clean the dataset and remove unwanted columns from dataset for faster computation.

## Exploratory Data Analysis (EDA)

The project report includes a basic Exploratory Data Analysis (EDA) and data cleaning using Jupiter Notebook. The EDA involves examining the dataset's structure, identifying missing values, and exploring the

distributions and relationships of variables. Data cleaning involves handling missing data, removing duplicates, and standardizing data formats. These processes ensure the data is prepared for further analysis and provide a foundation for extracting meaningful insights related to COVID-19.

# Data Visualization

The retrieved data from different sites can be used to track the status of the corona. we use the dataset and check the consistency of the

dataset by checking the values out of the dataset randomly. Then we do data visualization for better understanding of data by the use of various plots, graph and heatmaps. All this graphs and plots gets us an insight into huge datasets easily.

## ABOUT DATASET

Coronavirus disease (COVID-19) is an infectious disease caused by the SARSCoV-2 virus.

Most people infected with the virus will experience mild to moderate respiratory illness and recover without requiring special treatment. However,

some will become seriously ill and require medical attention. Older people and those with underlying medical conditions like cardiovascular disease, diabetes,

chronic respiratory disease, or cancer are more likely to develop serious illness. Anyone can get sick with COVID-19 and become seriously ill or die at any age.

The best way to prevent and slow down transmission is to be well informed about the disease and how the virus spreads. Protect yourself and others from infection by staying at least 1 metre apart from others, wearing a properly

fitted mask, and washing your hands or using an alcohol-based rub frequently. Get vaccinated when it is your turn and follow local guidance.

The virus can spread from an infected person’s mouth or nose in small liquid particles when they cough, sneeze, speak, sing, or breathe. These particles

range from larger respiratory droplets to smaller aerosols. It is important to

practice respiratory etiquette, for example by coughing into a flexed elbow, and to stay home and self-isolate until you recover if you feel unwell.

Where are cases still high?

Daily global cases fell after a spike in the spring but are now rising again, with the emergence of the BA.4 and BA.5 subvariants of the Omicron variant.

Studies suggest that Omicron - which quickly became dominant in numerous countries - is milder than the Delta variant, but far more contagious. The subvariants are even more contagious.

## Data visualization process flow

1. Determine the decision you want to make
2. Identify the metrics that inform the decision
3. Develop the story you want to tell
4. Select the appropriate visual
5. Add relevant elements to the visual
6. Clearly label and review the visual
7. Let a nonexpert review the visual

### 1. Determine the decision you want to make

“One of the biggest pitfalls in data visualization is people worrying too much about making the visuals look a certain way. The important work happens long before that point,” says Cook. In other words, don’t get wrapped up in colors and other aesthetics too soon.

Your first step is figuring out what decision you’re trying to make. You can have all the data in the world, but it won’t mean much if you’re not sure what to do with it. Cook recommends posing the decision in the form of a question so you’re clear on the answer you’re seeking. “If you aren’t clear on your decision, your visual won’t be either,” he explains.

Here’s an example of a clear decision question: During which fiscal quarter should we launch our new product?

#### 2. Identify the metrics that inform the decision

You likely have tons of data available to you, but only certain data points will be relevant to your decision. Before getting overwhelmed by data sets, consider which specific points would be most helpful for answering your decision question.

Once you identify the right metrics, determine whether you can collect them with any accuracy. You may find that some data points either aren’t available or are inaccurate. In this case, you typically have two alternatives: Kick off a project to collect the data (such as developing and distributing a survey) or revisit the first step and adjust your question.

#### 3. Develop the story you want to tell

Next up is developing a story from your data. Cook shares a few questions you can use to prepare your narrative:

* Is the data about comparison? You may be deciding based on metrics being bigger or smaller — or faster or slower.
* Is the data about changes over time? Your decision may concern entering a new market or tracking product launch performance over time.
* Is the data about categorization? You may have a cost-based decision that needs to identify where the business is losing money.

#### 4. Select the appropriate visual

This part of the data visualization process is fairly simple, as most visuals naturally follow the type of story you want to tell. Consider these examples:

* Comparison stories typically work best with bar graphs.
* Time-based stories pair well with line charts.
* Categorical stories typically necessitate tree charts.

#### 5. Add relevant elements to the visual

“Now is the point in the data visualization process when you can focus on aesthetics,” says Cook. The purpose of this step is to make choices about your visual that aid in not only its appeal but also fostering comprehension.

You may need to add callouts to your chart to emphasize certain data points or add important context. For instance, say you created a chart that was missing a week of sales data. The audience may assume you made a mistake, but you

didn’t include the data for good reason — a hurricane caused the business to close that week. A well-placed callout can prevent this confusion.

Color decisions can benefit from a designer’s eye — and some common sense.

For example, people often associate red with negativity (recall the saying about sales being “in the red”). So if your chart is sharing good news, you may want to avoid using that color.

#### 6. Clearly label and review the visual

Where the previous step was about choosing visualization elements, this step is about making note of the choices you made. Title the visual appropriately.

Make sure units are correct (e.g., dollars vs euros) and incremented consistently. Ensure there’s a legend to explain color meanings.

“Here you’re just making sure the audience doesn’t have unnecessary questions about what they’re viewing,” Cook explains.

#### 7. Let a nonexpert review the visual

“The last step of the data visualization process is quite important. You need a different set of eyes on the visual you’ve created — preferably eyes that don’t have the same knowledge or experience as your own,” says Cook.

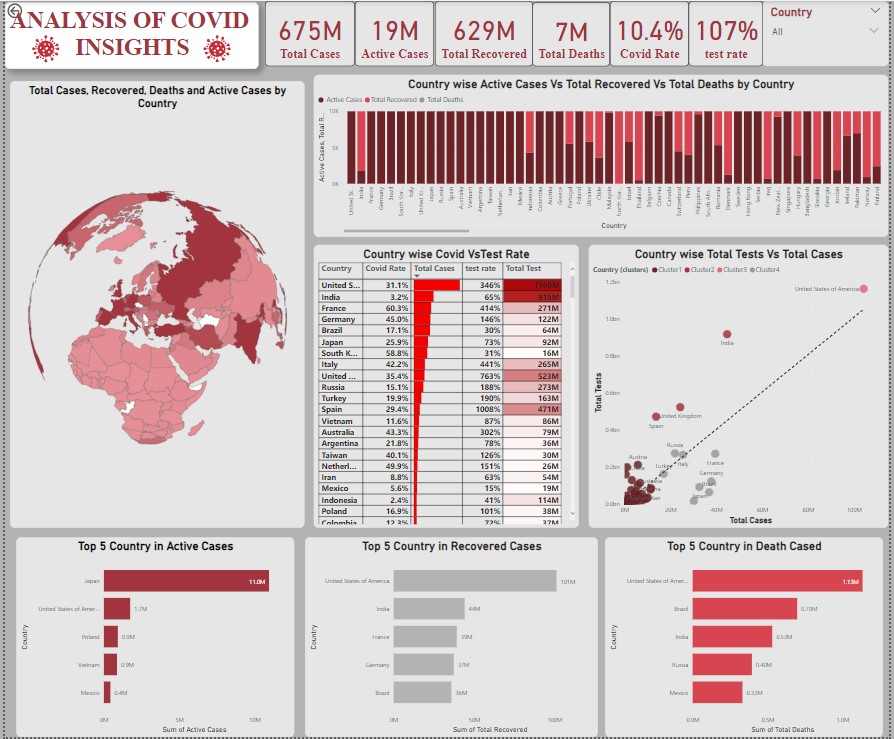
Giving your visual to someone else to review, especially someone who doesn’t know much about the subject matter or underlying data, is an important spot check. Ideally, they should be able to comprehend the story you’re trying to communicate without any issues.

If they have any trouble, Cook says you may need to go back a few steps. The most common problem is using the wrong type of chart for the data you’re presenting. Otherwise, you may just need to add a callout or two to fill in any blanks in the visual narrative.

“But if you’ve followed these steps carefully and dedicated a reasonable amount of time to the task, you should be set,” Cook says.

**RESULTS**

**Main Dashboard: -**





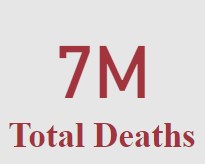
The first card visualization displays the total number of COVID-19 cases, which stands at a staggering 675 million.



The second card visualization displays the total active cases, which stands at 19 million.



The third card visualization represents the number of recovered cases, which has reached a total of 629 million.



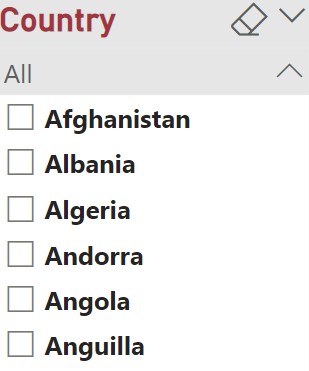
The fourth card visualization represents the total deaths due to COVID-19, which has reached 7 million cases.



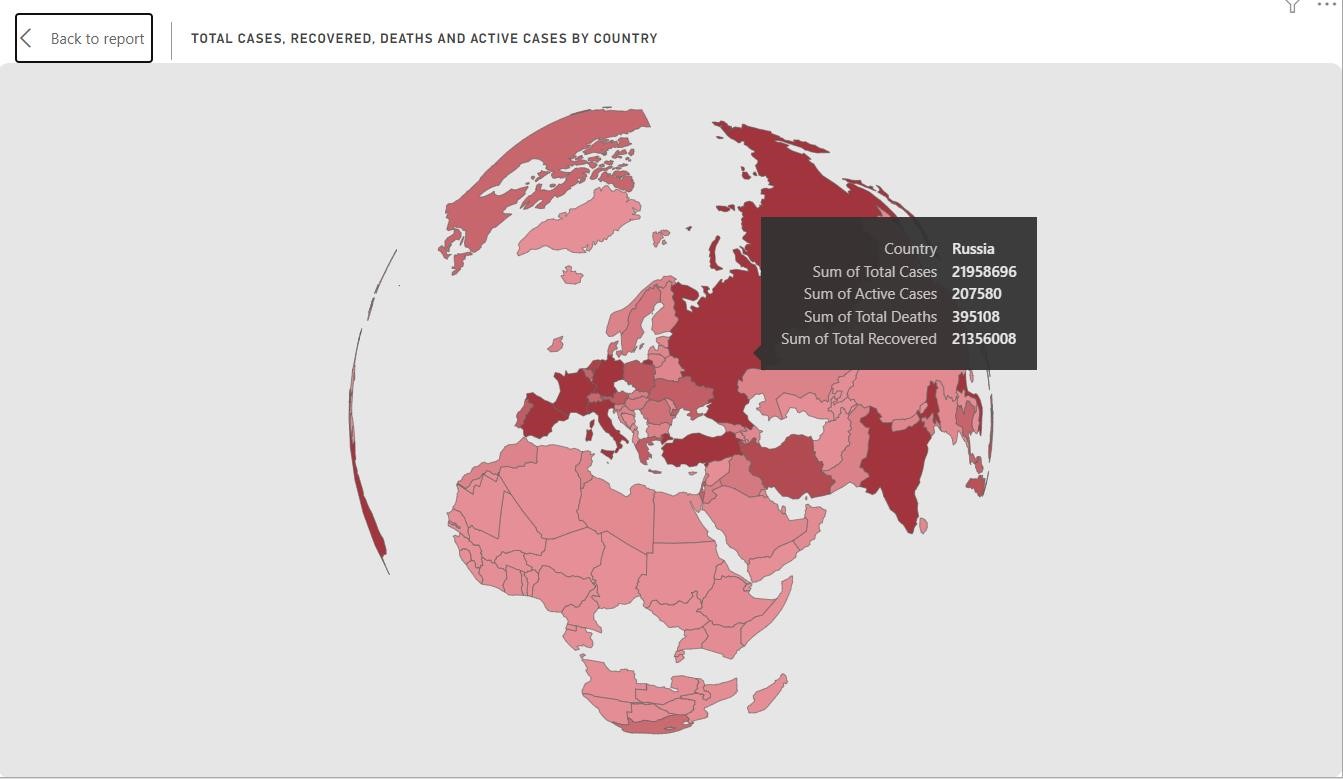
The fifth card visualization illustrates the COVID rate, which is currently at 10.4% worldwide. This indicates the proportion of the population affected by COVID-19.



The sixth card visualization represents the COVID test rate, which is currently recorded at 107% worldwide.



The Power BI slicer visualization is a powerful tool used for filtering and selecting specific data within a report. In this case, the field being displayed is "countries," indicating that the slicer allows users to choose and filter data based on different countries. The slicer style is set to "dropdown," which means that users can select a country from a dropdown menu. This provides a user-friendly experience, allowing users to easily navigate and choose specific countries of interest to analyse the data further. The dropdown style offers a compact and organized way to present multiple country options, making it convenient for users to filter data based on their preferences.



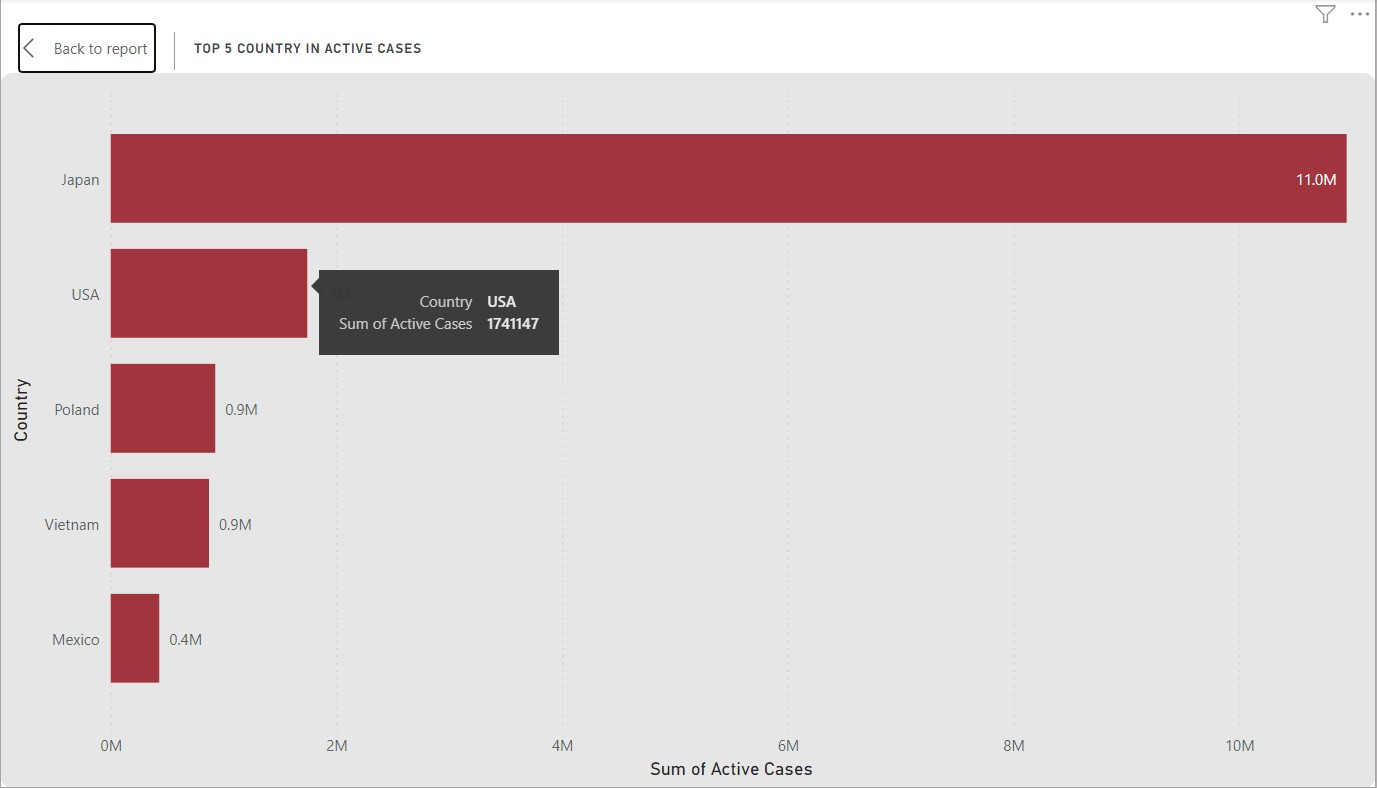
The Power BI shape map is a visualization that represents data on a map. In this case, it is utilized to analyse COVID-19 data. The location field is set to countries, allowing the data to be displayed based on countries' geographical boundaries.

To provide visual insights, the colour saturation is determined by the total cases, with darker shades of red indicating higher COVID-19 cases. The tooltips feature provides additional information, displaying the active cases, total deaths, and total recoveries when hovering over a specific country.

Analysing the map, it can be concluded that the United States of America has a significant number of COVID-19 cases, as it appears as a dark red country. The top five countries with the highest COVID-19 cases, based on colour saturation, are the United States of America, India, France, Germany, and Brazil.

This analysis indicates that these countries have been severely affected by the pandemic, experiencing many COVID-19 cases. The United States of America stands out as having the highest number of cases, followed by India, France, Germany, and Brazil.

By visualizing this information on the map, it becomes easier to comprehend the global distribution of COVID-19 cases and identify countries that require particular attention in terms of managing the pandemic.



The Power BI stacked bar chart visualizes COVID-19 data by representing countries on the y-axis and the sum of active cases on the x-axis. By applying a filter, the chart displays the top 5 countries with the highest active cases.

According to the current data, the top 5 countries with active COVID-19 cases are:

Japan: With approximately 11 million active cases.

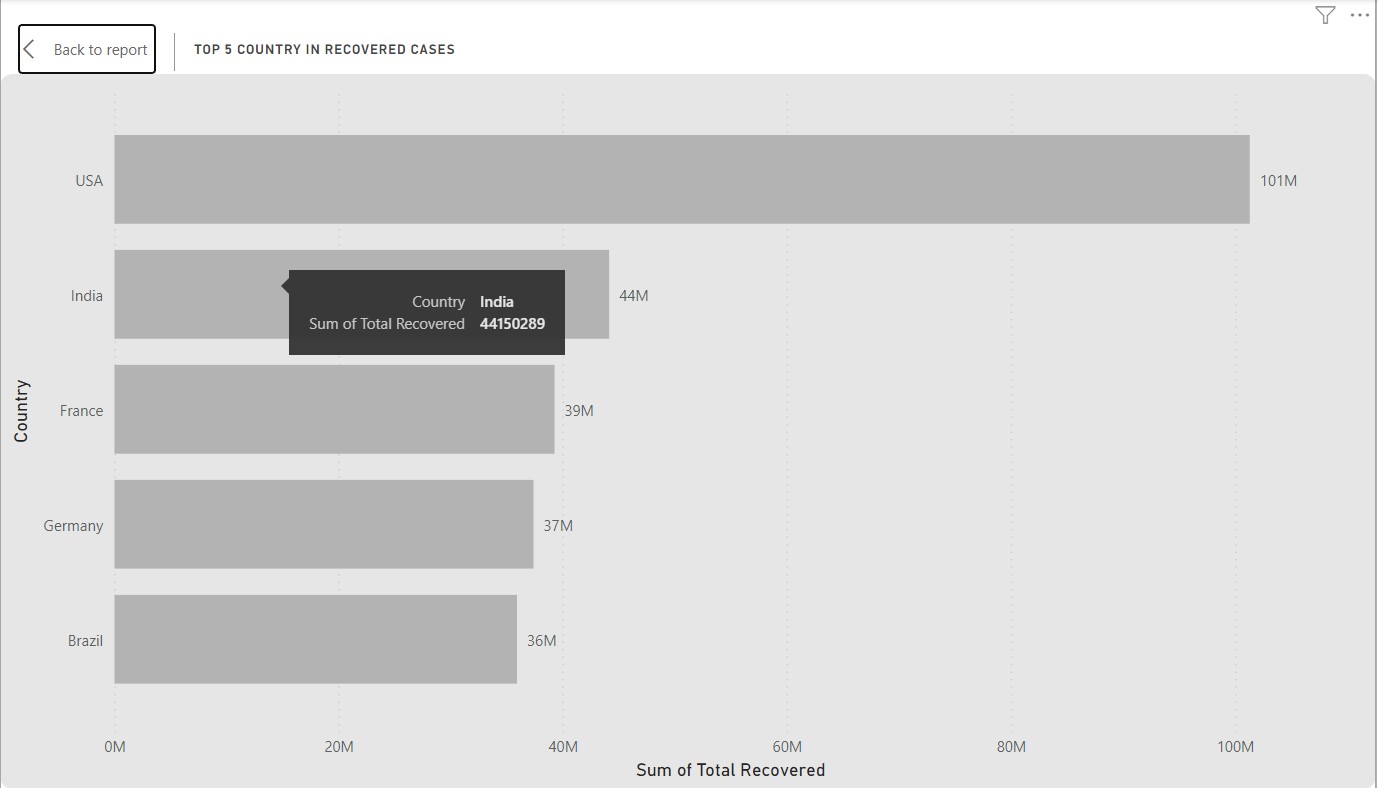
United States of America: Having around 1.7 million active cases.

Poland: Reporting approximately 0.9 million active cases.

Vietnam: Exhibiting approximately 0.8 million active cases.

Mexico: Reporting approximately 0.4 million active cases.

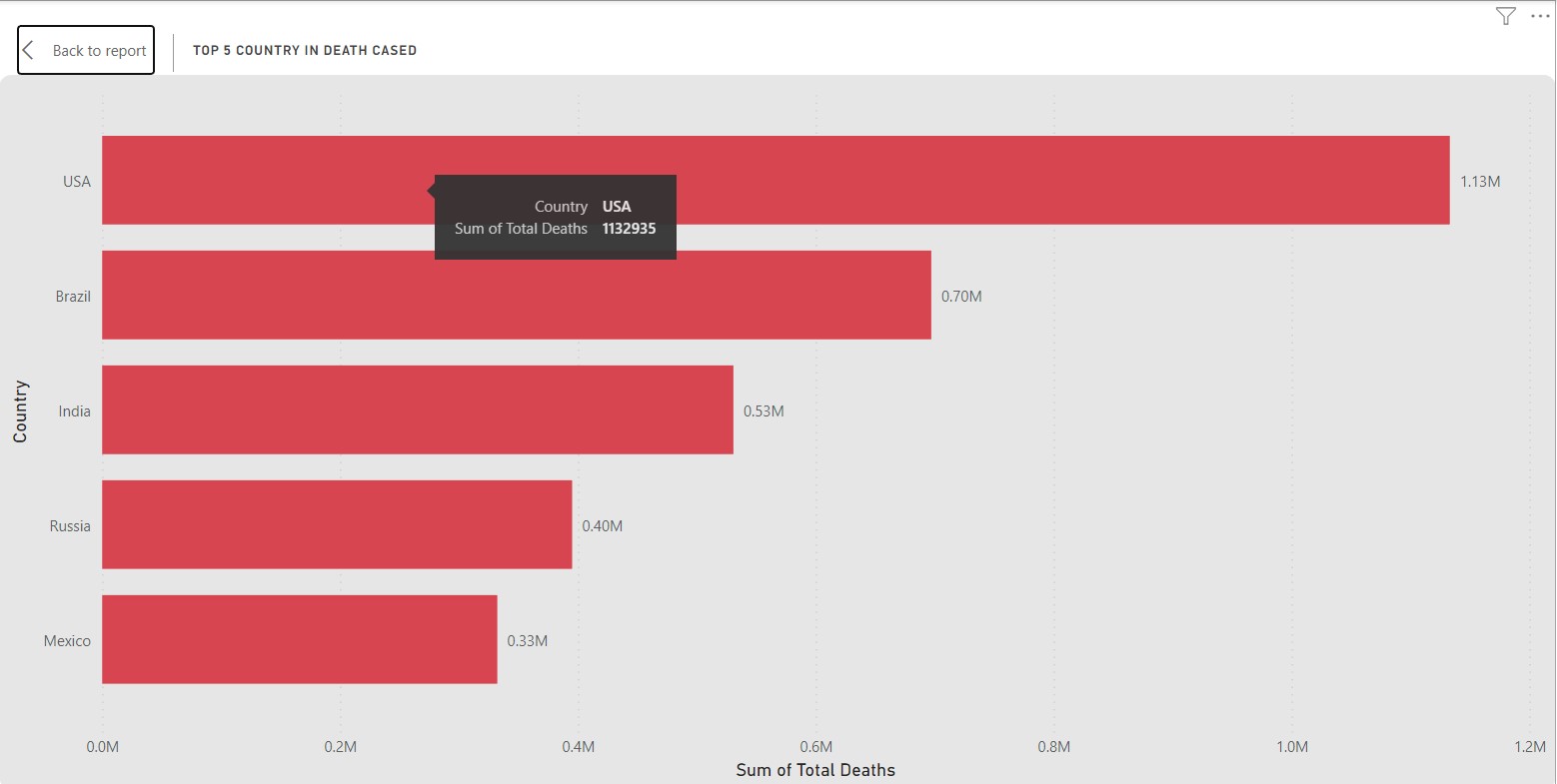
This stacked bar chart allows for a visual comparison of the active cases among these top 5 countries. The height of each stacked bar represents the total number of active cases, with different segments indicating the contribution of each country to the overall count.



The Power BI stacked bar chart is a visualization used to display the relationship between two variables by stacking bars on top of each other. In this case, COVID-19 data is used, with countries being placed on the y-axis and the sum of recovered cases on the x-axis.

A filter is applied to display only the top 5 countries with the highest number of recovered cases. The United States of America has the highest number of recovered cases at 101 million, followed by India at 44 million, France at 39 million, Germany at 37 million, and Brazil at 36 million.

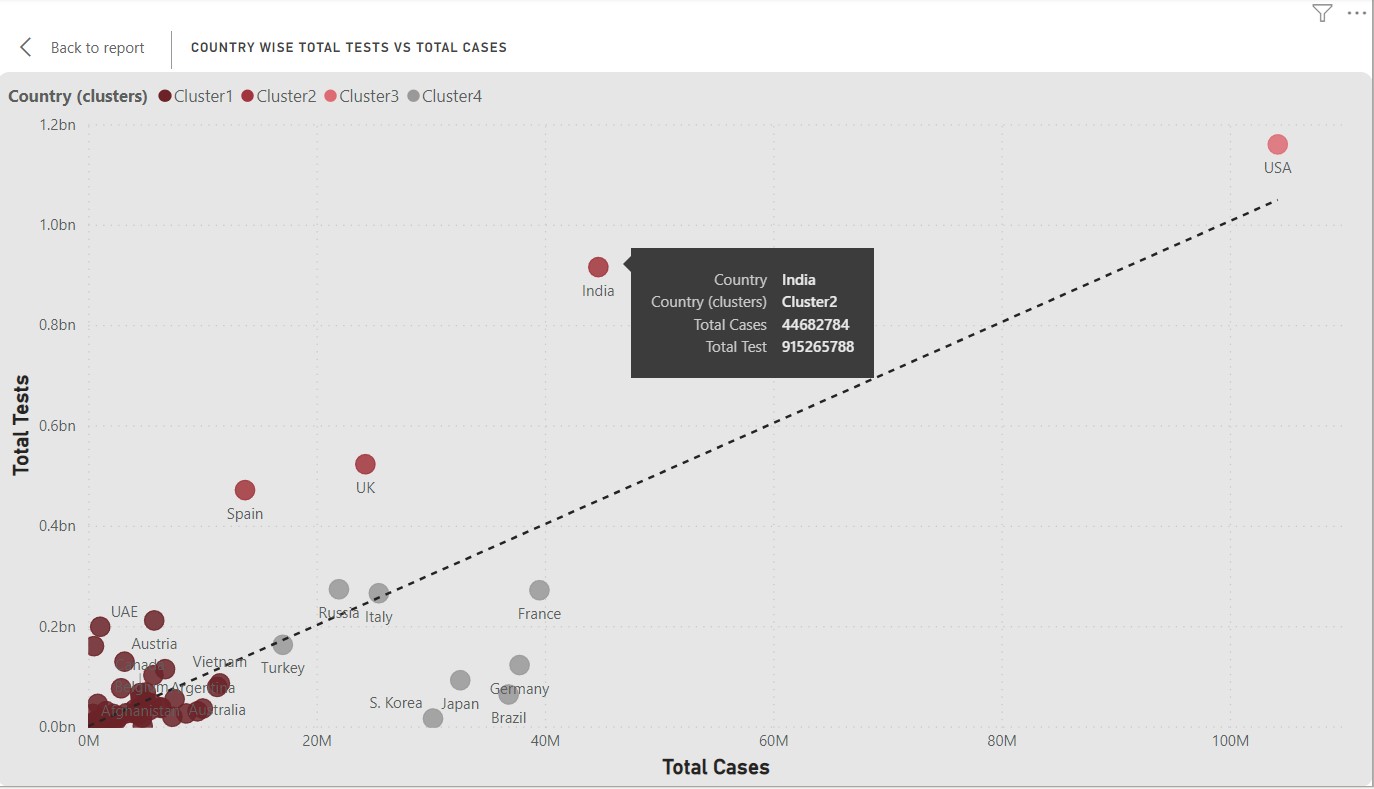
This visualization provides an effective way to compare the recovery status of different countries affected by COVID-19. It allows users to easily see the countries with the highest number of recovered cases and to analyse how different countries are responding to the pandemic. The stacked bar chart also highlights the gap between countries with high and low recovery rates, which is a critical factor in understanding the impact of the pandemic on different nations.



The Power BI stacked bar chart is a visualization that represents data using horizontal bars stacked on top of each other. In this case, the chart is based on COVID-19 data. The countries are placed on the y-axis, and the sum of total death cases is displayed on the x-axis.

A filter is applied to view the top 5 countries with the highest death cases. The United States of America has recorded 1.13 million death cases, making it the country with the highest death toll. Brazil follows with 0.70 million death cases, while India has 0.53 million death cases. Russia has reported 0.40 million death cases, and Mexico has recorded 0.33 million death cases.

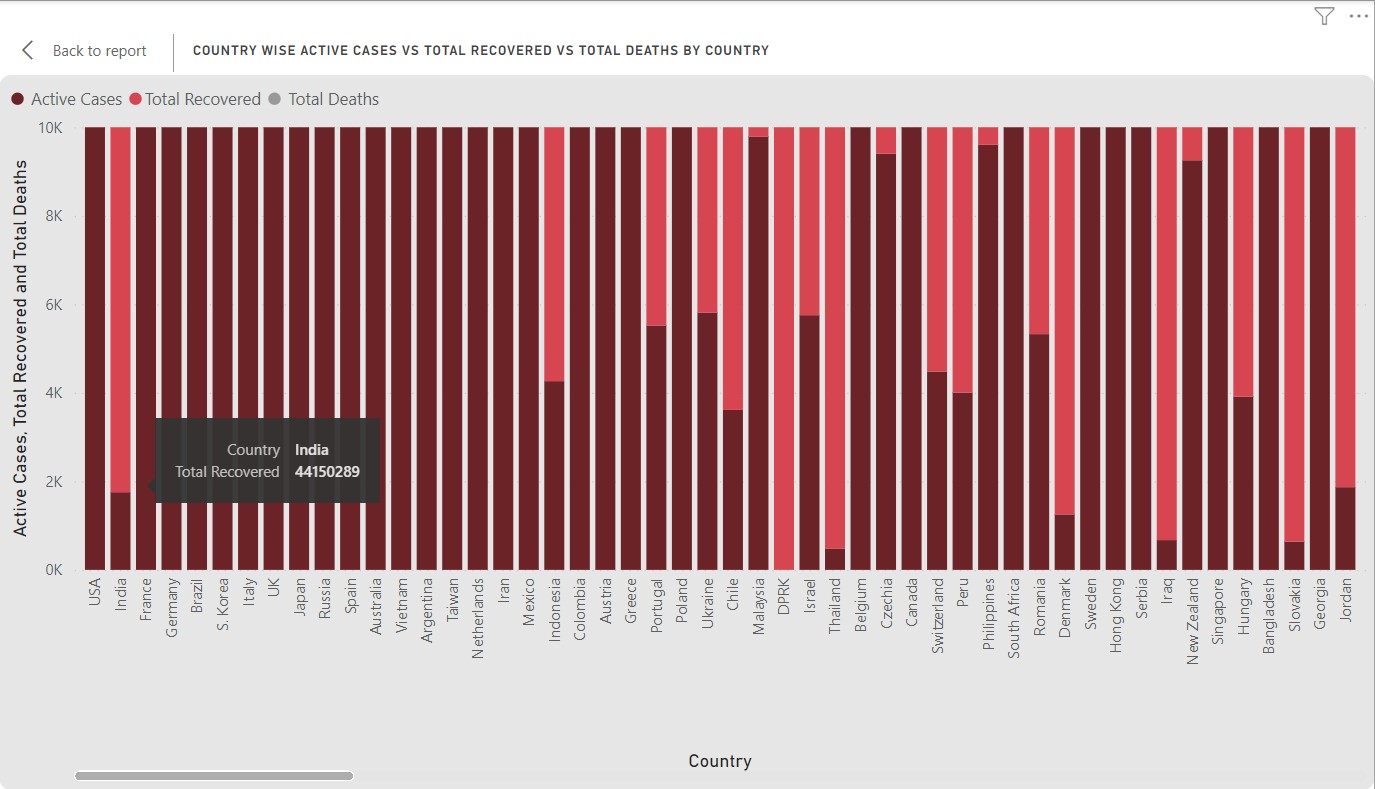
By utilizing the stacked bar chart, this visualization provides a clear comparison of the top 5 countries based on their death cases, allowing for easy identification of countries with the highest fatality rates due to COVID-19.



The Power BI scatter chart is a visual representation that uses dots to display data points on a two-dimensional plane. In this case, the COVID-19 data is used, with countries being assigned to the values field. The sum of total cases is plotted on the x-axis, and the total tests conducted is plotted on the y-axis.

By applying the auto cluster function, Power BI automatically groups similar data points together based on their proximity. This helps identify clusters or patterns in the data, such as countries with similar COVID-19 testing and case statistics. The auto cluster function helps users analyse the data more effectively by visually highlighting potential clusters or trends.

Additionally, a trend line can be applied to the scatter chart. The trend line represents the general direction and relationship between the total cases and total tests variables. By using the trend line, users can identify any correlation or pattern between these variables. In this case, with a trend line applied, it would help visualize the overall relationship between the total cases and total tests conducted across different countries.

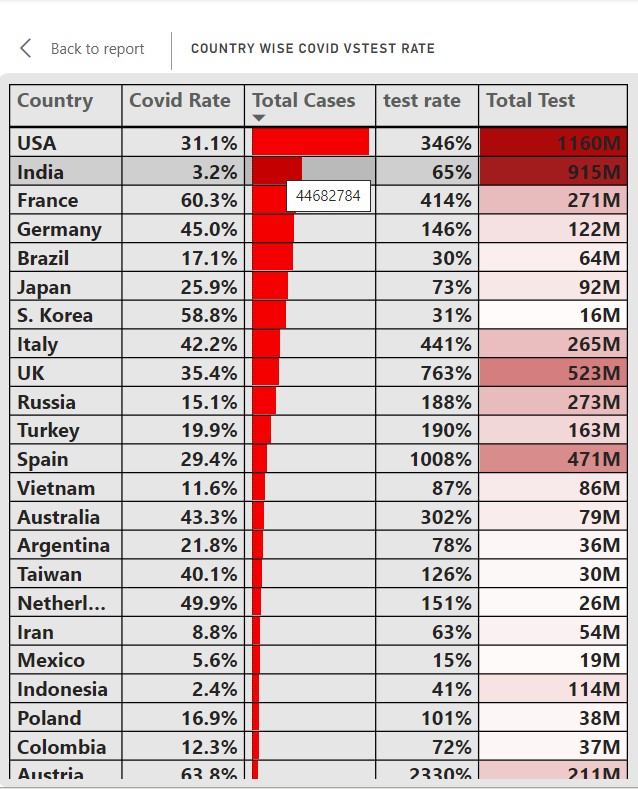


The Power BI stacked column chart is a visual representation that allows comparison of different categories on the x-axis, using vertical columns that are stacked on top of each other. In this scenario, COVID-19 data is utilized. The yaxis represents countries, while the x-axis displays three categories: total active cases, total recovered cases, and total deaths.

Each country is represented by a column, and the height of the column indicates the magnitude of the specific category. The total active cases, total recovered cases, and total deaths are depicted as segments within each column, with the segments stacked on top of one another.

This visualization enables easy comparison between countries and provides insights into the distribution of active cases, recoveries, and deaths. It allows for identifying patterns, trends, and relative proportions of the COVID-19 situation across different countries.

Users can interact with the chart by selecting specific countries or categories to focus on, which dynamically updates the visualization, facilitating a deeper analysis of the COVID-19 data. The stacked column chart is an effective tool for presenting and understanding the comparative COVID-19 statistics by country.



The table visualization in Power BI is a tabular representation of data that allows for organized display and analysis. In this case, the table is created using COVID-19 data. The columns consist of country data, COVID rate, total cases, test rate, and total tests.

Conditional formatting is applied to enhance the visual impact and highlight specific information. For instance, the total COVID cases column is formatted using a red colour bar, where the length of the bar corresponds to the magnitude of total COVID cases. This visual cue draws attention to countries with higher case counts.

Additionally, the total test column incorporates background colour based on the total test values. Dark red represents high total test numbers, indicating thorough testing, while lighter shades of red signify lower total test numbers, suggesting limited testing.

By leveraging conditional formatting, the table visualization offers a quick and intuitive overview of country-specific COVID-19 data, allowing users to identify countries with significant case numbers and variations in testing rates. It aids in identifying potential areas of concern and provides a comprehensive view of the COVID-19 situation across different countries.

**CONCLUSION**

In conclusion, the analysis of COVID-19 insights coupled with data visualization using power bi has provided a comprehensive and impactful understanding of the pandemic. Through the use of visual representations, key patterns, trends, and impacts of the virus have become more accessible and digestible.

The data visualizations have enabled us to witness the global spread of the virus, highlighting the significance of geographic location, population density patterns in its transmission. By visualizing case counts ,test rates and mortality rates over time, we have been able to grasp the scale and severity of the pandemic.

Furthermore, data visualization has allowed us to uncover demographic disparities in COVID-19 outcomes. By representing infection rates and mortality rates conditions, we have identified vulnerable populations and directed resources and interventions accordingly. Visualizations have also shed light on the unequal impact of the virus on different socioeconomic groups, emphasizing the need for equitable healthcare and support systems.

However, it is essential to acknowledge the challenges and limitations of data visualization in the context of COVID-19. The accuracy and reliability of data sources have varied, which can introduce biases and inaccuracies in the visual representations. Additionally, visualizations alone cannot provide a complete picture of the pandemic and should be complemented by rigorous analysis and interpretation.