## COVID VACINESS ANALYSIS

PROBLEM DEFINITION

The problem is to conduct an in-depth analysis of Covid-19 vaccine data, focusing on vaccine efficacy, distribution, and adverse effects. The goal is to provide insights that aid policymakers and health organizations in optimizing vaccine deployment strategies. This project involves data collection, data preprocessing, exploratory data analysis, statistical analysis, and visualization.

ABSTRACT

The COVID 19 pandemic caused due to the Corona virus devastated the world by causing several fatalities around the world. This virus originated in Wuhan, China in 2019 and was later spread throughout the world due to human contact in one way or the other. The disease showed symptoms as basic as mild fever and cold but also caused life threatening symptoms like breathing problems caused by damage to the lungs. As this virus was new to the world and there was no vaccine or cure to it at the initial period there were several deaths around the world.

This Project mainly aims to find out the vaccinations around the world for the prevention of the Covid 19 pandemic and how much has been achieved so far.

The countries around the world were forced to shut themselves to others in order to avoid the further spread of the virus and people were stuck inside their houses and faced many issues with their finances, mental health etc., and felt like animals in a cage. An effort was made to find a cure or vaccine by several health organizations to bring a stop to this pandemic

DATA COLLECTION

**Vaccine Distribution**: Tracking the quantity and location of vaccine doses distributed to healthcare providers and vaccination centers.

**Vaccine Administration**: Recording the number of doses administered, including details like date, location, and recipient demographics.

**Safety Monitoring**: Continuously monitoring and collecting data on adverse events following vaccination (AEFIs) to ensure vaccine safety.

**Efficacy Studies**: Conducting clinical trials and post-marketing surveillance to assess the effectiveness of vaccines in preventing COVID-19.

**Vaccine Coverage**: Determining the percentage of the population vaccinated to gauge progress toward herd immunity.

EXPLORATORY DATA ANALYSIS

**Vaccine Distribution**: Tracking the quantity and location of vaccine doses distributed to healthcare providers and vaccination centers.

**Vaccine Administration**: Recording the number of doses administered, including details like date, location, and recipient demographics.

**Safety Monitoring**: Continuously monitoring and collecting data on adverse events following vaccination (AEFIs) to ensure vaccine safety.

**Efficacy Studies**: Conducting clinical trials and post-marketing surveillance to assess the effectiveness of vaccines in preventing COVID-19.

**Vaccine Coverage**: Determining the percentage of the population vaccinated to gauge progress toward herd immunity.

**Variants Monitoring:** Collecting data on the prevalence and impact of COVID-19 variants on vaccine efficacy.

**Patient Records**: Ensuring proper record-keeping of vaccine recipients' information for follow-up doses and adverse event management.

Data collection is typically performed by healthcare agencies, vaccine manufacturers, research institutions, and government health departments, and it plays a crucial role in managing the COVID-19 pandemic.

STATSTICAL ANALAYISIS

**Data Collection**: Gather data on vaccine administration, infection rates, and outcomes. This data can come from clinical trials, real-world studies, and healthcare records.

**Define Metrics**: Identify what you want to measure, such as vaccine efficacy (preventing infection) or vaccine effectiveness (real-world impact on reducing illness and transmission).

**Study Design**: Choose the type of study - randomized controlled trials, observational studies, or ecological studies. Ensure the study design accounts for confounding variables.

**Comparative Analysis**: Compare infection rates, hospitalization rates, and mortality among vaccinated and unvaccinated groups.

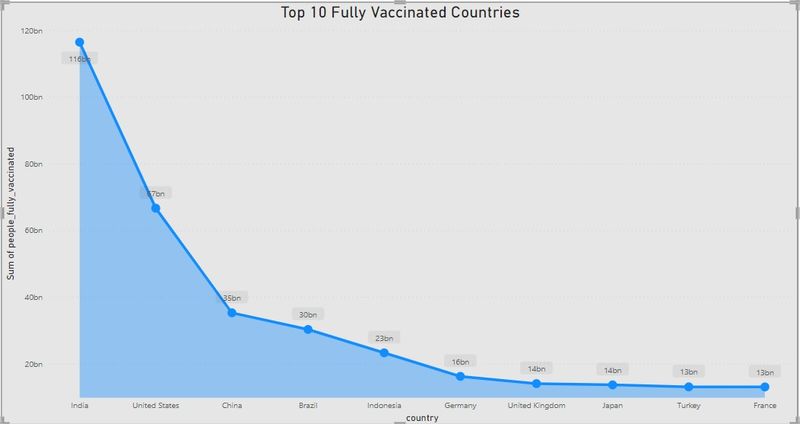
**Adjust for Confounders**: Account for variables like age, comorbidities, and geographic location that can influence outcomes.

**Time Analysis:** Evaluate vaccine effectiveness over time, considering variants and changes in vaccine distribution.

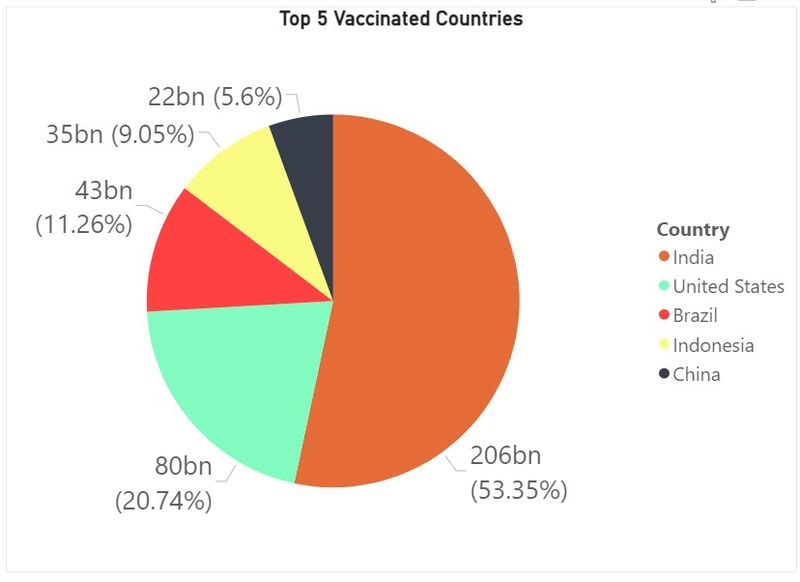
VISUALIZATION

In the analysis part first | have analyzed the top 10 fully vaccinated countries by using area chart and have used the filter option to find the top countries and the result obtained as below,

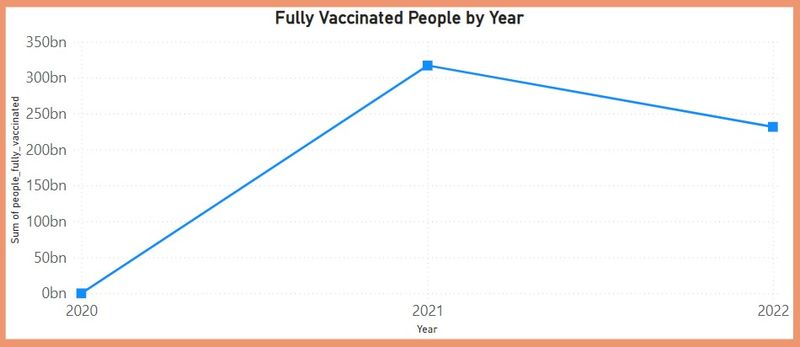
From the below image we can able to come to know that India is the top country in terms full vaccination with 116 billon , followed by united states of America and china with 67 billion and 35 billion respectively.



In the second analysis we have analyzed the top 5 vaccinated countries with the help of pie chart and used filter option to find the top countries and with that we came to know that India is the top country with more number of vaccinated peoples followed by United States of America and Brazil.



Fully vaccinated people by the year



INSIGHTS AND RECOMMENDATIONS

**Insight 1**: Vaccine Coverage Disparities

**Insight**: There are disparities in vaccination rates among different demographic groups, with some populations being under-vaccinated.

**Recommendation**: Implement targeted outreach and vaccination campaigns to address these disparities. This may include setting up vaccination sites in underserved areas, providing transportation assistance, and culturally sensitive education efforts.

**Insight 2:** Vaccine Hesitancy

**Insight:** Vaccine hesitancy remains a challenge, leading to lower vaccination rates.

**Recommendation**: Develop and disseminate evidence-based educational materials to address vaccine hesitancy concerns. Engage with community leaders, healthcare providers, and influencers to promote vaccine confidence.

**Insight 3**: Vaccine Efficacy

**Insight:** Ongoing monitoring shows variations in vaccine efficacy against new COVID-19 variants.

**Recommendation:** Continue genomic surveillance to track variant prevalence and adapt vaccination strategies as needed, such as booster doses or updated vaccines targeting emerging variants.

**Insight 4**: Adverse Events

**Insight**: Data shows a low rate of severe adverse events following vaccination.

Recommendation: Maintain robust adverse event monitoring systems, and ensure transparent reporting of vaccine safety data to build and maintain public trust.

**Insight 5**: Booster Doses

**Insight**: Evidence suggests that booster doses may be necessary to maintain long-term immunity.

**Recommendation**: Plan and implement booster shot campaigns for eligible populations, prioritizing those at higher risk of severe disease or waning immunity

**COVID VACINESS ANALYSIS**

**INNOVATION:**

**Step 1**: Define the Objectives

Clearly state the objectives of your innovation. What specific problem are you trying to solve in COVID vaccine analysis? Is it related to vaccine distribution, efficacy, or any other aspect?

**Step 2:** Research and Data Collection

Gather relevant data and information about COVID vaccines, including vaccine types, distribution, effectiveness, and side effects. Use credible sources like scientific journals, health organizations, and government reports.

**Step 3**: Ideation

Brainstorm innovative ideas for your analysis. Consider new approaches, technologies, or methodologies that could improve our understanding of COVID vaccine-related issues.

**Step 4**: Concept Development

Refine your ideas and develop detailed concepts. What methods or tools will you use for analysis? Are there any emerging technologies or trends you can leverage?

**Step 5**: Feasibility Assessment

Evaluate the feasibility of your innovation. Consider factors like resource availability, budget, and technology requirements. Assess whether your concept can realistically be implemented.

**Step 6**: Prototyping and Testing

Create a prototype or a proof-of-concept model for your innovation. Test it with a small-scale analysis to ensure it works as intended.

**Step 7**: Data Analysis and Interpretation

Apply your innovation to COVID vaccine data. Analyze the results and interpret findings. Identify trends, correlations, or insights that were not apparent before.

**Step 8:** Documentation

Create a detailed document that includes:

Introduction: Explain the background and purpose of the analysis.

Methodology: Describe the methods and tools used for the analysis.

Results: Present the findings, including data visualizations and key insights.

Discussion: Interpret the results, discuss implications, and suggest potential applications.

Conclusion: Summarize the innovation's impact and its relevance to COVID vaccine analysis.

**Step 9:** Peer Review

Share your document with experts in the field for peer review. Incorporate feedback and refine your analysis if needed.

**Step 10**: Presentation and Sharing

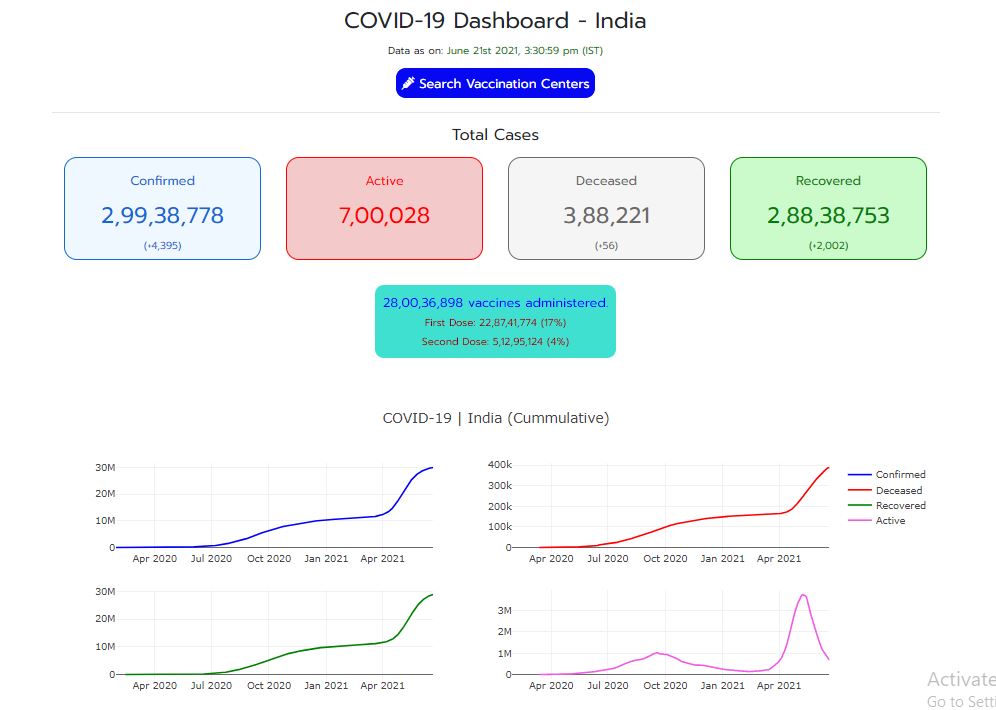
Present your innovation and findings to a relevant audience, such as researchers, healthcare professionals, or policymakers. Share your document and encourage discussion and collaboration.

**Step 11**: Assessment and Feedback

Collect feedback from your audience and assess the real-world impact of your innovation on COVID vaccine analysis.

**Step 12**: Iteration and Improvement

Use the feedback to make necessary improvements to your innovation. Continue refining your analysis and methodology for ongoing relevance.



**STEPS TO COLLECT AND PREPROCESS DATA**

**1. Data Collection:**

a. Data Sources:

Official Health Organizations: Websites of health organizations like the World Health Organization (WHO) or the Centers for Disease Control and Prevention (CDC) often provide reliable data.

Government Health Portals: Government health websites of different countries might provide region-specific data.

Kaggle Datasets: Kaggle is a platform where you can find a variety of datasets, including COVID-19 data.

**b. Data Variables**:

Vaccine Type: Specify which vaccines you are interested in (Pfizer, Moderna, AstraZeneca, etc.).

Vaccination Rates: Number of people vaccinated, both first and second doses.

Location Data: Breakdown by countries, states, or regions.

Time Period: Daily, weekly, or monthly data, depending on your analysis requirements.

2. Data Preprocessing:

a. Data Cleaning:

Handling Missing Values: Check for and handle missing or null values in the dataset. This might involve interpolation, removal, or replacement with appropriate values.

Data Validation: Ensure the data values fall within expected ranges.

**b. Data Transformation:**

Normalization/Standardization: Scale numerical features if necessary.

Encoding Categorical Variables: Convert categorical variables (like vaccine types) into numerical representations for analysis.

**c. Feature Engineering:**

Derive New Features: Create new features if necessary. For example, you might calculate vaccination rates based on the available data.

Date-Time Features: Extract features like day of the week, month, or year from the date if your dataset includes timestamps.

**d. Data Aggregation:**

Aggregate Data: Depending on your analysis, you might need to aggregate data at different levels (daily to weekly, country level to continent level).

**e. Data Visualization:**

Explore Data: Visualize the data using charts (line charts, bar charts, heatmaps) to understand patterns and trends.

Outlier Detection: Identify and handle outliers if they exist in the data.

**f. Data Splitting:**

Training and Testing Sets: If you're building predictive models, split the data into training and testing sets.

**g. Data Saving:**

Save Processed Data: Save the cleaned and preprocessed data into a separate file for further analysis. This ensures you can work from the same cleaned dataset for future analysis without repeating preprocessing steps.

**Python code**

import pandas as pd

# Load the data from CSV file

##### data = pd.read\_csv(' country\_vaccinations.csv')

# Data Cleaning

data = data.dropna()

# Drop rows with missing values

# Data Transformation

# Assuming 'Vaccine\_Type' is categorical, encode it into numerical values

data['Vaccine\_Type'] = pd.Categorical(data['Vaccine\_Type']).codes

# Data Aggregation (Example: Sum the number of vaccinated people per vaccine type)

aggregated\_data=data.groupby('Vaccine\_Type')['Number\_of\_Vaccinated'].sum().reset\_index()

# Data Saving (Optional)

aggregated\_data.to\_csv('aggregated\_vaccine\_data.csv', index=False)

# Print the preprocessed data

print(aggregated\_data)

**.What is EDA and why is this important?**

EDA is an approach to understand the various aspect of the data. The primary aim with EDA is to examine the distribution of the data, reveal an underlying structure, detect outliers, find the hidden correlation and relationships among the data, among others. EDA is an initial step in the data analysis process after data collection where the data is summarised, visualised, plotted to get insight into the dataset.

*The goal of EDA is to discover patterns in data.*

**How to perform EDA?**

1. *Summary Statistics*

These are measurements to describe data. We can use df.describe() function to get various summary statistics. However, this process excludes NaN values and categorical values. To include description for categorical data we need to use df.describe(include= "all").

2. *Visualise*

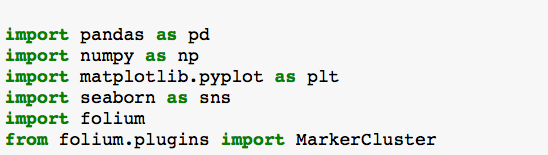
Different visualisation techniques such can histogram, scatter plots, box plot, distribution plots, maps, etc can be used to perform EDA.

In this article, we will do EDA using the COVID-19 dataset. We will use different commonly used EDA techniques to discover patterns in the COVID-19 dataset. The complete code can be found on my [github](https://github.com/pragpaudyal/covid" \t "_blank).

**Datasets**

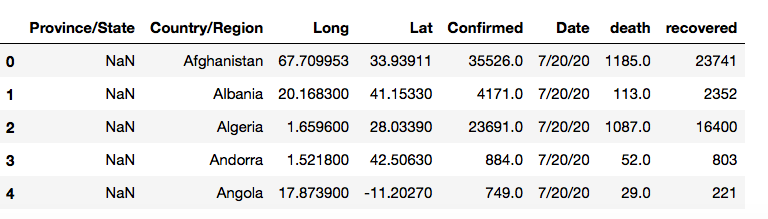
We will be working with the publicly available dataset from [JHU CSSE’s COVID-19 data repository](https://github.com/CSSEGISandData/COVID-19). The data repository consists of different datasets related to covid-19. I merged and concat the relevant data. You can find modified dataset on my [github](https://github.com/pragpaudyal/covid" \t "_blank).

**Import Libraries and loading data**

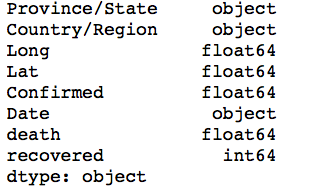


data=pd.read\_csv(“country\_vaccination.csv”)

We will use data.shape to check shape of the dataset. Using .head() function, we get a glance to the dataset. covid19\_csv consists of following row:



Using .dtypes() function, we can check the datatype. We can see data type are either object or float. When the data type is object, it is a categorical data.



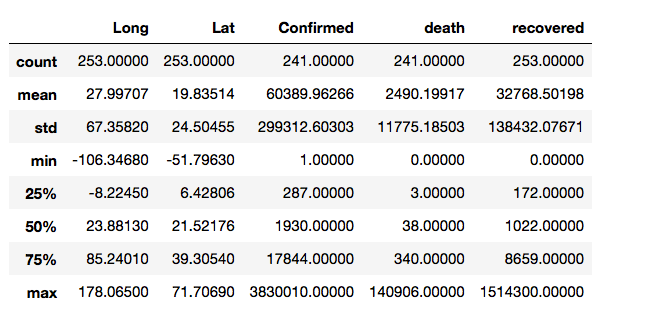
**Using EDA approach to get more insight of the data**

1. **Summary Statistic**

The describe functions give us descriptive statistics that summarise the count, mean, standard deviation, minimum, maximum, quartile (lower, middle and upper) values. Sometimes, you will get your result in scientific notation, to suppress scientific notation in Pandas we will use:

pd.set\_option (‘display. float format’, lambda x: ‘%.5f’ % x)

We can use data.describe() function to get summary statistic of the dataframe. This function will give us summary only for numerical datasets.

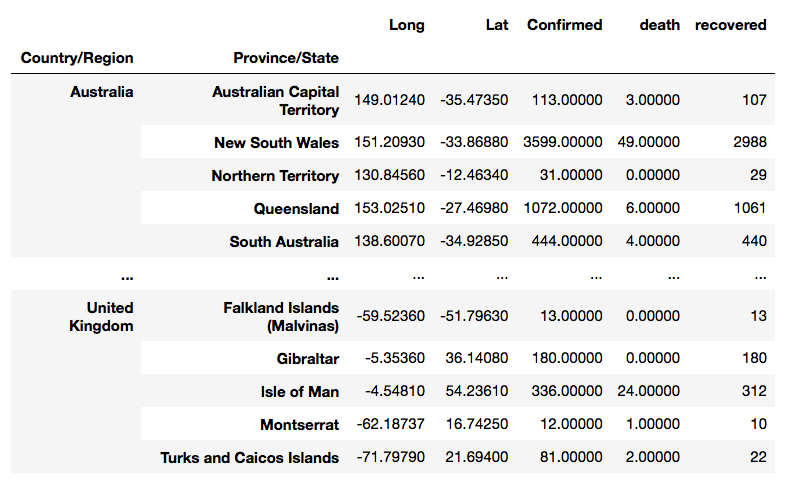


However, .describe(include=”all”) describes all columns of a dataframe regardless of data type. We can also observe if there are any missing values for any features. For instant, we can see **confirmed** and **death** have missing values. We can get more insight for confirmed, death and recovered data such as what is the maximum, minimum number for each feature.

**2. Visualisation**

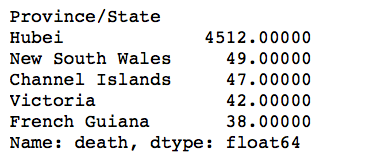
To group confirmed, death and recovery by Country/Region we use groupby function.

data. groupby([“Country/Region”, “Province/State”]).sum()



We can also group the data based on Province/State and view top 5 Province with highest death.

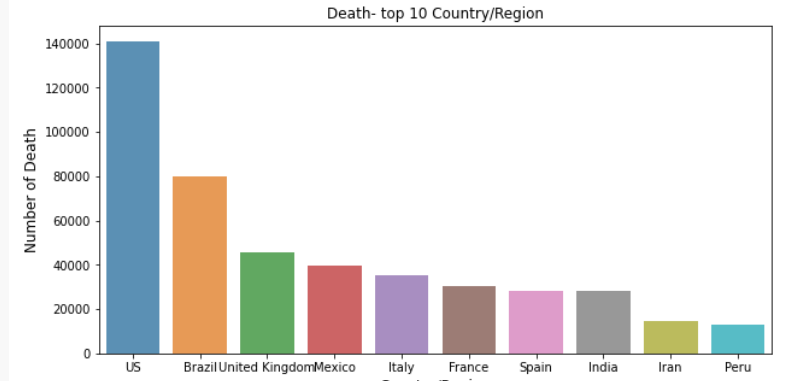
data. groupby(“Province/State”)[“death”]. sum().sort\_values(ascending=False) [:5]



provides lots of useful insight.

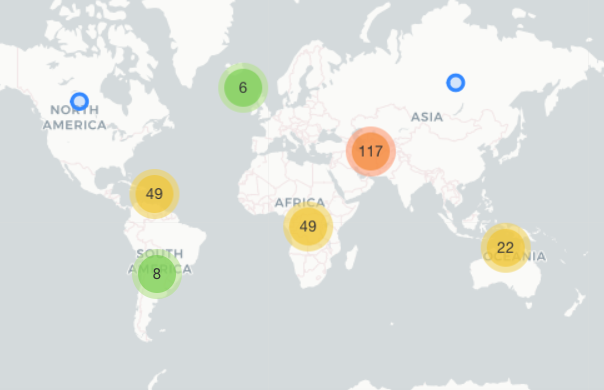
We can get a bar graph for This will show only top 5 Province with highest death. This information Province based on death rate.

Country=data.groupby("Country/Region")["death"].sum().sort\_values(ascending=False)[:10]  
plt.figure(figsize=(10,5))  
sns.barplot(Country.index, Country.values, alpha=0.8)  
plt.title('Death- top 10 Country/Region')  
plt.ylabel('Number of Death', fontsize=12)  
plt.xlabel('Country/Region', fontsize=12)  
plt.show()



In addition to these approaches, we also used map to explore geographical analysis on Covid-19. There is different python package to create visual and informative maps. Here we used folium to create a world map to show effect of covid-19 around the world. Here we used longitude and latitude data from the dataset.

world map= folium. Map(tiles="cartodbpositron")  
marker\_cluster = MarkerCluster().add\_to(world\_map)for i in range(len(data)):  
 Lat = data.iloc[i]['Lat']  
 Long = data.iloc[i]['Long']  
 radius=6  
 popup\_text = """Country/Region : {}<br>  
 Confirmed : {}<br>  
 death: {} <br>  
 recovered:{} <br>  
 Date:{} <br>"""  
 popup\_text = popup\_text.format(data.iloc[i]['Country/Region'],  
 data.iloc[i]['Confirmed'],  
 data.iloc[i]['death'],  
 data.iloc[i]['recovered'],  
 data.iloc[i]['Date']  
 )  
 folium.CircleMarker(location = [Lat, Long], radius=radius, popup= popup\_text, fill =True).add\_to(marker\_cluster)world\_map



When we click on the bubble, the map will zoom in to show more detail for country.

**Key Findings:**

COVID-19 vaccines have been instrumental in reducing the severity of the disease and preventing hospitalizations and deaths.

Vaccine effectiveness can vary based on the type of vaccine and emerging variants of the virus.

Mass vaccination campaigns have been successful in achieving high vaccination rates in many countries.

Vaccine distribution and accessibility disparities exist, with some regions facing challenges in vaccine access.

**Insights:**

Booster doses have become important for maintaining immunity, especially as the effectiveness of vaccines may decrease over time.

Vaccine hesitancy and misinformation continue to pose challenges in achieving global vaccination goals.

Variants like Delta and Omicron have raised concerns about the need for ongoing research and vaccine updates.

Vaccine equity is critical to ending the pandemic, as no one is safe until everyone is safe.

**Recommendations:**

Continue to promote and prioritize vaccination, including booster shots, to maintain immunity levels.

Combat vaccine hesitancy through education, community engagement, and targeted messaging.

Invest in research to monitor and adapt vaccines to emerging variants.

Support international efforts to ensure equitable vaccine distribution and access.

Maintain public health measures in parallel with vaccination efforts to control the spread of the virus.