BHARATIYA VIDYA BHAVAN'S
SARDAR PATEL INSTITUTE OF TECHNOLOGY
(Empowered Autonomous Institute Affiliated to University of Mumbai)
[Knowledge is Nectar]
Department of Computer Engineering

Name: Rohit Sanjay Phalke
UID: 2021300100
Batch: G

Aim:

Design Interactive Dashboards and Storytelling using Tableau / Power BI / R / Python / D3.js to be performed on the dataset - Disease spread / Healthcare

Theory:

Dataset Descriptions:

1. **case_time_series**: This dataset appears to track the COVID-19 pandemic over time on a national level. It captures daily COVID-19 statistics such as the number of confirmed cases, deaths, and recoveries. This time-series data allows for an analysis of how the pandemic evolved on a day-by-day basis over the observed period.
2. **raw_data_Indian_states**: This dataset contains COVID-19 data at the state level for India. It includes granular information for different states across the country, tracking daily confirmed cases, deaths, recoveries, and testing metrics for each state. This dataset allows for state-wise analysis of the pandemic's impact and progression over time.

Column Descriptions:

1. **case_time_series** columns:
    - **Country**: The country to which the data belongs.
    - **Daily Confirmed**: The number of new COVID-19 cases confirmed each day.
    - **Daily Deceased**: The number of COVID-19 deaths reported each day.
    - **Daily Recovered**: The number of recoveries from COVID-19 each day.
    - **Date**: Date of the record, indicating the day of the data point (formatted as month/day/year).
    - **Date_YMD**: Date in year-month-day (YMD) format, used for easier chronological sorting.

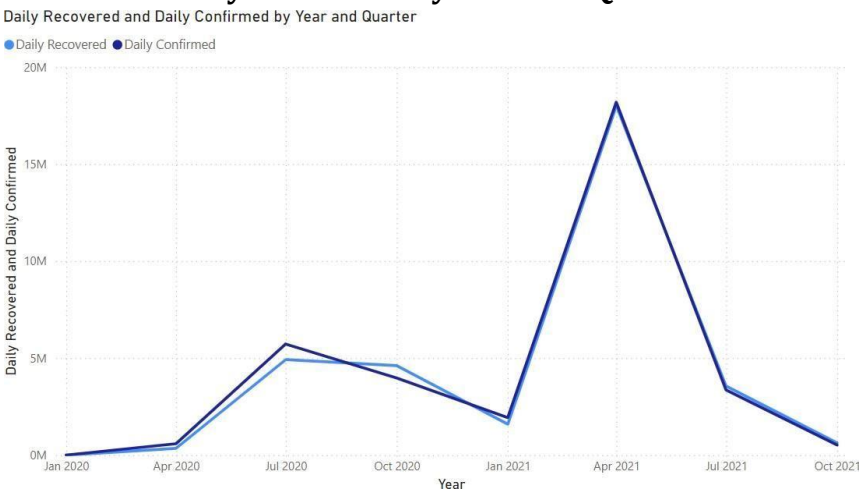
- **Total Confirmed**: The cumulative number of confirmed COVID-19 cases up to the given date.
- **Total Deceased**: The cumulative number of COVID-19 deaths reported up to the given date.
- **Total Recovered**: The cumulative number of recoveries from COVID-19 up to the given date.

2. **raw_data_Indian_states** columns:
   - **Confirmed**: The number of confirmed COVID-19 cases for a particular state on a given date.
   - **Country**: The country to which the data belongs ("India").
   - **Date**: Date of the record, indicating the specific day of the data point.
   - **Deceased**: The number of deaths due to COVID-19 reported for a particular state on a given date.
   - **Recovered**: The number of recoveries from COVID-19 reported for a particular state on a given date.
   - **State**: The state in India to which the data belongs.
   - **Tested**: The number of tests conducted in a particular state on a given date.

These datasets together can be used to assess the overall trends of the COVID-19 pandemic both nationally and at the state level in India during the specified time period.

Charts:

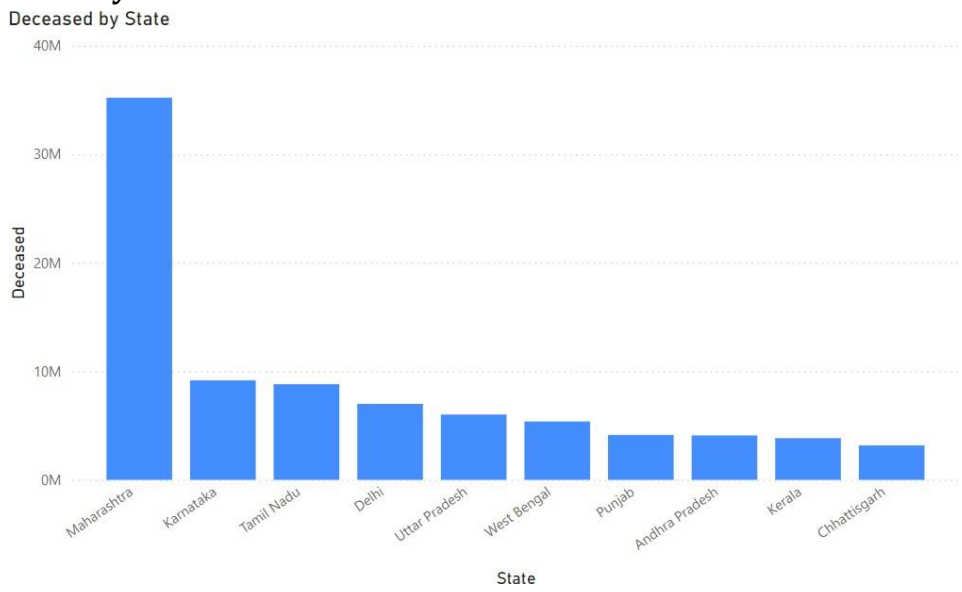
1. Daily Recovered and Daily Confirmed by Year and Quarter:



Observations:

- Peak in Q2 2021: The highest number of both daily confirmed and daily recovered cases occurred in the second quarter of 2021, with a sharp rise compared to previous quarters.
- Early Increase in 2020: There was a steady increase in both confirmed and recovered cases starting in the second quarter of 2020, followed by a gradual decline after Q3 2020.
- Post-peak Decline: After the peak in mid-2021, there is a significant drop in both daily confirmed and recovered cases, reaching low levels by the end of Q3 2021.
- Close Correlation: Throughout the observed period, the number of daily recovered cases closely tracks the number of daily confirmed cases, indicating a proportional recovery rate.

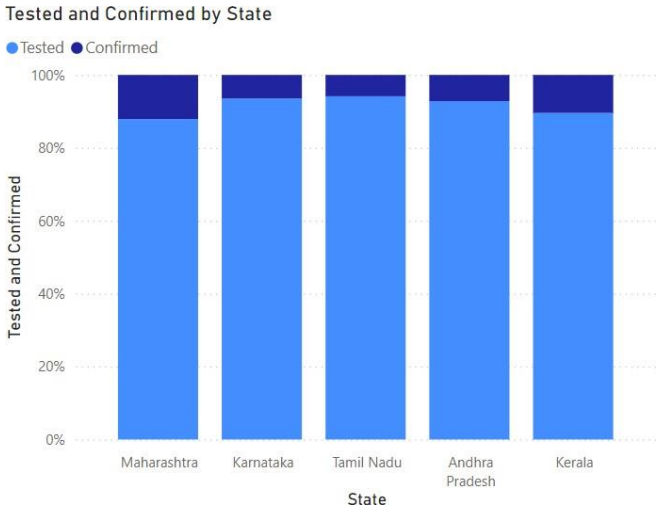
2. Deceased by State:



Observations:

- Maharashtra shows the highest number of deceased, significantly higher than all other states, with over 35M deaths.
- Karnataka and Tamil Nadu follow next, each with around 10M deceased.
- States like Delhi, Uttar Pradesh, and West Bengal have similar figures, ranging between 5M to 8M deceased.
- The lowest numbers of deceased are seen in states like Kerala and Chhattisgarh, both below 5M.

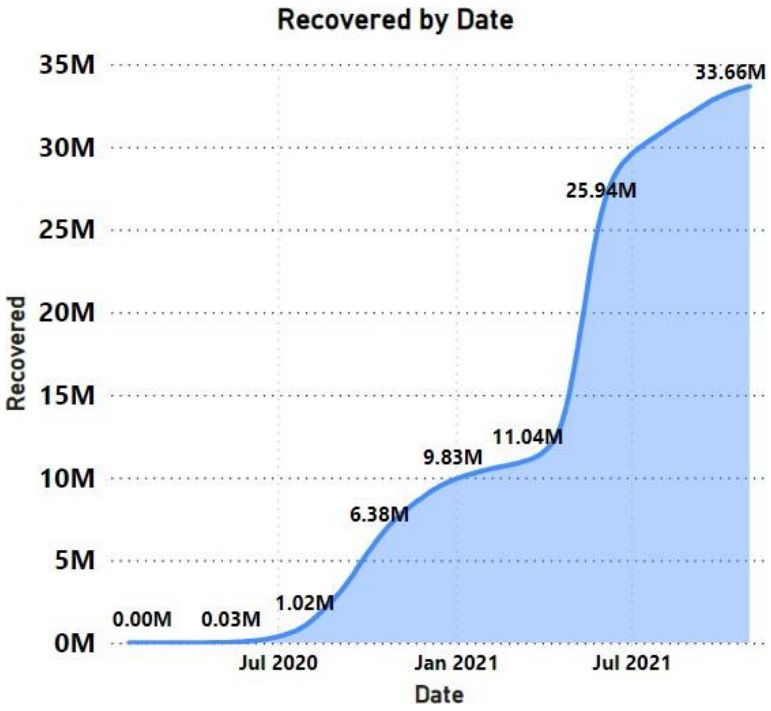
3. Tested and Confirmed by State:



Observations:

- All states have a very high percentage of confirmed cases compared to tested cases, with Maharashtra having a slightly larger gap between the two.
- Karnataka, Tamil Nadu, Andhra Pradesh, and Kerala show almost identical proportions of tested and confirmed cases.
- Maharashtra has the highest percentage of tested cases compared to the confirmed, indicating a slightly better testing-to-confirmation ratio than the other states.
- Overall, the percentage of confirmed cases across all the states is extremely high, exceeding 90% in all instances.

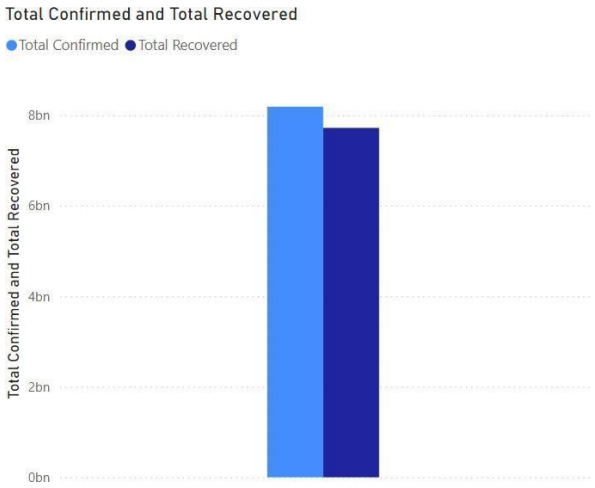
4. Recovered by State:



Observations:

- The recovery rate saw a steady increase from mid-2020, with significant acceleration starting in early 2021.
- By January 2021, recoveries reached approximately 11M, with a sharp rise thereafter.
- The most dramatic increase occurred between January 2021 and July 2021, where recoveries jumped from 11.04M to 33.66M.
- The graph indicates the highest recovery count of 33.66M by mid-2021, reflecting the impact of mass recovery efforts during this period.

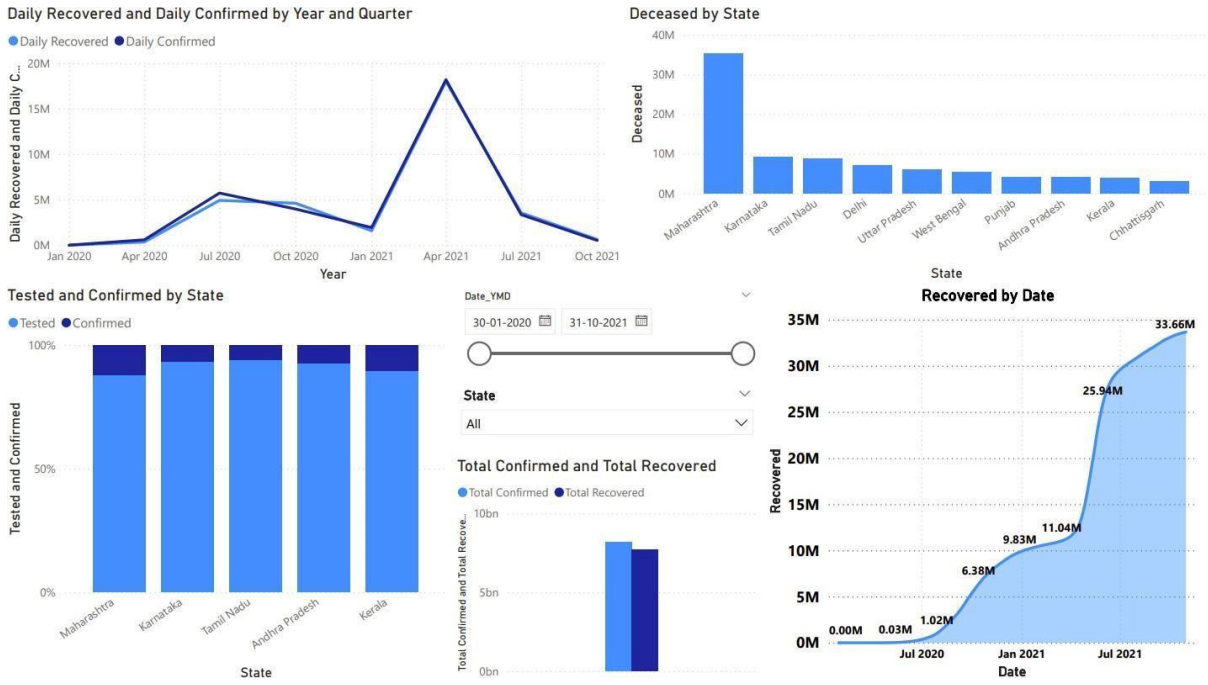
5. Total Confirmed and Total Recovered:



Observations:

- Total Confirmed Cases Exceed Total Recovered Cases: The blue bar representing total confirmed cases is taller than the dark blue bar representing total recovered cases. This indicates that there are more confirmed cases of COVID-19 than recovered cases. This also shows that the difference in the 2 charts is due to the deceased people.
- High Total Confirmed Cases: The total confirmed cases are very high, reaching around 8 billion. This suggests a widespread infection rate.
- Significant Recovery Rate: While the total confirmed cases are high, the total recovered cases are also substantial, reaching around 7 billion. This indicates a significant recovery rate.

Dashboard:



Conclusion:

From this experiment, I learned about the spread and impact of COVID-19 by visualizing key metrics such as deceased, tested, confirmed, and recovered cases using Power BI. Creating these charts and integrating them into a dashboard allowed me to analyze trends and patterns across different states and time periods. Additionally, I gained valuable experience in using Power BI for data analysis, improving my skills in designing interactive dashboards that help transform complex data into clear, actionable insights. This hands-on experience enhanced both my understanding of the pandemic and my proficiency in dashboard creation.