

COVID ANALYSIS IN INDIA USING POWER BI

by

Smruti Rajakumar	20BCE1114
Kartike Chadha	20BCE1147
Ishita Jalla	20BCE1821

A project report submitted to

Dr. PARVATHI R.

**SCHOOL OF COMPUTER SCIENCE AND
ENGINEERING**

in partial fulfilment of the requirements for the course of

CSE3020 – DATA VISUALIZATION

in

B. Tech. COMPUTER SCIENCE AND ENGINEERING



VIT[®]

Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Vandalur – Kelambakkam Road

Chennai – 600127

APRIL 2023

INDEX

Sr. No.	Topic	Pg. No.
1	Abstract	1
2	Introduction	3
3	Review of Literature	4
4	Materials and Methods	7
6	Proposed Works	14
7	Results and Discussion	15
8	Conclusion	20
9	References	21

ABSTRACT

The COVID-19 pandemic has affected millions of people worldwide, including India. The objective of this data visualization project using PowerBI is to analyse the impact of COVID-19 in India and to provide a top-down view of the situation. The project aims to provide insights into the spread of the virus, the effectiveness of preventive measures, and the impact on the economy and healthcare system.

To achieve this objective, we used PowerBI, a powerful data visualization tool, to create various types of charts and geospatial visualizations. The data used for this project was collected from reliable sources such as the World Health Organization (WHO), Ministry of Health and Family Welfare (MOHFW), and various other public sources and modelled to fit the projects use case.

The project uses a variety of charts, such as line charts, bar charts, and stacked bar charts, to visualize the trends in the number of cases, deaths, and recoveries over time. Additionally, the project uses a choropleth map to represent the COVID-19 cases in different states of India. We also created a heat map to visualize the concentration of cases across different regions in India.

The project also includes a dashboard that provides an overview of the COVID-19 situation in India. The dashboard displays the total number of cases, deaths, and recoveries, along with the percentage of active cases, deaths, and recoveries. It also displays the total number of tests conducted allowing the users to get an idea about the positivity ratio.

Another important aspect of the project is to analyse the impact of COVID-19 on the healthcare system and the economy. The project also goes a step further and shows the visualizations of the country's fight back against the virus by also incorporating vaccination data. In this way we have covered each and every aspect of how the country was affected by the deadly virus and how the country fought back by increasing testing and swiftly starting vaccination drives.

The outcome of the project is a comprehensive and interactive visualization of the COVID-19 situation in India. It provides insights into the spread of the virus, the effectiveness of preventive measures, and the impact on the healthcare system and economy. The project is beneficial for policymakers, researchers, and the general public in understanding the severity of the situation and making informed decisions.

The scope of the project can be extended by incorporating more data sources, such as hospitalization data and demographic data. The project can also be improved by

incorporating machine learning algorithms to predict the future trends of the virus and to identify the hotspots of the virus and how to fight them better and avoid a similar situation in the future. Overall, this project provides a useful tool for understanding the COVID-19 situation in India and for making informed decisions to combat the virus.

INTRODUCTION

During the COVID-19 pandemic the Indian government, in collaboration with various public health organizations, had implemented multiple measures to mitigate the spread of the virus and to minimize its impact on the healthcare system and the economy. However, to make informed decisions and to combat the pandemic effectively, it is essential to understand the COVID-19 situation in India comprehensively. One way to achieve this is through data visualization, which is an effective tool to analyse and communicate complex COVID-19 data. This project aims to use PowerBI, a powerful data visualization tool, to analyse and visualize the impact of COVID-19 in India, with the objective of providing a top-down view of the entire COVID-19 situation in India.

Data visualization is an effective tool to understand and communicate complex data related to the COVID-19 pandemic. PowerBI is a powerful data visualization tool that can help in creating insightful and interactive visualizations of COVID-19 data. This project aims to use PowerBI to analyse and visualize the impact of COVID-19 in India, with the objective of providing a top-down view of the entire COVID-19 situation in India.

The project will use various types of charts and geospatial visualizations to provide insights into the spread of the virus, the effectiveness of preventive measures, and the impact on the healthcare system and economy. The project will also include a dashboard that provides an overview of the COVID-19 situation in India, displaying the total number of cases, deaths, and recoveries, along with the percentage of active cases, deaths, and recoveries. Additionally, the project will analyse the impact of COVID-19 on the healthcare system and the economy using stacked bar charts, scatter plots, and other visualizations.

This project is essential to understanding the severity of the COVID-19 situation in India and making informed decisions to combat the virus. The project will be beneficial for policymakers, researchers, and the general public in understanding the spread of the virus, the effectiveness of preventive measures, and the impact on the healthcare system and economy. Furthermore, the project can be extended by incorporating more data sources, such as vaccination data and demographic data, and by incorporating machine learning algorithms to predict the future trends of the virus and identify hotspots of the virus.

REVIEW OF LITERATURE

SR. NO.	TITLE	AUTHOR / JOURNAL NAME / YEAR	TECHNIQUE	RESULT
1	Data analytics and visualization using Power BI and MS Excel for COVID-19 (Coronavirus)	Matthias, Daniel & Managwu, Chidozie. (2021).	PowerBI	The study presented different plots to show comparison among the attributes and global cases.
2	ANALYZING AND VISUALIZING IN POWER BI	Joshi, Miss & Kharade, Jyoti. (2022).	PowerBI	ETL converts raw data into structured data which is used in analysis.
3	Data Visualization Using Power BI, Orange and Excel	Roy, Shirshendu. (2021)	PowerBI, Orange	How visualization works in PowerBI and Excel
4	Learning Tableau: A data visualization tool. The Journal of Economic Education	Batt, Steven & Grealis, Tara & Harmon, Oskar & Tomolonis, Paul. (2020)	Tableau	How visualization works in PowerBI and Excel

5	High Impact Data Visualization with Power View, Power Map, and Power BI	Aspin, Adam. (2014)	PowerBI	Guides on how to use maps and Power Views using PowerBI.
6	Introduction to Power BI	Pearson, Mitchell & Knight, Brian & Knight, Devin & Quintana, Manuel. (2020)	PowerBI	Gives an overview of PowerBI
7	Analysis and Design of Visualization of Educational Institution Database using Power BI Tool	Mandava, Geetha Bhargava. (2018)	PowerApps, PowerBI	Creation of dashboards using Power BI tool can be viewed by means of Mobile Applications such as Power Apps and Mobile Power BI, Web such as by means of website or shareable link
8	Delivering a Comprehensive BI Solution with Microsoft Business Intelligence Stack	Ren, Zhijun. (2010)	PowerBI, SharePoint Server	Usage of different BI tools with PowerBI

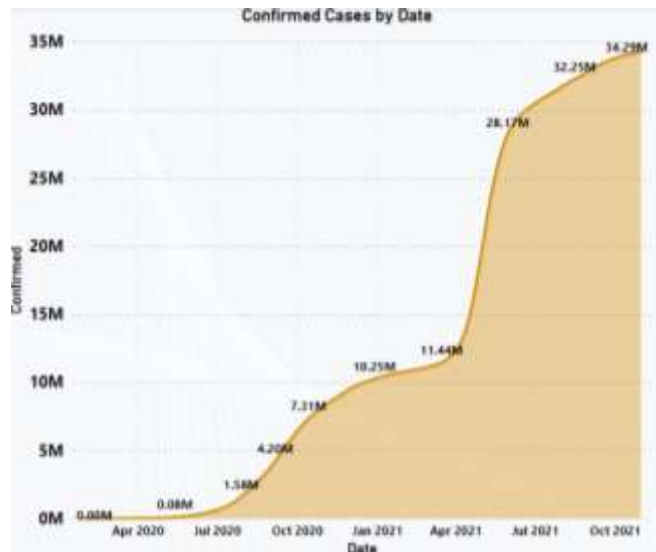
9	Power BI Data Visualization with Purpose: Communicating through Color, Shape, and Layout	Longoria, Meagan. (2019)	PowerBI	Showed how to effectively use the basics of data visualization in PowerBI
10	Maps in Power BI Desktop	Aspin, Adam. (2014)	Network Visualization and Analysis	Implementation of choropleth maps in PowerBI

MATERIALS AND METHODS

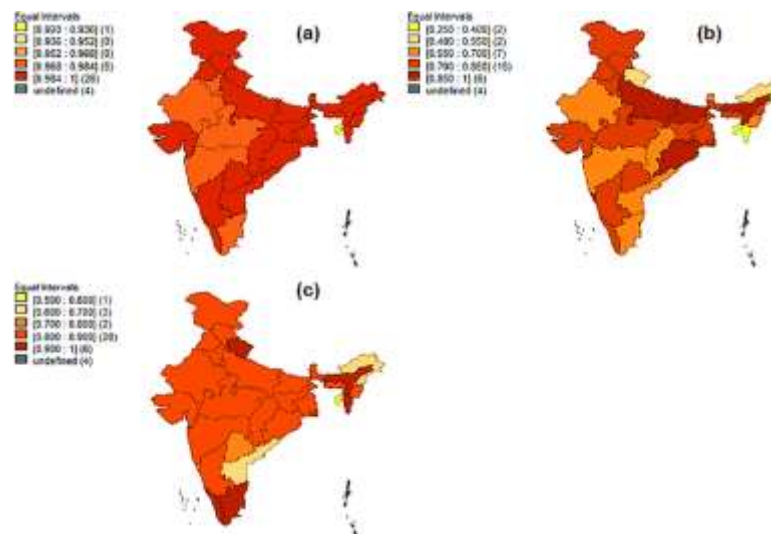
Information about Models

In this project we have used over many types of visualizations over 5 different reports. Some of these are –

- Area Chart – Area charts are a type of chart that displays quantitative data as shaded areas below or above a line. They are useful in depicting the magnitude and changes of data over time and can be effective in identifying trends and patterns in data.



- Choropleth Maps - Choropleth maps are a type of map that use colors or shading to represent the variation of a statistical variable over a geographical area, such as states or countries. They are effective in displaying patterns and trends in data and are commonly used in visualizing demographic, socioeconomic, and health-related data.



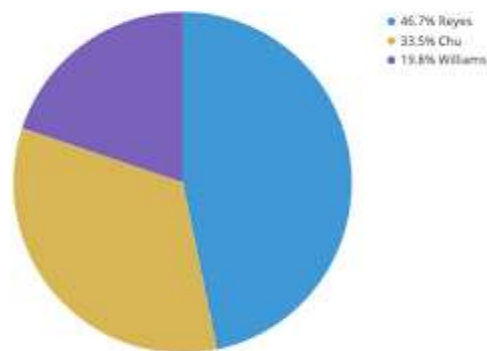
- Slicers – Slicers in Power BI are visual controls that allow users to filter and interact with data in a report or dashboard. They provide an easy way to select specific data points and quickly adjust the view of the report, making it more dynamic and user-friendly. Slicers can be used with various types of data, including dates, numbers, and text.



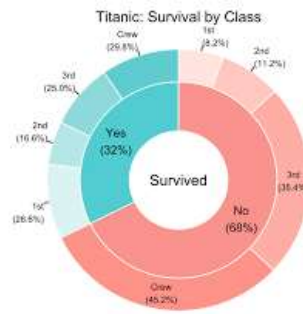
- Cards – Cards in Power BI are a type of visual that displays a single value or a small set of related values in a concise and easy-to-read format. They can be used to show key performance indicators (KPIs) or summary information from a data set. Cards can be customized with various formatting options to match the look and feel of the report or dashboard.



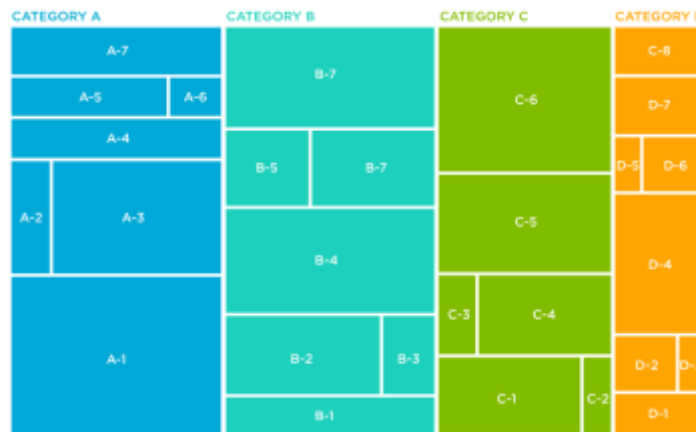
- Pie Charts - Pie charts in Power BI are circular charts that display data as a series of slices, with each slice representing a proportion of the whole. They are useful for showing the relative size of each category and can be customized with various colors and labels to make them more visually appealing and informative.



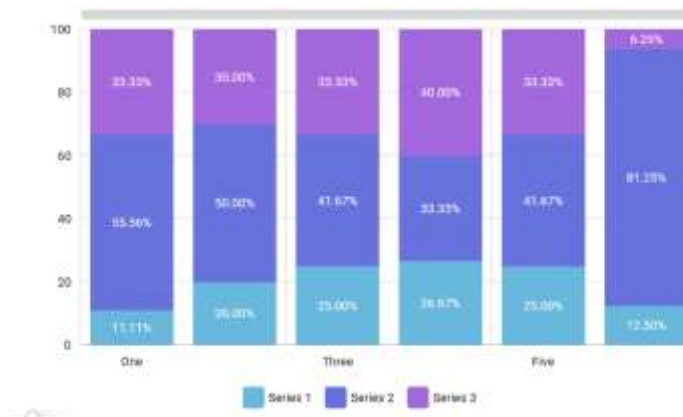
- Donut Charts – Donut charts in Power BI are a type of circular chart that displays data as a ring with a hole in the middle. They are similar to pie charts but allow for an additional level of comparison by showing multiple series of data. Donut charts are useful for displaying the relative size of each category and can be customized with various colors and labels to make them more visually appealing and informative.



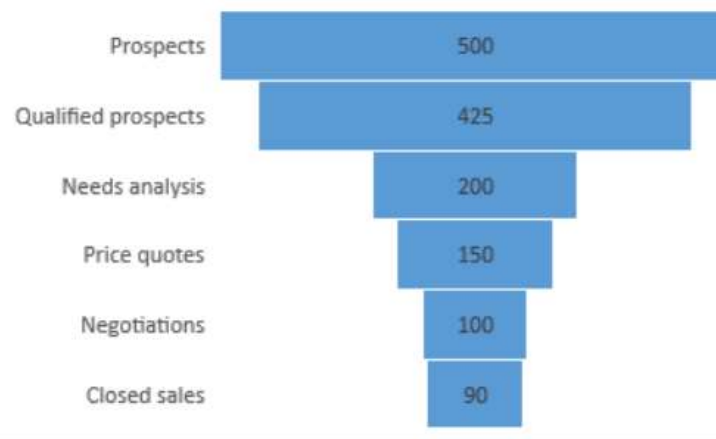
- **TreeMaps** – Tree maps in Power BI are a type of hierarchical chart that displays data as a series of nested rectangles, with each rectangle representing a category and its size proportional to its value. They are useful for visualizing the relative size and structure of a data set, and can be customized with various colors and labels to make them more visually appealing and informative. Tree maps can handle large amounts of data, but they can be difficult to read if the categories are too small or the values are too similar.



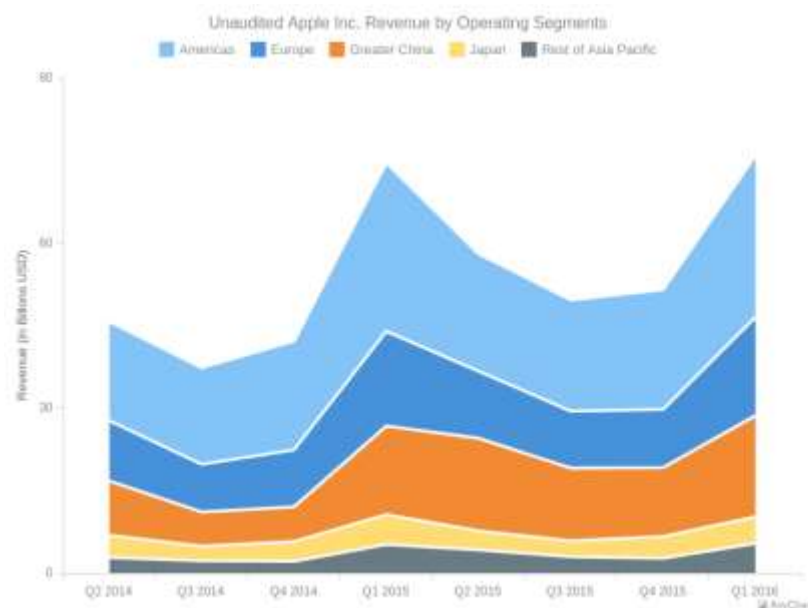
- **100% Stacked Charts** – 100% stacked charts in Power BI are a type of chart that displays multiple series of data as percentages of a whole. They are useful for showing how each category contributes to the total and can be used to compare the relative proportions of different categories across multiple data sets. 100% stacked charts are customizable with various colors and labels, and can be used to visualize trends and patterns in the data. However, they can be difficult to read if there are too many categories or the differences between categories are too small.



- **Funnel** – Funnel charts in Power BI are a type of chart that displays data as a series of steps, with each step representing a stage in a process or a funnel. They are useful for visualizing the progression of data through different stages and can be used to identify areas where the process may be experiencing bottlenecks or drop-offs. Funnel charts are customizable with various colors and labels, and can be used to visualize trends and patterns in the data. However, they should be used with caution as they can be difficult to read if there are too many steps or the differences between steps are too small.



- **Stacked Area Charts** – Stacked area charts in Power BI are a type of chart that displays multiple series of data as an area chart, with each series stacked on top of the previous one. They are useful for showing how each category contributes to the total over time and can be used to compare the relative proportions of different categories across multiple data sets. Stacked area charts are customizable with various colors and labels, and can be used to visualize trends and patterns in the data.



Dataset

- Covid19.xlsx

	A	B	C	D	E	F	G
1	Date	State	Confirmed	Recovered	Deceased	Tested	Country
2	12-Jun-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
3	30-May-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
4	31-May-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
5	02-Jun-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
6	16-May-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
7	09-May-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
8	12-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
9	13-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
10	14-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
11	17-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
12	18-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
13	19-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
14	20-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
15	24-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
16	28-Apr-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
17	02-May-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
18	13-Dec-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
19	15-Dec-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
20	21-Dec-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
21	23-Dec-20	Dadra and Nagar Haveli and Daman and Diu	0	0	0	0	India
21046	14-Oct-21	Maharashtra	2384	2343	35	136952	India
21047	15-Oct-21	Maharashtra	2149	1898	29	125530	India
21048	16-Oct-21	Maharashtra	1553	1682	26	100944	India
21049	17-Oct-21	Maharashtra	1715	2680	29	110465	India
21050	18-Oct-21	Maharashtra	1485	2078	27	95890	India
21051	19-Oct-21	Maharashtra	1638	2791	49	132467	India
21052	20-Oct-21	Maharashtra	1825	2879	21	121570	India
21053	21-Oct-21	Maharashtra	1573	2968	39	123700	India
21054	22-Oct-21	Maharashtra	1632	1744	40	132209	India
21055	23-Oct-21	Maharashtra	1701	1781	33	136664	India
21056	24-Oct-21	Maharashtra	1410	1520	18	130732	India
21057	25-Oct-21	Maharashtra	889	1586	12	84460	India
21058	26-Oct-21	Maharashtra	1201	1370	32	102048	India
21059	27-Oct-21	Maharashtra	1485	4947	38	122608	India
21060	28-Oct-21	Maharashtra	1418	2112	36	114099	India
21061	29-Oct-21	Maharashtra	1338	1584	36	122990	India
21062	30-Oct-21	Maharashtra	614	2148	26	119271	India
21063	31-Oct-21	Maharashtra	1172	1399	20	108040	India

This is the main dataset we are using in the project. This data contains Date, State, Confirmed, Recovered, Deceased and Tested columns along with Country. This dataset originally contained cumulative data of each day and was modified using a script to only contain data each particular day and not cumulative. It contains 21063 rows of data.

- raw_data_indian_states.xlsx

	A	B	C	D	E	F	G
1	Date	State	Confirmed	Recovered	Deceased	Tested	Country
2	30-Jan-20	Kerala	1	0	0		India
3	30-Jan-20	India	1	0	0		India
4	02-Mar-20	Delhi	1	0	0		India
5	02-Mar-20	Telangana	1	0	0		India
6	03-Mar-20	Delhi	1	0	0		India
7	03-Mar-20	Rajasthan	1	0	0		India
8	03-Mar-20	Telangana	1	0	0		India
9	04-Mar-20	Delhi	1	0	0		India
10	04-Mar-20	Telangana	1	0	0		India
11	05-Mar-20	Telangana	1	0	0		India
12	06-Mar-20	Telangana	1	0	0		India
13	07-Mar-20	Telangana	1	0	0		India
14	07-Mar-20	Tamil Nadu	1	0	0		India
15	08-Mar-20	Telangana	1	0	0		India
16	08-Mar-20	Tamil Nadu	1	0	0		India
17	09-Mar-20	Jammu and Kashmir	1	0	0		India
18	09-Mar-20	Karnataka	1	0	0		India
19	09-Mar-20	Punjab	1	0	0		India
20	09-Mar-20	Telangana	1	0	0		India
21	09-Mar-20	Tamil Nadu	1	0	0		India

This is a similar dataset but with cumulative data and contains Date, State, Confirmed, Recovered, Deceased and Tested columns along with Country with 21676 rows of data.

- vaccination-data.xlsx

	A	B	C	D	E	F	G
1	COUNTRY	DATE_UPDATED	TOTAL_VACCINATIONS	PERSONS_VACCINATED_1PLUS_DOSE	TOTAL_VACCINATIONS_PER100	PERSONS_VACCINATED_1PLUS_DOSE_PER100	PERSONS_FULLY_VACCINATED
2	India	07-Dec-21	1294608045	804768300	93.812	58.316	489839745
PERSONS_FULLY_VACCINATED_PER100 VACCINES_USED							FIRST_VACCINE_DATE
35.496 Bharat - Covaxin,Gamaleya - Gam-Covid-Vac,Janssen - Ad26.COV 2-5,Moderna - Spikevax,SII - Covishield							16-Jan-21
VACCINES_USED						FIRST_VACCINE_DATE	NUMBER_VACCINES_TYPES_USED
Bharat - Covaxin,Gamaleya - Gam-Covid-Vac,Janssen - Ad26.COV 2-5,Moderna - Spikevax,SII - Covishield						16-Jan-21	5

In the vaccination data we have the columns country, data updated, total vaccinations, persons vaccinated 1 plus dose, total vaccinations per 100, persons vaccinated 1 plus does per 100, persons full vaccinated, vaccines used, first vaccine data and number of vaccines types used.

Architecture and Implementation

In our project we have made 5 reports. These are Total, Testing and Vaccination, Confirmed, Deceased and Recovered.

The Total report gives the general overview of the project with area charts of confirmed cases and recovered cases with slicer for date and a drop down list to select any particular state. We can also see the deaths till the given date on a card.

In the Testing and Vaccination report we have an area chart depicting the Tested and Confirmed against time along with slicers for changing the time interval. This is along with cards for vaccination related data.

The next three reports of Confirmed, Recovered and Deceased follow a similar template to make it easier for the users to derive meaningful conclusions and analyse the reports easily and efficiently. They have a choropleth map of India depicting how many people from each state are in the dataset pertaining to each report title. We also have a pie chart or a donut chart depicting data by state. Treemaps and Area Charts are also used to make both categorical and time-series data easily analysable by the viewers.

PROPOSED WORKS

Novelty

The novelty of a data visualization project that combines COVID-19 data in India with vaccination data lies in the fact that there is currently no single resource that provides all this information in an organized and easily digestible manner. By aggregating and visualizing this data, the project would provide a comprehensive overview of the pandemic situation in India, including trends in the number of confirmed cases, deaths, and tests conducted, as well as progress in vaccination coverage.

The project could also potentially use innovative data visualization techniques to make the data more accessible and easier to understand. For example, interactive visualizations could allow users to explore the data and compare different regions or time periods. The use of maps, charts, and graphs could also help to highlight patterns and trends in the data, such as regional variations in case counts or changes in vaccination rates over time.

Overall, such a project could provide valuable insights into the pandemic situation in India, and could help policymakers and public health officials to make data-driven decisions about resource allocation, vaccination prioritization, and other key aspects of the pandemic response. Additionally, by making the data more accessible and understandable to the general public, the project could help to promote awareness and understanding of the pandemic situation, and encourage individuals to take appropriate precautions and follow public health guidelines.

Project Contributions

Smruti Rajakumar: Documentation, PowerBI Implementation

Kartike Chadha: Data Modelling, PowerBI Implementation

Ishita Jalla: Documentation, PowerBI Implementation

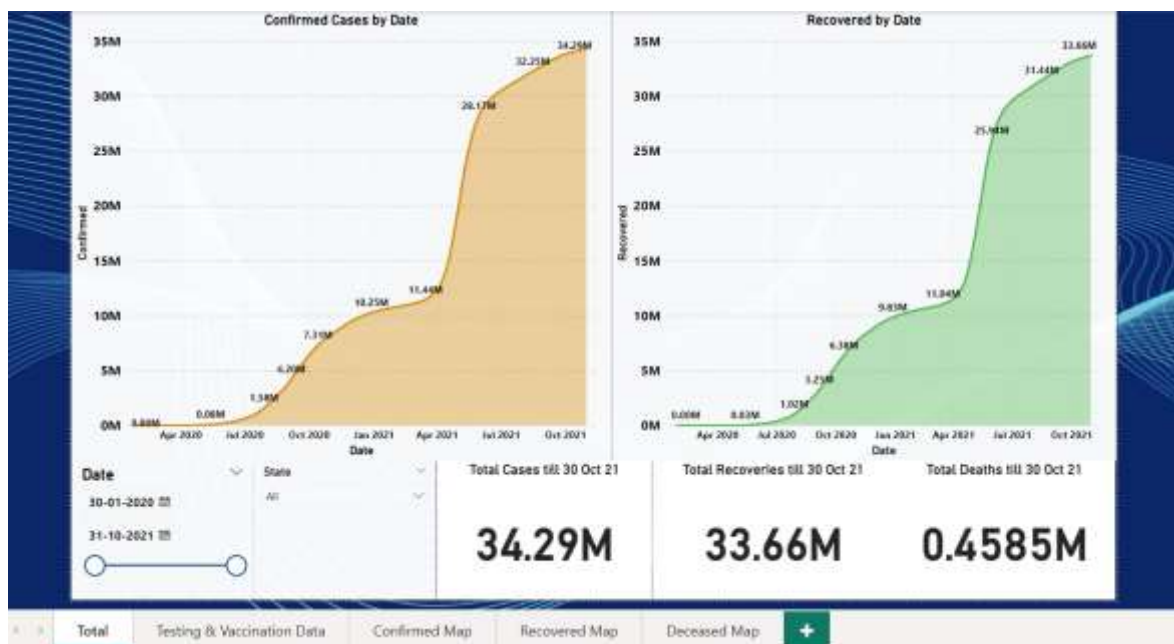
RESULTS AND DISCUSSIONS

Results

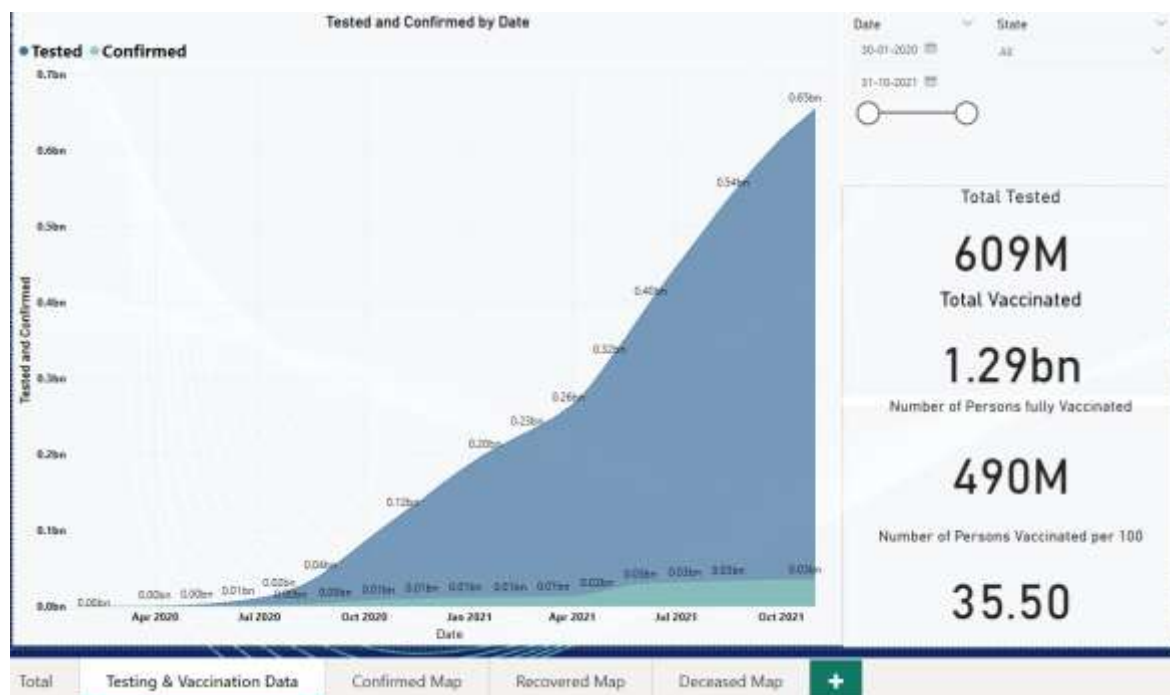
The data visualization project on Covid-19 in India using numerous different visualizations to create five reports, overview, testing and vaccination, confirmed cases, recovered cases, and deceased cases has provided valuable insights into the pandemic's impact on the country. The overview report presented a comprehensive summary of the situation in India, showcasing the total confirmed, active, recovered, and deceased cases through line graphs, pie charts, and maps. The testing and vaccination report presented the progress of testing and vaccination in India, highlighting the total number of tests conducted, the positivity rate, and the number of doses administered across different states through bar graphs and maps. The confirmed cases report provided an in-depth analysis of the distribution of confirmed cases across states and demographics through heat maps and bar graphs. The recovered cases report showcased the recovery rate across states and demographics through stacked bar graphs and heat maps. Finally, the deceased cases report presented the mortality rate across different states and demographics through stacked bar graphs and heat maps. Overall, these visualizations have enabled us to gain a better understanding of the Covid-19 situation in India and make informed decisions to combat the pandemic.

Figures

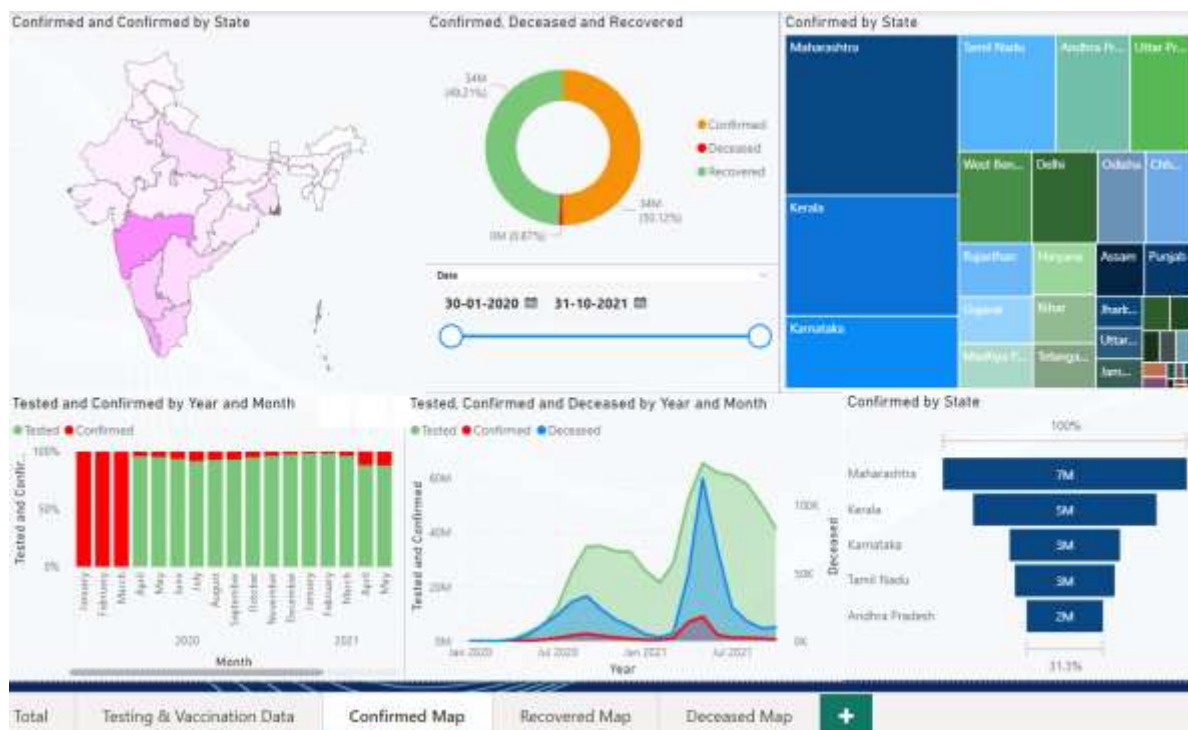
- Total Report (Overview)



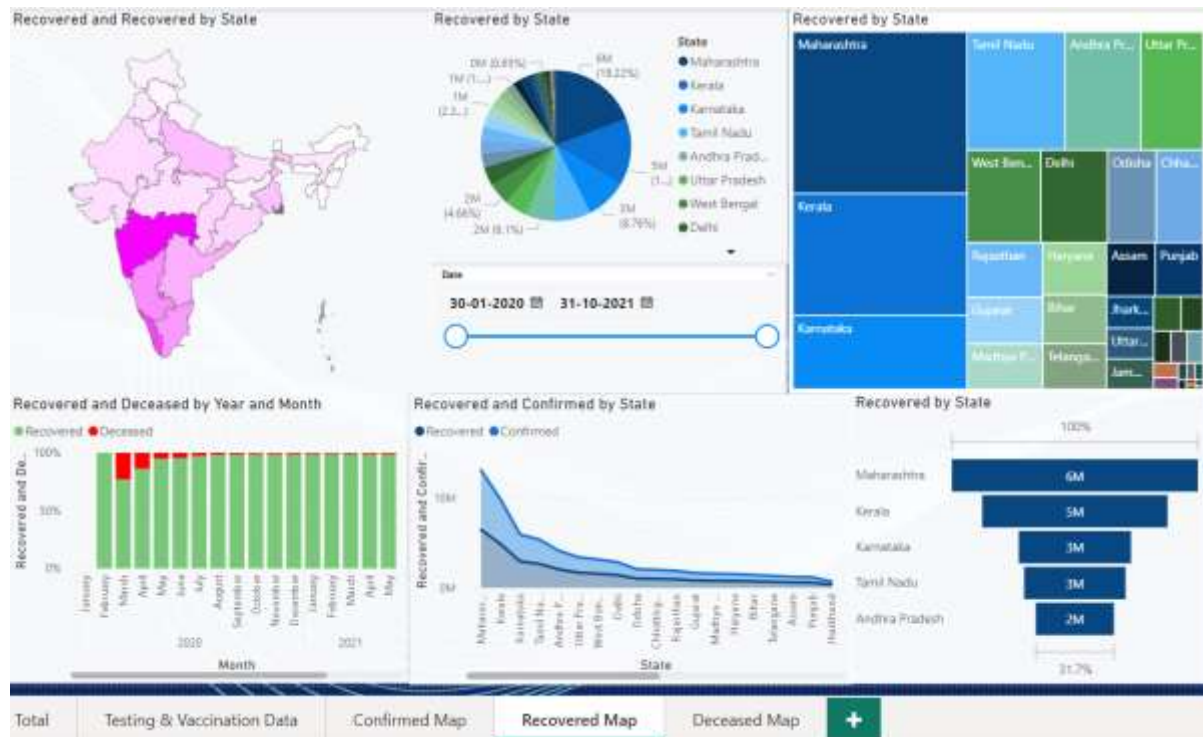
- Testing and Vaccination Data



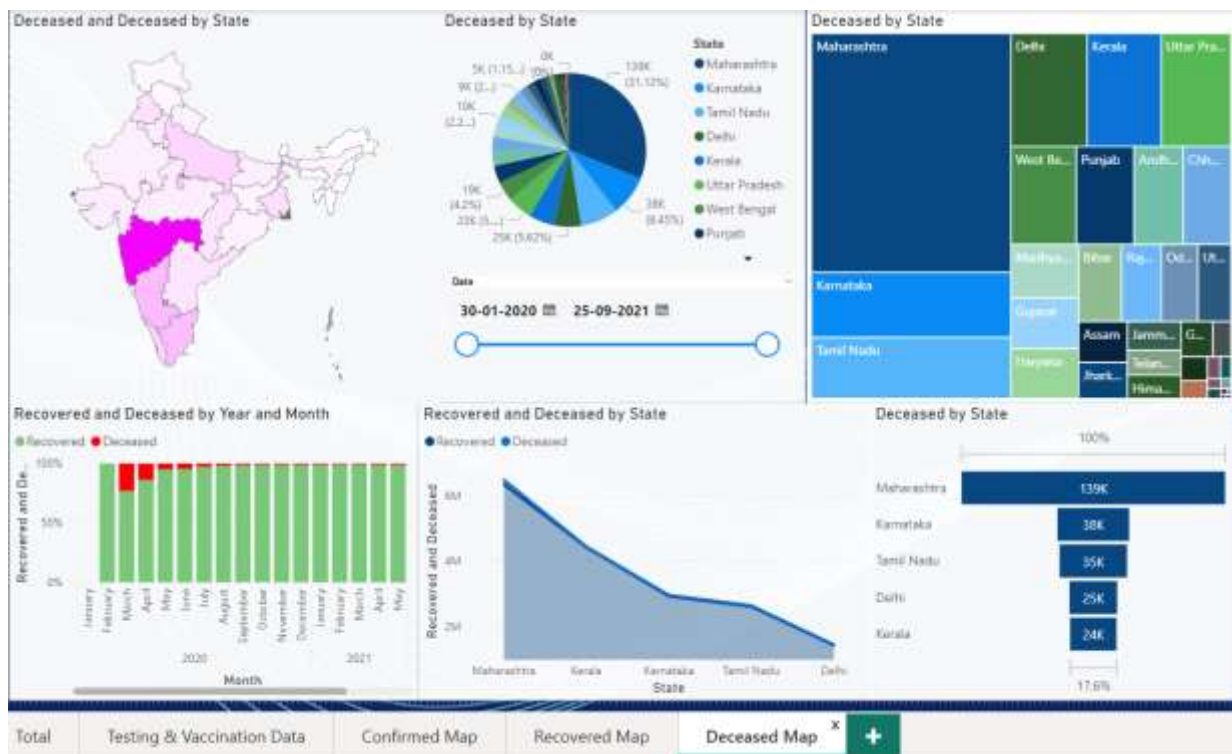
- Confirmed



- Recovered



- Deceased



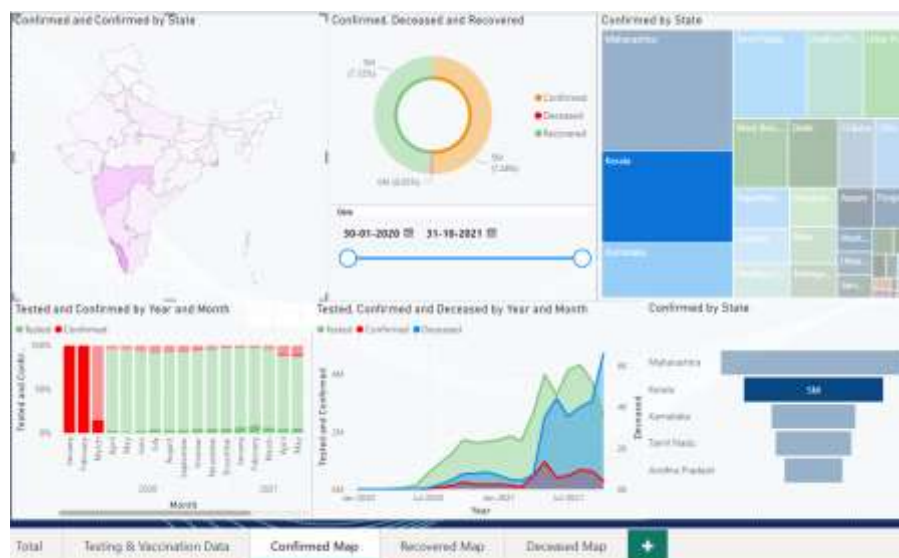
Explanation

From these reports meaningful conclusions can be drawn very easily and efficiently. Here are some explanations of these reports.



Deceased in Maharashtra

In this report we have visualized the situation of deceased people from Maharashtra from date 30-01-2020 to 25-08-2021. From the report we can see that the deceased in Maharashtra accounted for majority of the deaths across the country, but also had the most recovered people from COVID-19 in the country.



Confirmed in Kerala

In this report we have visualized the situation of confirmed cases from Kerala from date 30-01-2020 to 25-08-2021. From the report we can see that the Kerala was a major contributor in COVID-19 cases in India.



Confirmed Cases over a period of time

In this visualization we have the confirmed cases across the country but we are only showing the cases which occurred during the time period 21-06-2020 and 06-07-2021 using a slicer.

CONCLUSION

In conclusion, our data visualization project on COVID-19 in India has provided valuable insights into the impact of the pandemic on the country. Through our five Power BI reports - Total, Testing and Vaccination, Confirmed, Recovered, and Deceased - we have been able to analyse and present complex data in an intuitive and easy-to-understand manner.

In the process we have made a comprehensive resource for COVID-19 which can serve as a useful tool to study the patterns of the pandemic and how it has affected each and every state and UT across India over time. We can also use this tool in the current scenario where we are already seeing a slight uptick in the number of cases across the country and see how it corroborates from the patterns we have already seen.

We can also make this tool as a generic template for a future or past pandemic like situation by just modelling the data according to the tool and do studies on them. Overall, our data visualization project has highlighted the need for continued vigilance and efforts to combat the COVID-19 pandemic in India. By presenting data in a clear and concise manner, we hope to raise awareness about the severity of the situation and encourage people to take necessary precautions to stay safe.

Github Link

<https://github.com/KartikeChadha/DV-J-Comp>

REFERENCES

1. Matthias, Daniel & Managwu, Chidozie. (2021). Data analytics and visualization using Power BI and MS Excel for COVID-19 (Coronavirus). 10.13140/RG.2.2.25204.48001. Bengfort, Benjamin, Rebecca Bilbro, and Tony Ojeda. Applied text analysis with python: Enabling language-aware data products with machine learning. " O'Reilly Media, Inc.", 2018.
2. Joshi, Miss & Kharade, Jyoti. (2022). ANALYZING AND VISUALIZING IN POWER BI
3. Roy, Shirshendu. (2021). Data Visualization Using Power BI, Orange and Excel.
4. Batt, Steven & Grealis, Tara & Harmon, Oskar & Tomolonis, Paul. (2020). Learning Tableau: A data visualization tool. The Journal of Economic Education. 51. 1-12. 10.1080/00220485.2020.1804503.
5. Aspin, Adam. (2014). High Impact Data Visualization with Power View, Power Map, and Power BI. 10.1007/978-1-4302-6617-4.
6. Pearson, Mitchell & Knight, Brian & Knight, Devin & Quintana, Manuel. (2020). Introduction to Power BI. 10.1007/978-1-4842-6008-1_16.
7. Mandava, Geetha Bhargava. (2018). Analysis and Design of Visualization of Educational Institution Database using Power BI Tool.
8. Ren, Zhijun. (2010). Delivering a Comprehensive BI Solution with Microsoft Business Intelligence Stack. 278 - 281. 10.1109/CESCE.2010.183.
9. Longoria, Meagan. (2019). Power BI Data Visualization with Purpose: Communicating through Color, Shape, and Layout. 10.1007/978-1-4842-5081-5.
10. Aspin, Adam. (2018). Maps in Power BI Desktop. 10.1007/978-1-4842-3210-1_19.