International Journal of Technology Engineering Arts Mathematics Science
Vol. 3, No. 1, June2023, pp. 22~26
ISSN: 2583-1224

Understanding COVID-19 vaccination coverage in India

Anshika Singh[1], Ajinkya Sanas[2], Atharva Taras[3], Rishikesh Dound[4], Mayura Shelke[5]

[1,2,3,4,5] Artificial Intelligence & Data Science, AISSMS IOIT, Maharashtra, India

Corresponding Authors: Anshika Singh, (pksmailtome@gmail.com)

Article Information

Article history:

Received March 9, 2023
Revised June 12, 2023
Accepted June 19, 2023

ABSTRACT

The COVID-19 pandemic has unleashed a devastating impact on the world, causing unprecedented levels of disruption across the globe. The Indian subcontinent, known for its vibrant culture and bustling cities, has unfortunately not been spared from this scourge. The most effective way to tackle the pandemic is by vaccination, and the Indian government has launched a nationwide vaccination drive. In this study, analysis of the COVID-19 vaccination data for India, which is available in the "covid_vaccine_statewise.csv" dataset. The dataset includes information on the number of persons vaccinated for the first and second doses of the COVID-19 vaccine in different states of India and the number of males and females vaccinated. As of May 2023, the COVID-19 pandemic had a grieve impact on India, with more than 26 million confirmed cases and over 295,000 deaths reported. The aim is to analyze the vaccination data available in the dataset to understand the vaccination trends in different states and identify areas where improvements are necessary. Our findings from this research can help the government and healthcare officials to identify the vaccination trends in different states and take necessary actions to improve the vaccination drive's effectiveness, thereby mitigating the impact of the COVID-19 pandemic in India.

KEYWORDS: COVID-19, vaccination, Power BI, and policy makers.

1. INTRODUCTION

The COVID-19 pandemic has unleashed a global health crisis, necessitating urgent measures to combat the virus and safeguard public health. Vaccination has emerged as a vital tool in mitigating the spread and severity of the disease. In India, a country grappling with a colossal population and diverse healthcare landscape, analyzing the state-wise distribution of COVID-19 vaccinations has become an imperative endeavor. Additionally, addressing potential gender disparities in vaccination coverage is crucial to ensure equitable access to immunization and optimize public health outcomes.

This research paper aims to delve into the complex realm of COVID-19 vaccination data analysis in India, leveraging the dataset "covid_vaccine_statewise.csv" and harnessing the power of Power BI. By employing bioinformatics techniques, this study seeks to unravel meaningful insights that can inform evidence-based decision-making, refine vaccination strategies, and promote equitable vaccine distribution across states. The urgency and significance of this project stem from the intricate interplay between population dynamics, healthcare infrastructure, and vaccination coverage. India, with its immense population, presents unique challenges in terms of resource allocation, logistical considerations, and healthcare delivery. Analyzing the state-wise distribution of vaccinations allows us to identify geographical variations, decipher underlying causes, and formulate targeted interventions. This knowledge can aid policymakers and healthcare professionals in allocating resources efficiently, tailoring outreach programs, and ensuring that vulnerable populations, particularly in remote areas, receive adequate vaccine coverage.

Furthermore, this research paper recognizes the criticality of gender-disaggregated data analysis in public health research. The COVID-19 pandemic has highlighted existing gender disparities, with women often facing socio-cultural, economic, and healthcare access barriers. Incorporating gender as a lens in the analysis of vaccination data allows us to discern

potential disparities, uncover underlying factors, and design gender-responsive strategies. By addressing these disparities, can promote inclusivity, empower women, and enhance the overall effectiveness of the vaccination campaign.

Bioinformatics methodologies and leveraging the data analytics capabilities of Power BI, this research paper aspires to generate evidence-based insights into the state-wise distribution of COVID-19 vaccinations in India. Through a multidimensional analysis, encompassing demographic factors, geographical patterns, and gender dynamics, this study aims to contribute to the scientific understanding of vaccination campaigns, optimize resource allocation, and foster equitable access to vaccines. By synergizing bioinformatics and public health research, this endeavor seeks to play a pivotal role in India's journey towards controlling the pandemic and safeguarding the well-being of its population.

2. METHDOLOGY

To analyze the provided dataset "covid_vaccine_statewise.csv" using Power BI, the first step is to import and transform the dataset to prepare it for analysis. After downloading the dataset from the provided link, Power BI was launched and the dataset was loaded as a text/CSV file. Once loaded, various transformations are applied to ensure the data is in a suitable format for analysis.

The initial transformation step involves handling missing values. Power BI provides several options in handling missing data, such as replacing missing values with a default value, removing rows with missing values, or interpolating missing values based on neighboring values. The appropriate method is chosen based on the nature of the missing data and the analysis requirements.

Next, data types are adjusted as needed. Power BI automatically detects the data type for each column during the import process, but sometimes manual adjustments are necessary. For example, if a column containing numeric values was imported as text it needs to be converted to a numeric data type to perform mathematical calculations accurately.

In addition to handling missing values and adjusting data types, calculated columns were created if required. Calculated columns allow the creation of new columns based on existing data using formulas or expressions. For instance, a calculated column can be added to compute the vaccination coverage rate by dividing the number of vaccinated individuals by the total population for each state.

Once the data is ready, the focus shifts to data modeling. In the report view of Power BI, measures were created to derive the required analytics. Measures are calculations that aggregate or perform calculations on the data, such as summing up values, calculating averages, or finding maximum and minimum values. For this dataset, measures were created to calculate the total number of persons vaccinated for the first and second doses by using the SUM function on the respective columns and grouping the values by the state.

Furthermore, the dataset contains information about the gender of the vaccinated individuals. To analyze the gender distribution, use of another measure is created by summing up the values in the gender column based on their respective categories. This measure allows understanding the number of males and females vaccinated across different states.

To present the derived analytics effectively, use various visualization tools in Power BI. Visuals like bar charts, tables, and maps were created to showcase the number of vaccinated persons for the first and second doses in each state. A bar chart can display the vaccination numbers for each state, allowing easy comparison and identification of states with higher or lower vaccination rates. A table can provide a detailed breakdown of the vaccination numbers, including the coverage rate for each state. Additionally, a map visualization can provide a geographical representation of the vaccination status across different regions of India.

Another visual is dedicated to representing the gender distribution, providing an overview of the number of males and females vaccinated. It is achieved through a pie chart or a stacked bar chart illustrating the proportion of each gender category in the vaccinated population.

To enhance clarity and understanding, the visuals are customized with appropriate titles, labels, and legends. Colors and formatting are applied consistently throughout the dashboard to maintain a cohesive and professional appearance. The visuals are arranged on a report canvas to create a cohesive and informative dashboard, ensuring ease of navigation and facilitating insights.

In the interpretation and conclusion phase, the visuals are analyzed to gain insights into the COVID-19 vaccination status in India. The findings related to the number of vaccinated persons for the first and second doses are interpreted highlighting variations among states. This analysis can reveal which states have made significant progress in vaccination and identify areas that require more attention. By comparing the gender distribution of the vaccinated population, any disparities or imbalances can be identified and explored further.

The derived analytics and visualizations serve as the basis for drawing meaningful conclusions and providing recommendations or insights in the research paper's conclusion section. The interpretation of the data can shed light on the effectiveness of vaccination campaigns, identify potential challenges, and guide future strategies. It contributes to evidence-based decision-making and understanding the progress and challenges of COVID-19 vaccination in India, ultimately helping policymakers and healthcare professionals make informed choices and take appropriate actions to control the pandemic.

Figure 1 represents a block diagram with all the required steps for data analysis in Power BI.

Figure 1: Block Diagram showing the process of creating and analyzing the dashboard using Power BI

3. RESULTS AND DISCUSSION

As per earlier discussion, the aim was to build a user-friendly dashboard using data visualization tool. The dashboard made in the figures below, is made with Power BI, which shows very detailed review of the dataset used from Kaggle. It can be used to study the vaccination dataset and make required changes.

The dashboard shown in figure 2 represents the data analyzed for all the states in India. Being an interactive dashboard, it allows the user to hover over the map of India as given in figure 3, and choose the state they want the data to be explored from. It displays various parameters like average of Indian nationals and foreigners confirmed as covid positive, number of deaths and cured. Additionally, users can customize the timeframe to analyze the data according to their requirements.

Figure 3: Major cities in India used as landmark

Figure 4 showcases an interactive dashboard that offers a comprehensive analysis of vaccination data across all states in India. The dashboard provides valuable insights into the number of individuals who have received both the first and second doses of the vaccine. Moreover, it presents a detailed breakdown of the vaccination status, highlighting the distribution of vaccinated individuals by gender. To facilitate a better understanding of the vaccination coverage, the dashboard incorporates informative bar charts and funnel charts that illustrate the proportion of vaccinated individuals across various age groups. Notably, the dashboards have been thoughtfully designed to be inclusive of all genders, encompassing transgender individuals as well. Users can easily customize the timeframe of the displayed data to align with their specific requirements. A noteworthy achievement is that India has successfully administered a remarkable total of 70 million vaccination doses, as reported in the most recent survey and dataset. This milestone serves as a testament to the positive response and active participation of citizens in the nationwide vaccination campaign.

Figure 5 presents an interactive dashboard that provides a comprehensive overview of the state-wise testing status of COVID-19 in India. The dashboard showcases key metrics such as the number of positive tests and negative tests conducted in each state. By interacting with the dashboard, users can hover over the different orange marks on the map to access specific testing numbers for each state. This dashboard plays a crucial role in monitoring and evaluating the testing efforts across India, providing valuable insights

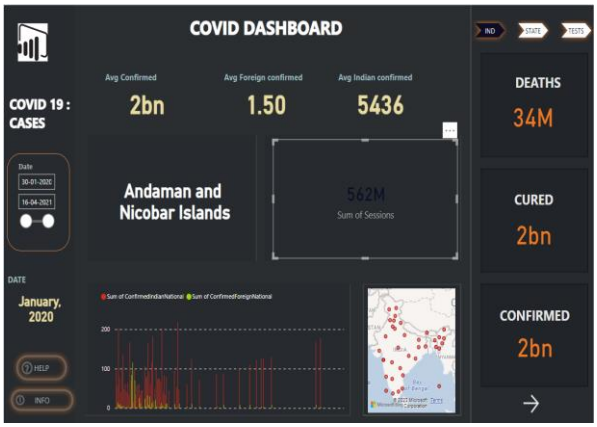

Figure 2: Dashboard for India-level vaccination data

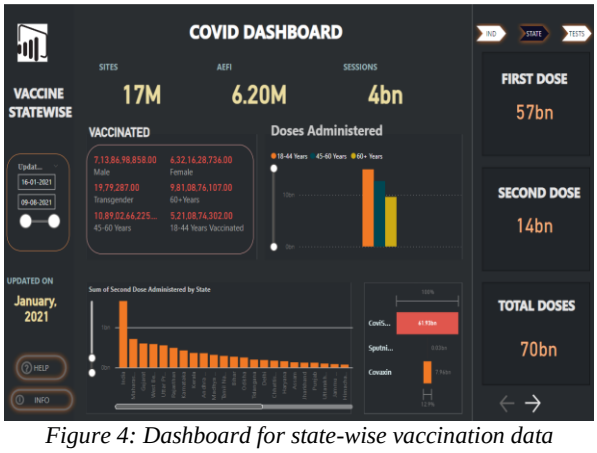

Figure 4: Dashboard for state-wise vaccination data

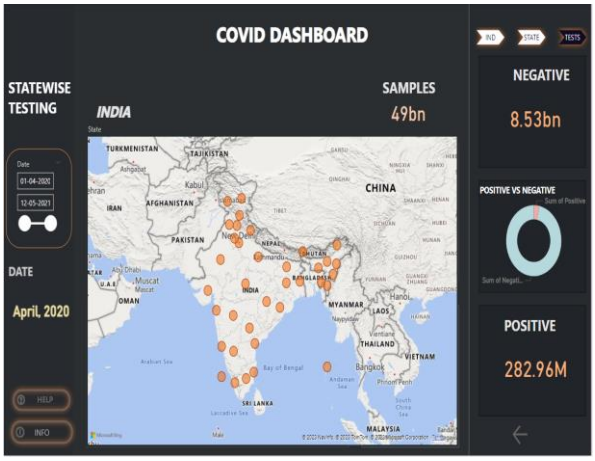

Figure 5: Dashboard for Covid-19 testing data

into the spread and containment of the virus. By examining the testing data, policymakers and healthcare officials can identify regions with high positivity rates, prioritize resource allocation, and implement targeted interventions to control the transmission of the virus.

Figure 6 presents a line graph illustrating the difference between confirmed cases of COVID-19 among Indian nationals and foreign nationals. The graph plots the data points using a line to demonstrate the trends over time. Line graphs are effective in showing the continuous variation of data points over a continuous interval or time period, making them suitable for analyzing patterns, trends, and relationships.

Based on the graph, it can be concluded that the highest number of confirmed cases among Indian nationals occurred at data point 55, with a total of 1,360.00% higher than the lowest number of confirmed cases at data point 16, which had only 15 cases. Data point 55 accounted for 4.03% of the total number of confirmed cases among Indian nationals. The divergence between the number of confirmed cases among Indian nationals and foreign nationals was most significant at data point 55, with a difference of 218 cases higher among Indian nationals compared to foreign nationals.

By utilizing a line graph, this visualization effectively displays the fluctuations and disparities in the number of COVID-19 cases between Indian nationals and foreign nationals over the analyzed time period. It enables the viewer to understand the magnitude of the differences and identify the points of greatest divergence in the data.

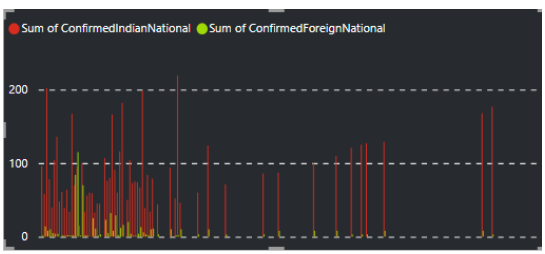

Figure 6 : Line graph expressing number of confirmed Covid-19 patients amongst Indian and Foreign Nationals

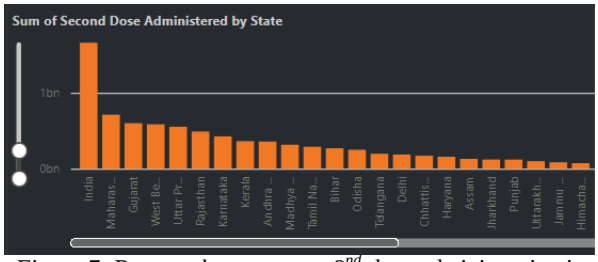

Figure 7: Bar graph to compare 2nd dose administration in different states

Figure 7 showcases a bar graph that provides a state-wise representation of the number of second doses administered after COVID-19 vaccination. Bar graphs, also known as bar charts, are widely used for comparing and presenting categorical or discrete data. Each rectangular bar in the graph corresponds to a specific category or data point, and the length or height of the bar represents the value or frequency associated with that category. According to the graph, Maharashtra had the highest sum of second doses administered, reaching 67,596,214. This value was 6,39,745.43% higher than the lowest sum of second doses administered in Lakshadweep, which stood at 1,056,446. The range of the sum of second doses administered across all 37 states and union territories varied from 1,056,446 to 6,759,621,451. By utilizing a bar graph, this visualization effectively presents a comparison of the number of second doses administered in different states, highlighting the significant variations in vaccination coverage. It allows viewers to quickly identify states with higher or lower numbers of second doses administered and provides a clear understanding of the magnitude of the differences.

Observe Figure 8, insights can be gained into the vaccination status of individuals based on their sexual orientation and age. The table provides a comprehensive overview of the total number of vaccinated individuals within each category. By analyzing the table, can identify the vaccination coverage for different sexual orientations and age groups. This information is crucial for understanding the inclusivity and effectiveness of the vaccination campaign in reaching diverse populations. The visualization allows us to assess any variations in vaccination rates among different sexual orientations and age groups. It provides valuable data on the number of vaccinated individuals in each category, enabling us to identify areas of success and potential areas that require more targeted efforts. Understanding the vaccination status based on sexual orientation and age is essential for ensuring equitable access to vaccines and addressing any disparities that may exist. By analyzing this data, policymakers and healthcare professionals can develop strategies to reach marginalized or underrepresented populations and ensure that vaccination efforts are comprehensive and inclusive.

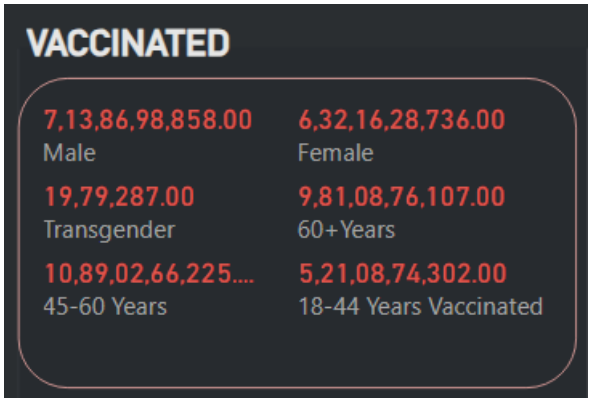

Figure 8: Table with vaccinated info acc. to gender & age

Figure 9 is a representation of number of users of different vaccines that were available in India. There were mainly three options, namely CoviShield, Sputnik V, and Covaxin. CoviShield, a widely used COVID-19 vaccine is based on the Oxford-AstraZeneca viral vector platform, offering effective protection against the virus with a two-dose regimen. Covaxin is an indigenous COVID-19 vaccine developed in India, while Sputnik V is a Russian-made vaccine. Both vaccines have been authorized for emergency use, contributing to global efforts in combating the pandemic. Use a funnel chart for the representation. A funnel chart, commonly known as a funnel diagram or sales funnel, is a visualization tool used to represent the progressive reduction of data or values at different stages of a process. It is particularly useful for illustrating the conversion rates or the step-by-step progression from a larger group to a smaller subset.

Observe carefully, the number of CoviShield vaccine users are the highest in numbers, approximately 61.93 billion. Next highest is Covaxin, with 7.9 billion users. Finally, Sputnik V, which is a foreign vaccine, was studied to be not too trusted by the Indian population with least number of users i.e., 0.03 billion.

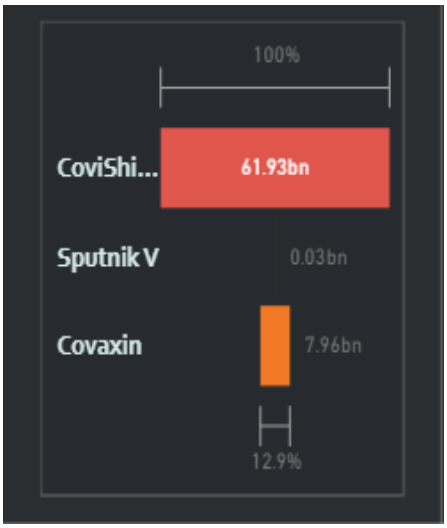

*Figure **Error! No text of specified style in document.**9: Funnel chart displaying data of different vaccine preferences*

4. CONCLUSION & SCOPE

In conclusion, this research study aimed to analyze the COVID-19 vaccination data for India, focusing on the dataset "covid_vaccine_statewise.csv." The analysis provided valuable insights into the vaccination trends in different states and highlighted areas where improvements are necessary. By understanding these trends, the government and healthcare officials can take actions to enhance the effectiveness of the vaccination drive and mitigate the impact of the COVID-19 pandemic in India.

The analysis of the vaccination numbers for the first and second doses in different states shed light on the progress and disparities among regions.

The examination of gender distribution in vaccination revealed valuable insights as well. While the dataset provided information on the number of males and females vaccinated, further analysis is necessary to assess any gender disparities.

The research findings serve as evidence for evidence-based decision-making. The conclusions drawn from this study can inform policy discussions, resource allocation, and targeted interventions to strengthen the COVID-19 vaccination drive in India. It is crucial for the government to prioritize areas with lower vaccination rates and address any barriers that hinder vaccination uptake, such as vaccine hesitancy, accessibility issues, or misinformation.

It is crucial to acknowledge the limitations of this study. The analysis was based on the provided dataset, which might not capture the complete picture of the vaccination efforts in India. Additionally, other factors influencing vaccination rates, such as socioeconomic factors or cultural beliefs, were not explored in detail. Future research can delve deeper into these aspects to gain a comprehensive understanding of the vaccination landscape.

REFERENCES

[1] https://www.analyticsvidhya.com/blog/2021/05/covid-19-vaccination-data-analysis-visualization

[2] S. Shaikh, J. Gala, A. Jain, S. Advani, S. Jaidhara and M. Roja Edinburgh, "Analysis and Prediction of COVID-19 using Regression Models and Time Series Forecasting," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021,

[3] Rimal, Y., Gochhait, S., & Bisht, A. (2021). Data interpretation and visualization of COVID-19 cases using R programming. Informatics in medicine unlocked, 26, 100705.

[4] Goel, I., Sharma, S., & Kashiramka, S. (2021). Effects of the COVID-19 pandemic in India: An analysis of policy and technological interventions. Health policy and technology, 10(1), 151–164.

[5] Ghosh, P., Ghosh, R., & Chakraborty, B. (2020). COVID-19 in India: Statewise Analysis and Prediction. JMIR public health and surveillance, 6(3), e20341.