ISSN 0974-763X

SOUTH ASIAN JOURNAL OF MANAGEMENT RESEARCH (SAJMR)

Volume 12 Number 1 **July, 2022**

Chhatrapati Shahu Institute of Business Education and Research (CSIBER)
(An Autonomous Institute)
University Road, Kolhpaur – 416 004 Maharashtra State.

SOUTH ASIAN JOURNAL OF MANAGEMENT RESEARCH (SAJMR)

ISSN 0974 – 763X

(An International Peer Reviewed Research Journal)

Published by

Chhatrapati Shahu Institute of Business Education & Research (CSIBER)

University Road, Kolhapur – 416 004, Maharashtra State

Ph: 91-231-2535706/07, Fax: 91-231-2535708, Website : www.siberindia.edu.in

Email: sajmr@siberindia.edu.in, sibersajmr@gmail.com

Insights of Exploratory Data Analysis (EDA) of Covid-19 Vaccine Administration in India

V. L Badadare, S. S Jamsandekar, R. S. Kamath
Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, MS – India

Abstract: The goal of the study is to analyze the vaccination administration in India with**a** keen interest in finding meaningful insights using exploratory data analysis. In this study dataset from Kaggle comprising of covid-19 vaccination records of 8 months from January 2021 to August 2021 is used Vaccination details of all 36 the states and Union Territories are recorded in the dataset, and Python libraries including *Pandas, Matplotlib, and Seaborn packagesare used to* compare state-wise vaccination progress, analyze the share of covaxin and covishield in total dosage administered and to find the acceptance rate of covid-19 vaccination in India. The finding of the analysis identified that Uttar Pradesh, Maharashtra, Rajasthan, and Gujarat are the top 4 states that administered the highest number of Covid-19 vaccine doses. The study also confirms that for all the states in India the proportion of Covishield vaccine dosage administered is higher as compared to Covaxin vaccine

Keywords: exploratory data analysis, Covid-19, vaccination, visualization, python

1.0 INTRODUCTION

The COVID-19 cases upsurge, officially known as the coronavirus disease epidemic, would be a continuing global public health problem of coronavirus disease 2019 (COVID-19). The outburst happened in Wuhan, China, in December 2019. The World Health Organization (WHO) classified COVID-19 as a Public Health Emergency of International Concern on January 30, 2020, and a pandemic on March 11, 2020. 79 231 893 cases globally, 10,207,871 cases in India had occurred as of 29th December 2020, with 1,754,574 fatalities globally and 147 901 in India caused due to COVID-19. By December 2021 globally the confirmed cases raised upto 280,119,931 and 5,403,662deaths and in India 34,799,691 confirmed cases and 480,290 total deaths, making it one of the worst pandemics in history.

A big step taken by WHO and supported by various countries in combating the COVID-19 pandemic was the acceptance of a vaccination campaign for SARS-CoV-2.The upsurge in various vaccines developed by different pharmaceutical companies spurred curiosity about learning more about ongoing vaccination administration.

Many researchers and statisticians have carried out different statistical and visual data analyses on vaccination data. Kamaraj, M. et. al. [1], studied types of covid vaccines available in India and the analyzed process of vaccination drive in India. They collected secondary data from the COWIN dashboard. They concluded that one must continue vaccination even if the vaccinations are less successful against new variants. Chakraborty, C. et. al. [2], have discussed the different parameters that lead to an increase in the Covid-19 second wave. They also illustrated barriers affecting covid vaccination drives in India Padma T. V. [3], stated that though India is one of the biggest suppliers of covid vaccine, it is struggling with lots of crises regarding covid vaccination. The author also stated that surging cases and vaccine

demand in India are leading to regional shortages

V.M. Kumar et al[4], discussed numerous aspects like, various vaccines available in India, and COVID-19 VACCINE DISTRIBUTION, The author emphasized the overall crux of vaccine development and vaccination strategies used during a pandemic in a densely populated country (India). Panda, D. S et. Al. [5], presented a survey on covid-19 vaccine acceptance and safety concerns from a public perspective for Odisha, India., They concluded that, some safety concerns in adults as well as in children like side effects of a vaccine, and effectiveness, the spectrum of activity and cost are irrespective of age, gender, marital status, and occupation.

With a keen interest in finding meaningful insights an exploratory data analysis on the scraped data from the Kaggle website using Python libraries including *Pandas, Matplotlib, and Seaborn packages is carried out* in this study. We have analyzed the vaccine administration data in India with the prime objectives of this study is to compare state-wise vaccination progress, analyze the share of covaxin and covishield in total dosage administered, and the acceptance rate of covid-19 vaccination in India

2.0 MATERIALS AND METHODS

In the previous Section, the authors reviewed several recent studies concerning covid vaccination in India, and it is interpreted that there is a need for proper analysis of Covid Vaccination status in India.

In this study data set used is Secondary data, collected from https://www.kaggle.com/ **and the tools used for analysis are** MS-Excel 2013, Google Python Collaboratory

Dataset Description: .csv format dataset comprising of State-wise Covid vaccination details in India incorporating 7633 records with 24 features is used for the analysis.

Dataset features: ['Updated On', 'State', 'Total Doses Administered', 'Sessions', 'Sites', 'First Dose Administered', 'Second Dose Administered', 'Male (Doses Administered)', 'Female (Doses Administered)', 'Transgender (Doses Administered)', 'Covaxin (Doses Administered)',' CoviShield (Doses Administered)', 'Sputnik V (Doses Administered)','AEFI', '18-44 Years (Doses Administered)', '45-60 Years (Doses Administered)', '60+ Years (Doses Administered)','18-44 Years(Individuals Vaccinated)', '45-60 Years(Individuals Vaccinated)','60+ Years(Individuals Vaccinated)', 'Total Individuals Vaccinated']
This dataset contains covid-19 vaccination records for 8 months from January 2021 to August 2021. Vaccination details of all 36 the states and Union Territories are recorded

The study is conducted in 3 phases, First phase is the Data Extraction phase in which the dataset from Kaggle.com was pulled and mounted on Google Colab using Panda libraries, followed by Data Wrangling Phase, carrying out different data preprocessing techniques to make data ready for EDA. The third phase is the Exploratory Data Analysis to attain the objectives of the study using python matplotlib and seaborn libraries.

3.0 RESULTS AND DISCUSSION

Data Extraction Secondary dataset downloaded from Kaggle.com is stored on google drive and the drive is mounted in google colab, snapshot of it is shown in Figure 1.

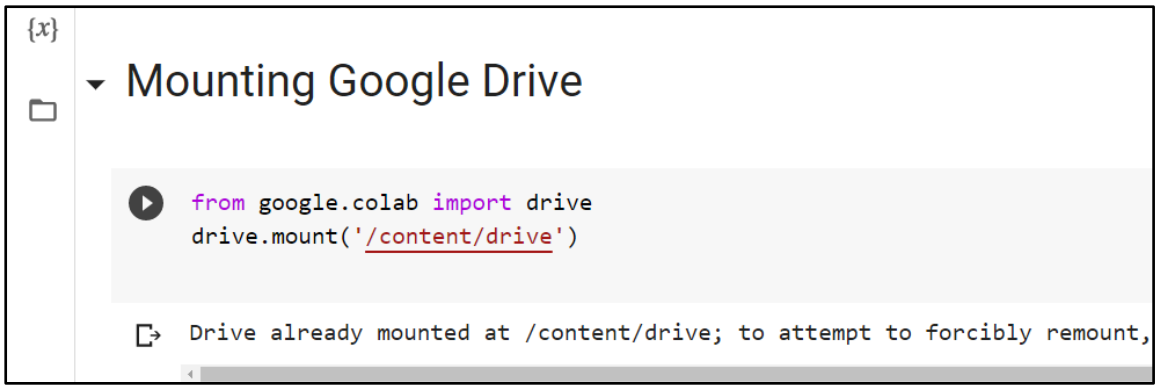
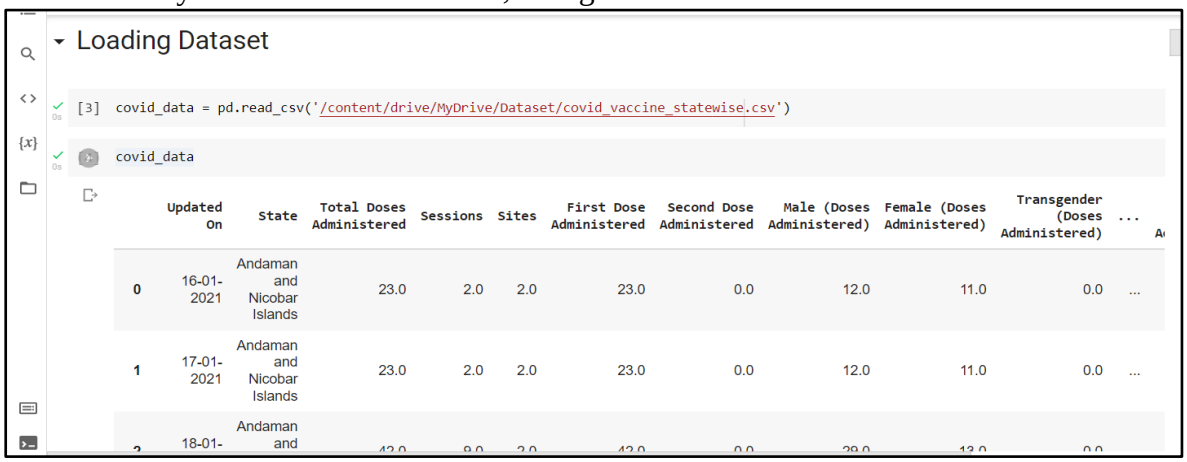


Figure 1: Google Colab Drive Mount

Pandas library is used to load a dataset, in Figure 2 shows the data set records



	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...
0	16-01-2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...
1	17-01-2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...
2	18-01-	Andaman and	43.0	8.0	3.0	43.0	0.0	29.0	13.0	0.0	

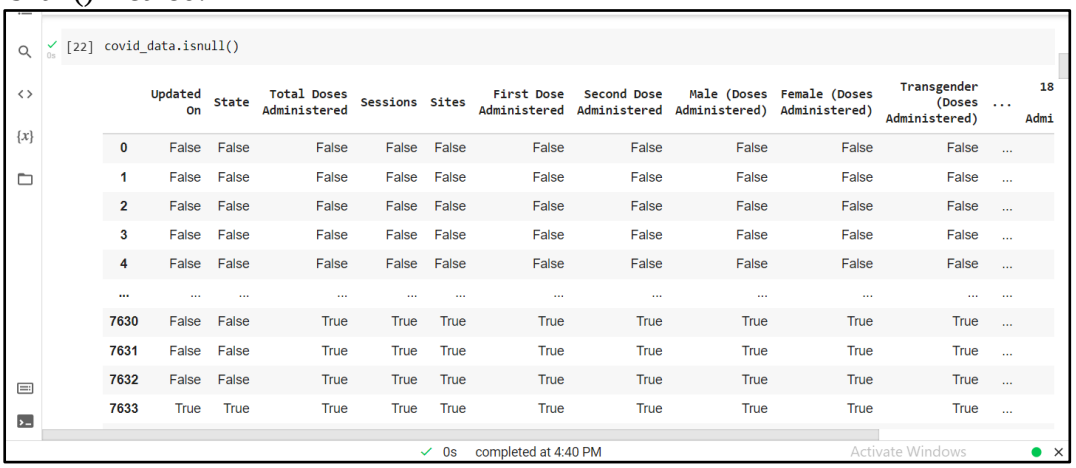
Figure 2: Snapshot of Covid vaccine statewise dataset loaded

Data Wrangling

The phase of Data wrangling contains numerous data pre-processing activities such as

- Checking Dataset for missing values
- Handling missing data
- Feature Selection

Checking Dataset for missing values: check for missing values in the dataset is done by isnull() method.



	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	18 Admi
0	False	False	False	False	False	False	False	False	False	False	...	
1	False	False	False	False	False	False	False	False	False	False	...	
2	False	False	False	False	False	False	False	False	False	False	...	
3	False	False	False	False	False	False	False	False	False	False	...	
4	False	False	False	False	False	False	False	False	False	False	...	
...	...	...	...	...	...	...	...	...	...	...	...	
7630	False	False	True	True	True	True	True	True	True	True	...	
7631	False	False	True	True	True	True	True	True	True	True	...	
7632	False	False	True	True	True	True	True	True	True	True	...	
7633	True	True	True	True	True	True	True	True	True	True	...	

Figure 3: Output of missing values check

The results in Figure 3 indicate that the dataset contains some null values. Records with missing values may lead to false data analysis

Handling missing data: In this study, the authors have attempted to omit all the rows containing the Null value and followed by omitting the columns containing the Null value. The resultant dataframe is the "Clean_Data1" as shown in Figure 4. And Figure 5 with no missing value. This pre-processed dataframe has 7633 records with 19 features

```
[5]  # Removing Null Records
     Clean_Data = covid_data.dropna(axis=0, how="all", thresh=None, subset=None, inplace=False)
     print(Clean_Data)
```

```
      Updated On                      State  Total Doses Administered  \
0     16-01-2021  Andaman and Nicobar Islands                      23.0
1     17-01-2021  Andaman and Nicobar Islands                      23.0
2     18-01-2021  Andaman and Nicobar Islands                      42.0
3     19-01-2021  Andaman and Nicobar Islands                      89.0
4     20-01-2021  Andaman and Nicobar Islands                     124.0
...          ...                          ...                       ...
7628  08-11-2021                West Bengal                       NaN
7629  08-12-2021                West Bengal                       NaN
7630  13-08-2021                West Bengal                       NaN
7631  14-08-2021                West Bengal                       NaN
7632  15-08-2021                West Bengal                       NaN
```

Figure 4: Dataframe with all missing row data omitted

```
[6]  Clean_Data1 = Clean_Data.dropna(axis=1, how="all", thresh=None, subset=None, inplace=False)
     Clean_Data1
```

	Updated On	State	Total Doses Administered	Sessions	Sites	First Dose Administered	Second Dose Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	...	A
0	16-01-2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	
1	17-01-2021	Andaman and Nicobar Islands	23.0	2.0	2.0	23.0	0.0	12.0	11.0	0.0	...	
2	18-01-2021	Andaman and Nicobar Islands	42.0	9.0	2.0	42.0	0.0	29.0	13.0	0.0	...	
		Andaman										

Figure 5: Dataframe with all missing column data omitted

Feature Selection

For desired data analysis some features are selected manually as below:

```
[9]  # Feature Selection
     Final_Data = Clean_Data1[['Updated On', 'State', 'Total Doses Administered','Male (Doses Administered)', 'Female (Doses Administered)',
           'Transgender (Doses Administered)', 'Covaxin (Doses Administered)','CoviShield (Doses Administered)','Total Individuals Vaccinated']]
```

4.0 Exploratory Data Analysis

In the previous phase,the authors have selected 9 features manually for further Exploratory Data Analysis. The current phase of Exploratory Data Analysis is divided into the following analysis activities:

- **State-wise vaccination details**
- **Share of Covaxin and Covishield in total dosage administered**
- **The acceptance rate of covid-19 vaccination in India**

State-wise vaccination details groupby() functionality is applied on Final_Data Dataframe. For this grouping,the 'State' feature is used as a grouping key, and sum() is used as the aggregation function shown in figure 6.

```
[ ] # Statewise Record Grouping
    grp_By_state = Final_Data.groupby(by=['State'], dropna=True).sum()
    grp_By_state
```

State	Total Doses Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	Covaxin (Doses Administered)	CoviShield (Doses Administered)	Total Individuals Vaccinated
Andaman and Nicobar Islands	2.054441e+07	1.045375e+07	8.964974e+06	1926.0	0.0	2.051143e+07	8.102125e+06
Andhra Pradesh	1.591678e+09	6.776694e+08	7.550539e+08	212810.0	263695249.0	1.323987e+09	5.645911e+08
Arunachal Pradesh	6.093730e+07	3.061580e+07	2.463594e+07	17345.0	1820.0	6.080052e+07	2.108156e+07
Assam	7.163890e+08	3.557939e+08	3.035113e+08	96969.0	118720291.0	5.970823e+08	2.397691e+08
Bihar	1.741293e+09	8.693946e+08	7.472145e+08	296154.0	156807520.0	1.582600e+09	6.608479e+08
Chandigarh	5.629684e+07	2.920781e+07	2.216730e+07	6599.0	117678.0	5.614530e+07	1.973150e+07
Chhattisgarh	9.681233e+08	4.381651e+08	4.507674e+08	96639.0	57309565.0	9.095543e+08	4.353092e+08
Dadra and Nagar Haveli and	3.818948e+07	2.378859e+07	1.263831e+07	6761.0	1072.0	3.817282e+07	1.135860e+07

Figure 6: groupby () functionality is applied on Final_Data Dataframe

It was observed that all the data values are of float type, so it is required to convert these values to the integer type. Figure 7 and figure 8 show the conversion of float to integer and its results respectively

```
[ ] grp_By_state.dtypes

Total Doses Administered            float64
Male (Doses Administered)           float64
Female (Doses Administered)         float64
Transgender (Doses Administered)    float64
Covaxin (Doses Administered)        float64
CoviShield (Doses Administered)     float64
Total Individuals Vaccinated        float64
dtype: object

# converting all columns to int type
grp_By_state = grp_By_state.astype(int)
grp_By_state.dtypes

Total Doses Administered            int64
Male (Doses Administered)           int64
Female (Doses Administered)         int64
Transgender (Doses Administered)    int64
Covaxin (Doses Administered)        int64
CoviShield (Doses Administered)     int64
Total Individuals Vaccinated        int64
dtype: object
```

Figure 7 : Conversion of data frame values from float to integer

```
[ ] grp_By_state
```

State	Total Doses Administered	Male (Doses Administered)	Female (Doses Administered)	Transgender (Doses Administered)	Covaxin (Doses Administered)	CoviShield (Doses Administered)	Total Individuals Vaccinated
Andaman and Nicobar Islands	20544408	10453750	8964974	1926	0	20511428	8102125
Andhra Pradesh	1591678440	677669372	755053860	212810	263695249	1323987110	564591113
Arunachal Pradesh	60937297	30615798	24635938	17345	1820	60800520	21081557
Assam	716389018	355793883	303511268	96969	118720291	597082285	239769091
Bihar	1741293449	869394635	747214483	296154	156807520	1582599853	660847900
Chandigarh	56296840	29207808	22167297	6599	117678	56145302	19731496
Chhattisgarh	968123302	438165070	450767363	96639	57309565	909554256	435309241
Dadra and Nagar Haveli and Daman and Diu	38189479	23788589	12638310	6761	1072	38172816	11358600
Delhi	812558419	423907445	305152174	148510	218692040	592068871	305737211
Goa	92189542	44847366	40607362	7244	1469705	90655057	32114784
Gujarat	2732064385	1326682808	1127534017	338188	307771021	2420791163	1078260544

Figure 8: Data frame values after conversion

Based on the above-mentioned outcome, visualization of State-wise vaccination details is presented by using a barplot as shown in figure 9. pyplot module of the Matplotlib package is used for the aforementioned visualization

Figure 9: State-wise vaccination visualization

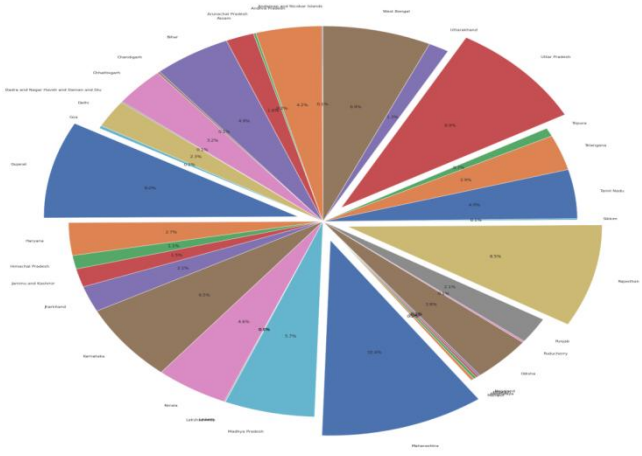


Figure 10: Pieplot visualization of state-wise vaccines administered

From a foreknown Exploratory Data Analysis, it is interpreted that,Maharashtra, Gujarat, Uttar Pradesh, and Rajasthan are the top 4 states that administered the highest number of Covid-19 vaccine doses from January 2021 to August 2021 shown in Figure 10.

Share of Covaxin and Covishield in total dosage administered

Most of the populace from India preferred either Covishield or Covaxin to fight against the plight of Covid-19. So, in the current phase, it is estimated that these two vaccines arethe most chosen. A Scatter

plot is used to present the state-wise proportion of Covaxin and Covishield in the total dosage administered

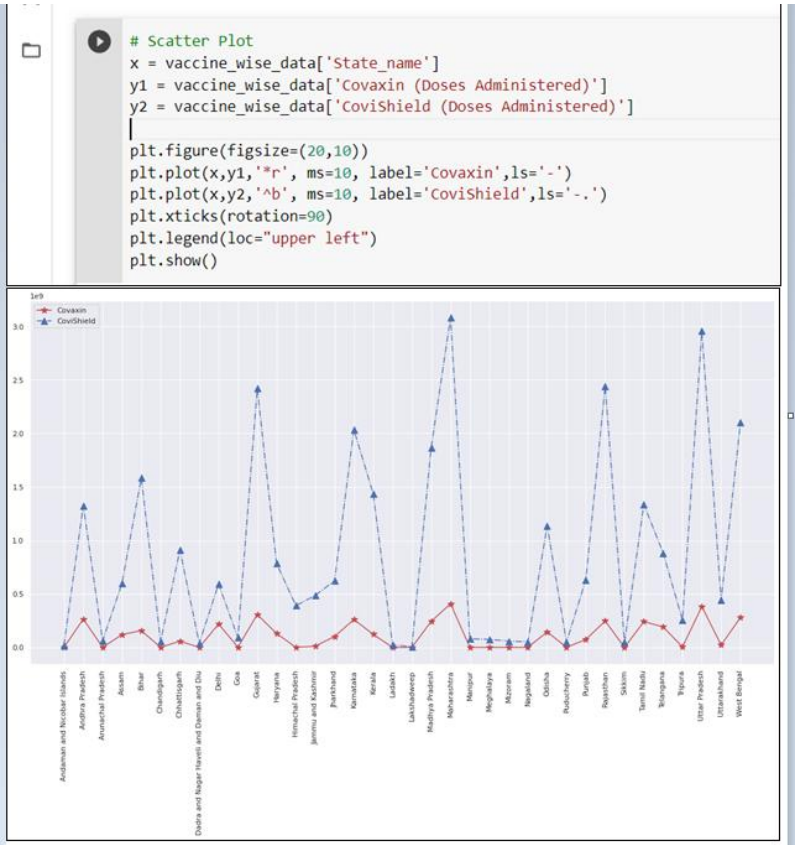


Figure 11: Scatterplot to visualize the state-wise proportion of Covaxin and Covishield in total dosage administered

It is observed from figure 11 that in almost all the states proportion of Covishield dosage administered is higher as compared to Covaxin

The acceptance rate of covid-19 vaccination in India

The rapid development of the covid-19 vaccine has given some hope to health practitioners around the world. The safety of the populace was dependent on vaccination,But the main challenge was acceptance of these vaccines among people.

To show the acceptance rate authors have plotted individuals vaccinated against time, the plot in Figure 12. plots x-axis (Date) and Y-axis (total individuals vaccinated), from the plot it is observed that there is a positive correlation. i.e. Vaccination rate is increasing day by day

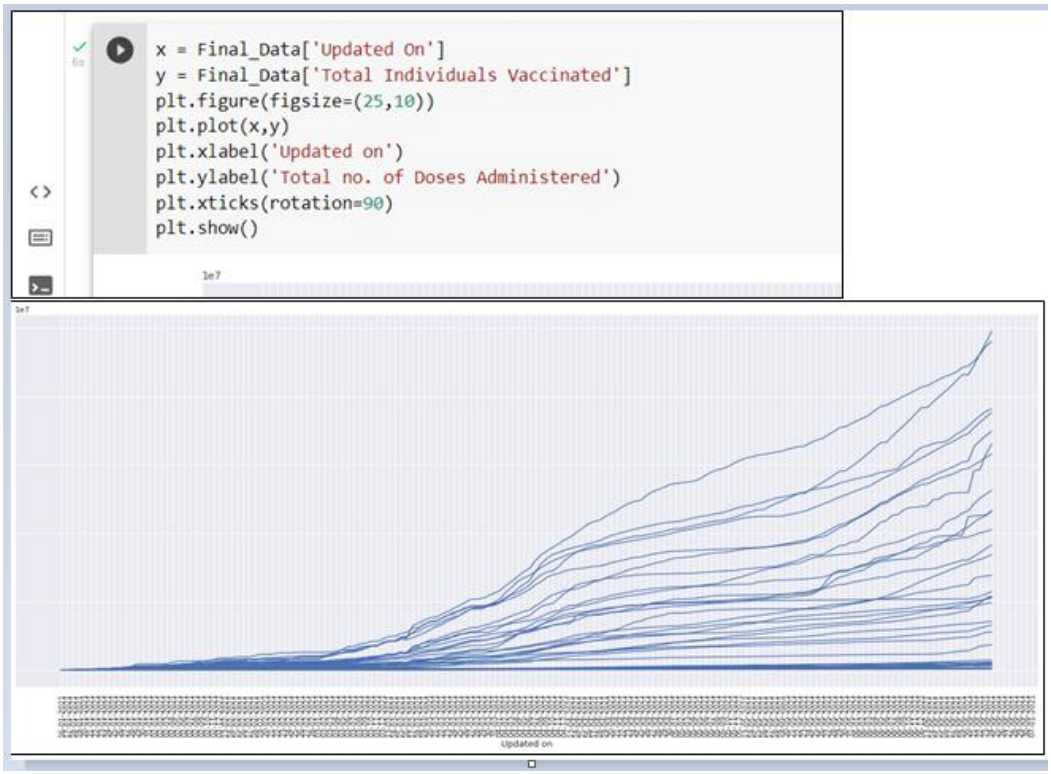


Figure 12: Line plot of date wise individuals vaccinated

5.0 CONCLUSION

In this article,the authors have visualized and debated the COVID-19 vaccination in India in terms of everyday vaccinations as per state, cumulative vaccinations per state, and share of Covaxin and Covishield in total dosage administered. The finding of the analysis identified that Uttar Pradesh, Maharashtra, Rajasthan, and Gujarat are the top 4 states that administered the highest number of Covid-19 vaccine doses. The study also confirms that for all the states in India the proportion of Covishield vaccine dosage administered is higher as compared to the Covaxin vaccine. The acceptance rate of vaccines is observed to be increasing.

6.0 REFERENCES

Kamraju, M., Ali, M. A., Mohd, *, & Ali, A. (2021). A Study on Covid-19 Vaccination Drive in India. BRICS Journal of Educational Research, 11(2), 76–79. https://www.researchgate.net/publication/352903714

Chakraborty, C., Sharma, A. R., Bhattacharya, M., Agoramoorthy, G., & Lee, S. S. (2021). The current second wave and COVID-19 vaccination status in India. In Brain, Behavior, and Immunity (Vol. 96, pp. 1–4). Academic Press Inc. https://doi.org/10.1016/j.bbi. 2021.05.018

Padma T. V. (2021). India's Vaccine Crunch.

Kumar, V. M., Pandi-Perumal, S. R., Trakht, I., & Thyagarajan, S. P. (2021). Strategy for COVID-19 vaccination in India: the country with the second highest population and number of cases. In npj **Vaccines (Vol. 6, Issue 1).** Nature Research. https://doi.org/10.1038/s41541-021-00327-2

Panda, D. S., Giri, R. K., Nagarajappa, A. K., & Basha, S. (2021). Covid-19 vaccine, acceptance, and concern of safety

from public perspective in the state of Odisha, India. Human Vaccines and Immunotherapeutics, 17(10), 3333–3337. https://doi.org/10.1080/21645515.2021.1924017

Kamath, Rajani & Jamsandekar, Shruti & Kharade, Kabir & Kamat, Rajanish. (2019). Data Analytics in R: A Case Study Based Approach.

Muthusami Rathinasamy Saritha Kandhasamy, An exploratory analysis of COVID‑19 vaccination administration, Proceedings of the Indian National Science Academy Springer Publications (2022) 88:43–52 https://doi.org/10.1007/s43538-022-00064-7