Introduction

Since December 2019 the world is experiencing a deadly disease caused by a novel coronavirus termed as severe acute respiratory syndrome coronavirus 2. The disease associated with this virus is known as COVID-19.It has been spreading since the end of 2019 and has infected more than 100 million people worldwide. In India the situation is worse than most of the countries and is the second most infected country globally. Due to such a large population the data collected for Covid19 is huge and it needs to be presented in ways which are understandable and used to make important predictions. Data Visualization is the first step towards getting an insight into a large data set in every data science project. The aim here is to extract useful information from the data. We have used Python and its few powerful libraries to achieve the task. This project focuses on data visualisation on COVID-19 based on freely available datasets including the ones in Kaggle repository. Data visualisation is provided on a number of aspects of COVID-19 including the symptoms of this disease, total recovered confirmed and fatal cases in India and globally, comparison of infections in males/females which shows that males are more prone to this disease and the older people are more at risk. The project helps the user to visualize the present condition globally through construction of geocode graphs, the project uses JavaScript charts JS library to achieve the same. The UI is constructed using the JavaScript library Bootstrap-4 which neatly does the job. The routes and the server are managed using Nodejs which is a server-side JavaScript run-time. Express,the famous Nodejs web application framework supports the project to work from various API's to collect the data from the web and display in a neat manner to the user. The user can use the web application and get information on various parameters and compare the data based on different conditions.

Problem Statement

The global outbreak of COVID-19 has had a strong impact on economic and social life in various countries. Everyday there are a huge amount of positive cases, recoveries and deaths in every country. The data collected on Covid19 is growing day by day due to the huge population of India. To deal with such huge data it is important to visualize it, data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend and therefore makes it easier to identify trends, patterns, and outliers within large data sets. Our project aims to propose different visual idioms based on the user requirements. The proposed method is able to visualize different conditions from the data collected. It compares different vaccinations available globally and which country has vaccinated the most percentage of its citizens. The project is integrated with a web-application which makes it easier for any user to use it. The dataset is collected from open Kaggle datasets and uses a Kaggle notebook to plot the graphs using various different libraries.

Data Collection

Datasets used are:

S No.	Name	Used For
1	Covid_19_india	Visualize Geo-plots and line charts for India and states
2	Age-group	% of each group affected due to covid
3	Global_Covid	Plotting global Geoplot showing covid in each country
4	Global_vaccines	Total global vaccines present for covid comparison
5	Indian_vaccine	Visualize vaccine process with respect to states and gender

The following datasets were used to make the visual idioms shown below. The snapshots of the datasets are included below and the attributes are divided on categorical and quantitative.

1. Covid_19_india dataset: It contains 15000+ entries

	A	B	C	D	E	F	G	H	I
1	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Deaths	Confirmed
2	1	30/01/20	6:00 PM	Kerala	1	0	0	0	1
3	2	31/01/20	6:00 PM	Kerala	1	0	0	0	1
4	3	1/2/2020	6:00 PM	Kerala	2	0	0	0	2
5	4	2/2/2020	6:00 PM	Kerala	3	0	0	0	3
6	5	3/2/2020	6:00 PM	Kerala	3	0	0	0	3
7	6	4/2/2020	6:00 PM	Kerala	3	0	0	0	3
8	7	5/2/2020	6:00 PM	Kerala	3	0	0	0	3
9	8	6/2/2020	6:00 PM	Kerala	3	0	0	0	3
10	9	7/2/2020	6:00 PM	Kerala	3	0	0	0	3
11	10	8/2/2020	6:00 PM	Kerala	3	0	0	0	3
12	11	9/2/2020	6:00 PM	Kerala	3	0	0	0	3

Categorical attributes present – State/Union Territory

Quantitative attributes present- Date, Time, Confirmed Indians,

Confirmed-Foreign-National, Confirmed, Cured, Deaths.

2.Global COVID-19:

date	country	# cumulative...	# daily_new_...	# active_cas...	# cumulative...
2020-2-15	Afghanistan	0.0		0.0	0.0
2020-2-16	Afghanistan	0.0		0.0	0.0
2020-2-17	Afghanistan	0.0		0.0	0.0
2020-2-18	Afghanistan	0.0		0.0	0.0
2020-2-19	Afghanistan	0.0		0.0	0.0
2020-2-20	Afghanistan	0.0		0.0	0.0
2020-2-21	Afghanistan	0.0		0.0	0.0
2020-2-22	Afghanistan	0.0		0.0	0.0
2020-2-23	Afghanistan	0.0		0.0	0.0
2020-2-24	Afghanistan	1.0		1.0	0.0
2020-2-25	Afghanistan	1.0	0.0	1.0	0.0
2020-2-26	Afghanistan	1.0	0.0	1.0	0.0

Categorical attributes present – Country.

Quantitative attributes present- Date, daily_new_ cases, active_cases, cumulative_total_cases, cumulative_total_deaths, daily_new_deaths .

3. Global Vaccine:

▲ location	▢ date	▲ vaccine	# total_vacci...
Chile	2020-12-24	Pfizer/BioNTech	420
Chile	2020-12-25	Pfizer/BioNTech	5198
Chile	2020-12-26	Pfizer/BioNTech	8338
Chile	2020-12-27	Pfizer/BioNTech	8649
Chile	2020-12-28	Pfizer/BioNTech	8649
Chile	2020-12-29	Pfizer/BioNTech	8649
Chile	2020-12-30	Pfizer/BioNTech	8649
Chile	2020-12-31	Pfizer/BioNTech	8649
Chile	2021-01-01	Pfizer/BioNTech	8649
Chile	2021-01-02	Pfizer/BioNTech	8649
Chile	2021-01-03	Pfizer/BioNTech	8649
Chile	2021-01-04	Pfizer/BioNTech	9054

Categorical attributes present – Location, Vaccine.

Quantitative attributes present- Date, Total Vaccine.

Tools Used

- **Numpy -** NumPy is a Python library used for working with arrays. It also has functions for working in domain of linear algebra, fourier transform, and matrices.

- **Pandas -** Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

- **Matplotlib-** Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications.

- **Seaborn-** Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

- **Plotly-** The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

- **OrderedDict-** OrderedDict is a dict subclass that preserves the order in which key-value pairs, commonly known as items, are inserted into the dictionary.

- **Datetime-** It is a Python library used for Time access and conversions.

- **Bar_chart_race-** Make animated bar chart races in Python with matplotlib. Img.

- **Nodejs-** Node.js is an open-source, cross-platform, back-end JavaScript runtime environment that runs on the V8 engine and executes JavaScript code outside a web browser.

- **Npm-** It is a package manager for the JavaScript programming language. npm, Inc. is a subsidiary of GitHub, that provides hosting for software development and version control with the usage of Git.

- **Chart.js-** Chart.js is a free open-source JavaScript library for data visualization, which supports 8 chart types: bar, line, area, pie, bubble, radar, polar, and scatter.

Tasks and Actions

The project has been broken down into various stages and the workflow of the same has been presented below. The user requirements addressed are mentioned below. The dataset used are public datasets which are available on Kaggle. The dataset used are up-to-date to get the best results. Furthermore, appropriate visual idioms are proposed for the defined user requirements such as pie charts, geocoded graphs to display cases in different geographical regions, histograms are used in case of visualising frequencies and many more different visual idioms are made. The plotting is done by popular python libraries including geopandas, seaborn ,matplotlib etc.

The project also consists of a web-application which uses API's to fetch data from reliable source and used the data collected in JSON format to visualise various idioms. The benefit of using API's is that the values keep on updating on regular intervals and there is no need to hard code the data. The web application is constructed by using NodeJs for backend and collecting the data from API. It is responsible for handling different routes. The front end of the web app is made using HTML and CSS.Javascript is used to add interactivity to the application and to plot the data which is sent by Nodejs. ChartJs library is used for the same. Lastly the application is deployed on Heroku a free web-based application hosting platform so that it can be accessed from anywhere.
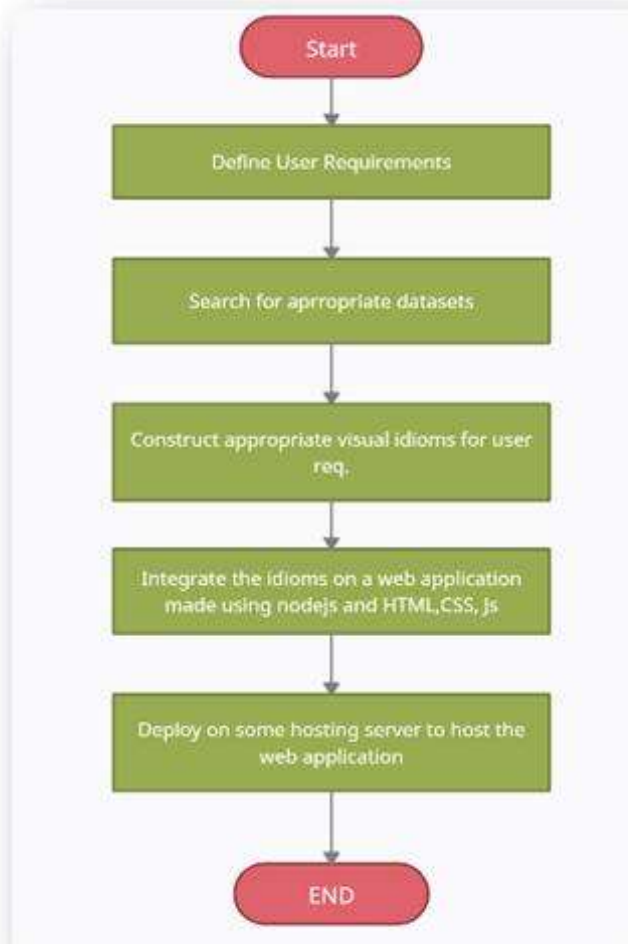
Fig: A workflow of the Project

User Requirements

- User can compare different countries on the basis of total confirmed Covid19 cases
- User should be able to get the latest updates about Covid count in different countries
- The model should analyse the patient data and give some insights based on gender and age parameters
- A visualization idiom to compare vaccination process should be done
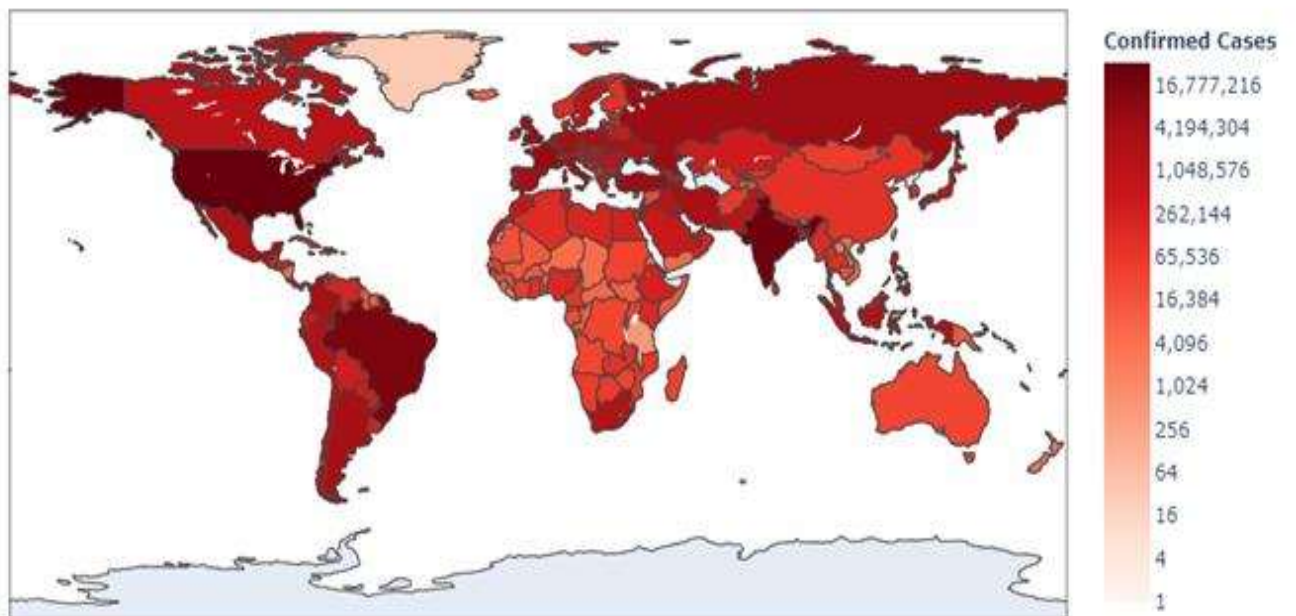- The user should be able to access a web-application hosted online for easy access

Visual Idioms used to address User Requirements

To visualize Covid19 situation on a global level

Visual Idioms Used:

Geo-Map since the data is distributed across different geographical regions globally

Coronavirus Confirmed Cases Around The Globe

Tree-map is primarily used to display data that is grouped and nested in a hierarchical (or tree-based) structure.
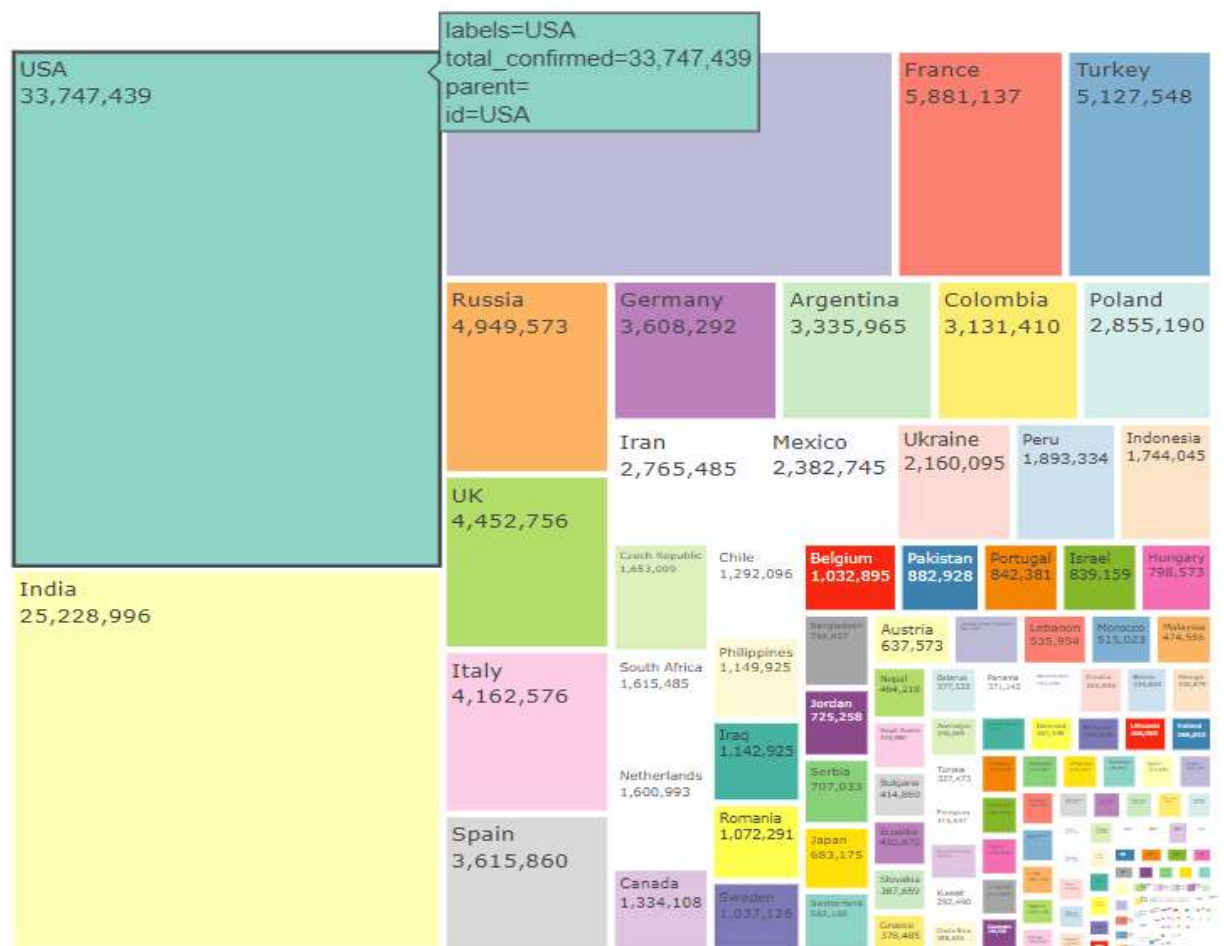
To Visualize Covid19 situation in India

Geomaps are used to visualize the same since the data is distributed among different states in India which are geographically separated

Tree-map is used to compare various Indian states on the basis of severity of Covid situation

Geomap showing state-wise confirmed cases
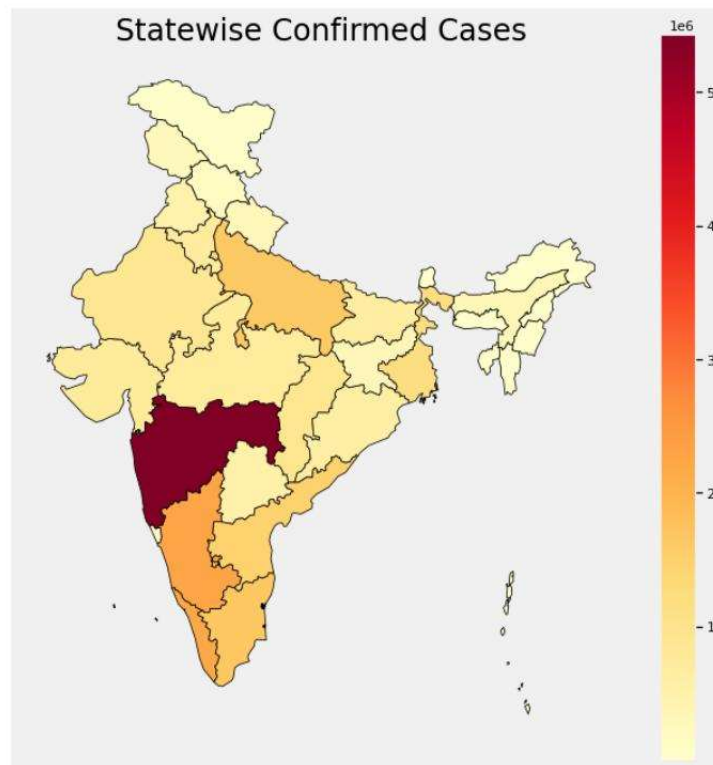
Geomap showing state-wise Recoveredcases
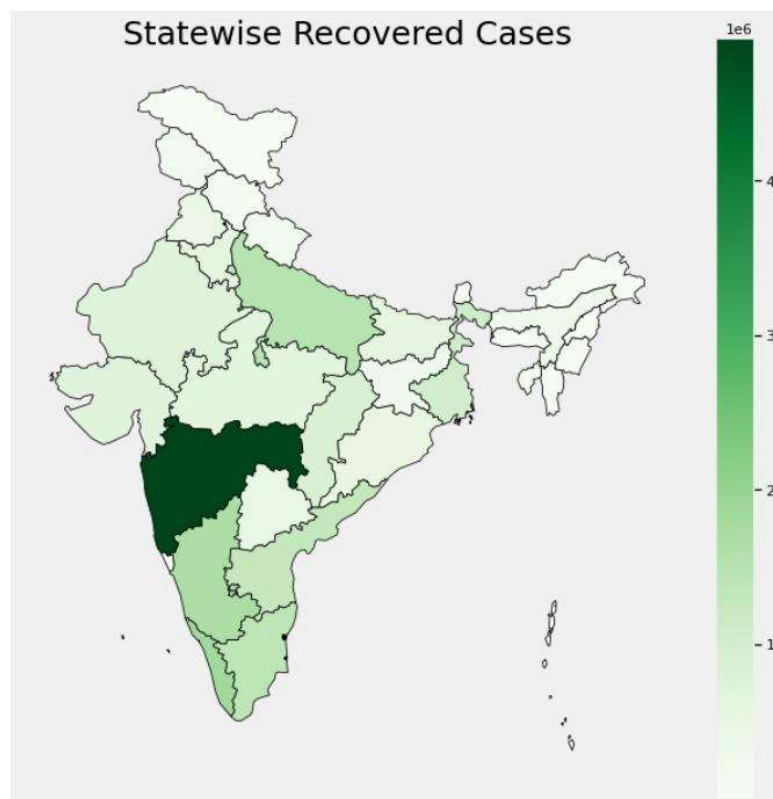
Geomap showing state-wise Fatal cases

Statewise Deaths
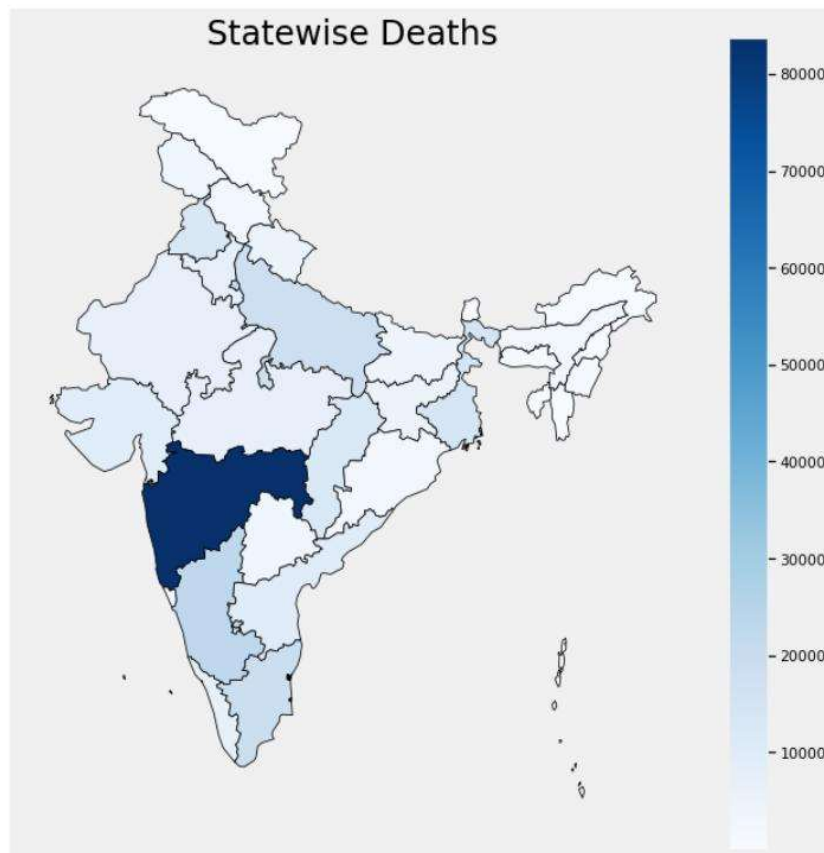
Tree-Map showing the overall Indian states comparison

Overall States Comparision of deaths
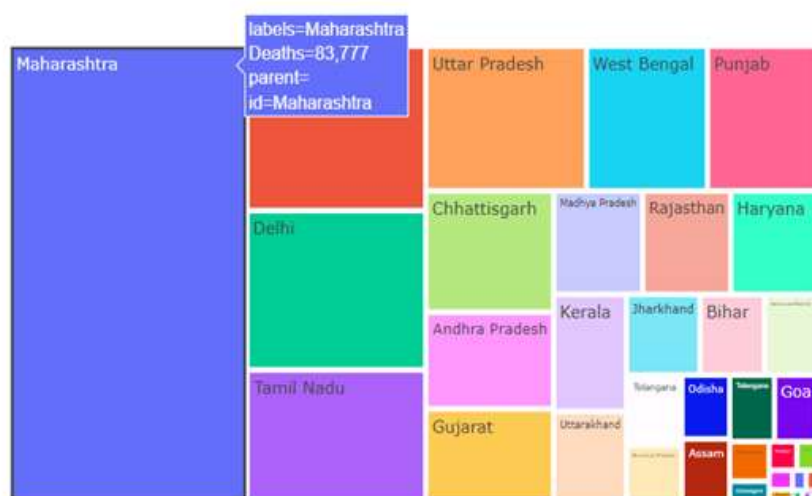
Visual Idiom to show the growth of the cases using Line chart since January 2020
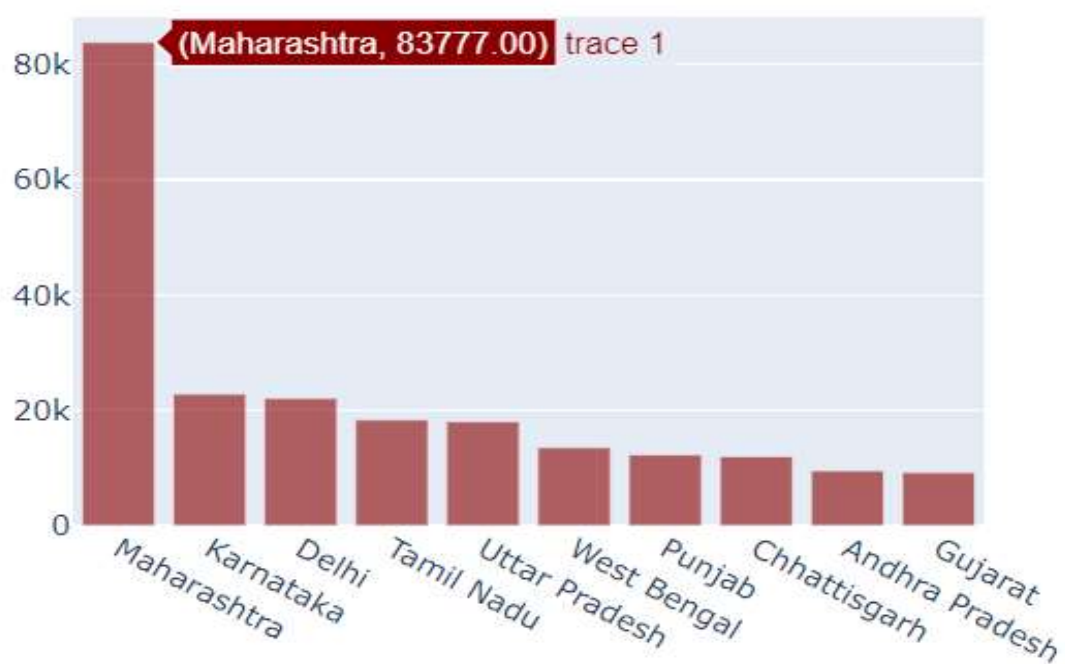
India COVID-19 Cases
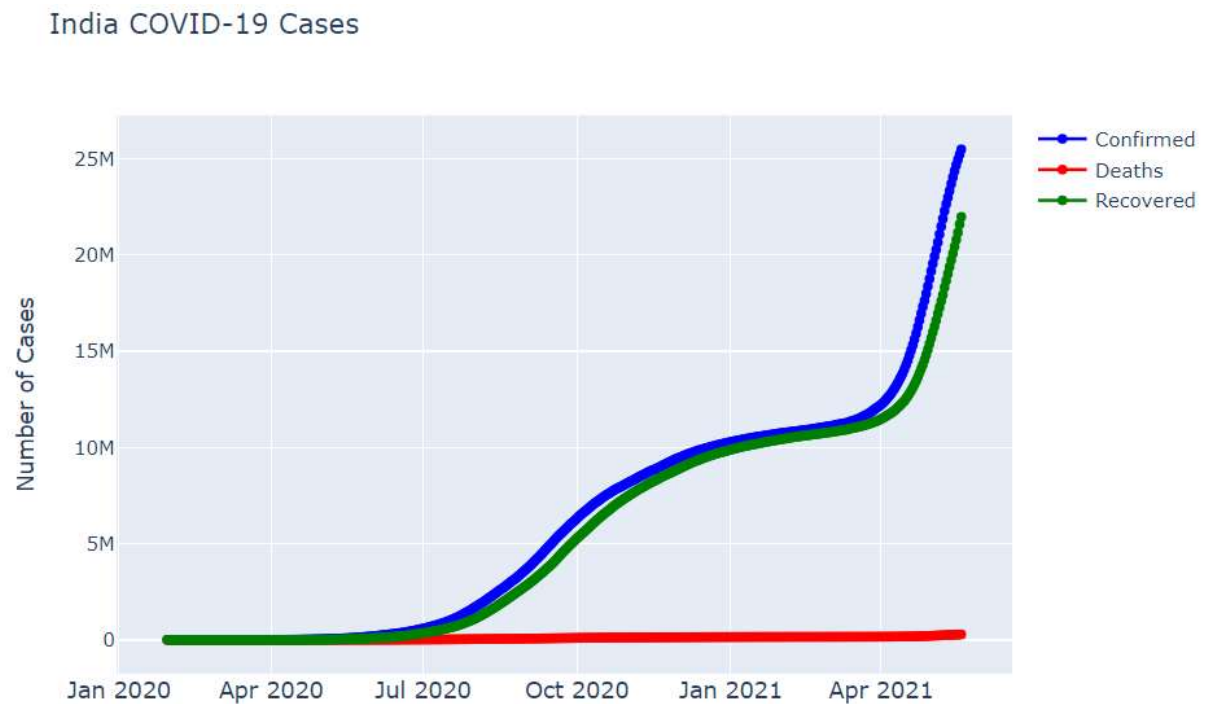
Bar chart is plotted for top 10 Indian states with most active cases with Maharashtra being the state with most active cases followed by Karnataka and Kerala
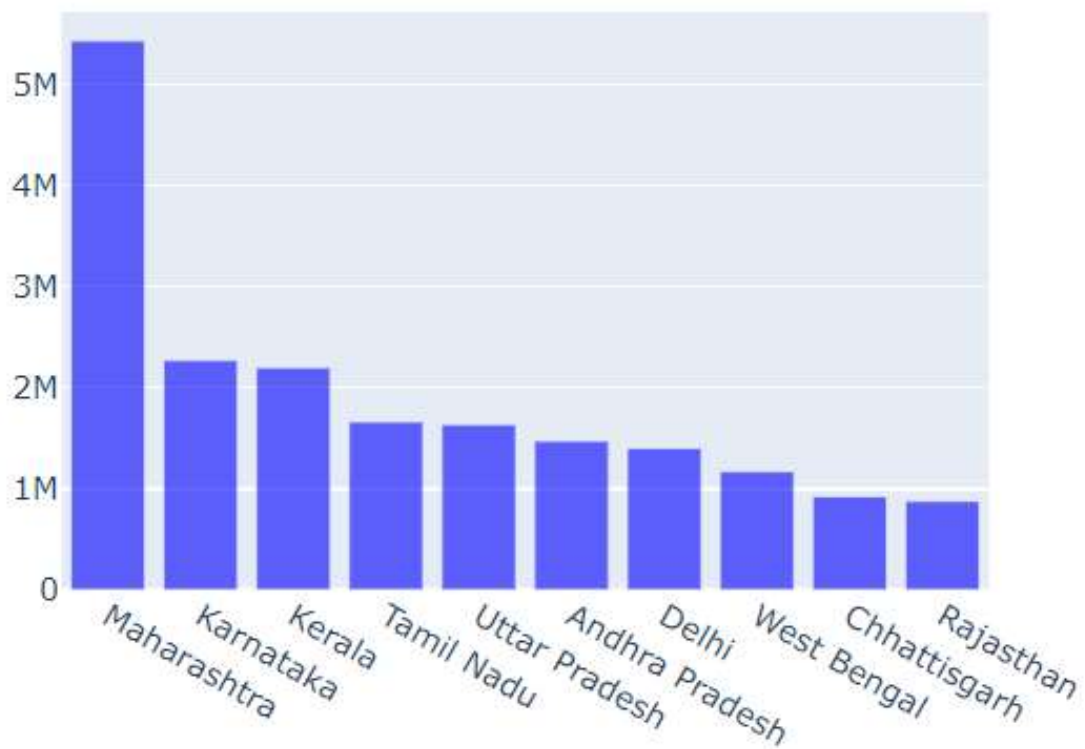
Similarly, Bar chart is plotted for top 10 Indian states with most Recovered Cases
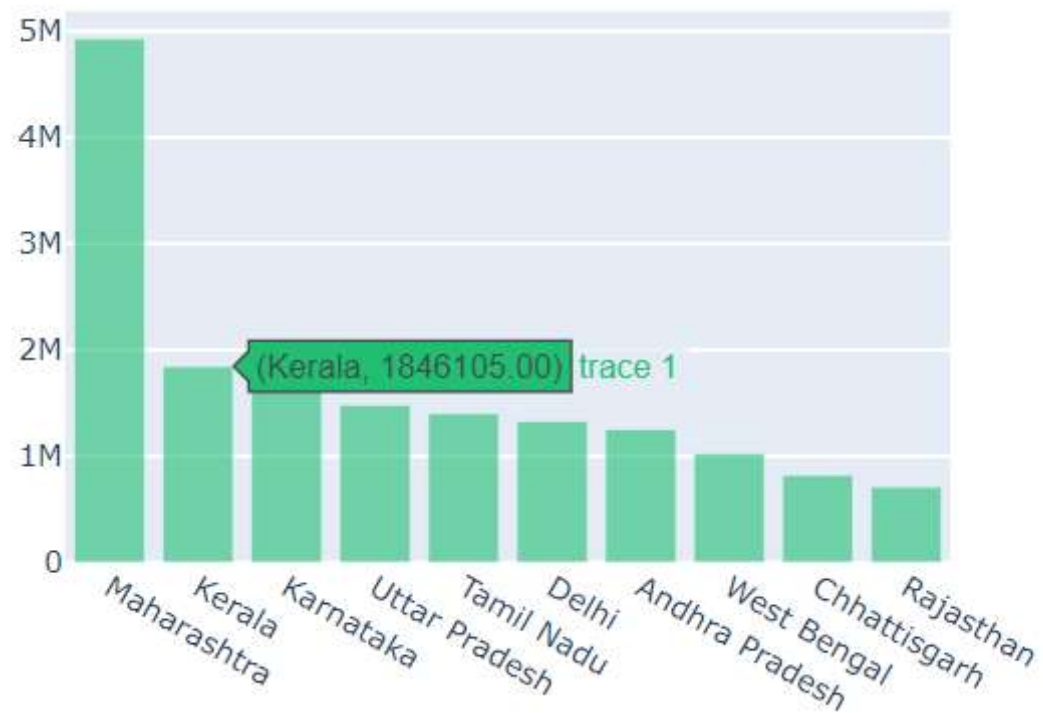
Bar chart is plotted for top 10 Indian states with most Fatal Cases

To Visualize Covid19 Testing situation in Indian States

The bar plot shows the testing in the major Indian states with UP being the state with most test conducted followed by Bihar and Tamil Nadu.
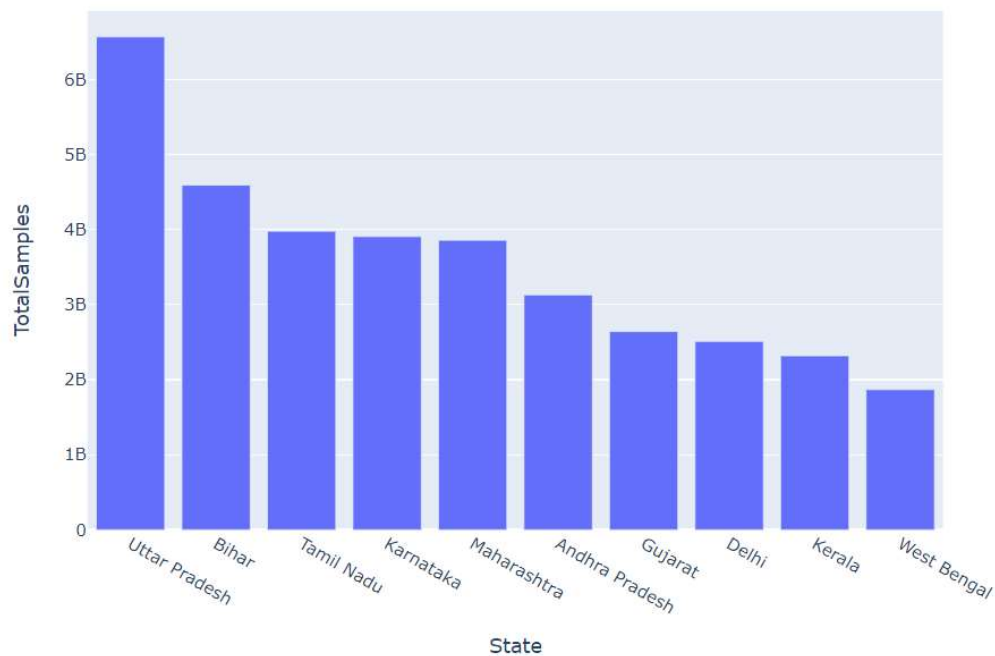

Summarised cases and death-per 100 and Cure rate per 100 is shown in a tabular format using pretty table library with gradient corresponding to values for cleared visualisation.

| State/UnionTerritory	Confirmed	Deaths	Cured	Active	Death Rate (per 100)	Cure Rate (per 100)
Maharashtra	5433506.000000	83777.000000	4927480.000000	422249.000000	1.540000	90.69000
Karnataka	2272374.000000	22838.000000	1674487.000000	575049.000000	1.010000	73.69000
Kerala	2200706.000000	6612.000000	1846105.000000	347989.000000	0.300000	83.89000
Tamil Nadu	1664350.000000	18369.000000	1403052.000000	242929.000000	1.100000	84.30000
Uttar Pradesh	1637663.000000	18072.000000	1483249.000000	136342.000000	1.100000	90.57000
Andhra Pradesh	1475372.000000	9580.000000	1254291.000000	211501.000000	0.650000	85.02000
Delhi	1402873.000000	22111.000000	1329899.000000	50863.000000	1.580000	94.80000
West Bengal	1171861.000000	13576.000000	1026492.000000	131793.000000	1.160000	87.60000
Chhattisgarh	925531.000000	12036.000000	823113.000000	90382.000000	1.300000	88.93000
Rajasthan	879664.000000	7080.000000	713129.000000	159455.000000	0.800000	81.07000
Gujarat	766201.000000	9269.000000	660489.000000	96443.000000	1.210000	86.20000
Madhya Pradesh	742718.000000	7139.000000	652612.000000	82967.000000	0.960000	87.87000
Haryana	709689.000000	6923.000000	626852.000000	75914.000000	0.980000	88.33000
Bihar	664115.000000	4039.000000	595377.000000	64699.000000	0.610000	89.65000
Odisha	633302.000000	2357.000000	536595.000000	94350.000000	0.370000	84.73000
Telangana	536766.000000	3012.000000	485644.000000	48110.000000	0.560000	90.48000
Punjab	511652.000000	12317.000000	427058.000000	72277.000000	2.410000	83.47000

To Visualize Covid19 Age Group Distribution in India

A bar chart is plotted to show how the cases is distributed in India on the age groups
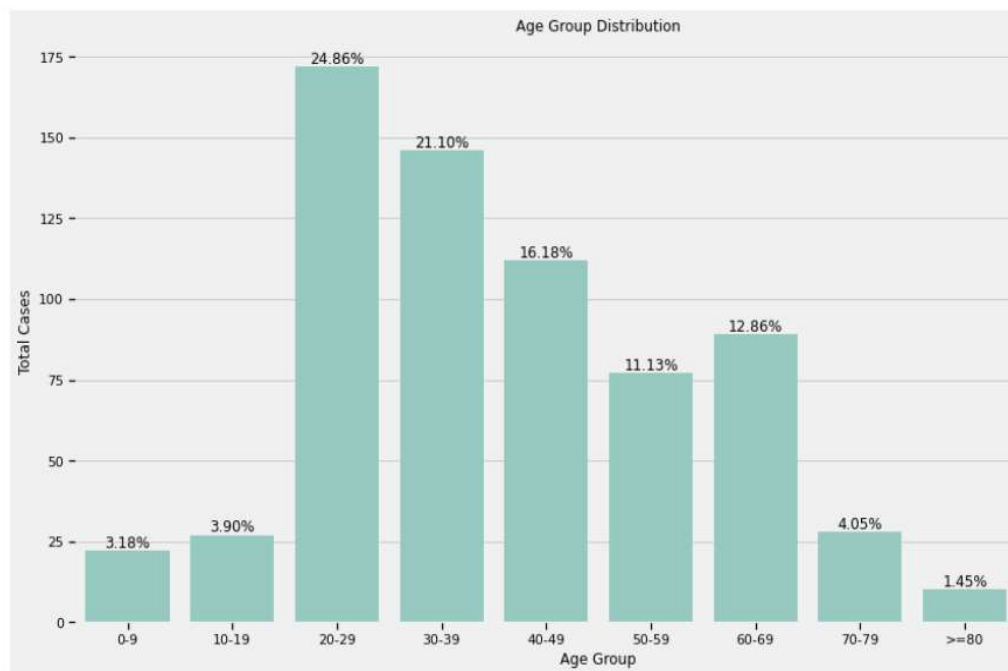
**To Visualize Covid19 Symptoms as per WHO website**
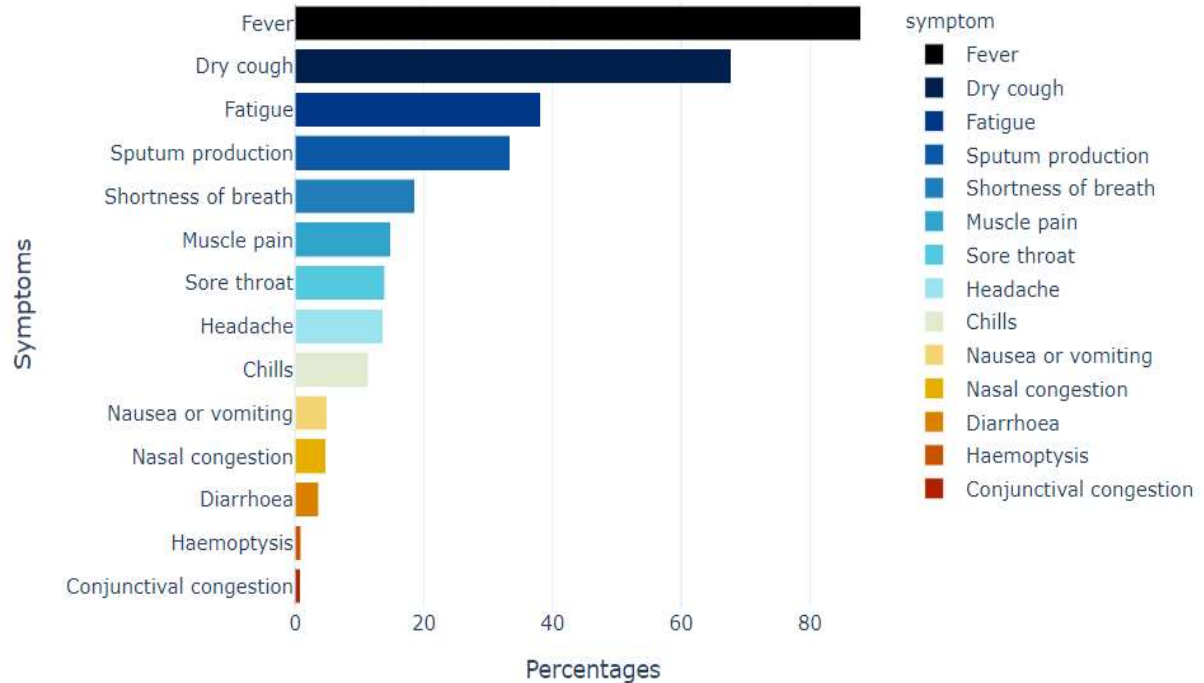
A horizontal bar chart displaying the common symptoms as per WHO organization

Vaccination situation on a Global Level

For visualization of countries with the most percentage of population vaccinated we have used bar plot as one is categorical attribute (countries) and the other is quantitative (% of population).
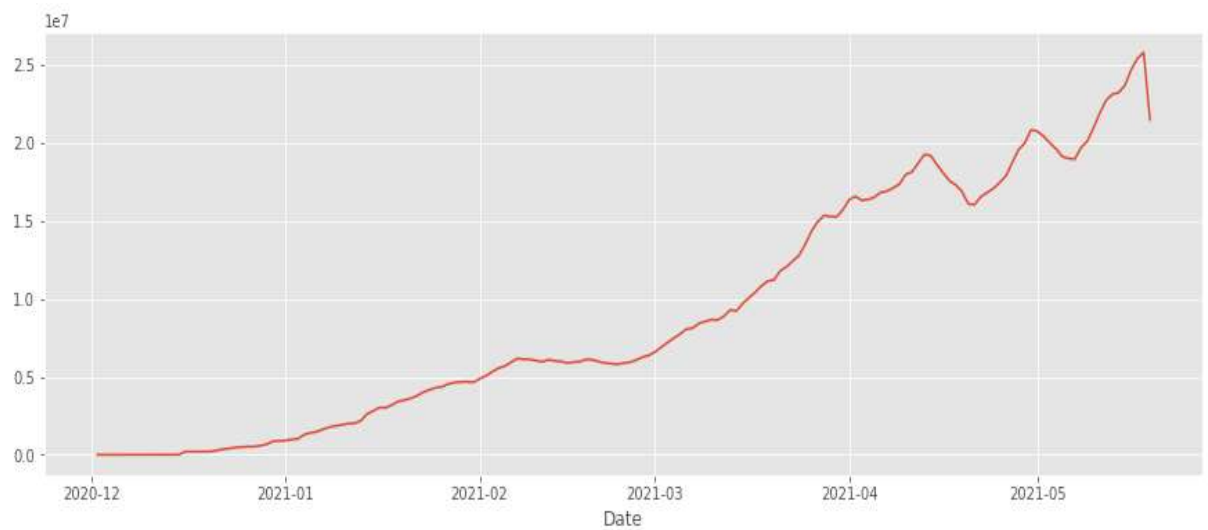
Top Countries by % People Fully vaccinated
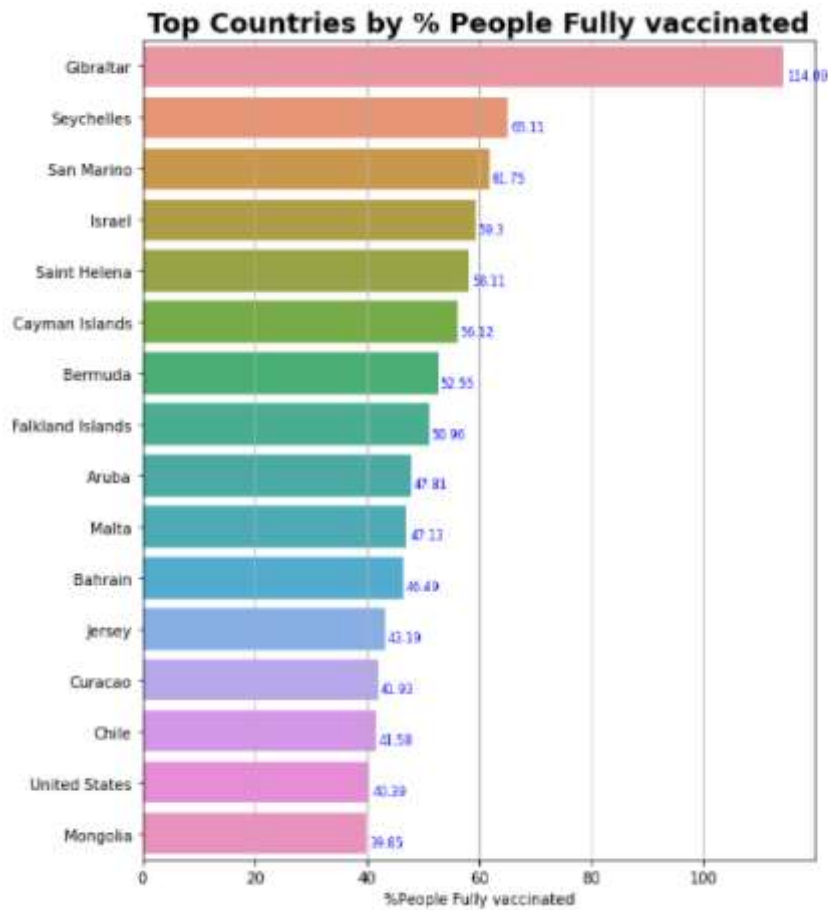
Second visualization consists of comparison of vaccines around the globe for that we have used barplot using seaborn.
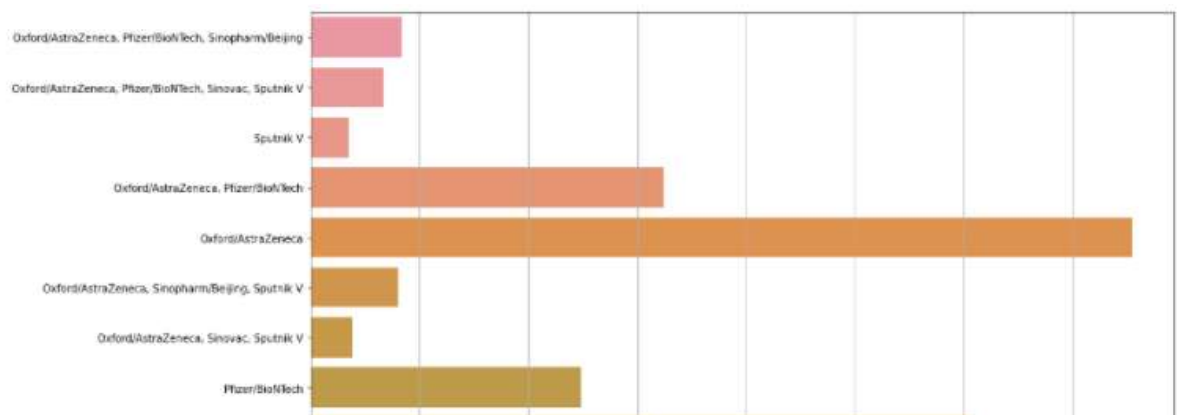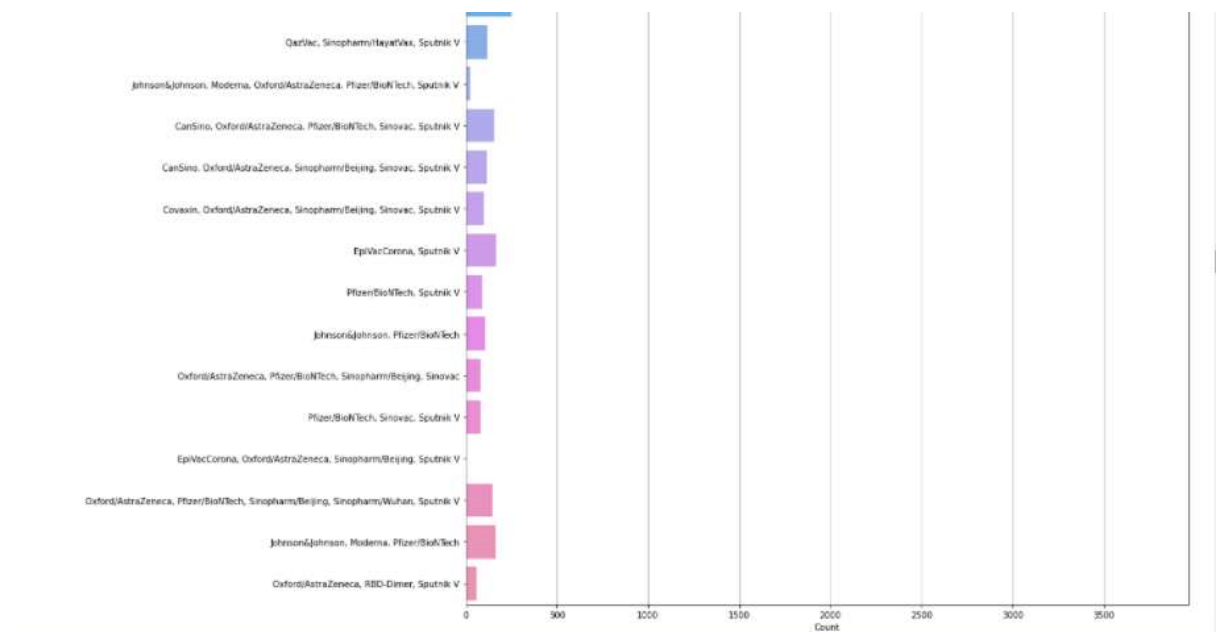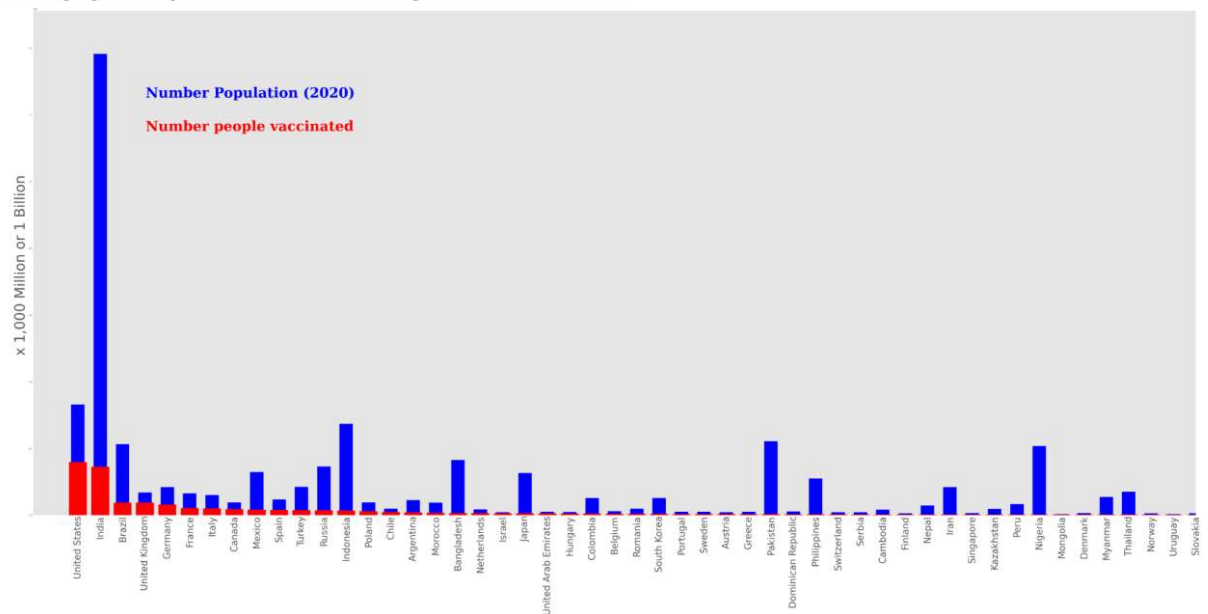
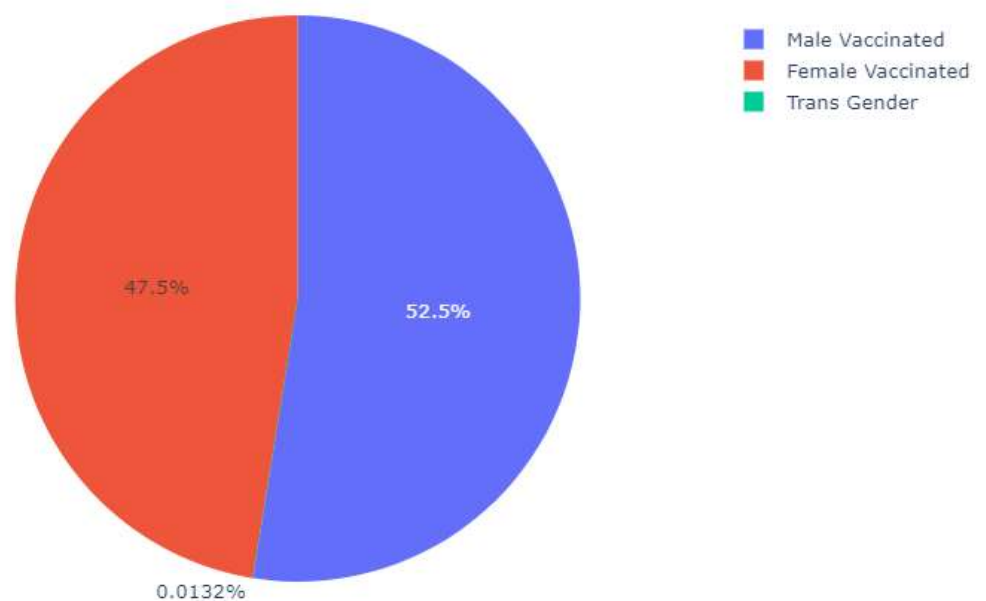To visualize the number of daily vaccinations around the globe we have used a line chart.

To visualize people vaccinated and the total population of each country we have used a stacked bar chart as there are two categorical attributes and one quantitative attribute.

Number people of Fully vaccinated and Number of Population of each Countries
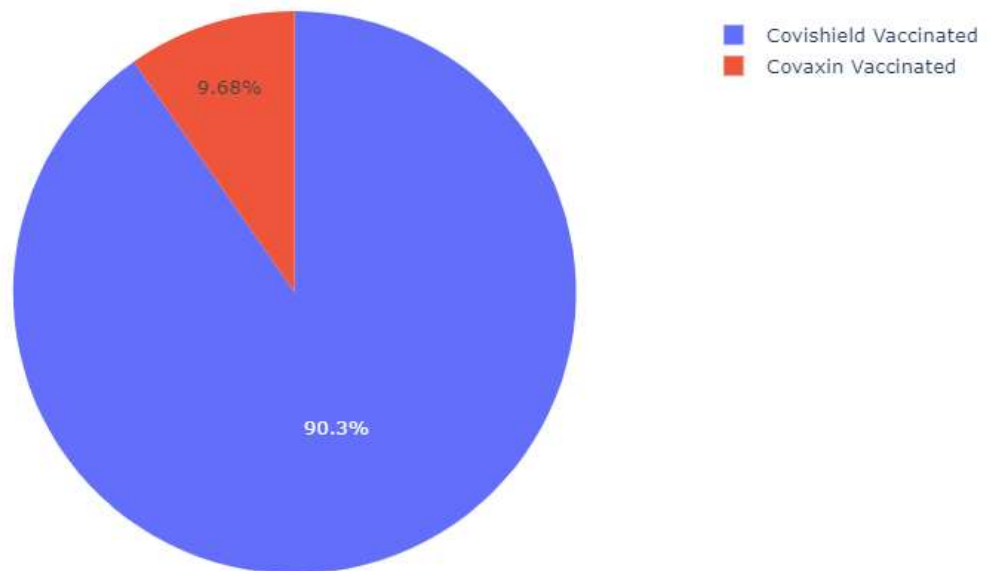
To visualize male, female and transgender ratio of vaccinated people in India we have used pie chart as a visual idiom.

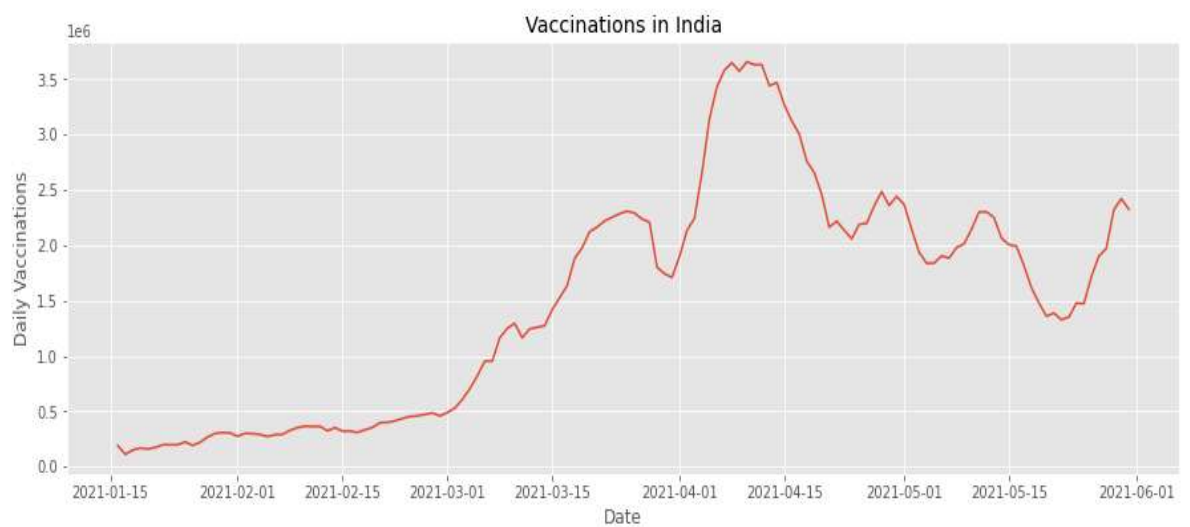Male and Female Vaccinated ratio for Covid19

To visualize the percentage of types of vaccines used in India we have pie chart.
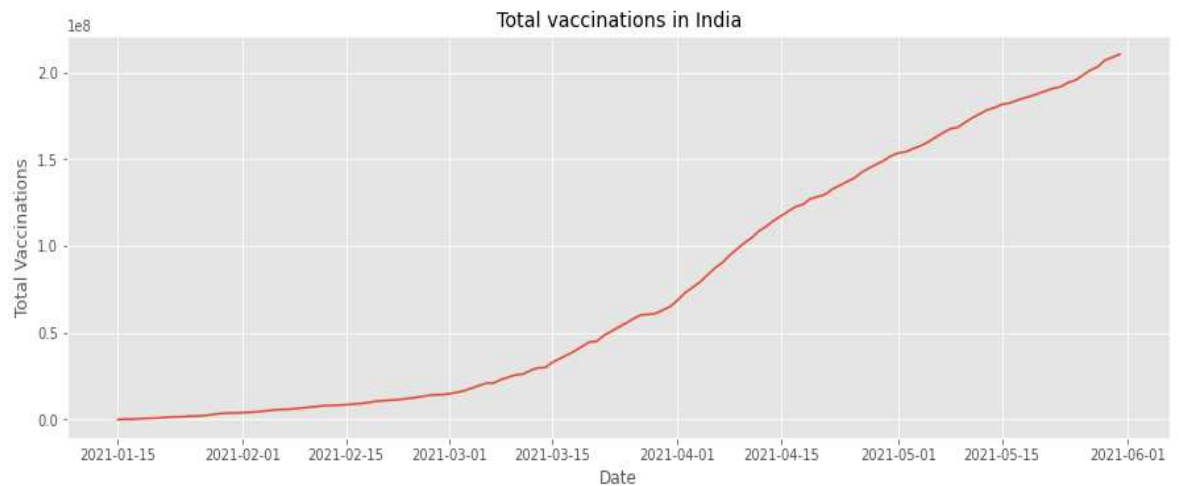
Covaxin and Covishield Vaccination

For the visualization of daily vaccination taking place in India we have used a Line Chart.
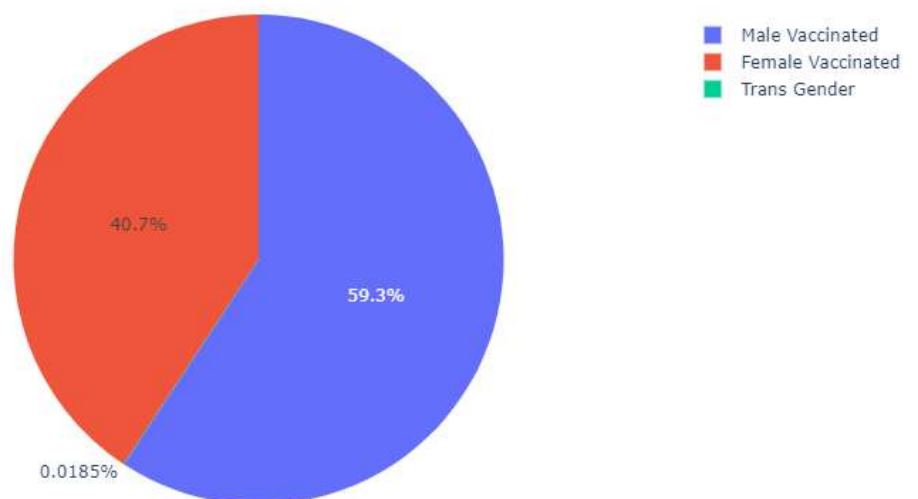
For the visualization of total vaccination taking place in India we have used a Line Chart.

We have also taken two major states which are Madhya Pradesh and Delhi for further visualization.
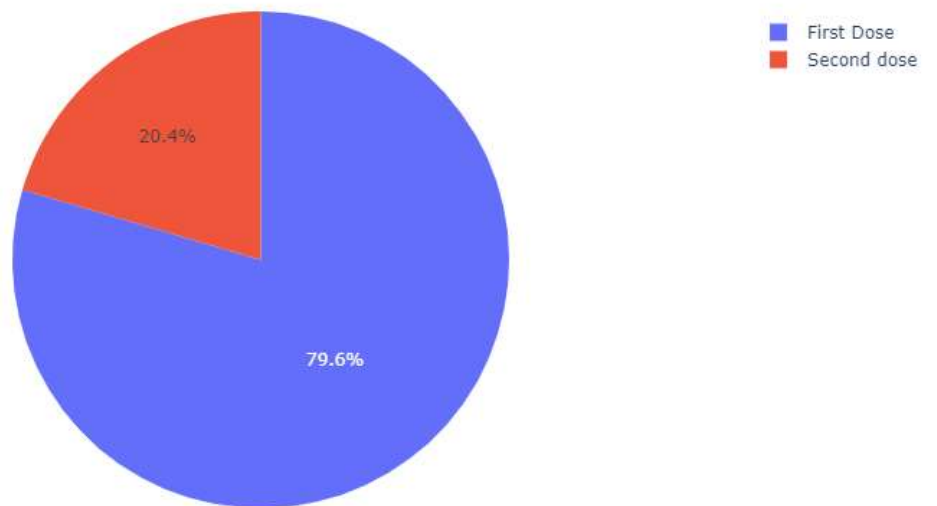
To visualize male, female and transgender ratio of vaccinated people in Delhi & Madhya Pradesh we have used pie chart as a visual idiom.
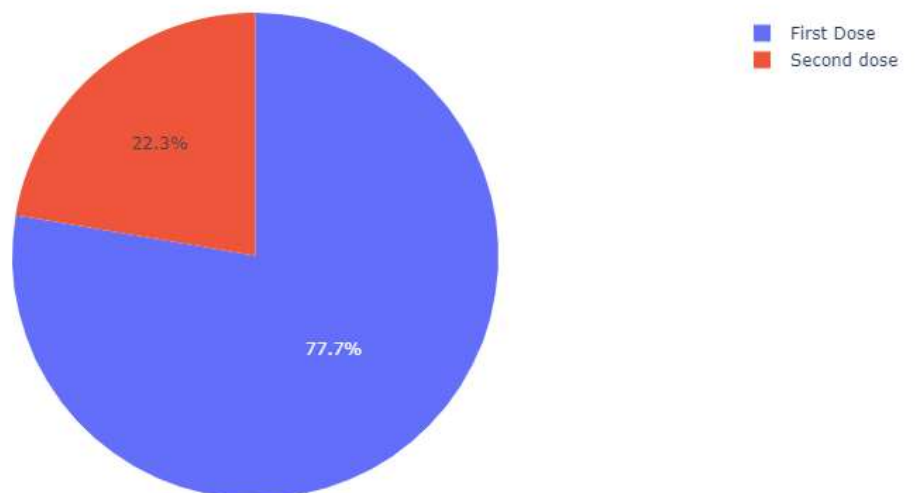
To visualize the First and Second ration for covid in Delhi and Andhra Pradesh we have used Pie Chart as a visual idiom.

First and Second dosage ratio for covid in Delhi

First and Second dosage ratio for covid in andhra pradesh

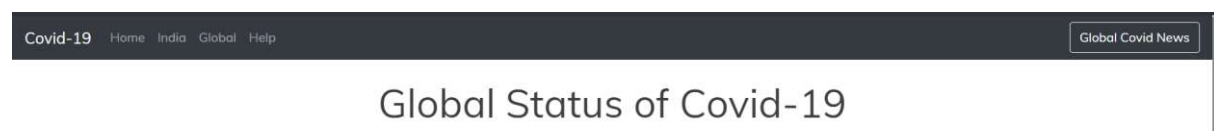## Web Portal for COVID19 Information on National and Inter-National Level

The web-application uses an open API's to fetch data from reliable sources and uses the data collected in JSON format to visualise various idioms. The benefit of using API's is that the values keep on updating on regular intervals and there is no need to hard code the data. The web application is constructed by using NodeJs for backend and collecting the data from API. It is responsible for handling different routes. The front end of the web app is made using HTML and CSS. Javascript is used to add interactivity to the application and to plot the data which is sent by Nodejs. ChartJs library is used for the same. Lastly the application is deployed on Heroku a free web-based application hosting platform so that it can be accessed from anywhere.

Link to the deployed application:

http://covid-trackerapi.herokuapp.com/

Navbar to view different pages and a link to google news to get the global covid situation updates
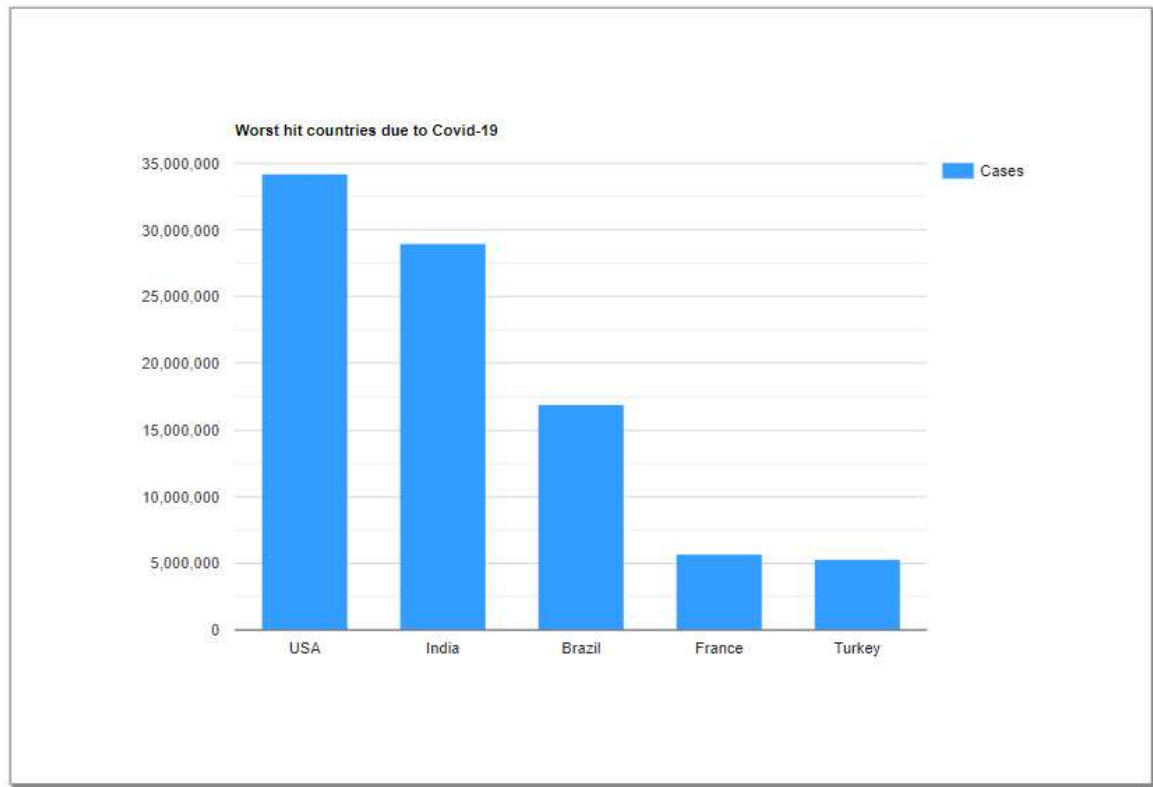
Navbar Screenshot

Dashboard Screenshots

Global Status of Covid-19

Confirmed	Recovered	Deceased
173053278	111118329	3726347

Global Geo Graph

Worst affected nations

Worst hit countries due to Covid-19

Stay Safe & Stay Home | Created by Navneet

Indian Page Screenshot

The recent updates are collected from https://www.covid19india.org/ open API which is free for public use

India on Covid-19

Confirmed	Recovered	Deceased
28975431	27285331	350657

Recent Updates

1279 new cases, 3309 recoveries and 26 deaths in Chhattisgarh

977 new cases, 4178 recoveries and 16 deaths in Jammu and Kashmir

11958 new cases, 27299 recoveries and 340 deaths in Karnataka 13 new cases and 86 recoveries in Madhya Pradesh

4872 new cases, 13702 recoveries and 86 deaths in Andhra Pradesh 1 new case and 4 recoveries in Dadra and Nagar

India Geocoded Graph

0 ▢▢▢▢▢▢ 5,831,781

Daily new cases since March 1st

Daily New Cases from 1st March

— new cases
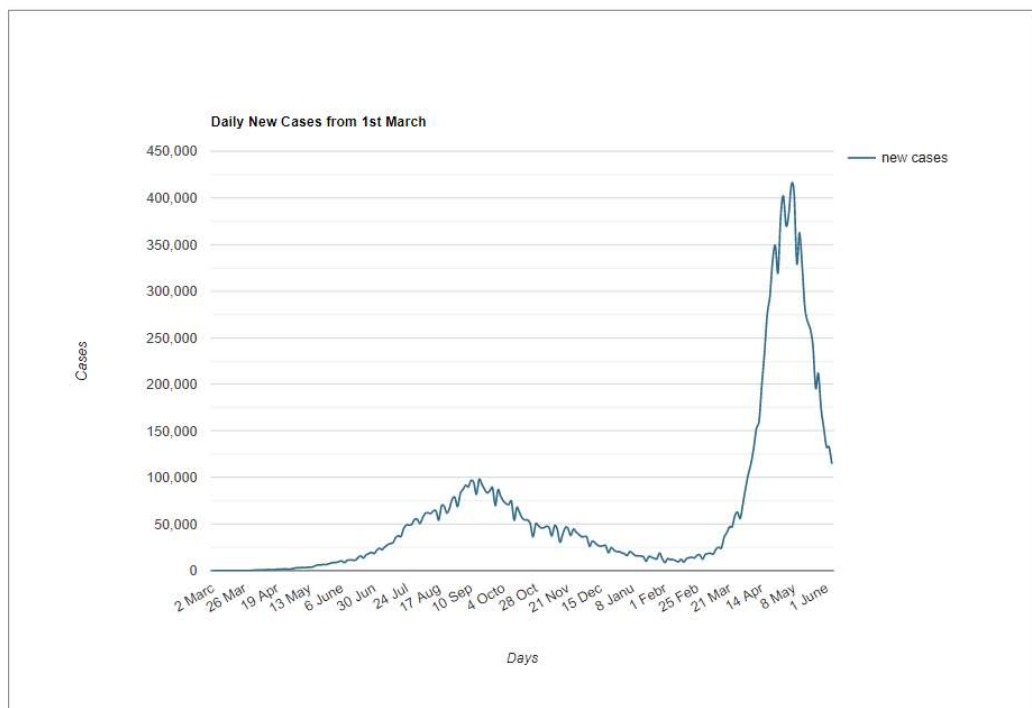
Global Statistics Page

Consist of a scrollable table which shows the information for all the countries and updated in real time from the open Covid API

## Global Statistics

#	Country	New Confirmed	Confirmed	Recovered	Deaths
84	Italy	2275	4232428	3913633	126523
85	Jamaica	0	48810	25936	960
86	Japan	2027	762980	703369	13531
87	Jordan	0	739847	721016	9530
88	Kazakhstan	0	450868	418803	3473
89	Kenya	0	172491	117595	3287
90	Kiribati	0	2	0	0
91	Korea (South)	0	144637	134861	1974
92	Kuwait	0	317197	301137	1795
93	Kyrgyzstan	0	106973	100447	1847

Global Cases Time Series

Total world to date

covid-19 powered by ⊓ DataHub

🏠 Stay Safe & Stay Home | Created by Navneet
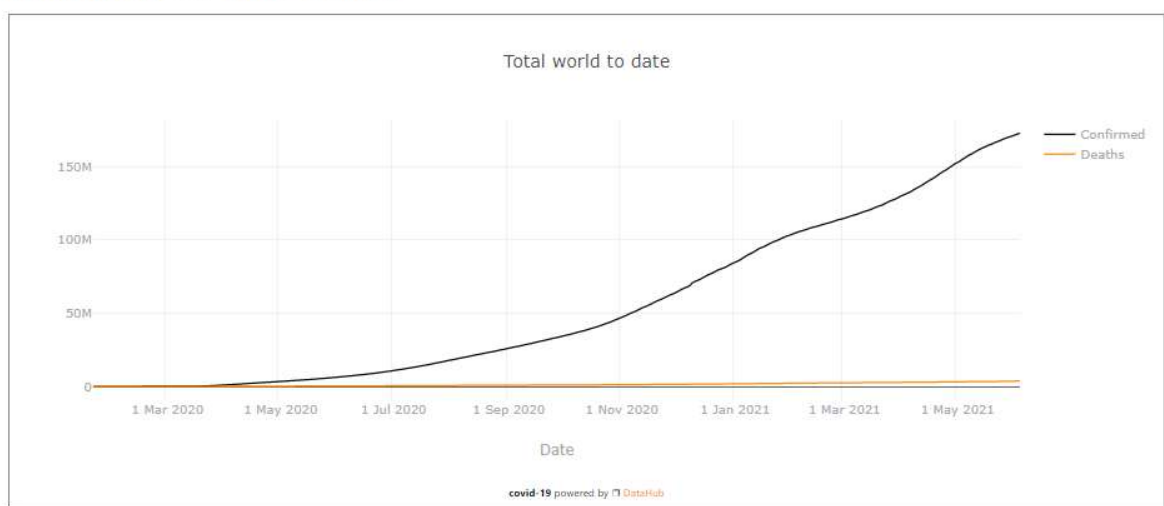
Error Page

The web-portal consist of the error page which is displayed if the user tries to access the pages not present in the app or enters the wrong URL. It navigates the user to the dashboard hence all URL navigation has been taken care of.
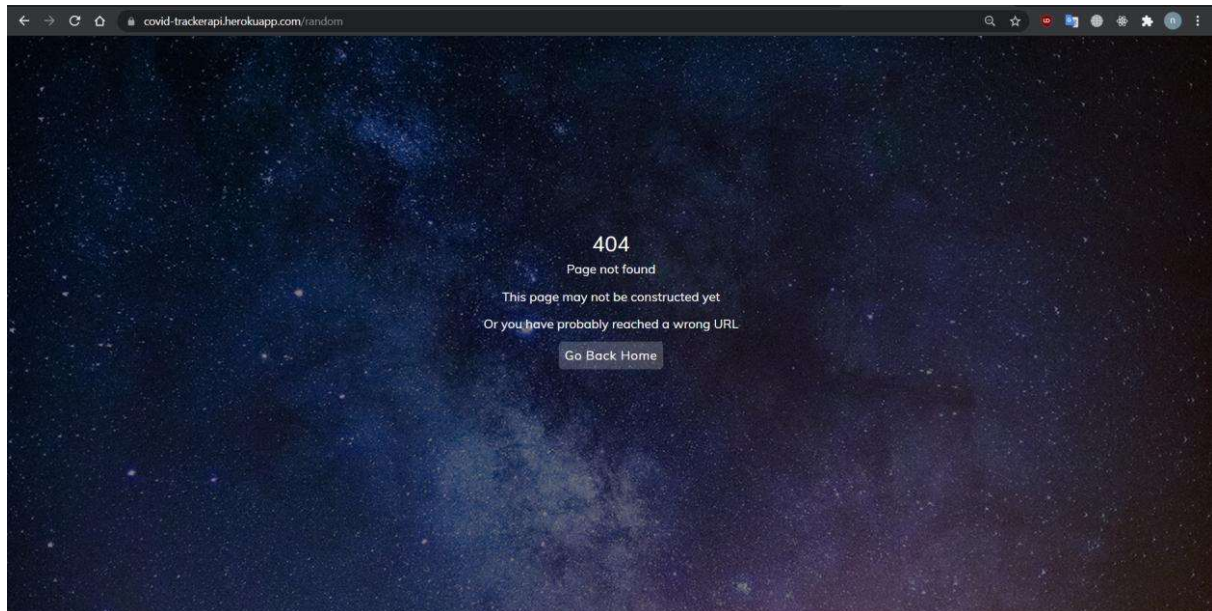
The web app portal can be accessed at: https://covid-trackerapi.herokuapp.com/

Validation Comparison

We have selected all the visual idioms according to the rules by considering all the attribute types (quantitative, categorical, and sequential), dataset type, Multidimensional or single dimensional table, dataset with keys or no keys.

- With no keys, and data having two quantitative attributes we have used Scatterplot.

- With one key, one value and one categorical, one quantitative attribute we have used a Bar Plot.

- With one key, one value and two categorical, one quantitative attributes we have used a Stacked Bar Plot.

- With one key, one value and two quantitative attributes we have used Line Chart as a visual idiom.

- With two keys, one value and two categorical, one quantitative attributes we have used Heat map as a visual idiom.

- With two quantitative attributes present and data arranged in hierarchical order we have used Tree Map.

- With two quantitative attributes and when data is present in the form of frequency we have used histogram.

Conclusion

The project covers all the user requirement stated at the beginning of the work. All the visual idioms used are justified and is an optimal solution to the user requirement. The topics covered for the visualization include global covid cases, Indian covid situation based on cases , testing and vaccination. The symptoms listed out by the WHO official have been plotted based on the frequency of occurrence among individuals. The project is integrated with a full-fledged working web application which is deployed (Link provided above) on Heroku host. The web application collects the data in real time by use of open API's which keeps the user updated about the situation. Lastly all the screenshots of the visual Idioms plotted along with the dataset used have been included for better understanding of the reader.