

# Contents

<b>1</b>	<b>Acknowledgement</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
<b>3</b>	<b>Objective</b>	<b>3</b>
<b>4</b>	<b>Methodology</b>	<b>4</b>
4.1	Exploratory Data Analysis . . . . .	4
4.2	Feature Engineering . . . . .	4
4.3	K means clustering . . . . .	4
4.4	Formula used : . . . . .	5
<b>5</b>	<b>Results</b>	<b>6</b>
	<b>References</b>	<b>13</b>

# 1 Acknowledgement

We would like to thank Dr. Sarita Azad Maa'm for her support and guidance in completing our project on the topic (Covid-19 Vaccination Cases and Mobility Analysis). It was a great learning experience. I would like to take this opportunity to express my gratitude to all of my group members Mohit Kumar Jain, Himanshu, Pawan. This project would not have been successful without their cooperation.

## 2 Introduction

The new coronavirus, which was first identified in Wuhan, China, in December 2019, is currently inflicting an unparalleled catastrophe all over the world. The World Health Organization (WHO) describes the coronavirus family as include the common cold, Middle East respiratory illness coronavirus, and severe acute respiratory syndrome coronavirus. Coronaviruses can be spread from some wild animals to people since they are present in some of them. These viruses can cause respiratory issues in addition to the standard symptoms of a cold and fever. There is currently no known treatment method for coronaviruses. One can reduce their risk of infection by maintaining proper personal hygiene and avoiding close contact with others who are ill.

The coronavirus illness was classified as a global pandemic by the WHO on March 11, 2020. Over 3 million confirmed cases of the illness have been reported globally, in 212 nations and territories. On January 30, 2020, a student who had just returned from Wuhan was found to have the sickness in Kerala, India. As of May 3, 2020, there have been more than 37,000 confirmed cases of infection throughout India. This displays the COVID-19 instances in India's daily increase. There were no additional verified covid cases for around a month after the initial three cases from January 30 to February 3, 2020. Beginning on March 2, 2020, more covid cases started to emerge. These cases are related to people who have been evacuated or have arrived from covid-affected countries. From March 20, 2020, onwards, there is an exponential growth in the daily number of covid cases at the pan-India level.

COVID-19 goes through three stages, depending on how the virus spreads. A country or region experiences imported infected cases with a history of travel from virus-affected countries during the initial stage. In the second stage, individuals who have no prior travel history but have come into touch with those categorised in Stage 1 spread new illnesses throughout a nation or region. Community transmission is Stage 2, during which a person who has never been in contact with an infected person or anyone who has travelled to a country where the virus has been present becomes newly infected. At Stage 3, the virus propagation is essentially unstoppable, and the nation may experience numerous significant infection clusters. The covid pandemic has led to a dramatic loss of human life worldwide and presented an unexpected challenge to public health. As of now, vaccination is the only means to slow down the spread of the virus in the communit.

### 3 Objective

Firstly, we are collecting the data from different sources which are available on websites like kaggle. In which we are taking statewise data of covid cases, and after that we are removing the null values from the data. Also, drop the unimportant features of the data. We are finding the positive rate for all states and UT. Then, we are plotting the positive rate on india map to indicates which state has greater postive rate and which state has lesser positive rate and also categorised into different categories using k-means clustering. This postive rate is also shown in bar plot.

Secondly, we taken the data of covid -19 vaccinations state wise in which we have a data of Total doses administered, First Dose administered, Second Dose Administered, Male (Doses Administered), Female (Doses Administered), 18-44 Years (Doses Administered), 45-60 Years (Doses Administered), 60+ Years (Doses Administered). After that we are handling the null data and plotting the first and second doses taking date as x-axis and Total dose as y-axis. We are also, plotting Male and Female dose date wise. We are taking the total doses state wise and finding the vaccine percentage wise for all states and UTs. And Plotting the bar plot then, we are using k-means clustering for categorised the data (percentage wise) into different categories by forming a clusters on india map. Similarly, we are following the above procedure for 18-44 Years (individual vaccinated), Male (individual vaccinated), Female(individual vaccinated).

Now, we are taking the data of mobility which is transport mobility, pharma mobility and plotting it on the india map using clustering and check the relationship between mobility and vaccinations.

## 4 Methodology

### 4.1 Exploratory Data Analysis

Exploratory data analysis is a fundamental early step after data collection and pre-processing, where the data is simply visualized, plotted, manipulated, without any assumptions, in order to help assessing the quality of the data and building models. “Most Exploratory data analysis techniques are graphical in nature with a few quantitative techniques. The reason for the heavy reliance on graphics is that by its very nature the main role of Exploratory data analysis is to explore, and graphics gives the analysts unparalleled power to do so, while being ready to gain insight into the data. There are many ways to categorize the many Exploratory data analysis techniques”.

Exploratory data analysis can be summarized as follows:

- (1) Maximize insight into the database/understand the database structure.
- (2) Detect outliers and anomalies (values that are significantly different from the other observations).
- (3) Extract and create clinically relevant variables.

### 4.2 Feature Engineering

Feature engineering is a machine learning technique that leverages data to create new variables that are not in the training set. It can produce new features for both supervised and unsupervised learning, with the goal of simplifying and speeding up data transformations while also enhancing model accuracy. Feature engineering is required when working with machine learning models. Regardless of the data or architecture, a terrible feature will have a direct impact on your model.

Feature Engineering is a very important step in machine learning. Feature engineering refers to the process of designing artificial features into an algorithm. These artificial features are then used by that algorithm in order to improve its performance, or in other words reap better results.

### 4.3 K means clustering

K-Means Clustering is an Unsupervised Learning algorithm, which groups the unlabeled dataset into different clusters. Here K defines the number of pre-defined clusters that need to be created in the process

It allows us to cluster the data into different groups and a convenient way to discover the categories of groups in the unlabeled dataset on its own without the need for any training. It is a centroid-based algorithm, where each cluster is associated with a centroid. The main aim of this algorithm is to minimize the sum of distances between the data point and their corresponding clusters.

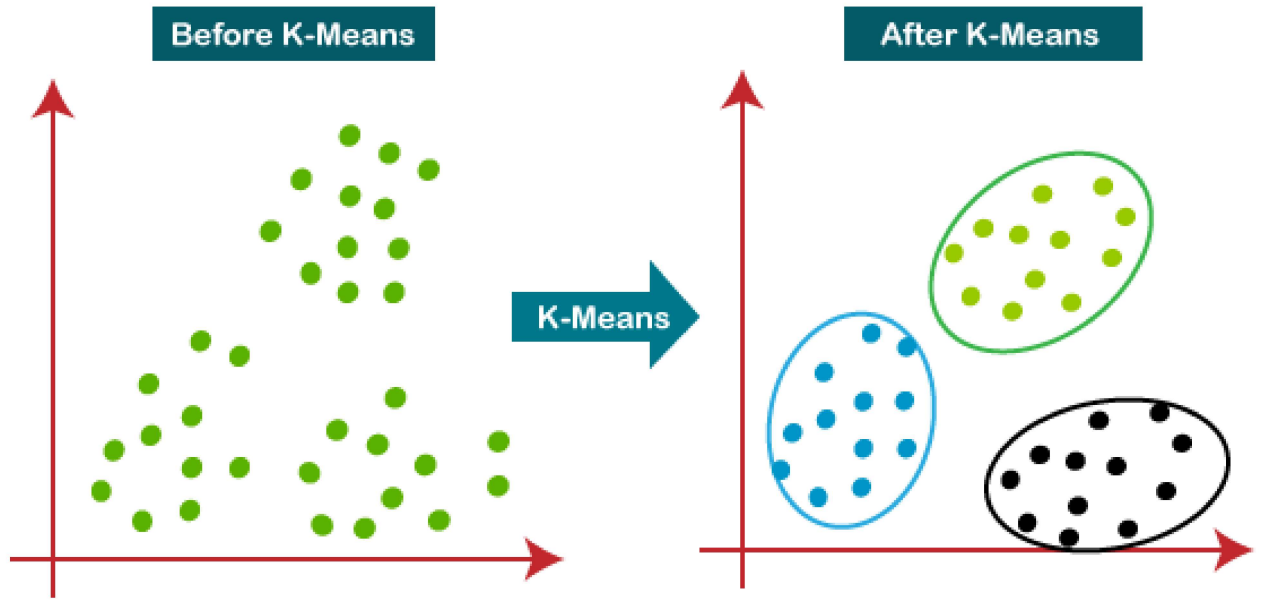


Figure 1: K-means Clustering algorithm

#### 4.4 Formula used :

To finding the positive rate we are using following formula:

$$\text{Positive Rate} = \frac{\text{Number of Positive Case}}{\text{Total number of samples}}$$

To convert the number of vaccinations into percentage wise we are using following formula:

$$\text{Percentage wise vaccinations} = \frac{\text{Total number of vaccinations administered}}{\text{Population of the state}}$$

## 5 Results

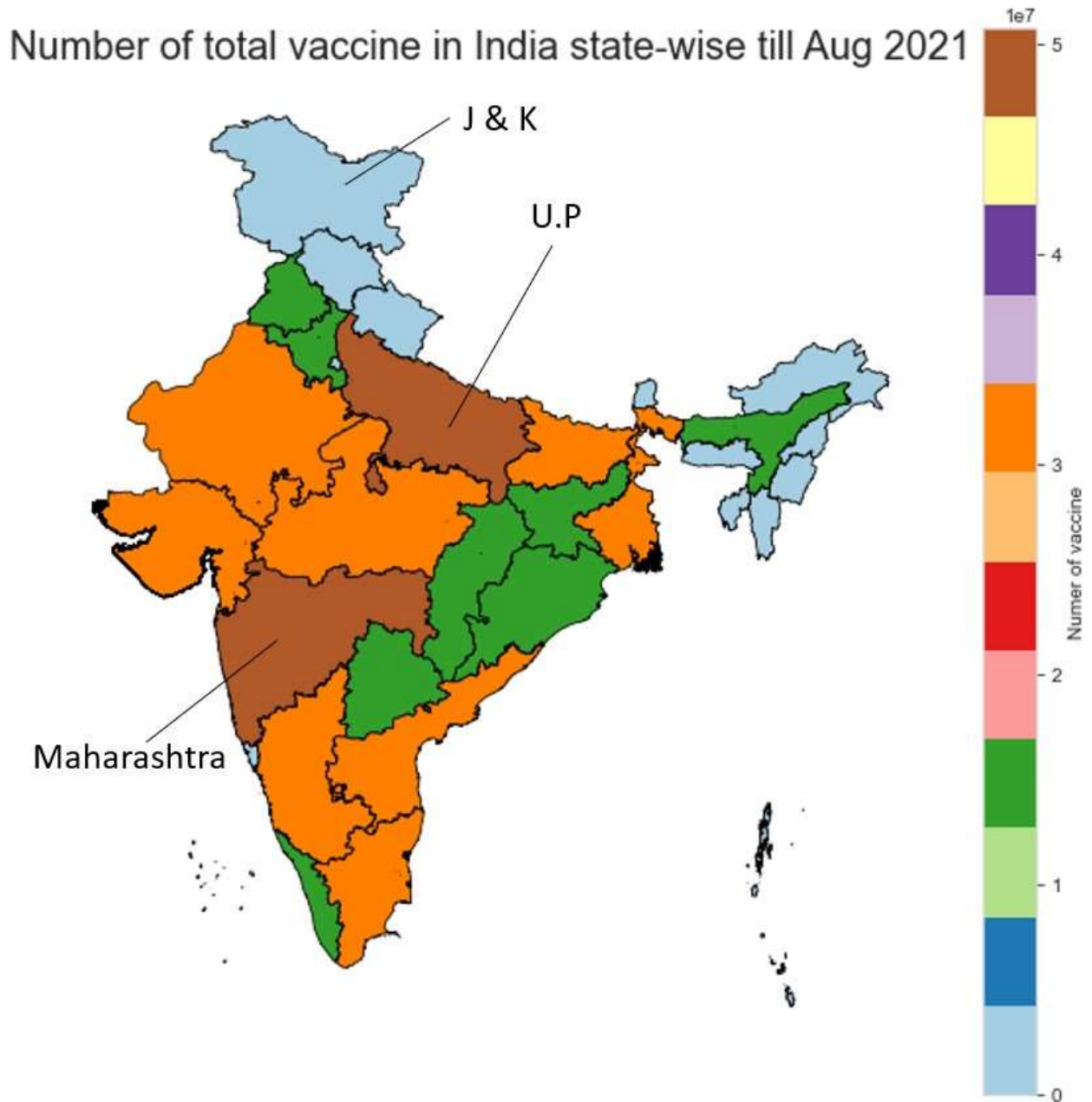


Figure 2: Number of total vaccinations in India till Aug-2021

This plot shows the total no. of vaccinations in India till Aug 2021 this shows UP and Maharashtra are the state which does highest vaccinations in counting and J K is the state which are in the list of low vaccinations states in counting.

## Percentage vaccine India state-wise till Aug 2021

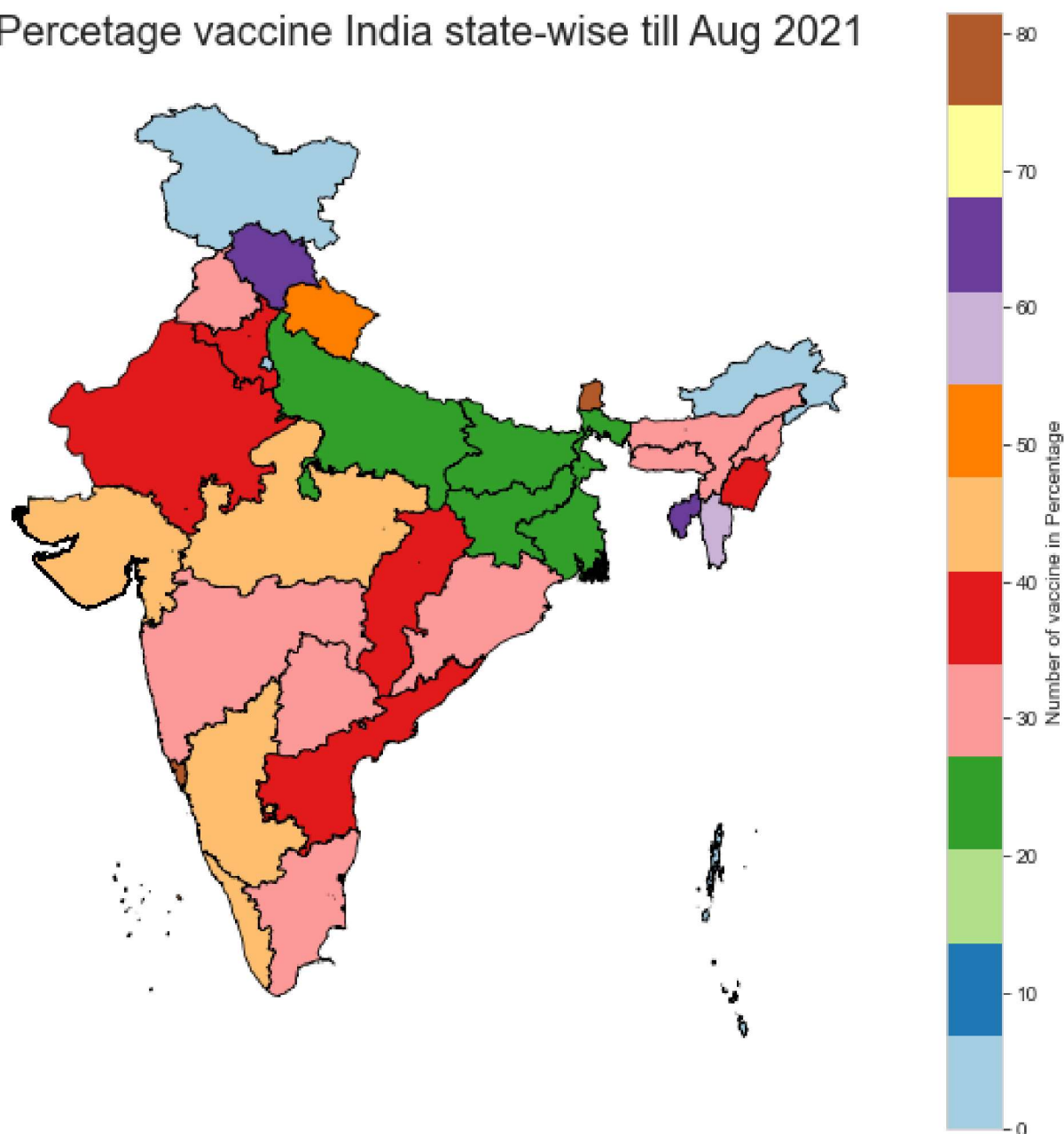


Figure 3: Number of vaccinations percentagewise in India till Aug-2021

This Plot shows the No. of vaccinations population wise in states is highest in Sikkim that is 81.53 % and after that Goa (75.03 %) does highest vaccinations. And, Bihar (22 %) and Uttar Pradesh (22 %) does lowest vaccinations population wise in states till Aug 2021.

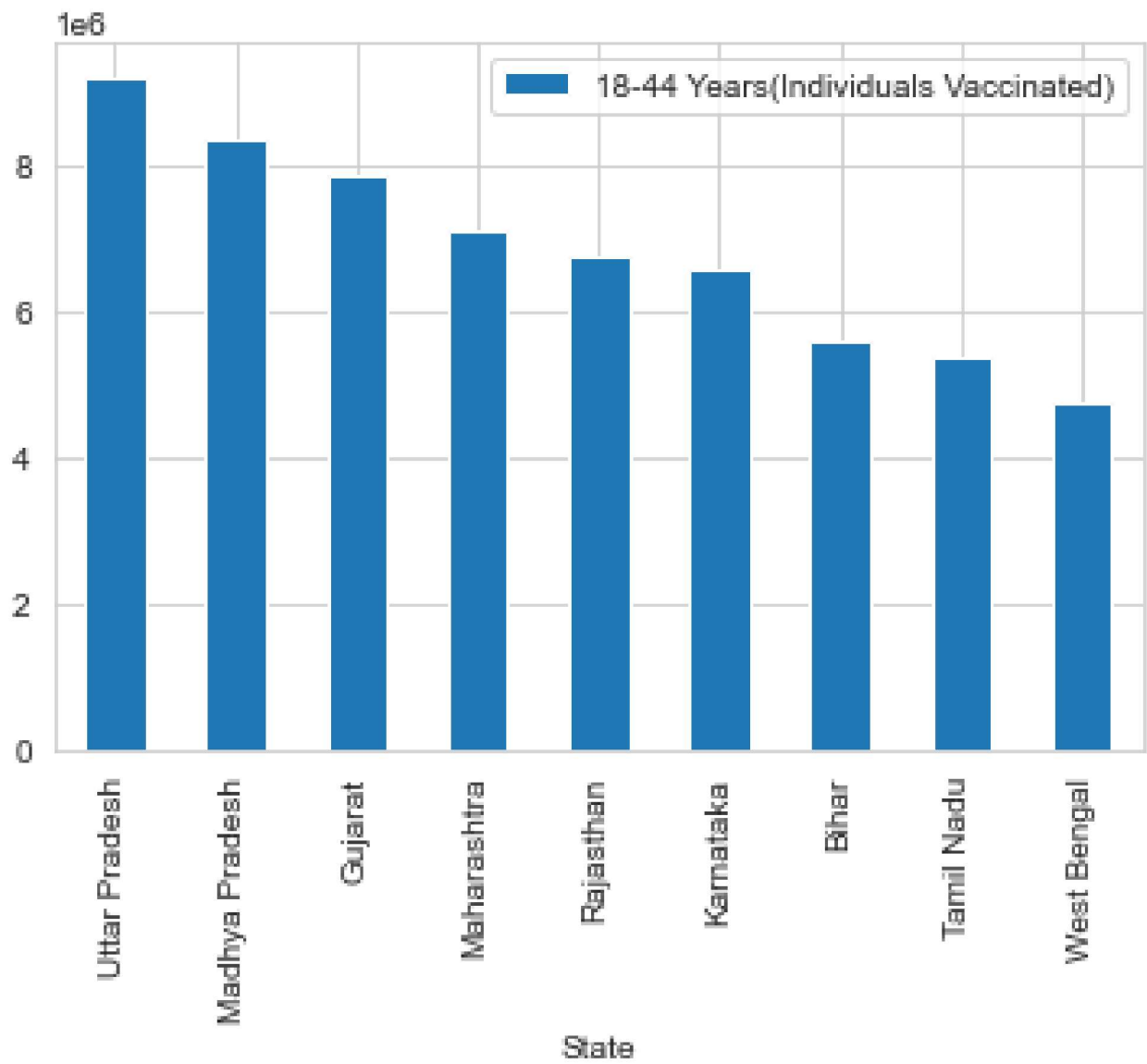


Figure 4: Age 18-44 vaccinations bar plot



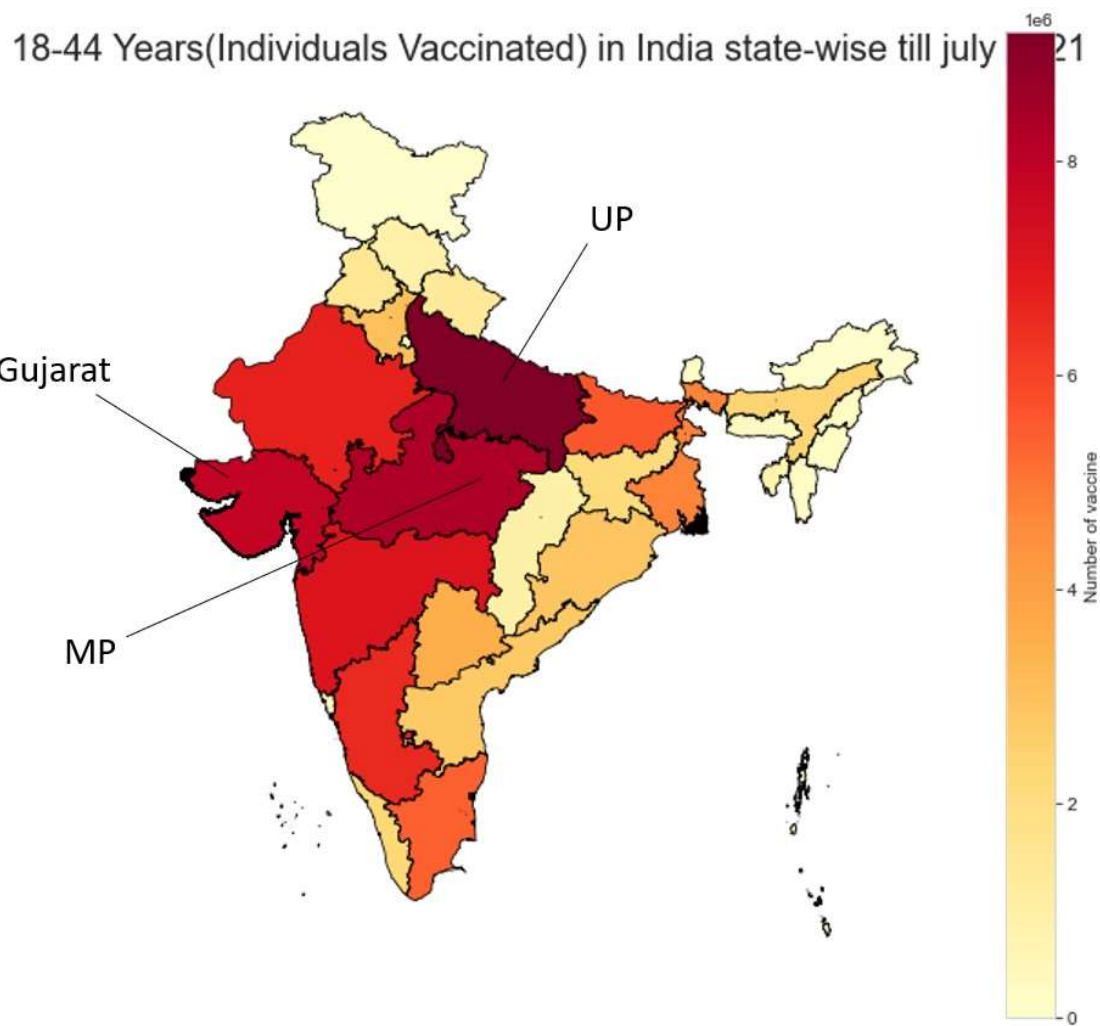


Figure 5: Age 18-44 vaccinations plot on India Map

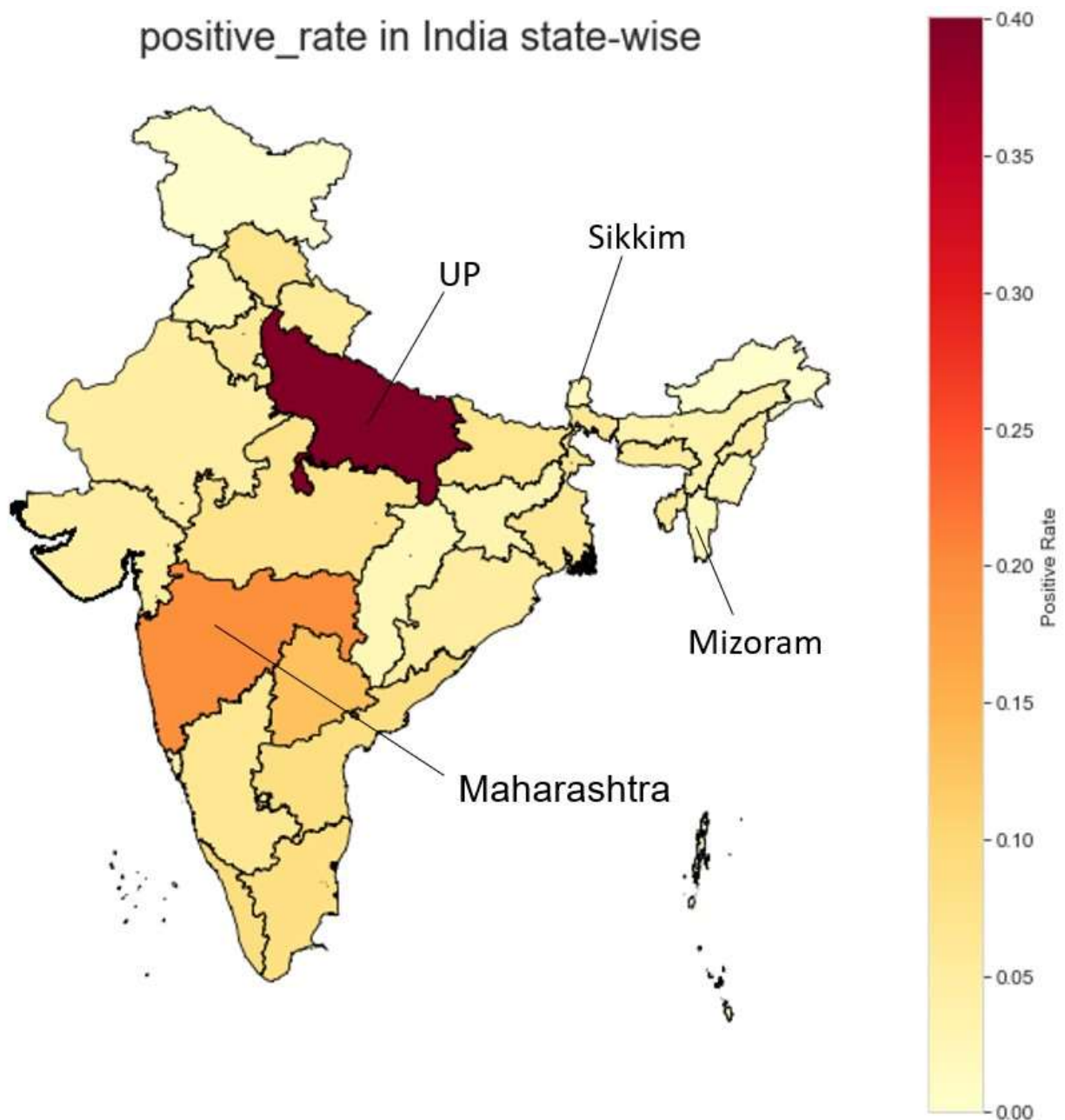


Figure 6: Positive rate on India Map

This Map shows that UP is the state which has the highest Positive rate state-wise i.e. 0.40 after that Maharashtra which has positive rate is 0.19. Mizoram (0.019) and Sikkim (0.018) is the states which has lowest positive rate.

## transport mobility in India state-wise Till Aug 2021

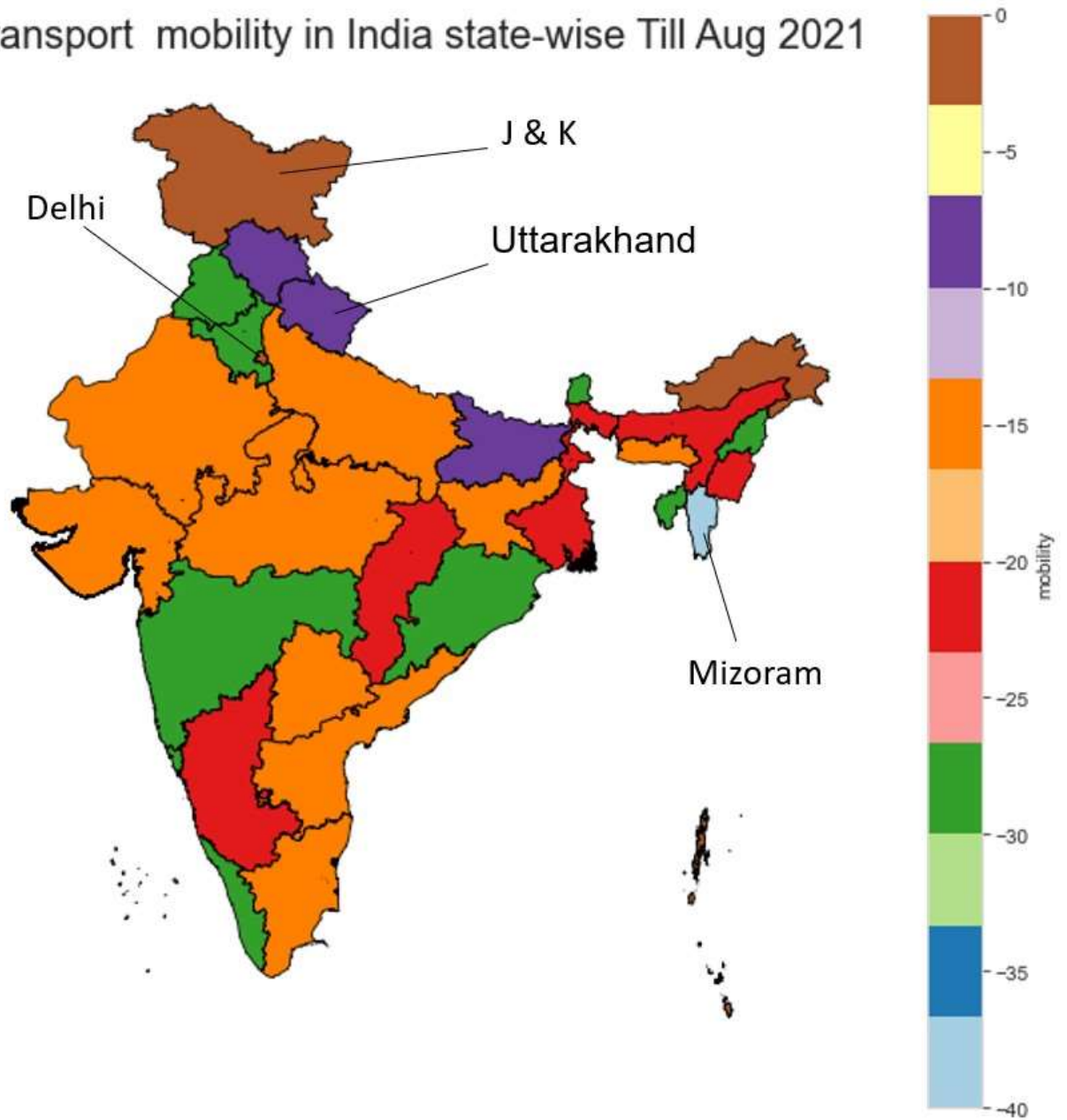


Figure 7: Transport mobility statewise in India till Aug-2021

This Map shows that Jammu and Kashmir has the highest Transport mobility after that Uttarakh-  
hand is the state which has highest Transport mobility state-wise.  
Mizoram and Delhi is the states which has lowest Transport mobility.

## pharma mobility in India state-wise Till Aug 2021

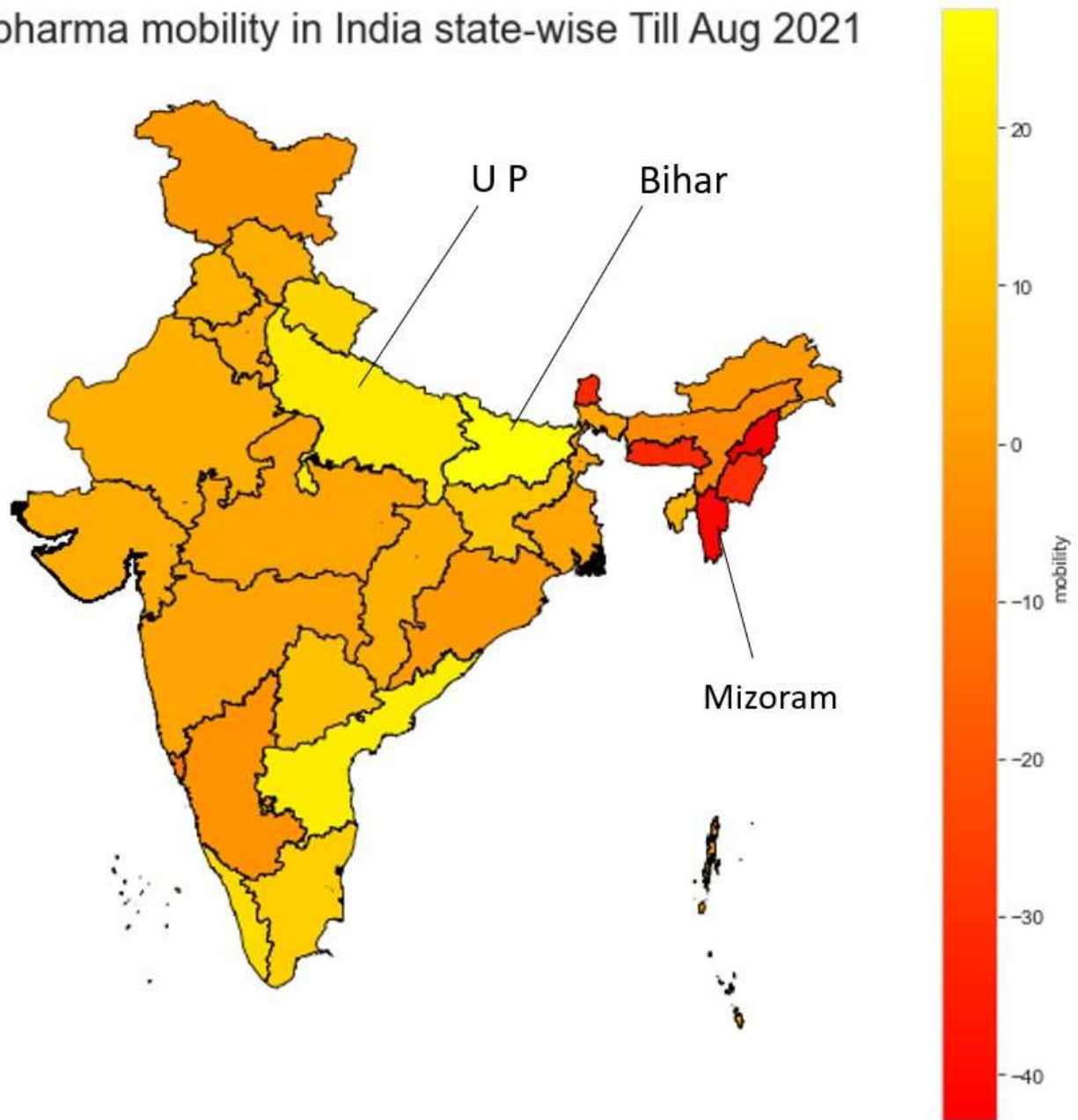


Figure 8: Pharma mobility statewise in India till Aug-2021

This Map shows that UP and Bihar states has the highest Pharma mobility . Mizoram, Nagaland and Sikkim states has lowest Pharma mobility