Data Science Project Report

(Project Semester August-December 2020)

Covid-19 Analysis in India

Submitted by
M. Lakshmi Narasa Reddy
Registration No. 11902757

Programme and Section: P132 & KM002

Course Code: INT217

Under the Guidance of

Maneet Kaur (UID:15709)

Discipline of CSE/IT

Lovely School of Technology and Science

Lovely Professional University, Phagwara



CERTIFICATE

This is to certify that **M. Lakshmi Narasa Reddy** bearing Registration no. **11902757** has completed **INT217** project titled, "Covid-19 Analysis in India" under my guidance and supervision. To the best of my knowledge, the present work is the result of his/her original development, effort, and study.

Signature and Name of the Supervisor

Designation of the Supervisor

School of Computer Science and Engineering

Lovely Professional University

Phagwara, Punjab.

Date:

DECLARATION

I, M. Lakshmi Narasa Reddy, student of P132 under CSE/IT Discipline at, Lovely Professional University, Punjab, hereby declare that all the information furnished in this project report is based on my own intensive work and is genuine.

Name of the student: M. Lakshmi Narasa

Date:

Signature

Registration No. 11902757

Reddy

TABLE OF CONTENTS

Sr. No	Торіс	Page No
1	Introduction	<u>5</u>
2	Objectives/Scope of the Analysis	<u>6</u>
3	Source of dataset	<u>8</u>
4	ETL Process	9
5	Joining Data Sets	<u>15</u>
6	Analysis on Data	<u>16</u>
7	Results	<u>26</u>
8	References	<u>32</u>
9	Bibliography	<u>33</u>

INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a contagious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The first known case was identified in Wuhan, China, in December 2019. The disease has since spread worldwide, leading to an ongoing pandemic. More than a year has passed since and the virus is still wreaking havoc in many nations including especially the US and India.

Symptoms of COVID-19 are variable, but often include fever, cough, headache, fatigue, breathing difficulties, and loss of smell and taste. Symptoms may begin one to fourteen days after exposure to the virus. At least a third of people who are infected do not develop noticeable symptoms. Of those people who develop symptoms noticeable enough to be classed as patients, most (81%) develop mild to moderate symptoms (up to mild pneumonia), while 14% develop severe symptoms (dyspnoea, hypoxia, or more than 50% lung involvement on imaging), and 5% suffer critical symptoms (respiratory failure, shock, or multiorgan dysfunction). Older people are at a higher risk of developing severe symptoms. Some people continue to experience a range of effects (long COVID) for months after recovery, and damage to organs has been observed. Multi-year studies are underway to further investigate the long-term effects of the disease

COVID-19 transmits when people breathe in air contaminated by droplets and small airborne particles containing the virus. The risk of breathing these in is highest when people are in close proximity, but they can be inhaled over longer distances, particularly indoors. Transmission can also occur if splashed or sprayed with contaminated fluids in the eyes, nose, or mouth, and, rarely, via contaminated surfaces. People remain contagious for up to 20 days and can spread the virus even if they do not develop symptoms.

Several testing methods have been developed to diagnose the disease. The standard diagnostic method is by detection of the virus' nucleic acid by real-time reverse transcription polymerase chain reaction (rprt-PCR), transcription-mediated amplification (TMA), or by reverse transcription loop-mediated isothermal amplification (RT-LAMP) from a nasopharyngeal swab.

Several COVID-19 vaccines have been approved and distributed in various countries, which have initiated mass vaccination campaigns. Other preventive measures include physical or social distancing, quarantining, ventilation of indoor spaces, covering coughs and sneezes, hand washing, and keeping unwashed hands away from the face. The use of face masks or coverings has been recommended in public settings to minimize the risk of transmissions. While work is underway to develop drugs that inhibit the virus, the primary treatment is symptomatic. Management involves the treatment of symptoms, supportive care, isolation, and experimental measures.

OBJECTIVES

The objectives of the analysis presented in this Report are as follows:

- In this study we cluster the states of India in terms of the spread of COVID-19 and related variables such as active cases, total Positives, Deaths etc and the Population of the state.
- Comparing positivity rates of different states
- Building a Search Board
- Dashboard of Covid 19 across India
- Visual Representation of Covid 19 across states
- Vaccination Tracker
- Visual representation of vaccination across states
- Population Analysis

SCOPE OF ANALYSIS

Data analytics is a process through which data is cleaned analysed and modelled using tools. This data is then used to derive insights. The insights are then used for decision-making purposes.

The world is becoming more tech-driven and fast-paced now, data analysis professional is already playing an increasingly crucial role in businesses. Both start-ups and tech giants in India are adding data analysts to their workforce who are able of recording, analysing and dissecting data to drive efficient decision-making.

CAREER PROSPECTS AND SCOPE OF DATA ANALYTICS:

It has been estimated that data is expected to grow up to 50 times by 2020. Companies must stay updated with the requirements of huge volumes of data so that they don't become obsolete. Therefore, experts who are well-versed in advanced analytics are crucial for organizations to adjust their business models and stay ahead of the competition. To master Data analytics, one can join data analytics training course.

The scope of data analytics In India includes companies in policing, banking, healthcare, fraud detection, e-commerce, energy, telecommunications, and risk management.

JPMorgan, Accenture, Microsoft, Adobe, Flipkart, AIG, Ernst & Young, Wipro, Vodafone & Deloitte are the organizations that had the greatest number of openings for data analysts last year.

Thus, if you have the required skillset and are ready to keep yourself updated, your career as a Data Analyst is expected to keep growing onwards and upwards. This line stands true especially when we consider that a data analyst salary in India is directly or indirectly dependent on how upskilled and updated, they are.

The scope of the project contains Dashboards and visual charts that provide us an insight to the situation of ongoing covid 19 in India across all states. The comparison between the states provides us the idea of action that respective states took against covid 19.Population analysis provides us how the heavily populated states are performing in this time of pandemic.

SOURCE OF DATASET

• https://www.covid19india.org/

It is an open-source website where anyone can get the data and use it for the educational purposes.

I have downloaded my covid 19 data from this website.

• https://www.mohfw.gov.in/

It is managed by ministry of MINISTRY OF HEALTH AND FAMILY WELFARE.

I have downloaded my Vaccination data from this website.

• https://en.wikipedia.org/wiki/List of states and union territories of India by population

It is the Wikipedia page where the population data of various states is available from 2001 census.

I have downloaded state wise population data from this website.

ETL PROCESS

1. ETL FOR STATE_WISE COVID DATA:

Extraction:

I took the data from covid19India website where the data available in the form csv file updated regularly.

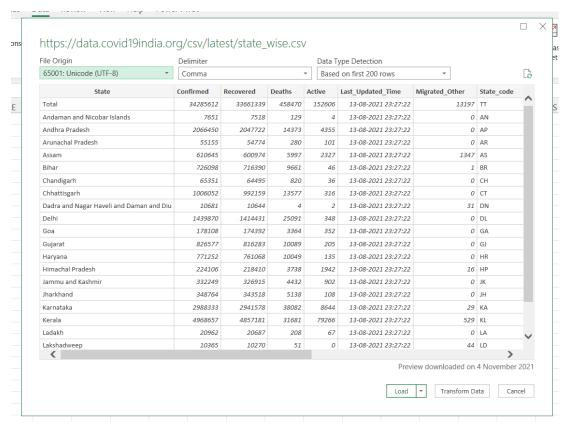
Link for the file: https://data.covid19india.org/csv/latest/state_wise.csv

Then go to data tab in excel and select get data from web and paste the above csv file link in the **URL** place and press ok.



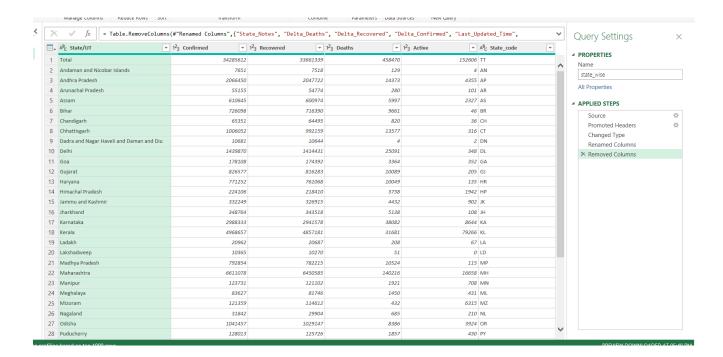
Transformation:

Press TRANSFORM DATA in the given table



Then transform the data as required.

Applied changes can be seen under applied steps section.



LOAD:

After performing the necessary changes click CLOSE AND LOAD option to load the data in to the required excel sheet.

2. ETL FOR VACCINE DATA:

Extraction:

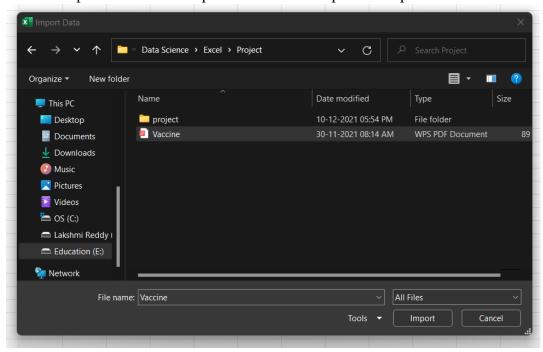
I took the vaccination data of states from MOHFW website where vaccination data of all the states are available in the form of pdf.

Download the pdf

Link to Pdf:

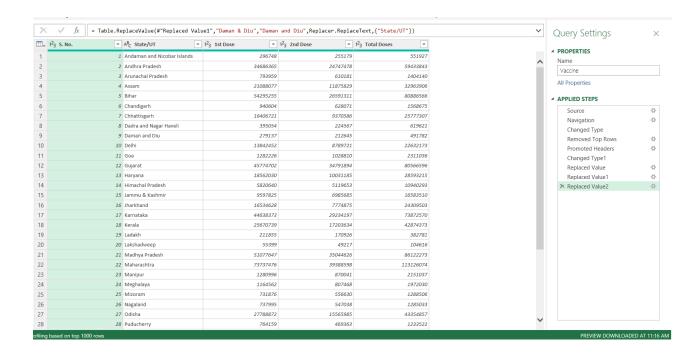
https://www.mohfw.gov.in/pdf/CummulativeCovidVaccinationReport09december202 1.pdf

Then go to data tab in excel and select data from file option and select pdf. Select the path of the vaccine pdf file and click import data option.



Transformation:

Applied changes can be seen under applied steps section.



LOAD:

Click CLOSE AND LOAD to get the data in to the required sheet.

3. ETL FOR STATE_WISE POPULATION DATA:

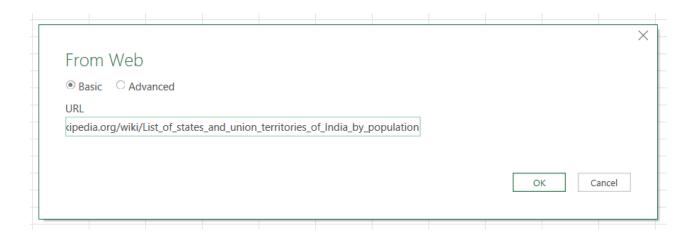
EXTRACTION:

I took the population data of all states across India from Wikipedia.

Link to download data:

 $\underline{https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_popu_\\ \underline{lation}$

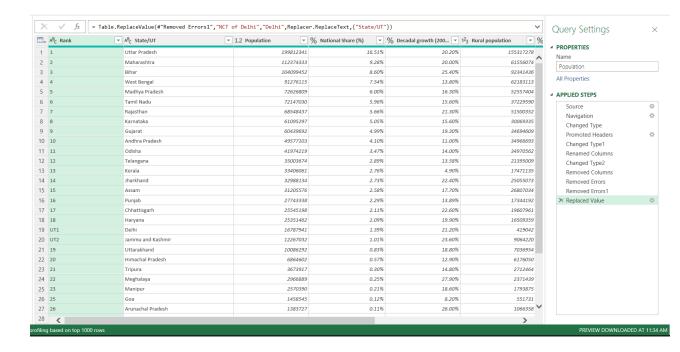
Then go to data tab in excel and select data from web option and paste above given URL and press ok.



TRANSFORMATION:

Perform the necessary changes

Applied changes are visible under applied steps section.



LOAD:

After the necessary changes are made click CLOSE AND LOAD to get the data to the required sheet

JOINING DATASETS

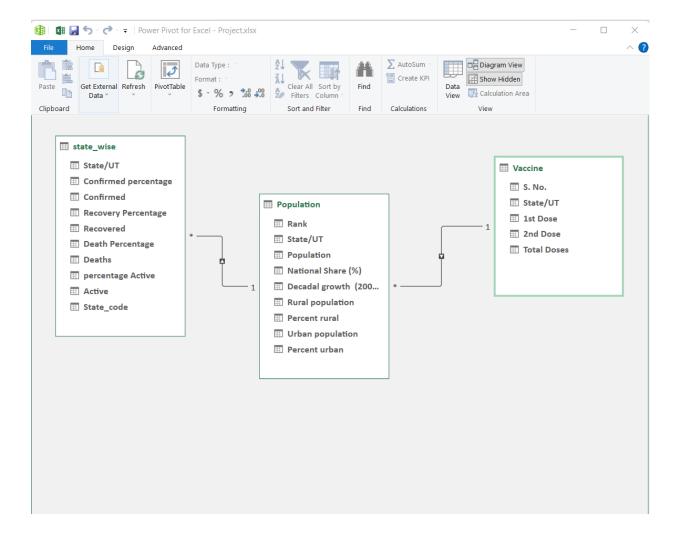
Before performing the analysis on our data lets join our datasets to ne another.

I have joined my all three datasets based on state name column: STATE/UT

STEPS TO JOIN:

To join these data sets we have to addon power pivot from microsoft excel add ons After adding power pivot go to manage data option and click it.

Then join the data sets as shown below.



ANALYSIS ON DATA

State wise covid analysis

Analyze the data from all states to know about the severity in spread of covid-19

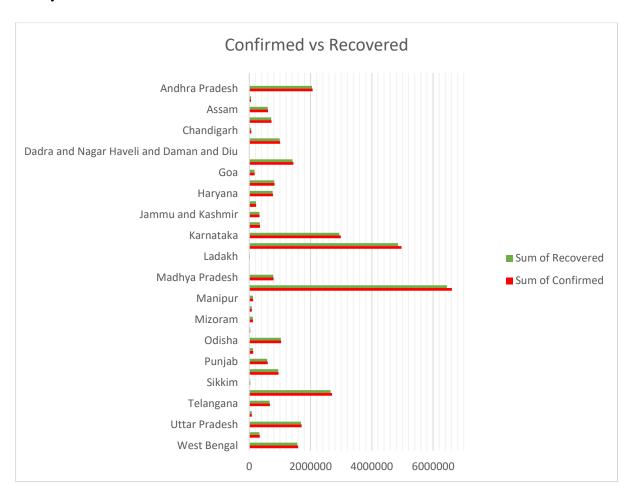
1. Confirmed Vs Recovered:

Description:

Get the data into a pivot Table Select states as columns and get the no.of confirmed vs no of recovered cases of covid-19.

Requirements:

Pivot table skills, graphs



2.Top 10 States with Covid-19

Description:

Visual Representation of top 10 states with covid 19.

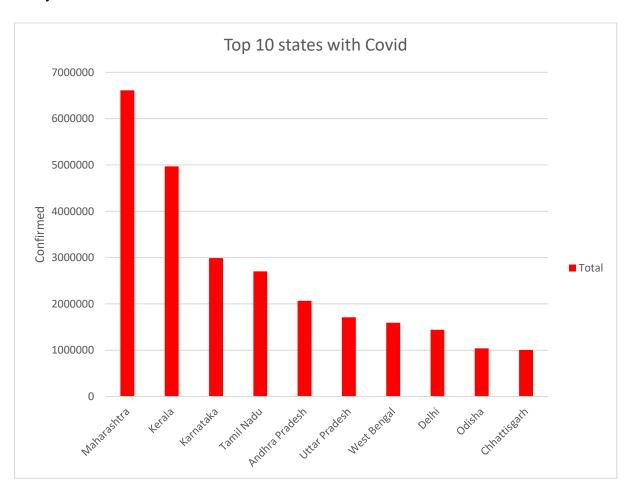
Project a graph with total no of confirmed cases across states.

Sort the confirmed cases column as largest to smallest

Filter the data to show top 10 states.

Requirements:

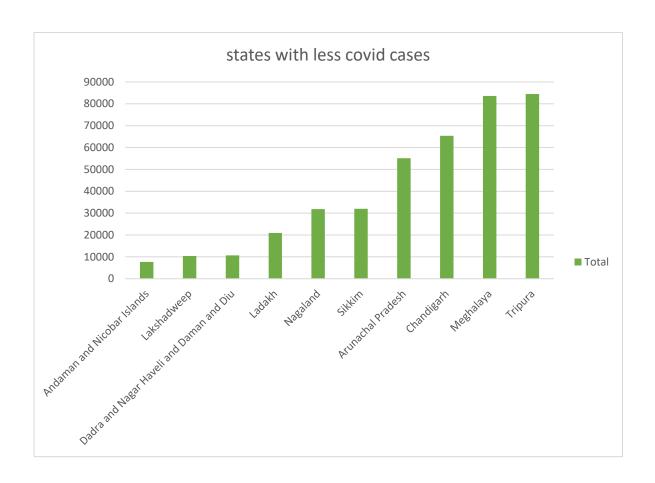
Pivot table skills, sorting, applying filters, graphs



3.States with less covid

Process:

Follow the same as above but this time filter the data to get last 10 states Analysis Results:



4. States with high Recovery rate:

Description:

Get the data into a pivot Table

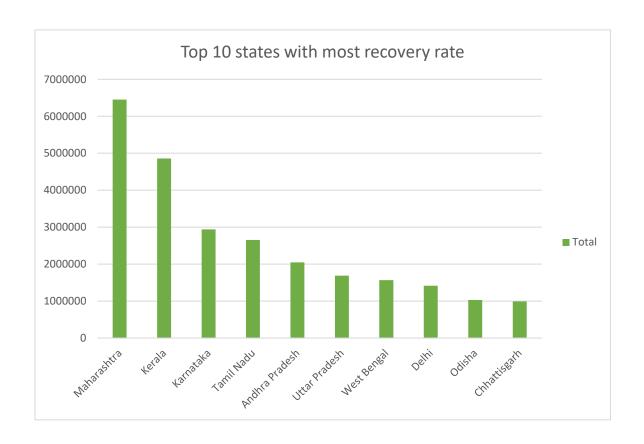
Select states as columns and get the no.of recovered cases of covid-19.

Sort the recovered cases column as largest to smallest

Filter the data to show top 10 states

Requirements:

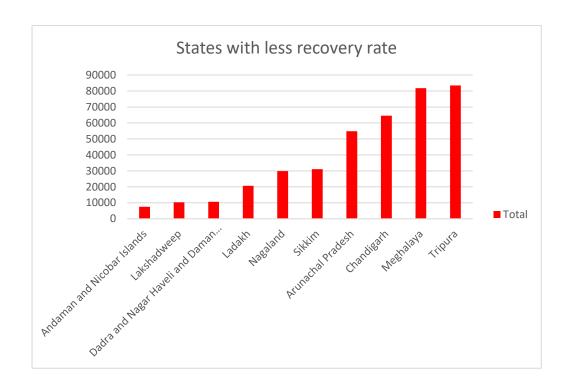
Pivot table skills, sort, filter, graphs



5.States with less recovery rate:

Process:

Follow the same as above but this time filter the data to get last 10 states Analysis Results:



Vaccine data Analysis

Analyze the data from all states to know about the Vaccination against covid-19

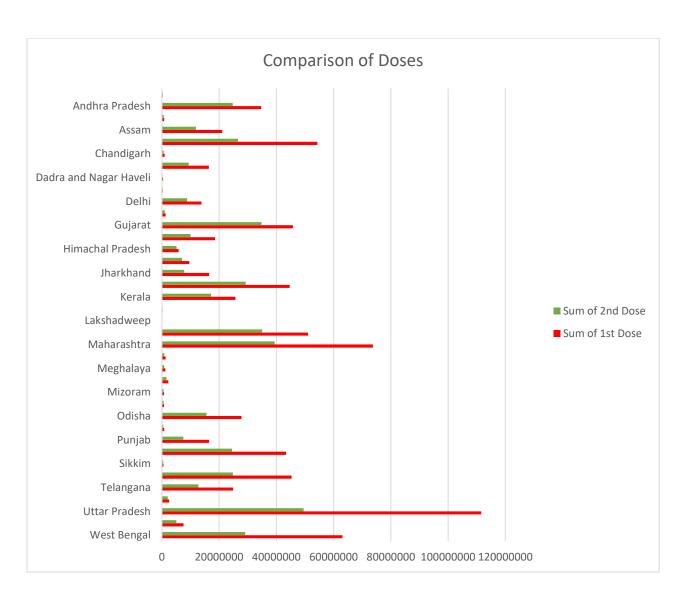
1. Comparison of Doses:

Description:

Get the data into a pivot Table Select states as columns and get the no.of first Dose vs no.of second Dose vaccine of covid-19.

Requirements:

Pivot table skills, graphs



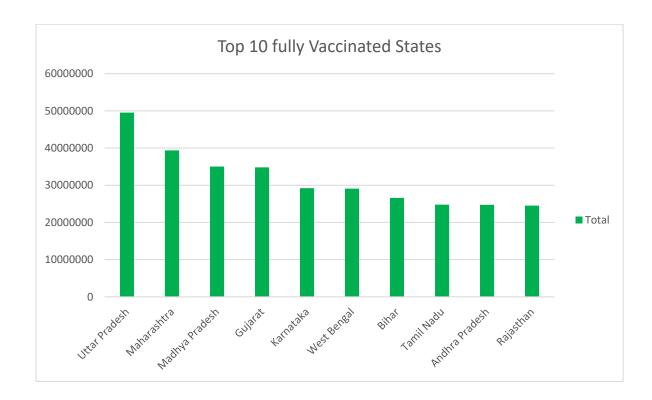
2.Top 10 fully Vaccinated States:

Description:

Get the data into a pivot Table Select states as columns and get the no.of second Dose vaccine of covid-19. Sort the data as largest to smallest and filter top 10 states.

Requirements:

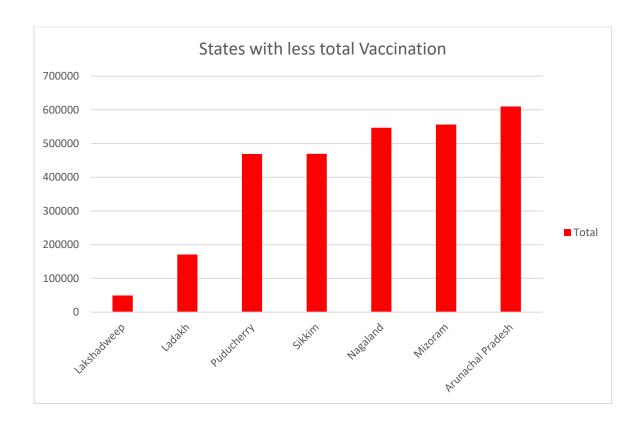
Pivot table skills, sorting, filters, graphs



3. States with less total Vaccinations

Process:

Follow the same as above but filter to get last 10 states.



Population Analysis

Using the data from the population dataset to bring the population of each states in to picture and analyzing the data keeping the population in view to know how the heavily populated states are performing compared to less populated states.

1. Population vs Confirmed cases:

Description:

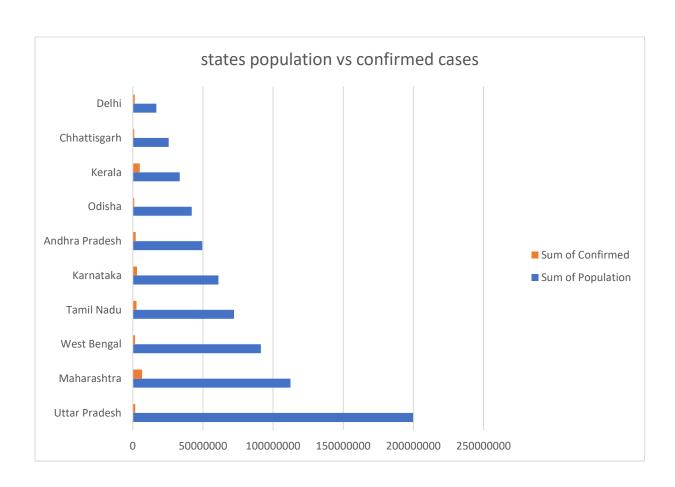
Get the data into a pivot Table

Select population of states as columns from population dataset and get the no.of confirmed cases from state_wise dataset.

Sort the population as largest to smallest and filter top 10 states.

Requirements:

Pivot table skills, sorting, filters, graphs



2.Top 10 Populated states vs Vaccination:

Description:

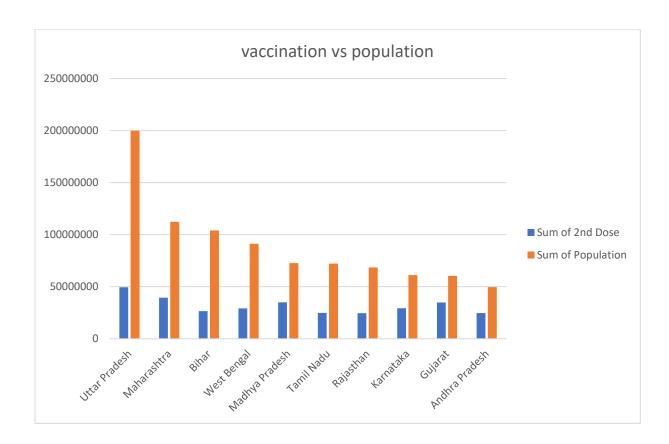
Get the data into a pivot Table

Select population of states as columns from population dataset and get the no.of second dose cases from vaccine dataset.

Sort the population as largest to smallest and filter top 10 states.

Requirements:

Pivot table skills, sorting, filters, graphs



RESULTS

DASHBOARD PREPARATION

Covid Dashboard Preparation

DASHBOARD:

			Covid -19 Das	hBoard o	f India			
Total Cases		Total Active		Total Recovered		Total Deaths		
No	3,42,85,612	No	1,52,606	No	3,36,61,339	No	4,58,470	
%	100.00%	%	0.45%	%	98.18%	%	1.34%	
No ?	Highest Total Case 66,11,078 Maharashtra	No ?	Highest Active 79,266 Kerala	No ?	lighest Recovered 64,50,585 Maharashtra	No ?	Highest Death 1,40,216 Maharashtra	
Lowest Total Cases			Lowest Active		Lowest Recovered		Lowest Deaths	
No	7,651 Andaman and	No	Dadra and Nagar Haveli and Daman	No	7,518 Andaman and	No	4 Dadra and Nagar Haveli and Daman	
?	Nicobar Islands	?	and Diu	?	Nicobar Islands	?	and Diu	

1.TOTAL CASES:

No = {Total no of confirmed cases across all states}

Formula= {=MAX (state_wise [Confirmed])}

Percentage= {percentage of total cases}

Formula= {no of cases/total cases}

FOR TOTAL ACTIVE, TOTAL RECOVERED, TOTAL DEATHS:

Follow the same as above and select respective columns as required.

- Active cases column for total active
- Recovered cases column for total recovered
- Deaths column for total deaths

2.HIGHEST TOTAL CASES:

No = {Maximum no of cases state wise}

Formula= {=MAX (state_wise! C3:C38)}

? is used to denote from which state the above-mentioned cases are taken

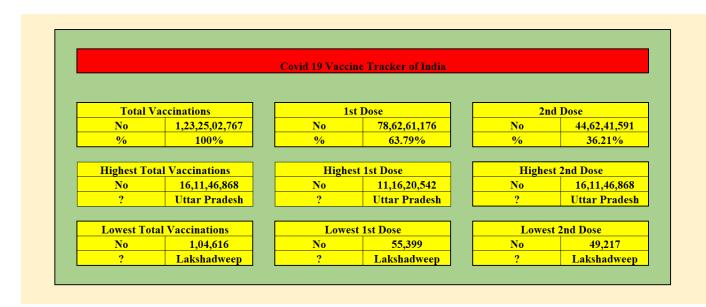
Formula= {=INDEX(state_wise!A3:A39,MATCH('Covid DashBoard'!G14,state_wise!C3:C38,0))}

FOR ALL THE REMAINING SECTIONS:

Follow the same as above but change the respective column as required

Vaccine Dashboard Preparation

DASHBOARD:



1.TOTAL VACCINATIONS:

No = {Total no of vaccinations across all states}

Formula= {=SUM (Vaccine[Total Doses])}

Percentage= {percentage of total vaccination}

Formula= {no of vaccinations/total vaccinations}

FOR FIRST DOSE AND SECOND DOSE SECTIONS:

Follow the same process as above and change the respective colums

- Use first Dose column for first dose
- Use second dose column for second dose

2.HIGHEST TOTAL VACCINATION:

No = {Maximum no of Total vaccinations}

Formula= {=MAX (Vaccine [Total Doses])}

? which state

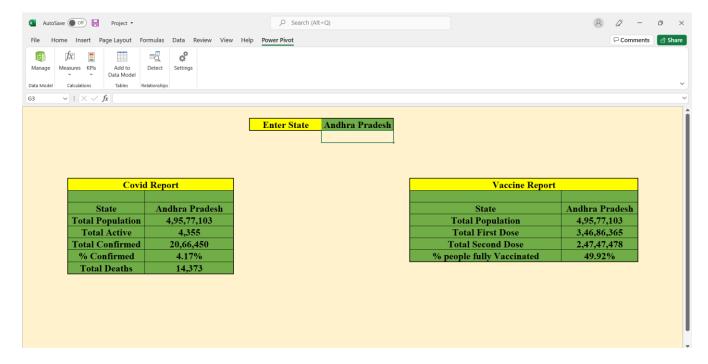
Formula= {=INDEX (Vaccine [State/UT], MATCH('vaccine DashBoard'!I12,Vaccine[Total Doses],0))}

FOR ALL THE REMAINING SECTIONS

Follow the same process and change respective columns

SEARCH BOARD

BOARD:



Usage:

Just type the name of the state in the enter state cell to get all the info about that state as shown above.

Formulae used in Vaccine Report Table:

Item	Formula
State	=\$G\$2 {getting from the state typed by user}
Total Population	=INDEX(Population[Population],MATCH(Search!D9,Population[State/UT],0))
Total First Dose	=INDEX(Vaccine[1st Dose],MATCH(Search!J9,Vaccine[State/UT],0))
Total Second Dose	=INDEX(Vaccine[2nd Dose],MATCH(Search!J9,Vaccine[State/UT],0))
% fully Vaccinated	=J12/J10

Formulae used in Covid Report Table:

Item	Formula
State	=\$G\$2 {getting from the state typed by user}
Total Population	=INDEX(Population[Population],MATCH(Search!D9,Population[State/UT],0))
Total Active	=INDEX(state_wise!I3:I39,MATCH(Search!D9,state_wise!A3:A39,0))
Total Confirmed	=INDEX(state_wise!C3:C39,MATCH(Search!D9,state_wise!A3:A39,0))
% Confirmed	=D12/D10
Total Deaths	=INDEX(state_wise!G3:G39,MATCH(Search!D9,state_wise!A3:A39,0))

REFERENCES

- https://www.covid19india.org/
- https://www.mohfw.gov.in/
- https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population
- https://data.covid19india.org/csv/latest/state_wise.csv
- $\bullet \quad \underline{https://www.mohfw.gov.in/pdf/CummulativeCovidVaccinationReport09december 202} \\ \underline{1.pdf}$

BIBLIOGRAPHY

- https://www.covid19india.org/
- https://www.mohfw.gov.in/
- https://en.wikipedia.org/wiki/List_of_states_and_union_territories_of_India_by_population
- https://data.covid19india.org/csv/latest/state_wise.csv
- $\bullet \quad \underline{https://www.mohfw.gov.in/pdf/CummulativeCovidVaccinationReport09december 202} \\ \underline{1.pdf}$