1. **Populations that have the highest risk of contracting Covid-19:**

* Datasets that can be considered:
* Covid Dataset having the information reg the symptoms with which a person is affected with covid,what are the chronic diseases a patient is suffering with,age,which region he belongs to
* Mask-covid affected and not affected
* Datasets related to :
* Age

<https://www.kaggle.com/saurabhshahane/covid19-age-risk-factor>

[https://www.kaggle.co](https://www.kaggle.com/zohaib123/provisional-covid19-deaths-counts-by-age-in-years)[Covid-19 Age Risk Factor](https://www.kaggle.com/saurabhshahane/covid19-age-risk-factor)[/zohaib123/provisional-covid19-deaths-counts-by-age-in-years](https://www.kaggle.com/zohaib123/provisional-covid19-deaths-counts-by-age-in-years)

* Region

[https://www.kaggle.com/lin0li/covid19testing9](https://www.kaggle.com/lin0li/covid19testing9world)

<https://www.kaggle.com/sudalairajkumar/covid19-in-india>

\*\*\* https://www.kaggle.com/zeus75/distribution-of-covid19-cases-worldwide

https://www.kaggle.com/pranabpandit/covid-19

* Symptoms

https://www.kaggle.com/hafeezuddinshaik/covid-19

* Mask wear
* Month -Season
* What are the possible situations under which a person is affected with covid:
* How can he possibly get affected with covid?
* How well is his mental health?
* What is his income level?
* Gender
* Any specific chronic diseases a person is suffering
* How long has it been for him to contact a covid +ve patient before he contracted the disease?
* Is he vulnerable to common colds and coughs frequently?

1. **Possible severity of illness in a particular patient/demographic:**

* Post Covid Scenarios:
* Can consider datasets related to the diseases that people affect post covid.
* List of the diseases that are suffered by covid patients.

<https://www.kaggle.com/madan44/covid19-patient-precondition-dataset-cleaned>

(post condition)

<https://www.kaggle.com/tanmoyx/covid19-patient-precondition-dataset?select=covid.csv>

(severity)

* Study them in depth: What is the specific reason for the occurrence of that disease?Is it due to disease and troubles that existed before or have some habits like consuming alcohol or smoking?(symptoms)
* Apply these understandings and predict the possible diseases that one can face post covid
* Can also give cautionary measures that a person can take given a particular symptom. We need to be able to predict what can be a precautionary measure stating the severity of the illness based on the given symptoms.

1. **Incidence of Covid-19 infection among cancer patients:**

* Need to collect information on:
* What is the possibility of cancer patients contracting covid
* Can give a relation between covid and cancer
* Identify the symptoms of cancer and identify the symptoms of covid.
* If a person is having these common symptoms, some conclusion can be

made out of this

https://www.kaggle.com/rishidamarla/cancer-patients-data

* Pics of the lungs can be taken and then conclusion can be drawn

1. **How the implementation of strategies impacts rates of Covid-19 infection?**

**Mask**

**Social distancing**

**Sanitization**

**Vaccination**

1. **Vulnerability**

**\*HDI**

**\*POVERTY**

**\*Development\_and\_Deprivation**

**https://www.kaggle.com/mpwolke/cusersmarildownloadsvulnerabilitycsv**

**https://www.kaggle.com/mariaren/covid19-healthy-diet-dataset?select=Supply\_Food\_Data\_Descriptions.csv**

**Steps to be followed for the project:..Kaggle.com**

1. Appropriate dataset collection based on the above requirements.

Question making on each parameter

Preprocessing-Cleaning the data(remove unnecessary data)

1. Having the data handy, now we need to perform data cleaning
2. For the cleaned data, we can now perform basic data analysis using pandas to understand the trends in the records

Prediction-using pandas library

Classification algorithm

1. Data Visualization using (Matplotlib………………………….) to understand the visuals in the data
2. Accuracy of the Model

Ml model prediction

**Datasets reg covid:**

**7 to 9**

1. [**https://www.kaggle.com/hgunraj/covidxct**](https://www.kaggle.com/hgunraj/covidxct)
2. [**https://www.kaggle.com/ravichaubey1506/covid19-india**](https://www.kaggle.com/ravichaubey1506/covid19-india)
3. [**https://www.kaggle.com/tunguz/data-on-covid19-coronavirus**](https://www.kaggle.com/tunguz/data-on-covid19-coronavirus)
4. [**https://www.kaggle.com/mpwolke/cusersmarildownloadshumanitariancsv**](https://www.kaggle.com/mpwolke/cusersmarildownloadshumanitariancsv)
5. [**https://www.kaggle.com/kunal28chaturvedi/covid19-and-its-impact-on-students**](https://www.kaggle.com/kunal28chaturvedi/covid19-and-its-impact-on-students)
6. [**https://www.kaggle.com/mpwolke/cusersmarildownloadscovidriskscsv**](https://www.kaggle.com/mpwolke/cusersmarildownloadscovidriskscsv)
7. [**https://www.kaggle.com/duttadebadri/covid19-india-complete-data**](https://www.kaggle.com/duttadebadri/covid19-india-complete-data)
8. [**https://www.kaggle.com/ruchi798/covid19-pulmonary-abnormalities**](https://www.kaggle.com/ruchi798/covid19-pulmonary-abnormalities)
9. [**https://www.kaggle.com/mpwolke/cusersmarildownloadssecondarycsv**](https://www.kaggle.com/mpwolke/cusersmarildownloadssecondarycsv)
10. **===========================================================**

**Datasets related to age:**

**#\*Different countries covid cases analysis**

**https://www.kaggle.com/aestheteaman01/covcsd-covid19-countries-statistical-dataset**

**==================================**

Age:

Age vs Deaths count-got it

Max death for which age- done

Max death for female and for which age

Max death for male and for which age

Ratio of total deaths vs covid deaths

Symptoms:

1. Value counts of yes or no for the following diseases:

* Eye pain
* Sore throat
* Chest pain
* Runny nose
* Weakness
* Aches
* Headache
* Cough
* Difficulty in breathing
* Change in sleep cycle

1. Do you personally know anyone in your local community who is ill with a fever and either a cough or difficulty breathing?

Value count for Yes

**=======================================================================**

**How many questions can you ask?**

**List of the datasets available:**

1. <https://www.ecdc.europa.eu/en/publications-data/data-daily-new-cases-covid-19-eueea-country>

* Number of deaths in the country?
* Maximum number of deaths per day in the country?
* Did the number of cases increase/decrease each day in a country?
* Which country has the lowest number of cases?
* Which country has the lowest number of deaths?
* In which month of the year there is a peak in the number of cases in the (specified)country?
* Is there any increase in the number of cases in the (specified)country?
* Which country has the highest number of deaths in a particular month?

1. <https://www.ecdc.europa.eu/en/publications-data/covid-19-testing>

* What is the population in the (specified)country?
* Is there any increase in new covid cases everyday in the (specified)country?
* How many covid-19 tests are done daily in the (specified)country?
* What is the rate of covid-19 testing in the (specified)country?
* How many positive cases are being reported in the (specified)country ?
* Is there increase or decrease in positive cases?
* What is the spread of covid-19 in the country, country wide (or) subnation?
* Which region of the country does the person belong to?
* Population vs tests done
* What is the avg rate(mean) of tests done in a country?
* Which country has the minimum rate of testing?
* Which has the max testing rate & positivity rate?
* Which week of the year has registered the highest no.of positive cases?
* What are the modes/sources of testing and how far are they efficient?
* Which country is more actively testing its people?
* Positive cases-- 2020 vs 2021-- Is there any decrease?
* Is there any increase in testing rate?

1. <https://www.kaggle.com/hafeezuddinshaik/covid-19>

* How many people are above 18 and how many are not above 18
* How many people are suffering with eye pain?
* How many people are suffering from chest pain?
* How many people are suffering with sore throats?
* How many people are suffering with eye pain?
* How many people are suffering with a runny nose?
* How many people are suffering with eye pain?
* How many people are suffering with weakness?
* How many people are suffering with aches?
* How many people are suffering with eye pain?
* How many people are suffering with headaches?
* How many people are suffering with difficulty in breathing?
* How many people are suffering from a change in sleep cycle?
* How many people are suffering from coughing?
* How many people have met someone with similar symptoms before?
* How many are such that if any of the symptoms is yes and if the person has met anyone with symptoms before,
* What is the max (give the top 5 sorted list) of th symptoms
* Give the count of the days a person had these symptoms
* Give the count of the record having these symptoms more than 5 days
* Count of number of people who have been tested for coronavirus
* How many people had been diagnosed with coron +ve
* How many people had at least 3 symptoms and has been tested positive
* How many people had symptoms but is tested negative
* What are the list of the diseases that are suffered by the people
* How often do you use face cover or mask while you travel out
* What is gender? How many males and how many females?
* How many smoke and how many don't?
* How many suffered with covid who smoke?
* How many consume alcohol?
* How many suffered with covid who consumed alcohol?
* Are people sanitizing their hands often?
* Give the age band list
* Which day has the maximum no.of cases registered?
* How many tests/surveys are done within 24hrs?
* Which region is more likely to have max. Cases?
* Is there any correlation b/w a person having more symptoms and his locality?
* Are the people tested positive before again prone to COVID for the second time?
* What is the rate of people having symptoms but being tested negative & viceversa.
* What is the rate of risk(severity) among the people tested positive?
* How efficiently is the vaccination drive going?
* What are the post covid complications?
* Rate of people liking to be jabbed?
* What is the no.of people tested positive with diseases/conditions like BP, diabetes..
* Are the people tested positive being quarantined?
* Rate of people taking precautions.
* Are consumption of alcohol, smoking risk factors?

1. [Factor](https://www.kaggle.com/saurabhshahane/covid19-age-risk-factor)[/zohaib123/provisional-covid19-deaths-counts-by-age-in-years](https://www.kaggle.com/zohaib123/provisional-covid19-deaths-counts-by-age-in-years)

* Give the number of males and females
* Give the how many number of age groups are present
* Sum up all the covid deaths and give its number
* For which age group, did the max covid death cases occur
* For males of which age group did the max covid deaths occur
* For females of which age group did the max covid deaths occur
* Give the top 5 age groups for which the max covid deaths occur(in a sorted order)

1. <https://www.kaggle.com/sudalairajkumar/covid19-in-india>

* How many states are there in total?
* What is the time period for each state
* ‘What are the positive cases for each date corresponding to a particular state
* In a particular time period of time, what are the total number of positive cases for a state
* When did the max positive cases occur for a state?
* What is the max positive case in total for all states?(“need to give the total number)
* When in the least positive case occur
* Where are the least positive cases-in which state
* Give mean for all the positive cases for each state
* Give the max of the mean and its corresponding state
* Identify the number of blank spaces
* How many negative cases are there for each state
* Give the ratio of positive cases vs negative cases for each state
* From the given ratio, identify which state is leading in increasing the spread of covid.

1. <https://www.kaggle.com/hafeezuddinshaik/covid-19> (repeated symptoms data only)

* Find how many max no of symptoms results covid(we need to process the data)
* What symptom mostly cause covid(using pi chart)
* How many of them suffered from covid but got negative result
* Higher immunity people in our locality(not got covid even though there is contact with locality)
* Low immunity people in our locality
* How many of them suffered from covid by contact of locality
* What bad habits results fast catch of covid(low immunity)(smoking,alcohol consumption)
* How the implementation strategy impact the rate of covid(people who taken care but suffered from covid)
* Does sanitizing helped people to spread or control covid(using data of how many people sanitized vs +ve covid case)

1. <https://www.kaggle.com/madan44/covid19-patient-precondition-dataset-cleaned>

* How many people are suffering from pneumonia
* How many people are suffering from diabetes
* How many people are suffering from copd
* How many people are suffering from asthma
* How many people are suffering from hypertension
* How many people are suffering from cardiovascular
* How many people are suffering from obesity
* How many people are suffering from renal\_chronic
* How many people are suffering from tobacco
* How many people are suffering from other diseases
* How many people suffered with inmsupr
* How many had diseases(any of them) and had covid
* How many had covid and died
* How many had covid and did not die
* How many had diseases and had covid and died
* How many had diseases, did not have covid and died
* How many had diseases, and had covid and did not die
* How many did not have any of the diseases and had covid
* Which age group had max number of diseases
* Which age group had max number of covid cases
* What is the total number of deaths
* What is the total number of covid cases
* What is the ratio of number of covid cases vs deaths
* Which gender suffered with maximum number of diseases
* Which age group suffered with maximum number of diseases

1. <https://www.kaggle.com/tanmoyx/covid19-patient-precondition-dataset?select=covid.csv>

* How many people condition is seviar who suffered from covid
* How many seviar people died
* How pregnant women condition severity
* How different diseased people severity
* How intubated patients severity
* Which disease mostly caused stviarity(+ve case+respective disease vs severity)

1. [Factor](https://www.kaggle.com/saurabhshahane/covid19-age-risk-factor)[/zohaib123/provisional-covid19-deaths-counts-by-age-in-years](https://www.kaggle.com/zohaib123/provisional-covid19-deaths-counts-by-age-in-years)

**(repeated)**

* Age vs Deaths count
* Max death for which age-yet
* Max death for female and for which age
* Max death for male and for which age
* Ratio of total deaths vs covid deaths

1. https://www.kaggle.com/pranabpandit/covid-19

* Which country health facility is good (use recovery,death data)
* Which country is in high risk of covid(use no of covid cases data) [use bar graph]

1. [**https://www.kaggle.com/sudalairajkumar/covid19-in-india**](https://www.kaggle.com/sudalairajkumar/covid19-in-india)

* Rate Spread of covid cases in a particular country (graph [time vs covid cases])
* Which country took more care on people (more no of tests, less spread)

1. **https://www.kaggle.com/anandhuh/latest-covid19-india-statewise-data**

* **What is spread of covid among india(bar graph)**
* **What ratio of people who got covid died**
* **Which state took more care(according to death rate)**
* **Which state have more spread of covid**

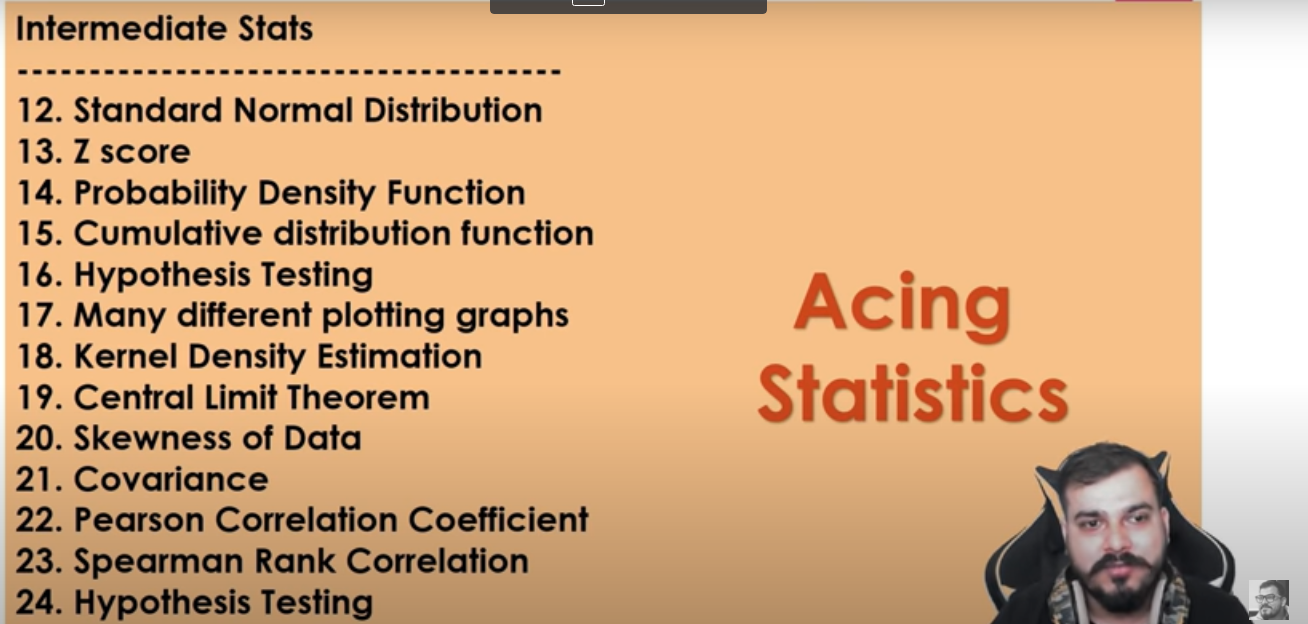
**Statistics for Data science:**

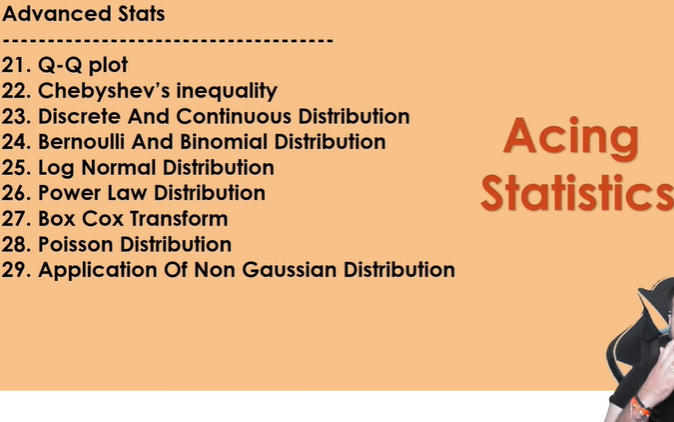
**Basic Stats for Beginners:**

Probability--is the base. Basic knowledge about probability is essential

1. Intro to basic terms
2. Variables
3. Random variables
4. Population ,sample,population mean,sample mean
5. Population distribution,sample distribution
6. Mean,median,mode
7. Range
8. Measure of dispersion
9. Variance
10. standard deviation
11. Gaussian distribution

**Intermediate stats:**





Feature Engineering Playlist: [https://www.youtube.com/playlist?list...](https://www.youtube.com/playlist?list=PLZoTAELRMXVPwYGE2PXD3x0bfKnR0cJjN)

Stats Playlist: [https://www.youtube.com/playlist?list...](https://www.youtube.com/playlist?list=PLZoTAELRMXVMhVyr3Ri9IQ-t5QPBtxzJO)

Complete ML Playlist: [https://www.youtube.com/playlist?list...](https://www.youtube.com/playlist?list=PLZoTAELRMXVPBTrWtJkn3wWQxZkmTXGwe)

Stats Syllabus:[https://drive.google.com/file/d/1gVUd…](https://drive.google.com/file/d/1gVUd%E2%80%A6)

=========================

* Z Score for outliers.
  + Age
  + Deaths
  + Symptoms
  + etc..
* Before that you can create normal distribution curves to important datasets
  + For many data sets

<http://www.elfhs.ssru.ac.th/morakot_wo/file.php/1/9781788290678-STATISTICS_FOR_DATA_SCIENCE.pdf>

The above stat book can give us direction for data science using stat

====================================================================================================================================

Exploratory Data Analysis:

1. State wise dataset for one day:

Key observations:

* Recovery rate vs death rate
* Gaussian distribution for:

i)Recovered cases

ii)Deaths

iii)Active cases

iv)Confirmed cases

* Outlier detection----Highest confirmed cases, active cases,death cases for which state?----represent through a **box plot**
* **Percentiles** ---1)confirmed cases 2)active cases 3)recovery cases
* Variance--a graph can be drawn representing variance of confirmed cases,active cases and deaths.. So that we can know how much far each state is, from the mean cases
* Max and min cases for top 5 states
* confirmed ,active,deaths bar charts
* Correlation between:and give the coefficients of them and also the matrix

i)any relation between confirmed cases and deaths?

ii)relation b/w confirmed and active

iii)relation b/w recovered and active

iv)relation b/w confirmed and recovered

* Give the Pearson and spearman coefficients
* log distribution for confirmed cases,active cases and recovered cases

1. District dataset:

Key observations:

* Gaussian distribution for active cases
* Boxplot for confirmed cases to show the spread in districts
* Z-score for active cases
* QQ-plot for active cases to compare it to gaussian distribution
* Histogram for Districts vs Recovered cases
* Mean for confirmed cases district wise and overall too
* Correlation b/w i)recovered and confirmed cases ii)recovered and active
* Correlation coefficient b/w i)recovered and confirmed ii)recovered and active
* Correlation matrix for above correlations
* Log normal distribution for confirmed, active and death cases
* Scatter plot for districts in a particular state vs active cases(can be done for a few states, but it might be difficult for all districts at the same time) and districts in a particular state vs confirmed cases

3.AGE DATASET

Key observations:

* age and covid deaths scatter plot
* Bar graph of age vs covid deaths
* Bar graph of age vs covid deaths and normal deaths
* Box plot of age vs deaths
* Outliners of different age groups deaths
* Histograms of different size group of deaths

4.Covid\_19\_india Dataset

Key Observations:

* Mean - Average confirmed cases in a particular state/union territory
* Std. deviation - spread of the state from the “mean” cases
* Z- score - state/union territory graph and outlier detection
* PDF - graphs regarding confirmed cases in different states and declaring the region based on the intersection(for eg., if pdfsome x-value== xyz state)
* CDF - graphs declaring the percentage of cases(either of confirmed, death, recovered) in a particular region
* Correlation - representing correlation between all the field using “heatmap”
* Correlation coefficient - finding correlation coefficient between “confirmed” and “cured” cases and if it is not such a good relationship, we suggest methods...
* correlation matrix - to find all the correlations
* Trimmed mean - average cured after removing outliers
* Relplot using seaborn between confirmed and cured cases & hue='State/UnionTerritory'

5.severiaty data set

* Pi chart of different diseases(who died)
* Pi chart of diseases suffered by icu patients(which shows that which disease cause severity)
* Bar plot of different diseases deaths
* Bar plot of male and female who got covid and covid deaths
* Pi plot of male and female who died due to covid
* Pi plot of male and female who got in to icu(who were resistive)
* Mode of disease caused who were died

6.DATASET\_symptoms:

Key Observations:

* Frequency table - count of yes in each symptom

-- bar plot of freq table

* Pie charts - percentage of people above 18 or so having each symptom
* P value - graph based on no.of symptoms a person has-- such that a person having more than 5 lies in the medieval region and has much probability to be prone to covid-- person having less than 5 lies in the outliers and has less chance to be prone
* Mode - the symptom having the highest no.of yes
* Percentiles - percentile matrix for different symptoms

===============================================================================================

EDA Sample projects/useful links

1.<https://medium.com/codex/covid-19-analysis-with-python-b898181ea627>

2. [An End-to-End Project on Time Series Analysis and Forecasting with Python | by Susan Li | Towards Data Science](https://towardsdatascience.com/an-end-to-end-project-on-time-series-analysis-and-forecasting-with-python-4835e6bf050b)

These are the 2 end to end EDA project

3.<https://pandas.pydata.org/pandas-docs/stable/getting_started/intro_tutorials/05_add_columns.html>

Mentioned about how to add a new column to the existing dataset. And how to find the ratio b/w two columns

4.<https://www.kite.com/python/answers/how-to-find-the-mean-of-a-pandas-dataframe-column-in-python>

To find the mean of a particular column

5. [How do I select multiple rows and columns from a pandas DataFrame?](https://www.youtube.com/watch?v=xvpNA7bC8cs)

Very helpful to select rows and columns to calculate mean for desired rows and columns

6.<https://stackoverflow.com/questions/41768629/normal-distribution-plot-by-name-from-pandas-dataframe>

Refer this when you are doing z-score

7.<https://mode.com/python-tutorial/python-histograms-boxplots-and-distributions/>

Refer for histograms and box plots

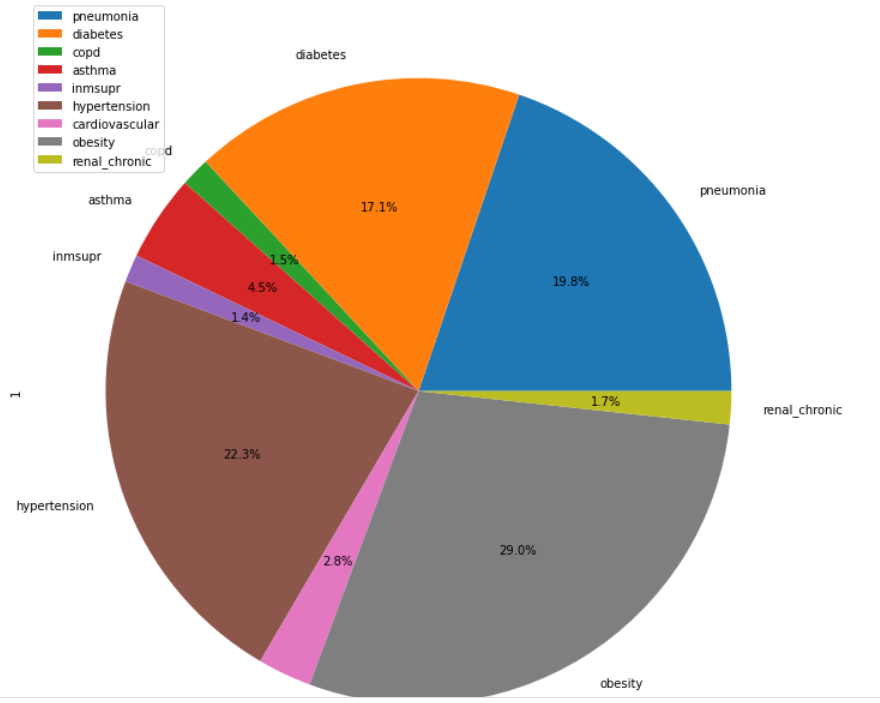
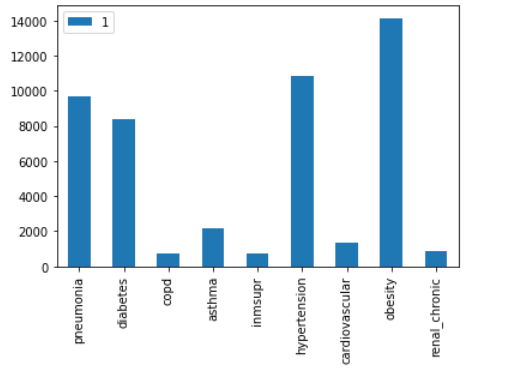
8.[How to Convert a Pandas DataFrame to a NumPy Array (marsja.se)](https://www.marsja.se/how-to-convert-a-pandas-dataframe-to-a-numpy-array/)

To convert a column into numpy array

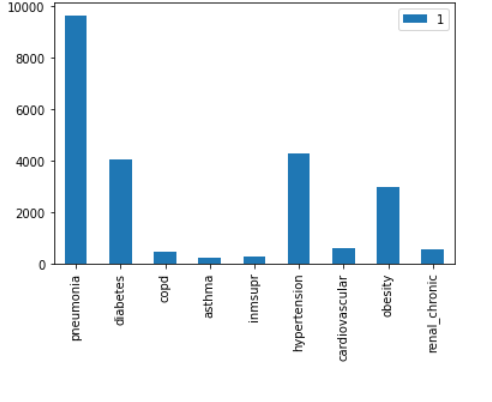
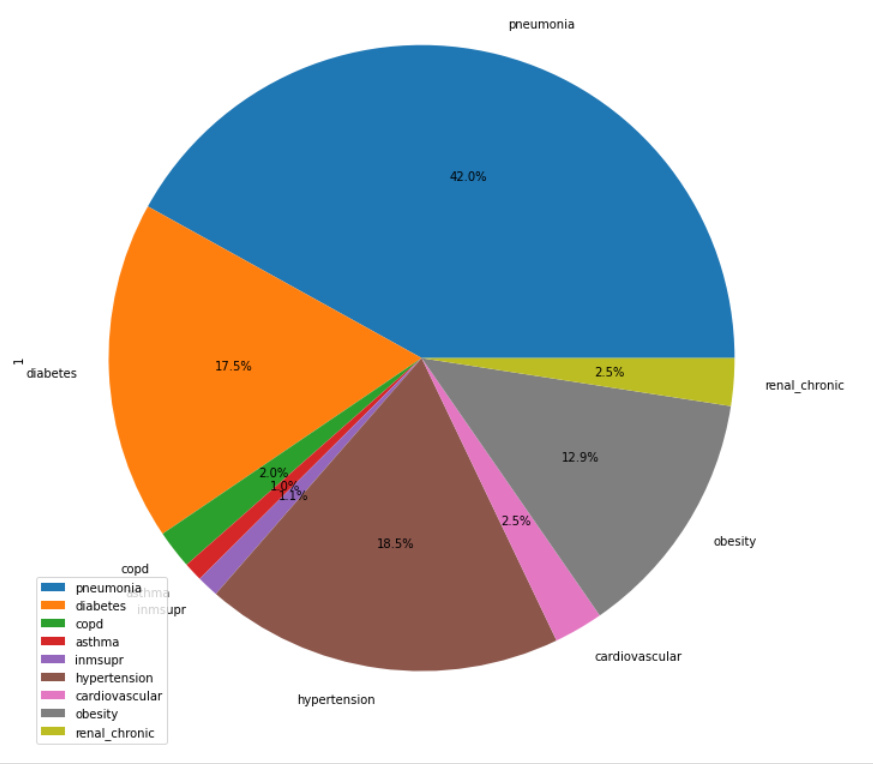
==========================================================================================================================================================================

Inferences from EDA:

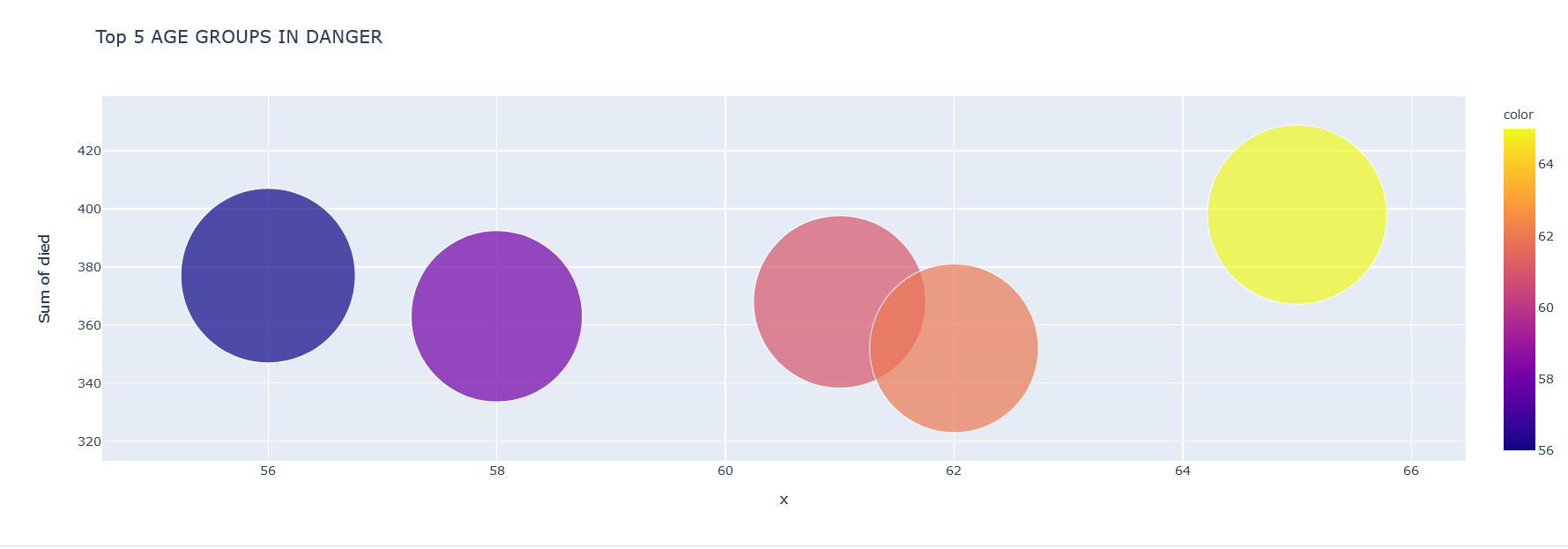
1)Severity Dataset:

1.This pie chart shows that which disease group people had high risk of COVID\_1(which means that the immunity of the people having respective disease is low and there is a possibility of getting covid due to contact of other covid patients is high) 

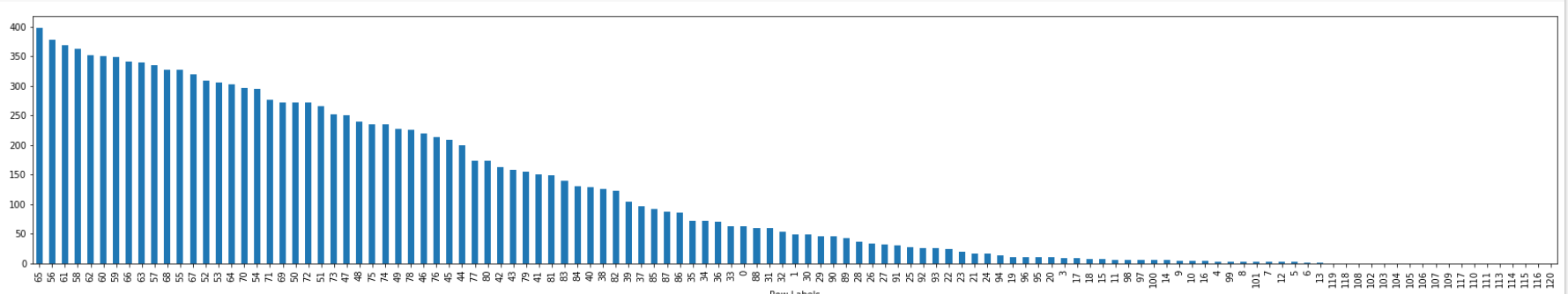
2.The above pie chart shows that which disease group people will have high risk of death in case of getting covid(which means that the people who are suffering from the respective disease should be more careful)



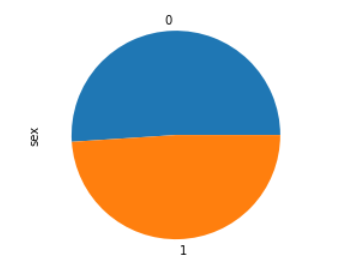
3.The top-5 age group of people who had high risk of covid



4.This graph shows different age groups who got COVID-19

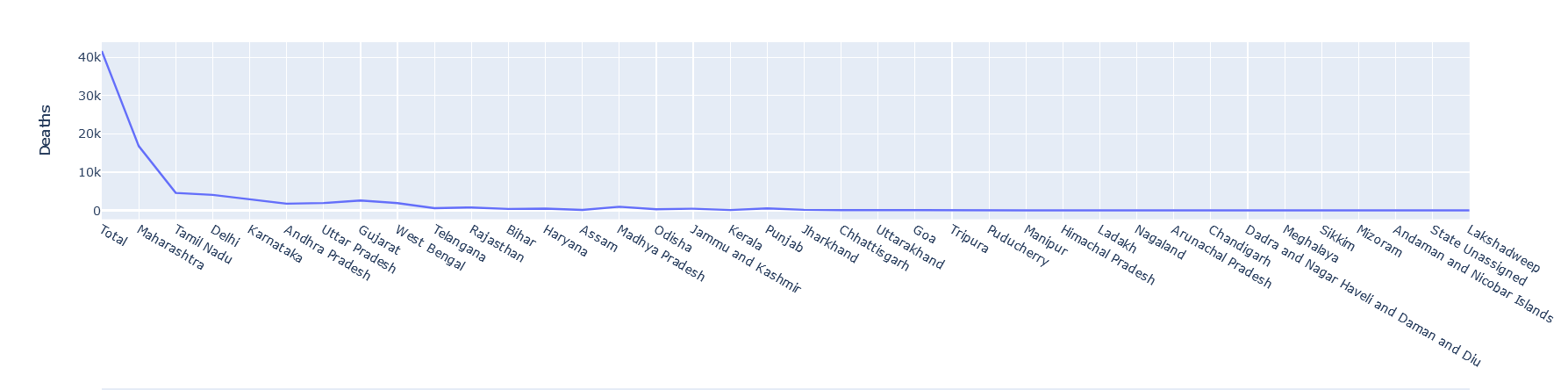


5.above plot shows that which gender has high COVID-19 deaths

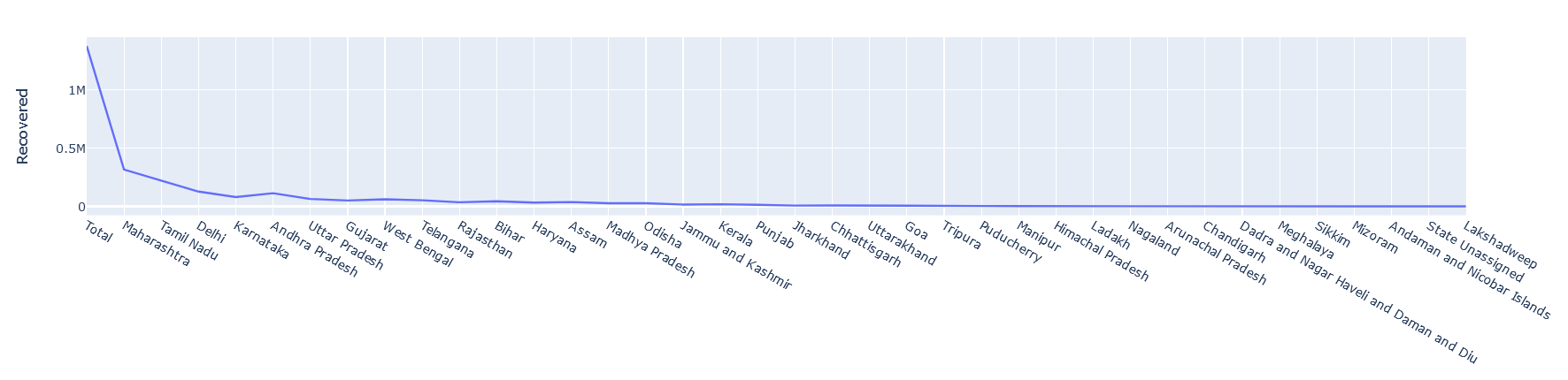


2)State Wise dataset:

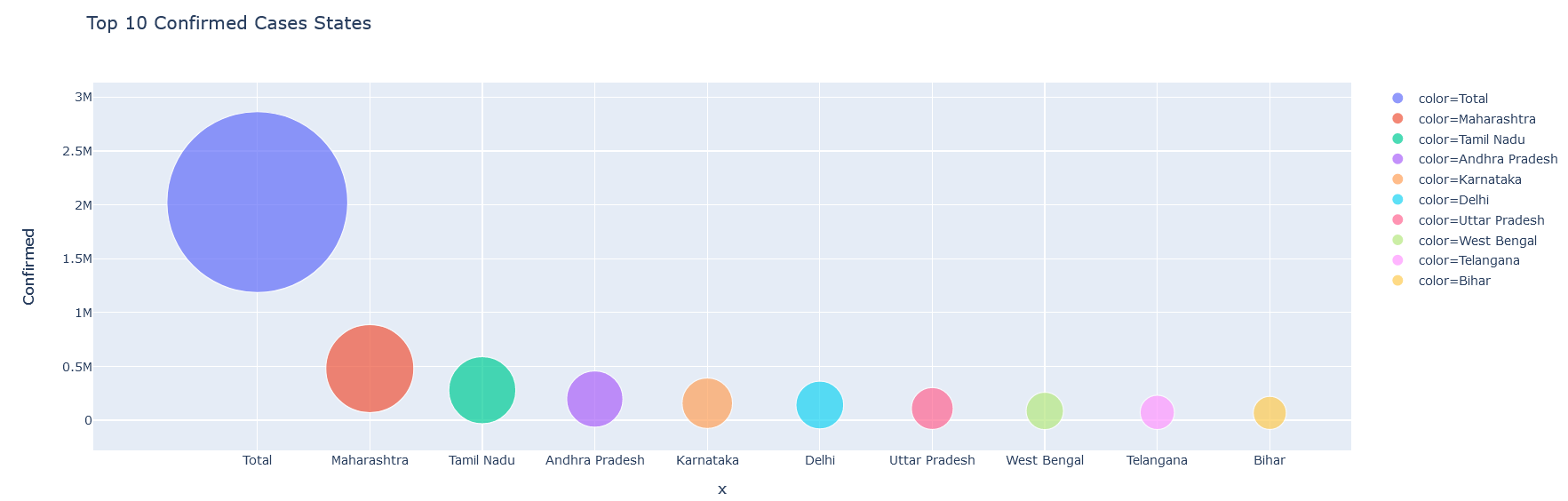
1.deaths of different states of india



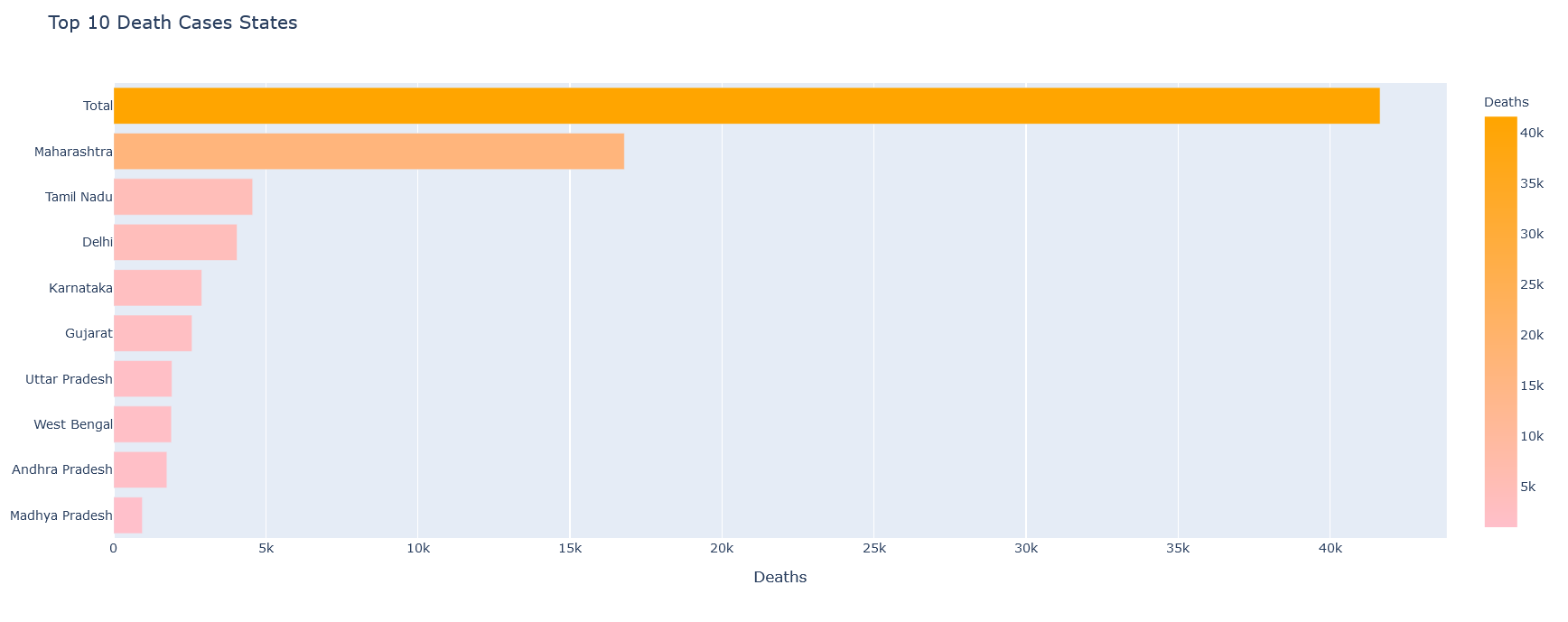
2.recovery of different states of india



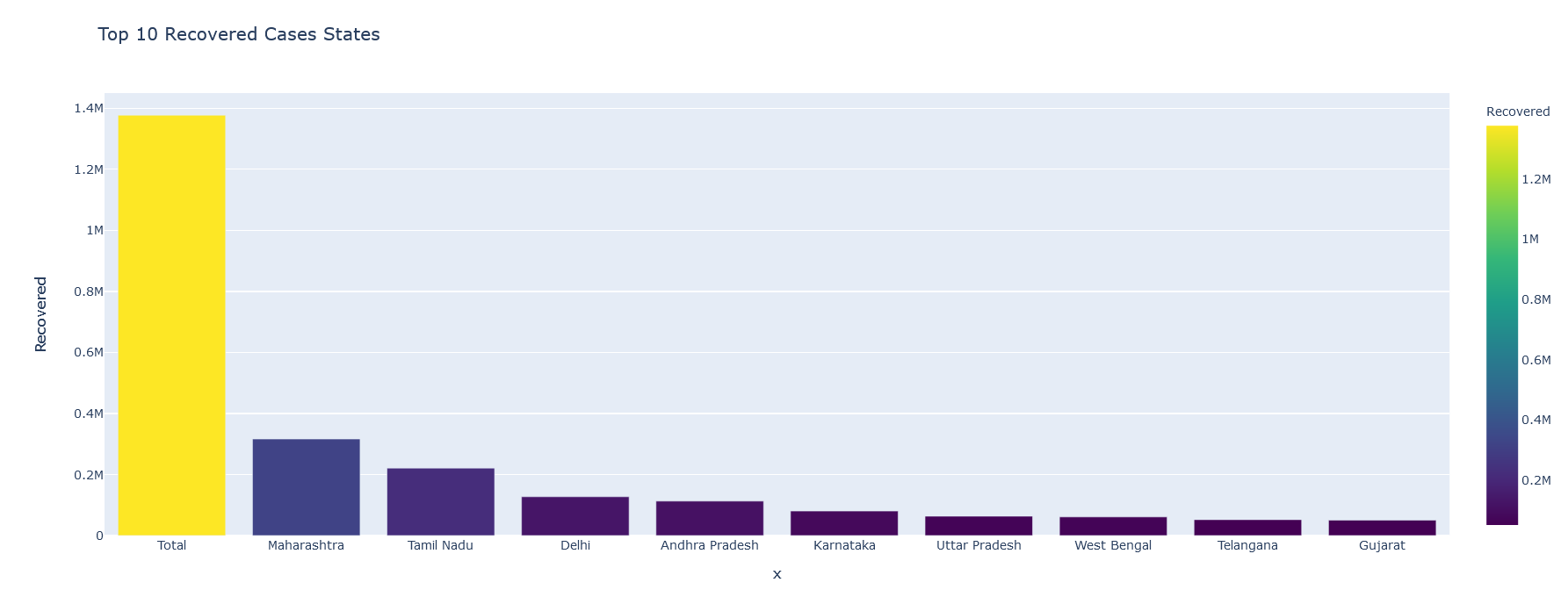
3.the above plot shows that top ten states in which had high covid cases



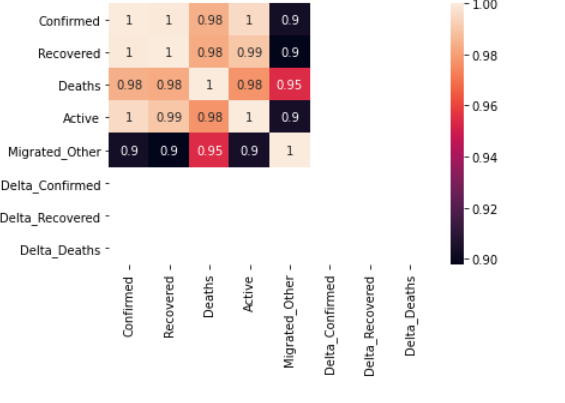
4.above bar graph shows tha top ten states with high deaths



5.above bar graph shows that top-10 states with high recovery

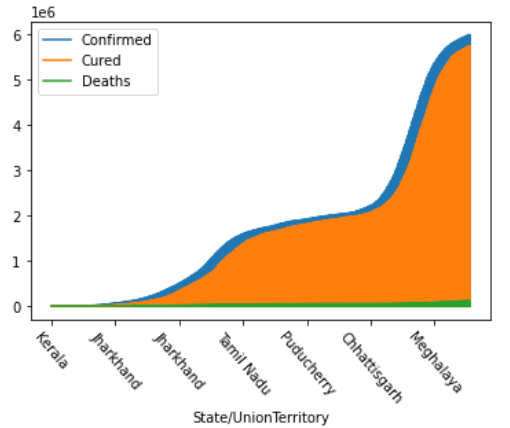


6.this shows the correlation between confirmed,recovered,deaths,active cases

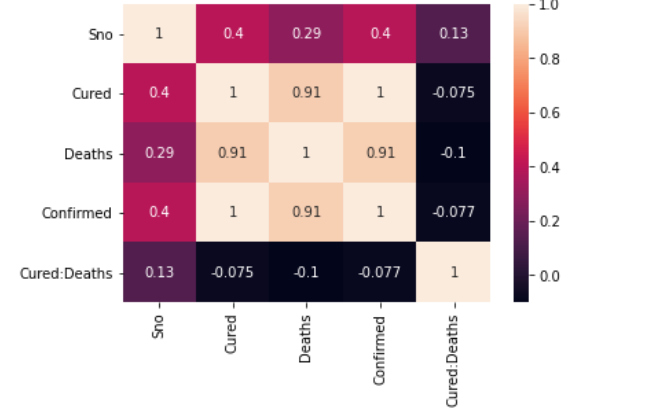


**3)Covid\_19 dataset:**

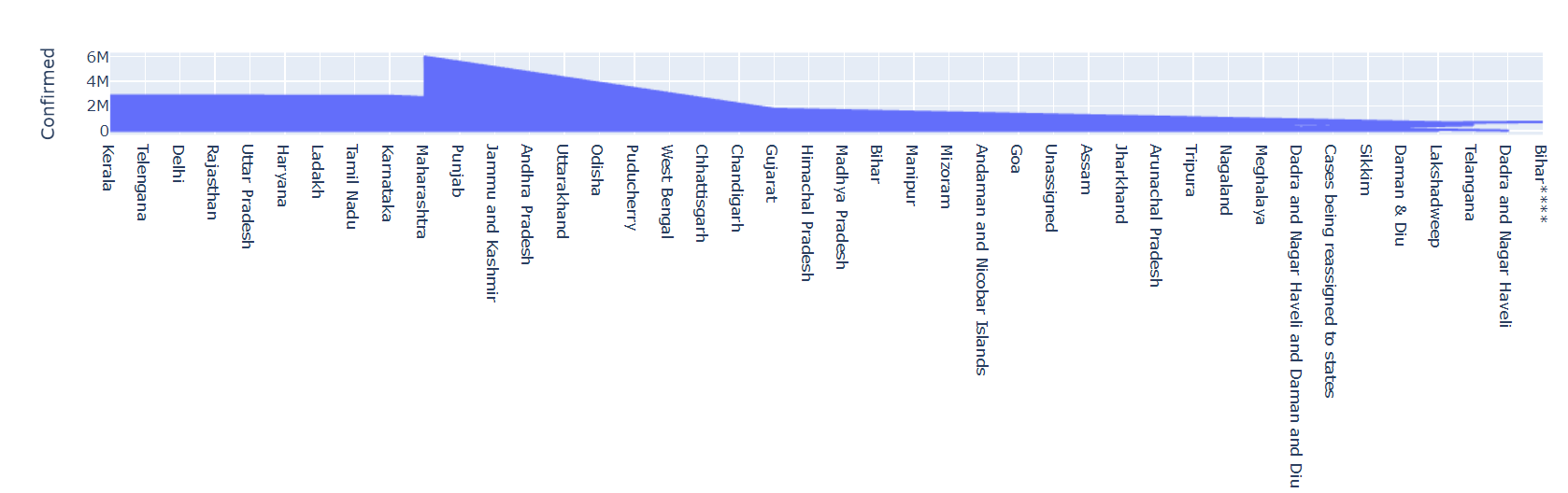
1. The below graph shows the basic plot of confirmed, cured and deaths in India.



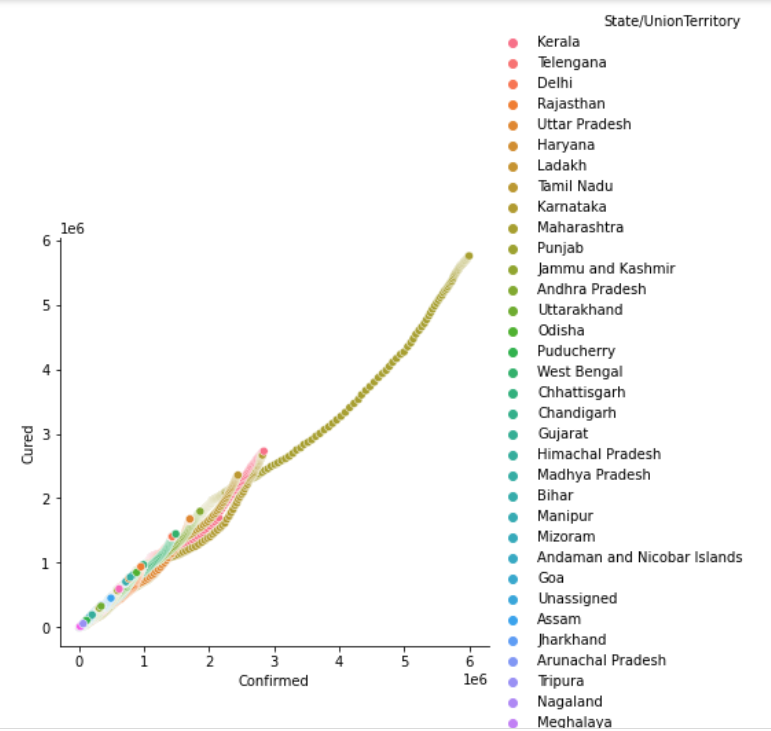
1. The correlation coefficients between attributes of our dataset. Here, we can clearly see that Cured cases and Confirmed cases are in a perfect relationship. So we can infer that there is a good balance and hope that people can be recovered by taking perfect measures.



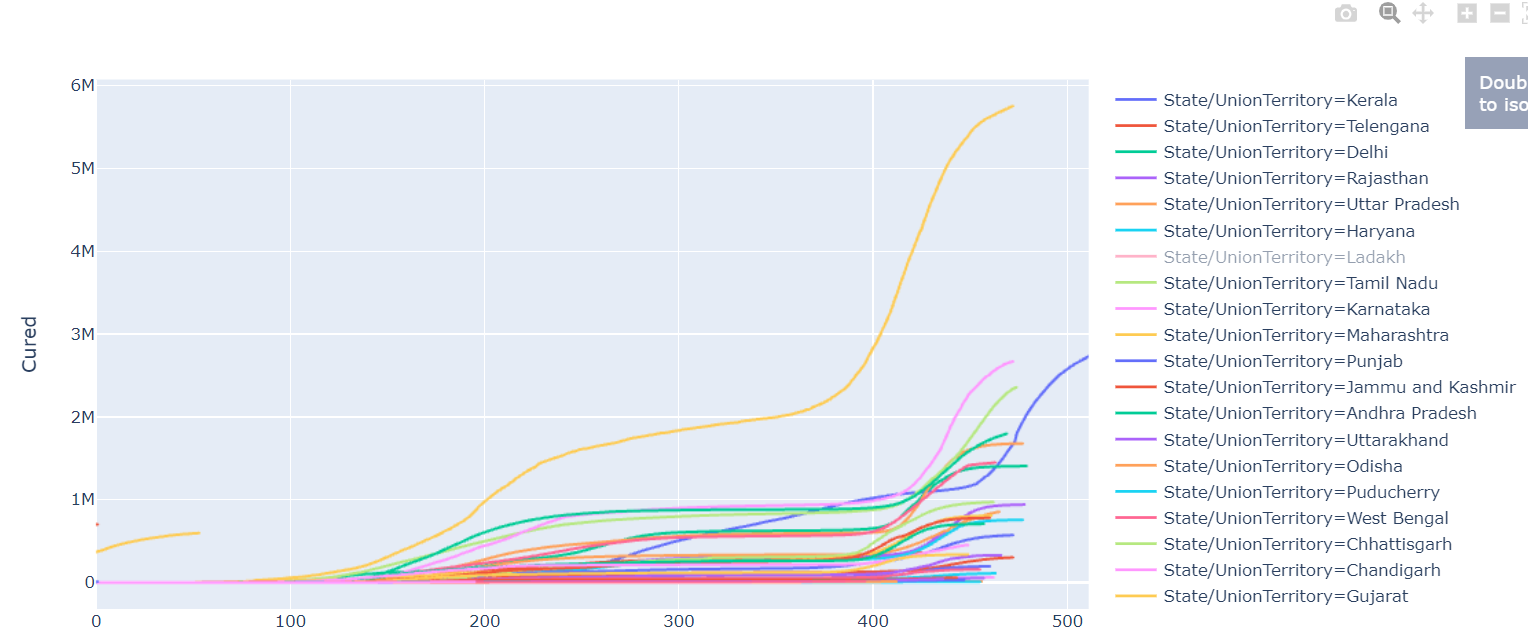
1. The graph is indicating us the spread of confirmed cases among different states across India. Maharashtra has the highest number of cases hitting 5.9M.

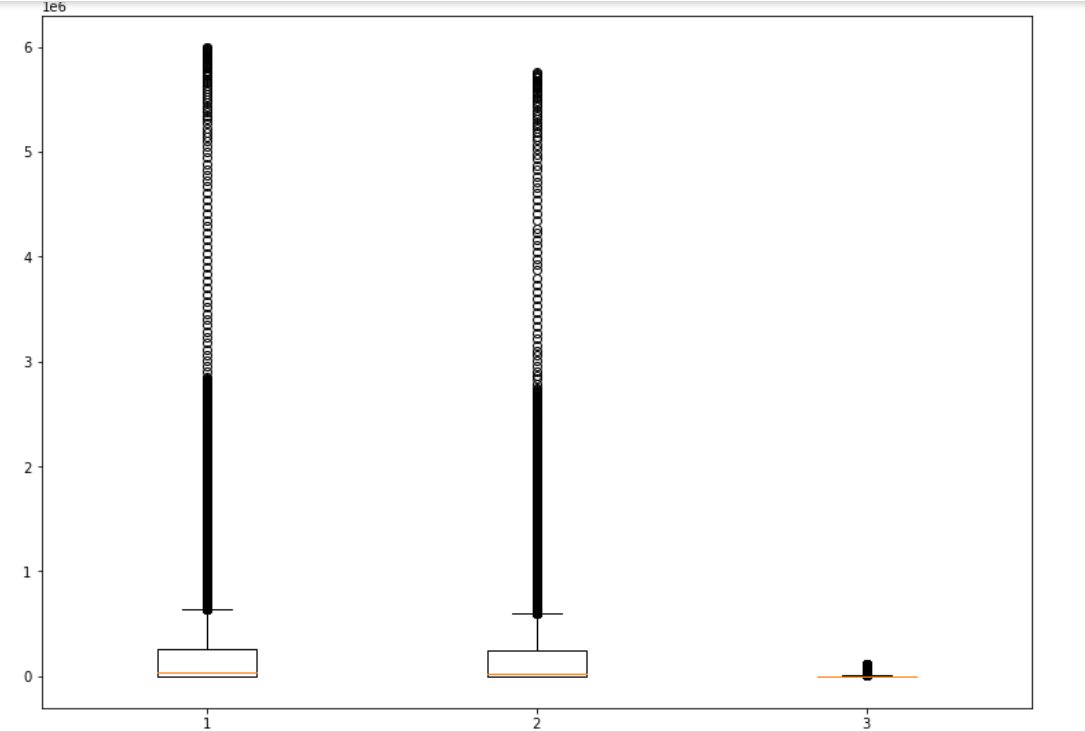
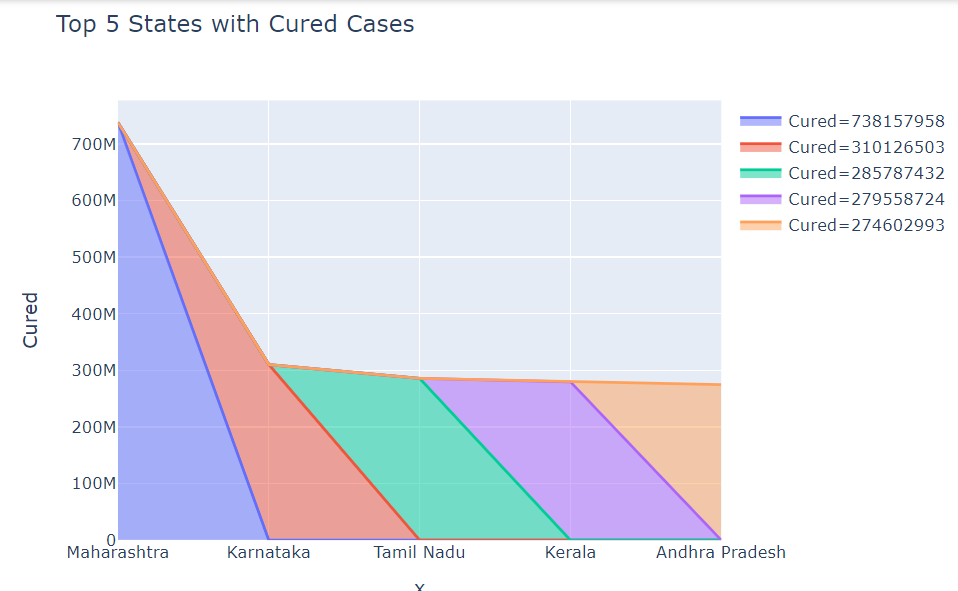
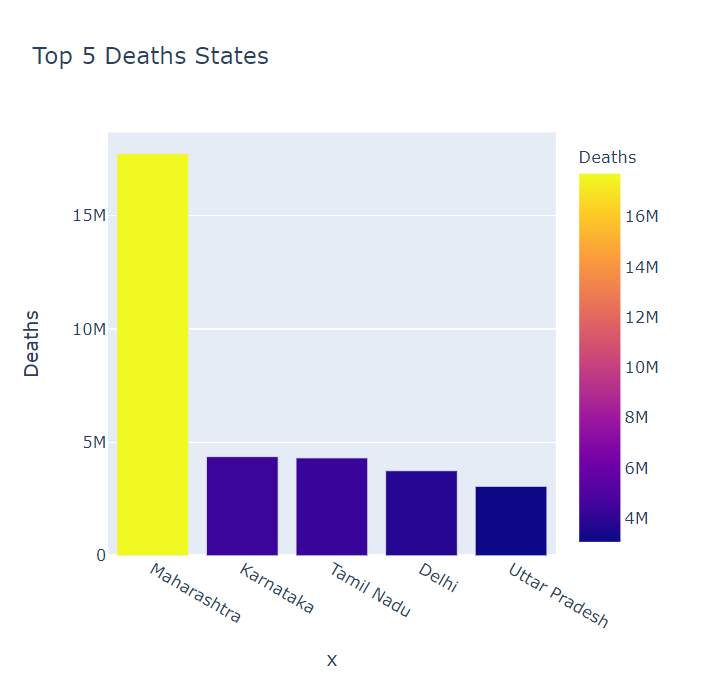
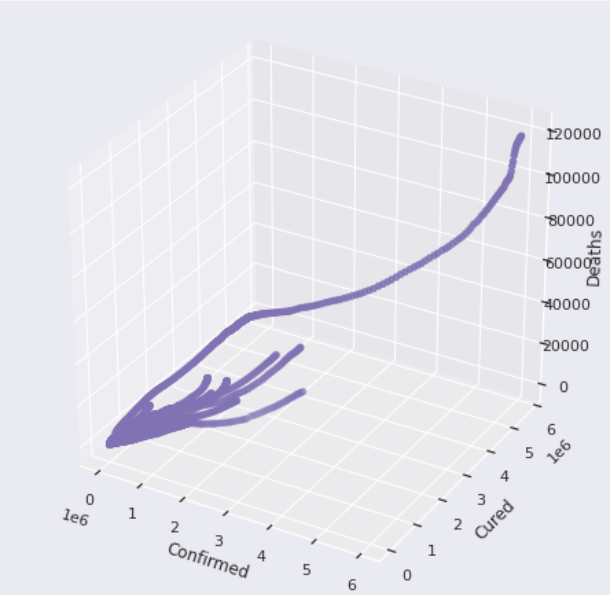


1. The below is the relation plot between the confirmed and cured cases of different states across India using “seaborn.”



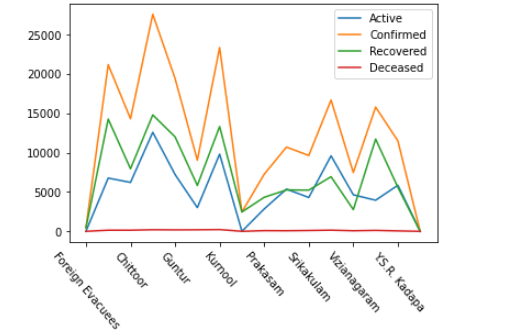
1. The below graph shows the plot of cured cases across different states from which we can infer that Maharashtra has the highest number of cured cases .

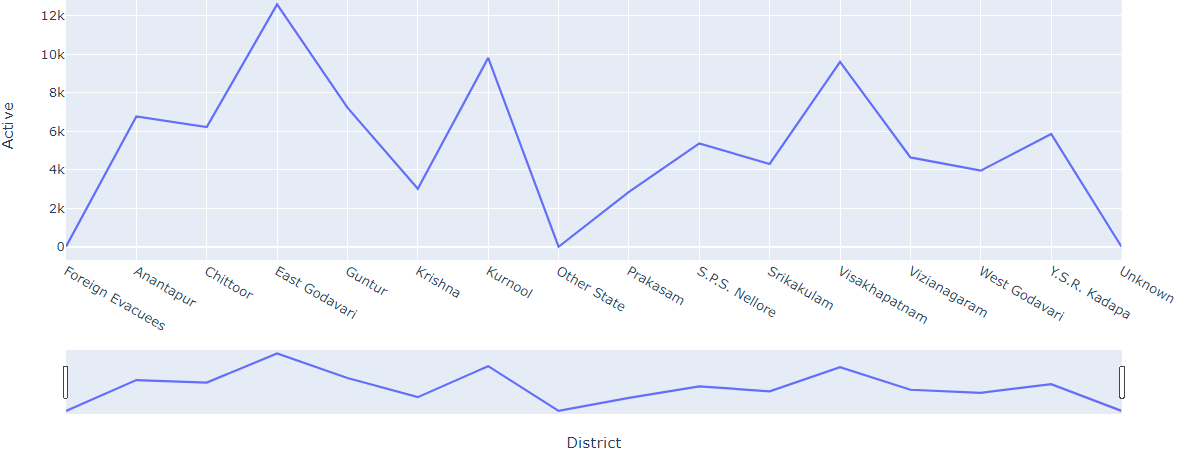


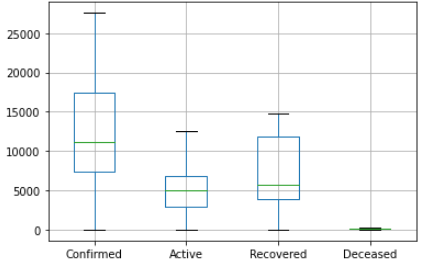
1. Below, we have a box plot of Confirmed, Cured and Death cases respectively. As we can see clearly, confirmed and cured has almost the same mean i.e., confirmed cases and cured cases are approximately proportional.
2. The below graph is a contour plot using plotly.express which gives us the top 5 states of confirmed cases.
3. The below graph shows us the top 5 states of cured cases. As we can see, in these two graph the same 5 states has the more confirmed and cured cases. So, across India, we have a perfect relationship between confirmed and cured cases.
4. The graph below is a basic bar graph using plotly.express depicting the top 5 states with deaths.
5. Below is the 3D graph between the attributes Confirmed, Cured and Death cases.(use if needed)

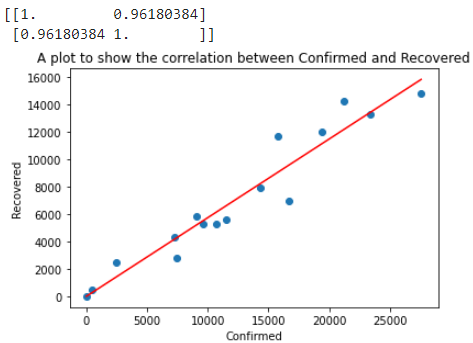
4)District wise dataset:

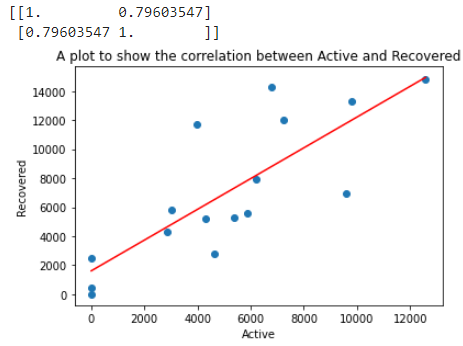
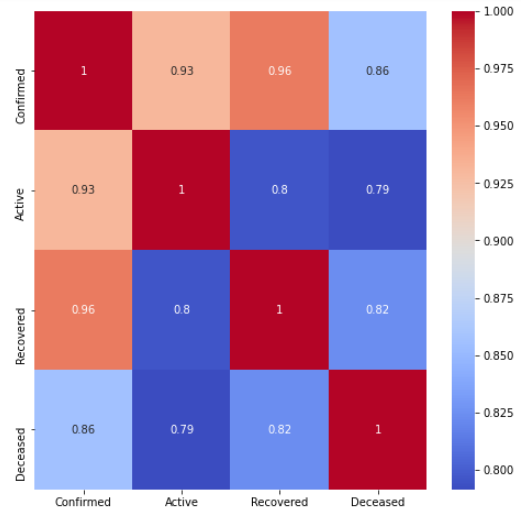
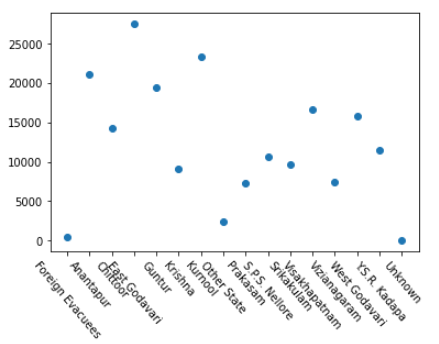
1.Below graph shows the confirmed,active,recovered and deceased cases of districts in Andhra Pradesh

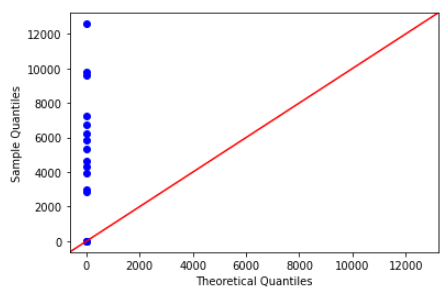


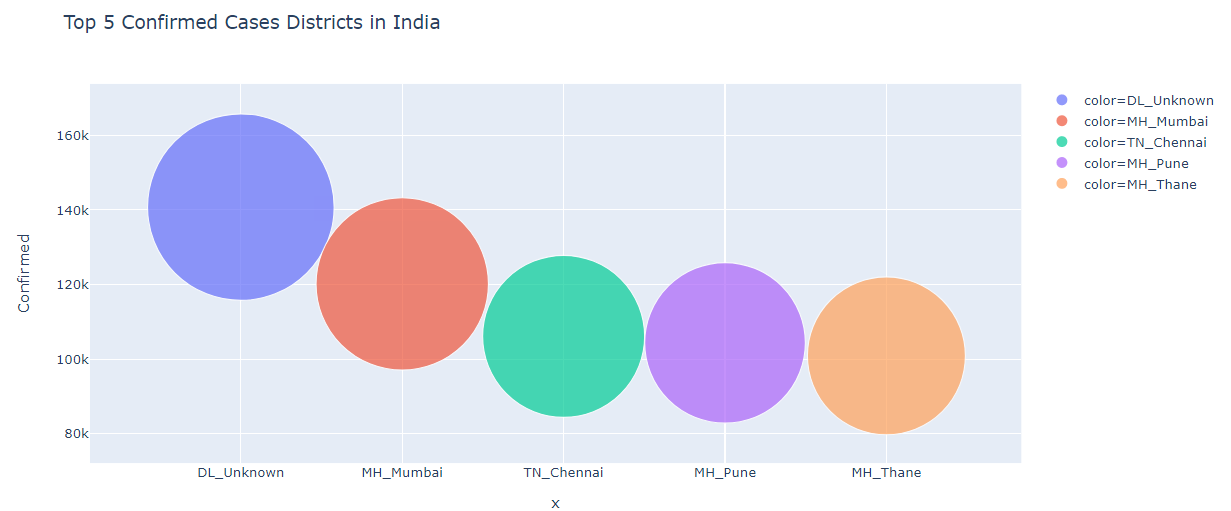
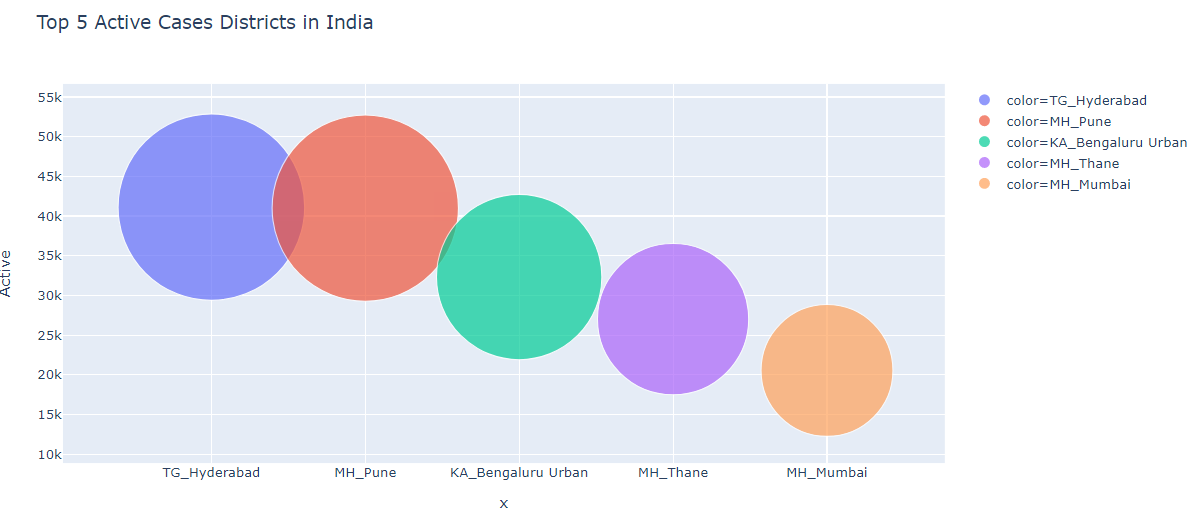
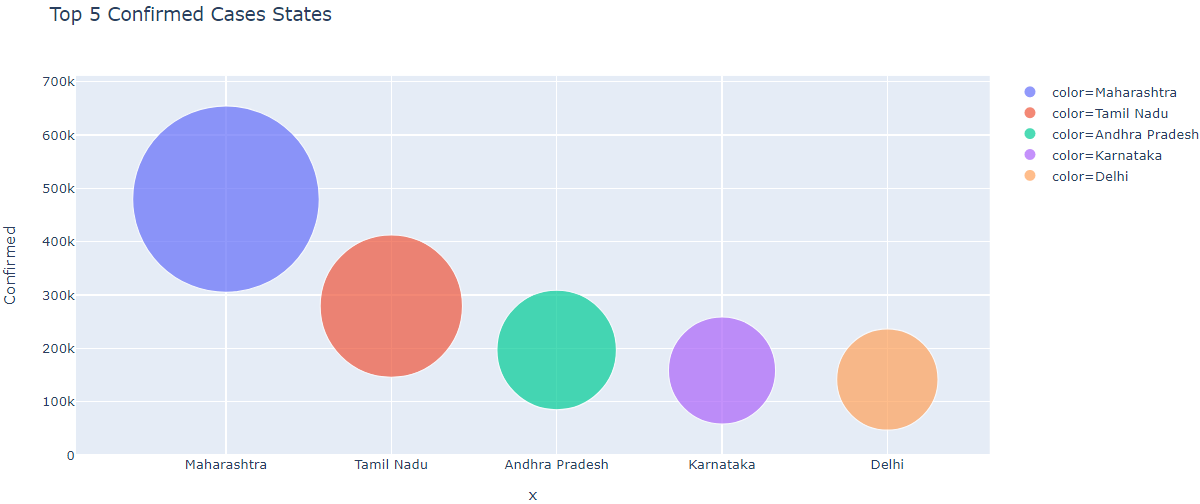
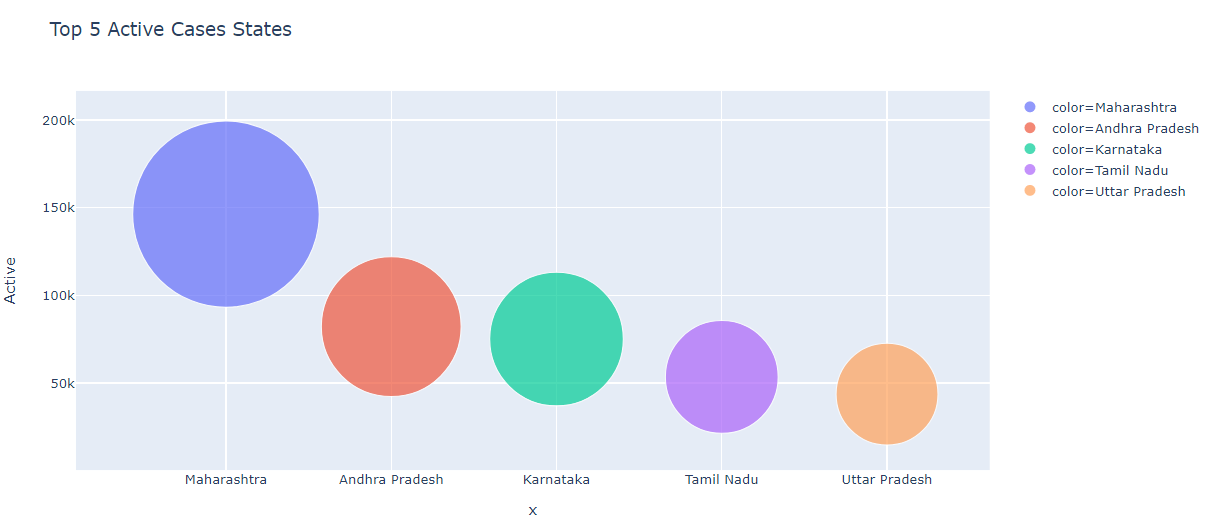
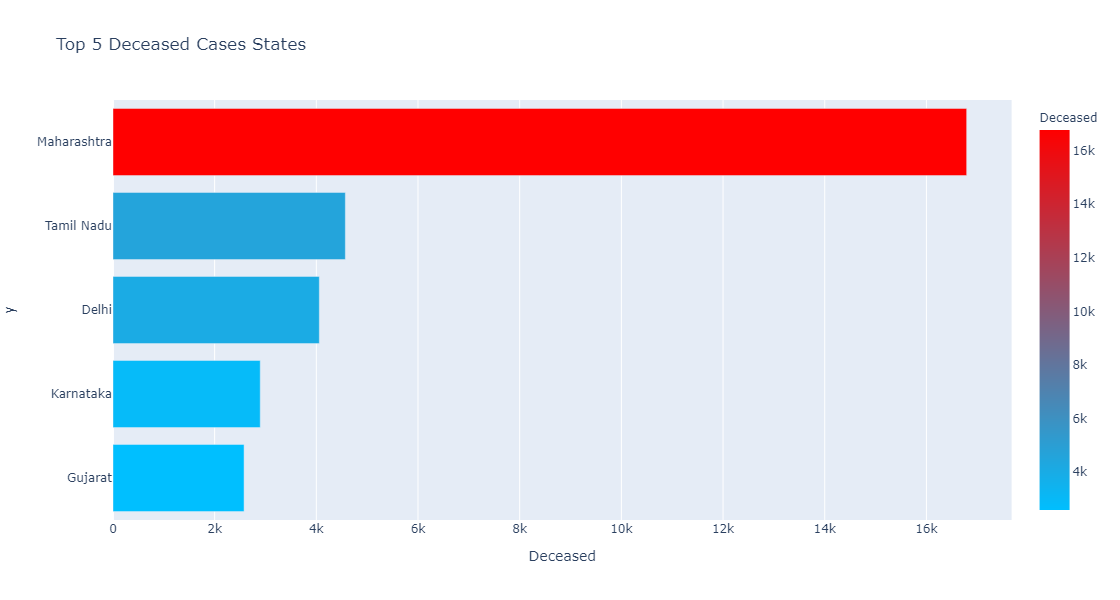
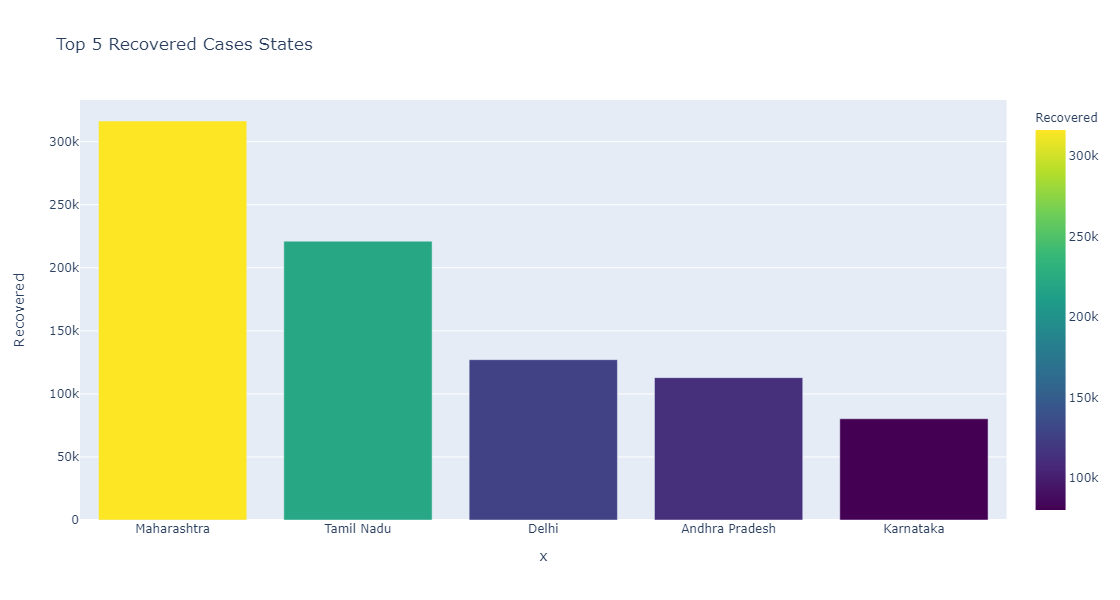
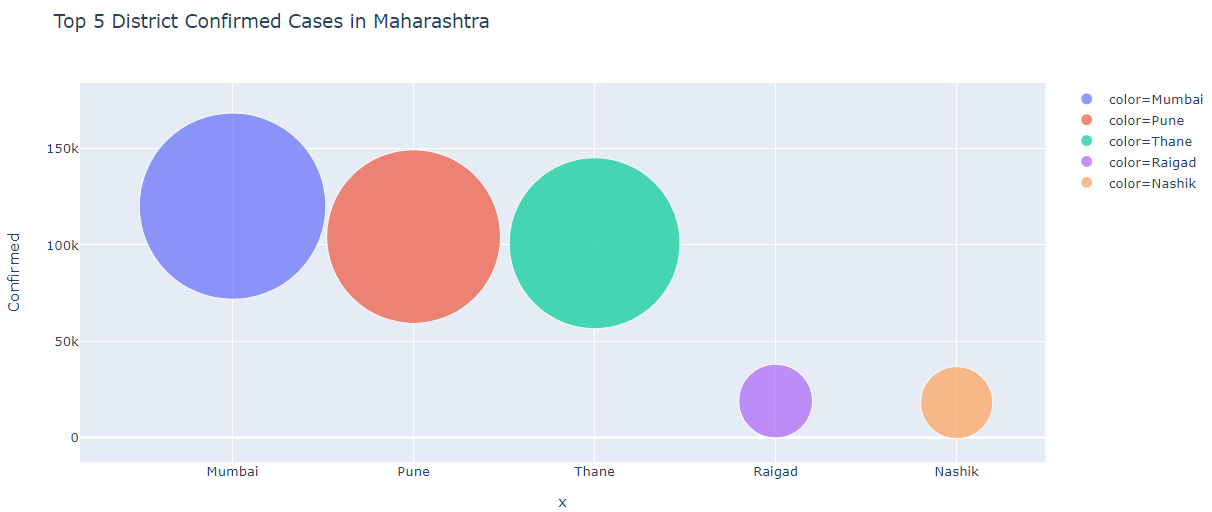
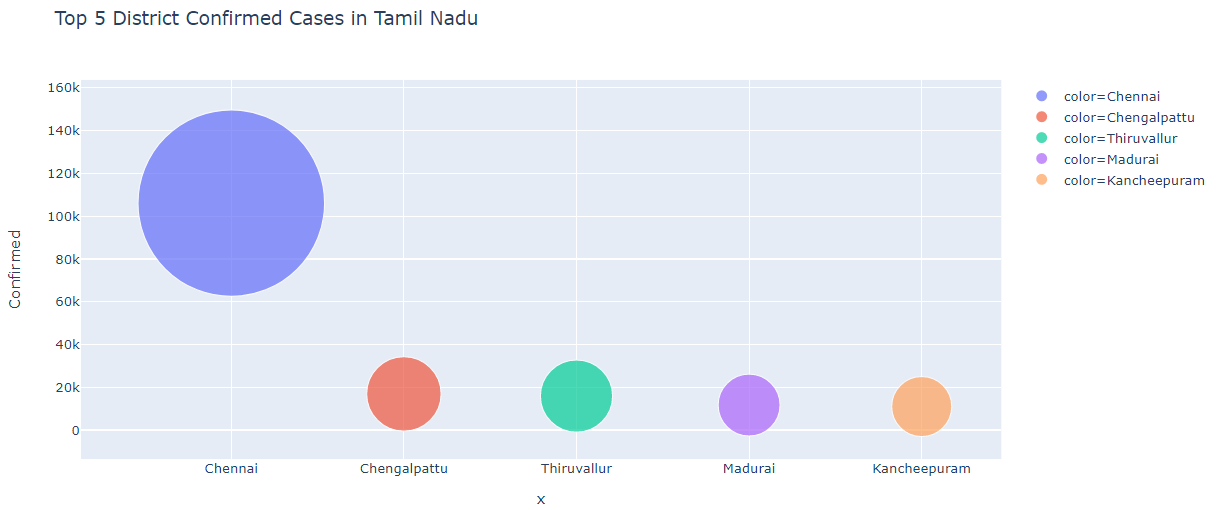
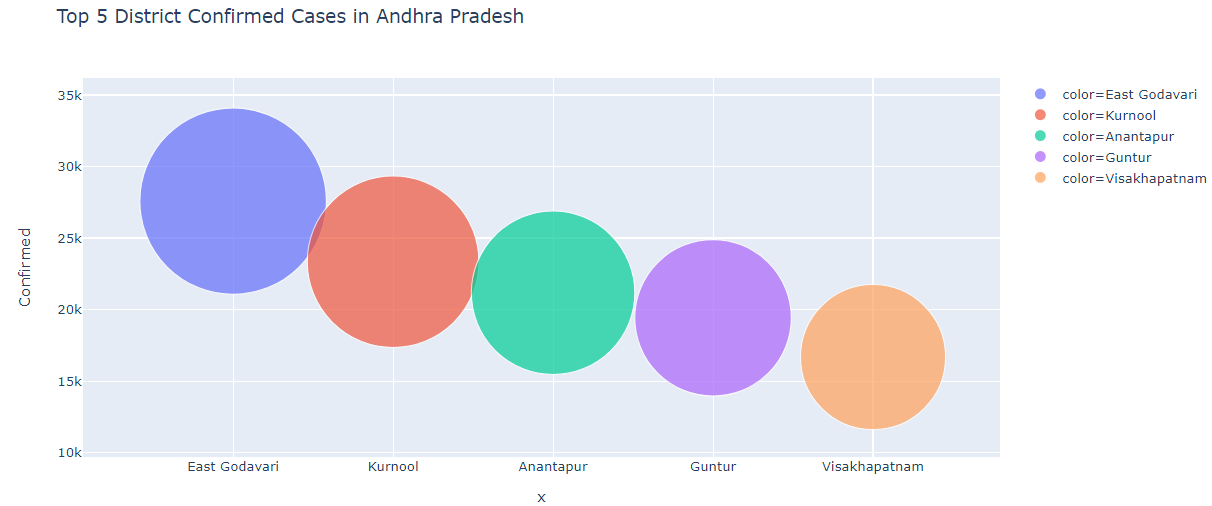
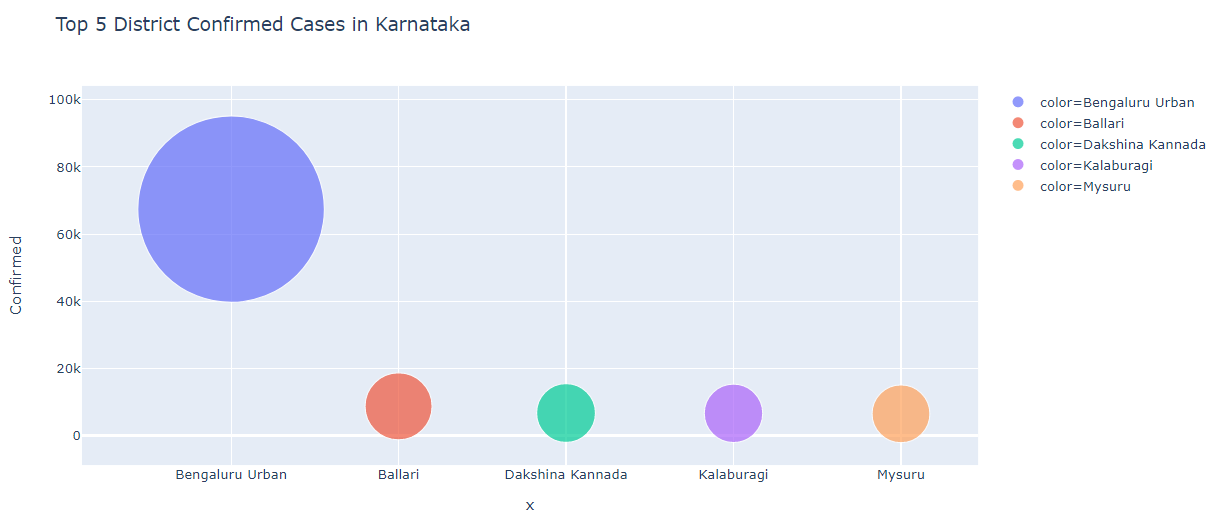
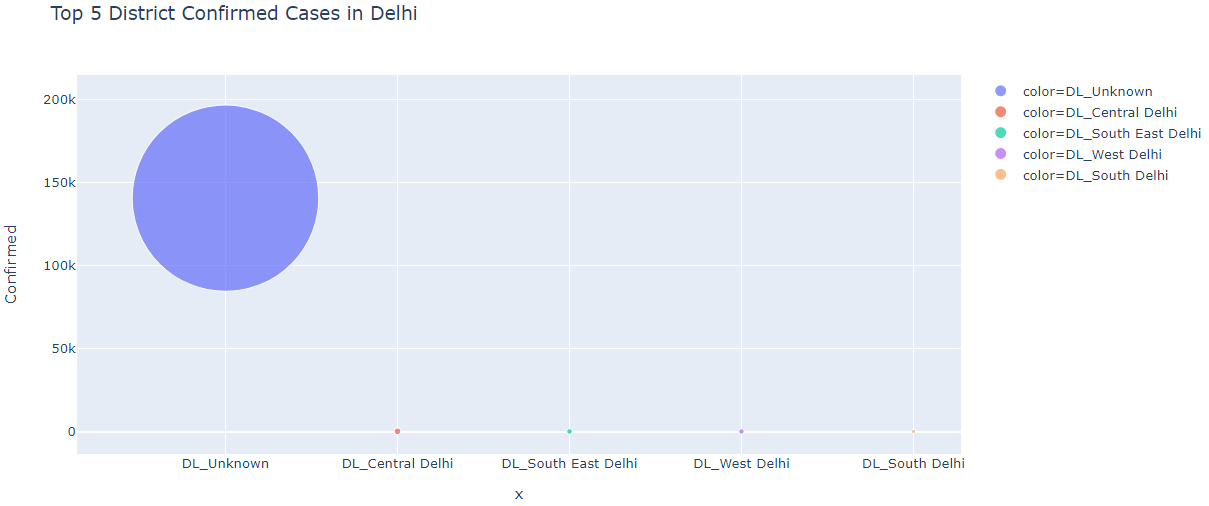
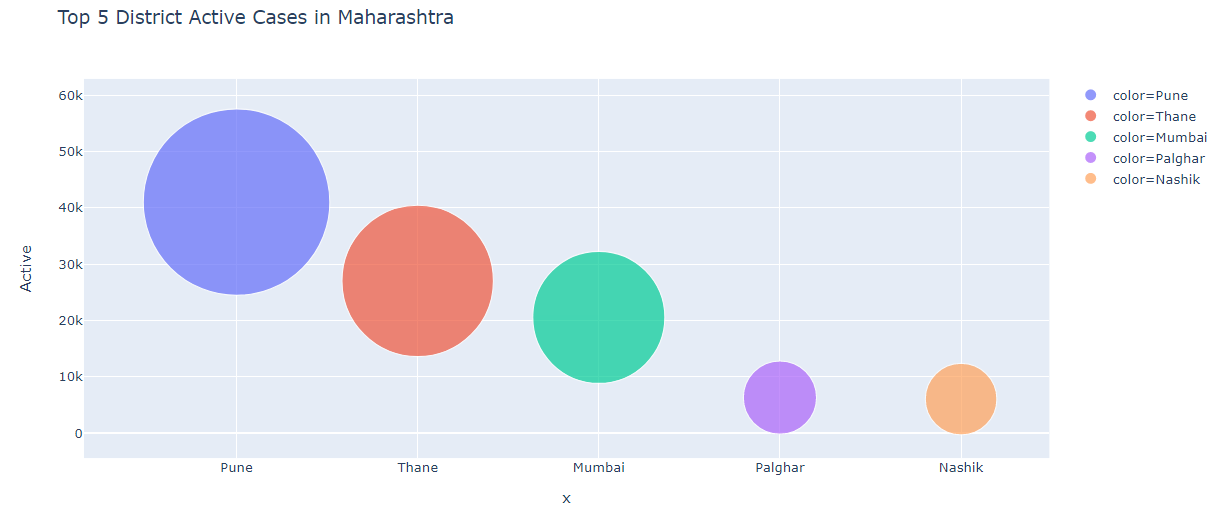
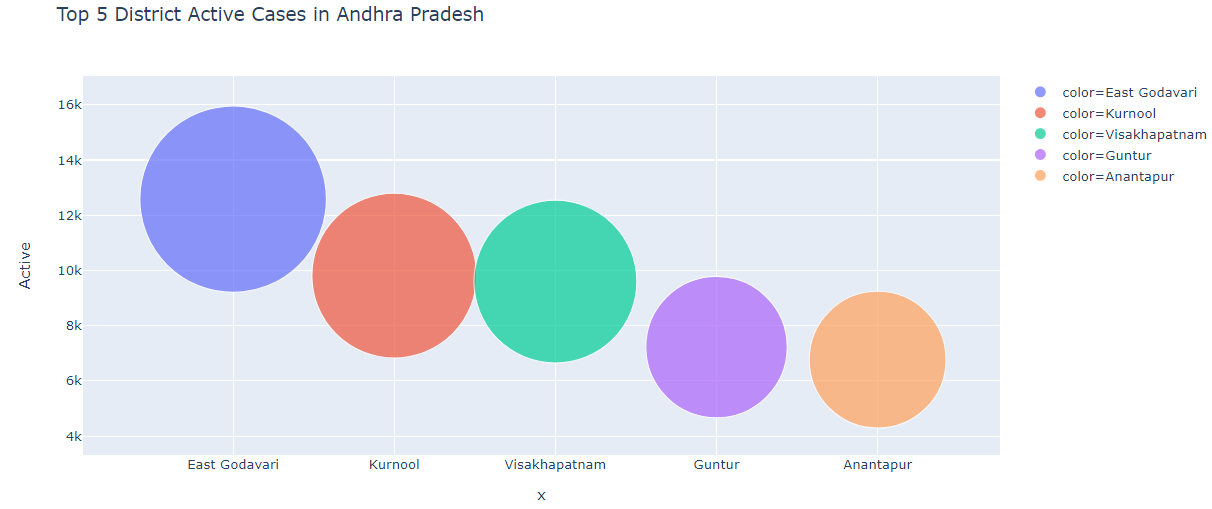
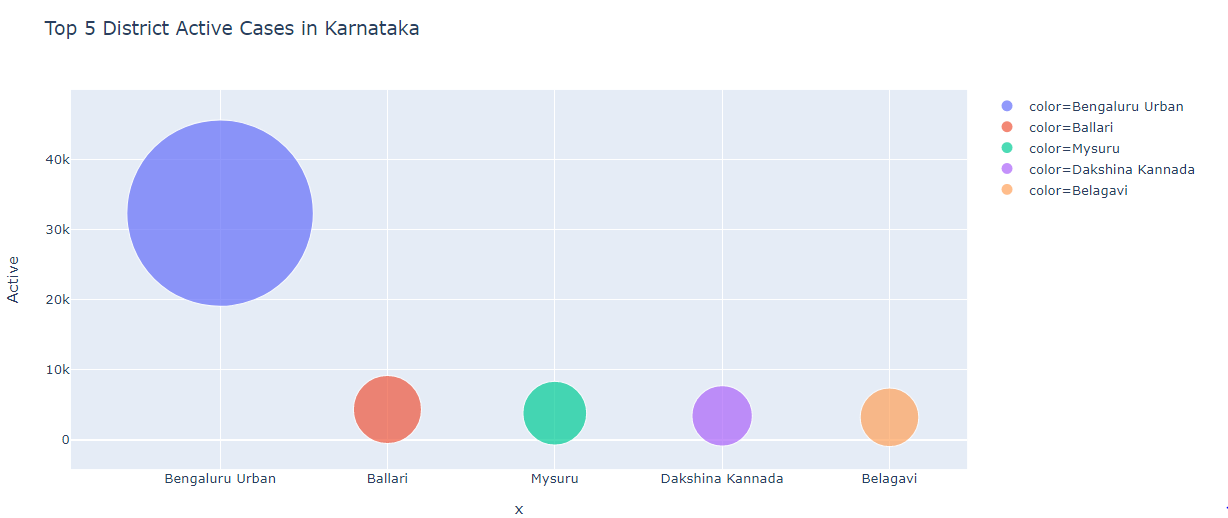
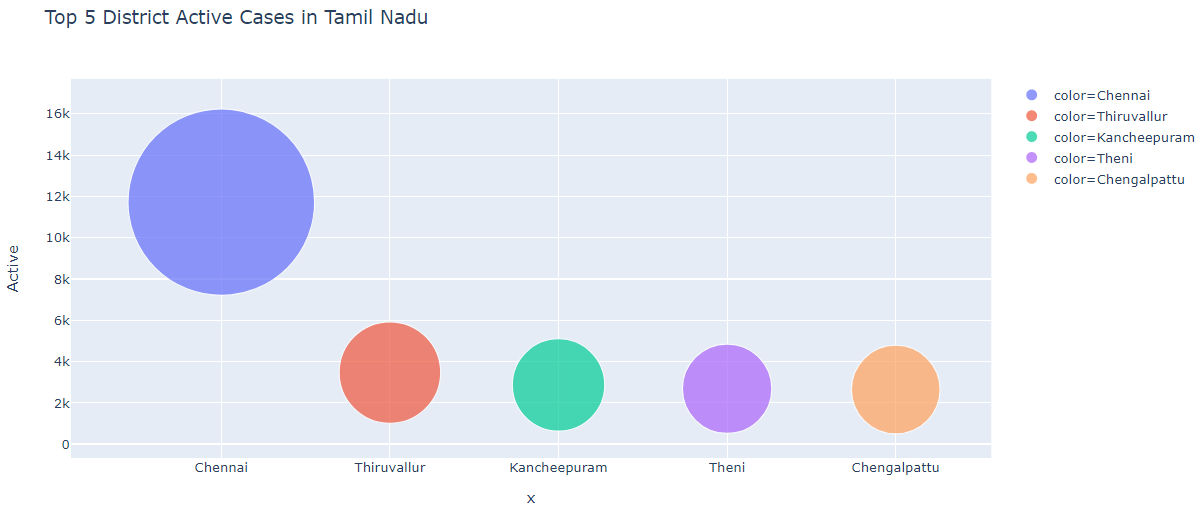
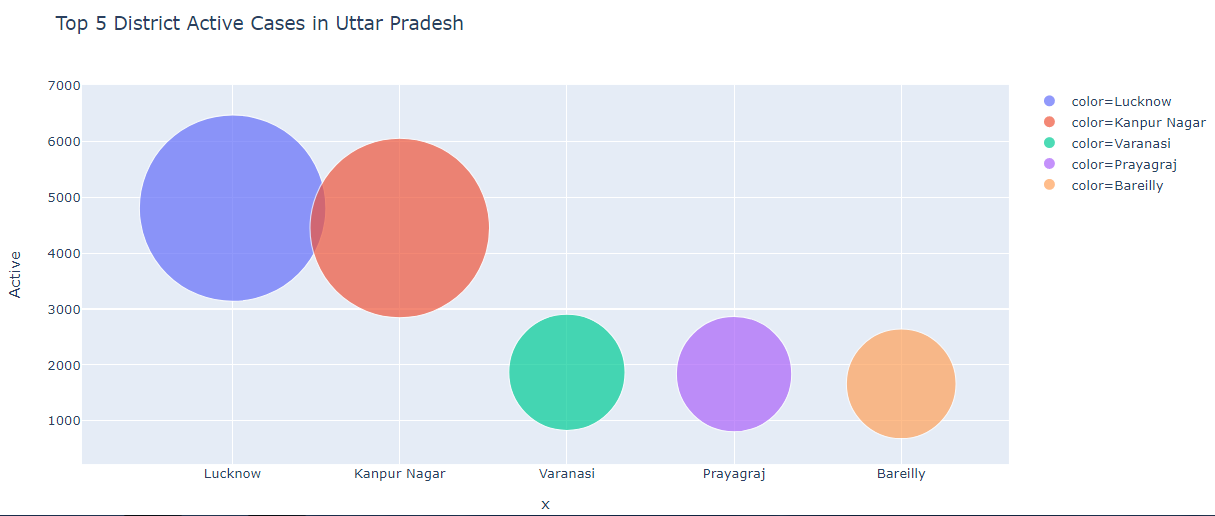
2. Below graph shows active cases of districts in Andhra Pradesh

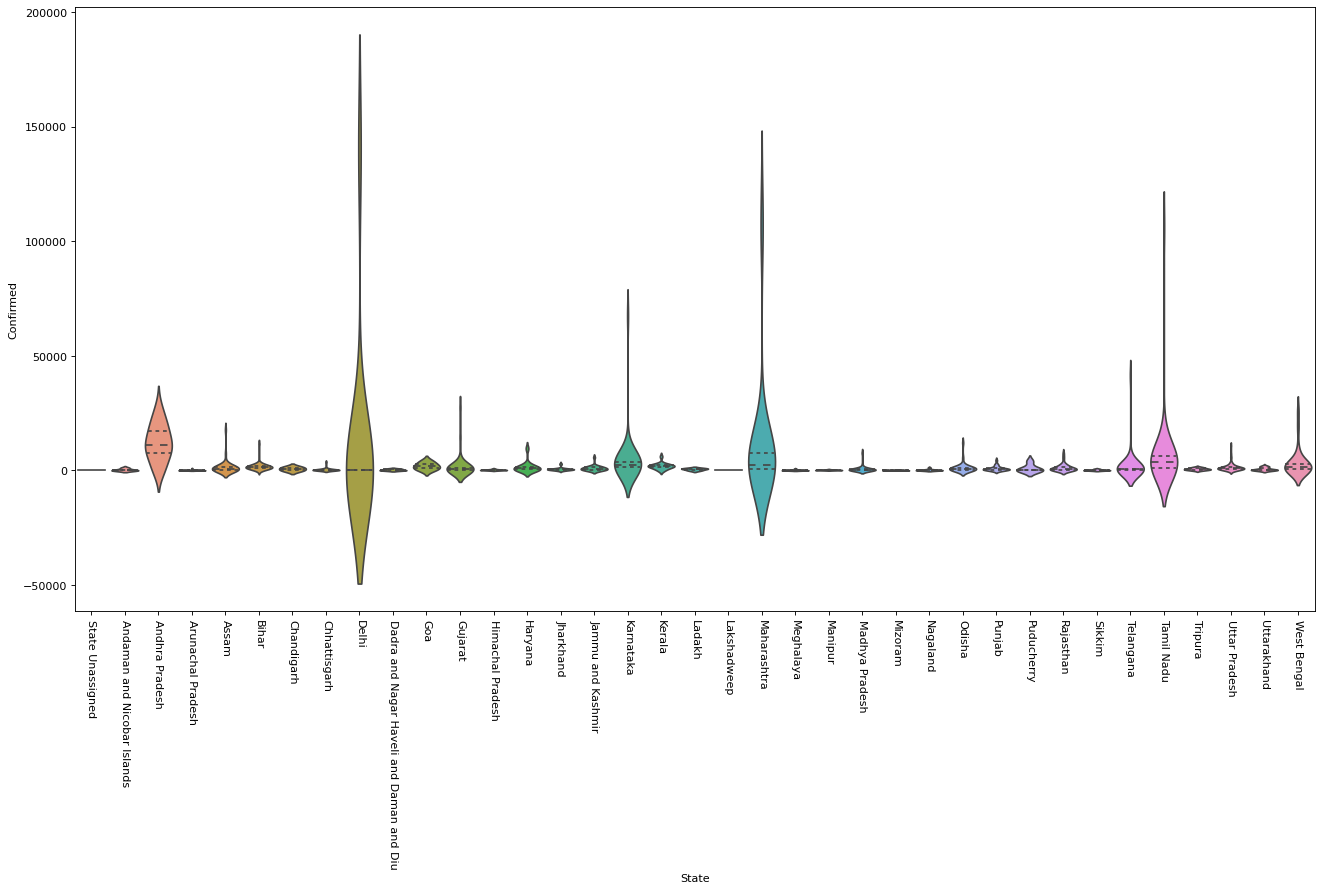
1. The lower quartile represents 25% of the dataset. For confirmed cases, 25% of the data is less than 7,500. For active cases, 25% of the data is less than 2,500. For recovered cases, 25% of the data is less than 4,000. The upper quartile represents 75% of the dataset. For confirmed cases, 75% of the data is less than 17,500. For active cases, 75% of the data is less than 6,000. For recovered cases, 75% of the data is less than 13,000. The green line represents the median of the data. The blue box represents the inter-quartile range. This dataset has no outliers. Confirmed cases, active cases box-plots are positively skewed, recovered cases is slightly negatively skewed.
2. Correlation coefficients;Below plot shows fairly strong positive relationship between confirmed and recovered cases



1. Correlation coefficients; Below plot shows moderate positive relationship between active and recovered cases
2. **HEATMAP**; each square shows the correlation between the variables on each axis. Correlation ranges from -1 to +1. Values closer to 0 means there is no linear trend between the two variables. Values closer to 1 means the variables are more positively correlated i.e., as one variable increases the other variable also increases;closer to 1, stronger the relationship is. Values closer to -1 means when one variable is increasing the other decreases. The diagonals are all 1(dark), because the variables are correlating to themselves. Larger the value , darker the colour ,higher the correlation between the two variables. This plot is symmetrical.
3. Below scatter plot tells us that the confirmed cases are scattered among the districts of Andhra Pradesh
4. The values are more spread out than supposed from the 2nd quantile

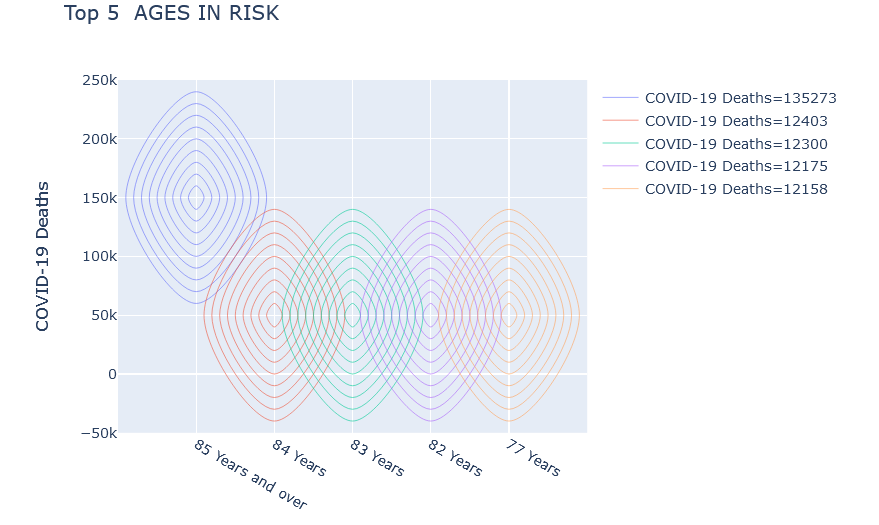


1. This plot shows the top 5 districts in overall India regarding confirmed cases
2. This plot shows the top 5 districts in overall India regarding active cases
3. This plot shows the top 5 states in India regarding confirmed cases
4. This plot shows the top 5 states in India regarding active cases
5. This plot shows the top 5 states in India regarding deceased cases
6. This plot shows the top 5 states in India regarding recovered cases
7. This plot shows the top 5 districts in Maharashtra regarding confirmed cases
8. This plot shows the top 5 districts in Tamil Nadu regarding confirmed cases
9. This plot shows the top 5 districts in Andhra Pradesh regarding confirmed cases
10. This plot shows the top 5 districts in Karnataka regarding confirmed cases
11. This plot shows the top 5 districts in Delhi regarding confirmed cases
12. This plot shows the top 5 districts in Maharashtra regarding active cases
13. This plot shows the top 5 districts in Andhra Pradesh regarding active cases
14. This plot shows the top 5 districts in Karnataka regarding active cases
15. This plot shows the top 5 districts in Tamil Nadu regarding active cases
16. This plot shows the top 5 districts in Uttar Pradesh regarding active cases
17. The shape of the distribution(extremely skinny at the each end and wide at the middle) indicates that the confirmed cases are highly concentrated at the median.

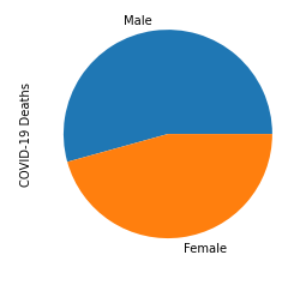


5)Age Dataset:

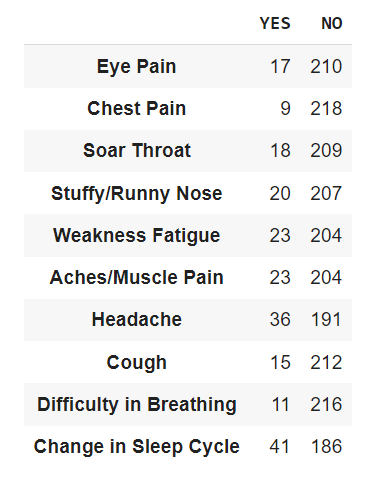
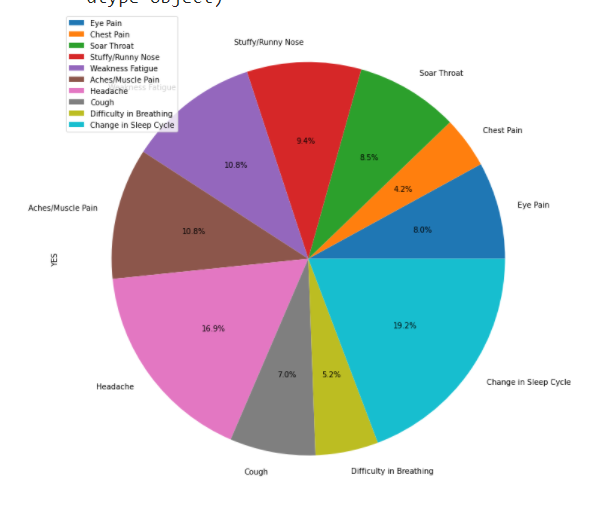
1. The top five highest number of deaths occured in the age groups-

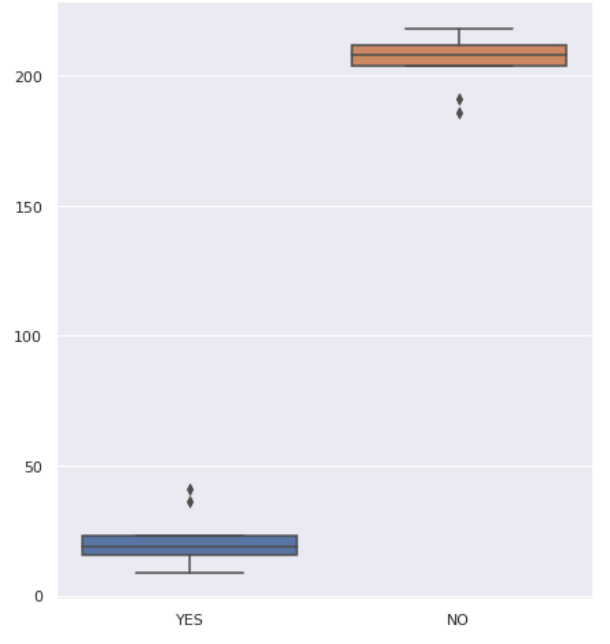


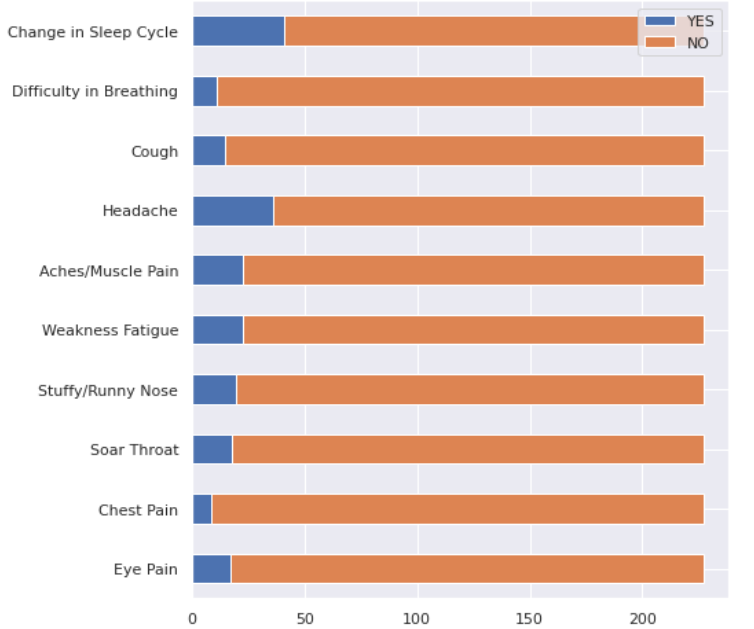
2.above plot shows that which gender has high COVID-19 deaths



**6) *Symptoms Dataset:***

1. Basic frequency table of counts of symptoms.
2. Pie chart showing the percentage of population having a specific symptom. From the pie, we can infer that change in sleep cycle is the most observed symptom.
3. Boxplot of yes/no of overall symptoms.



1. Stacked bar graph of count of yes/no of individual symptoms.

====================================================================================================

1.Covid --age vs deaths(bar chart)

2. Gender composition-pie

3. Scorecard--Confirmed,Deaths, Recovered(India wise)

4. Future Dates---numbers

5. Top 5 states--graph

6. Top 5 districts--graph

7. Symptoms--list

8.Future dates(ARIMA)