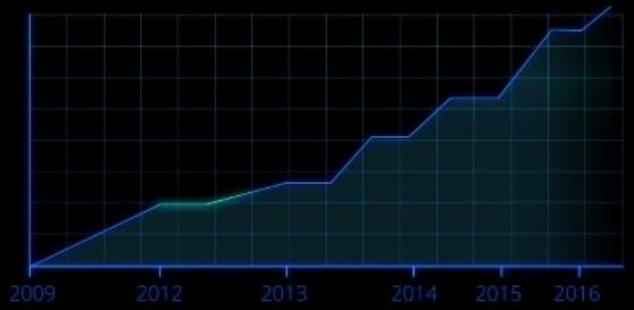
DA Project Group 11



Covid 19 Analysis through visualizations and Hypothesis Testing

Group Members

SR. NO.	NAME	ROLL NUMBER
1.	Khushi Yagnik	190030022
2.	Atharva Swami	190010008
3.	Saloni Singh	190010041
4.	Rishika Patel	190030037
5.	Adarsh Mall	190010003
6.	Dalpat Choudhary	190030010

Objective

The objective of this project was to visualize the effect of COVID-19 worldwide through various plots and to make hypothesis and prove them using various statistical tests.

Procedure used for the project:

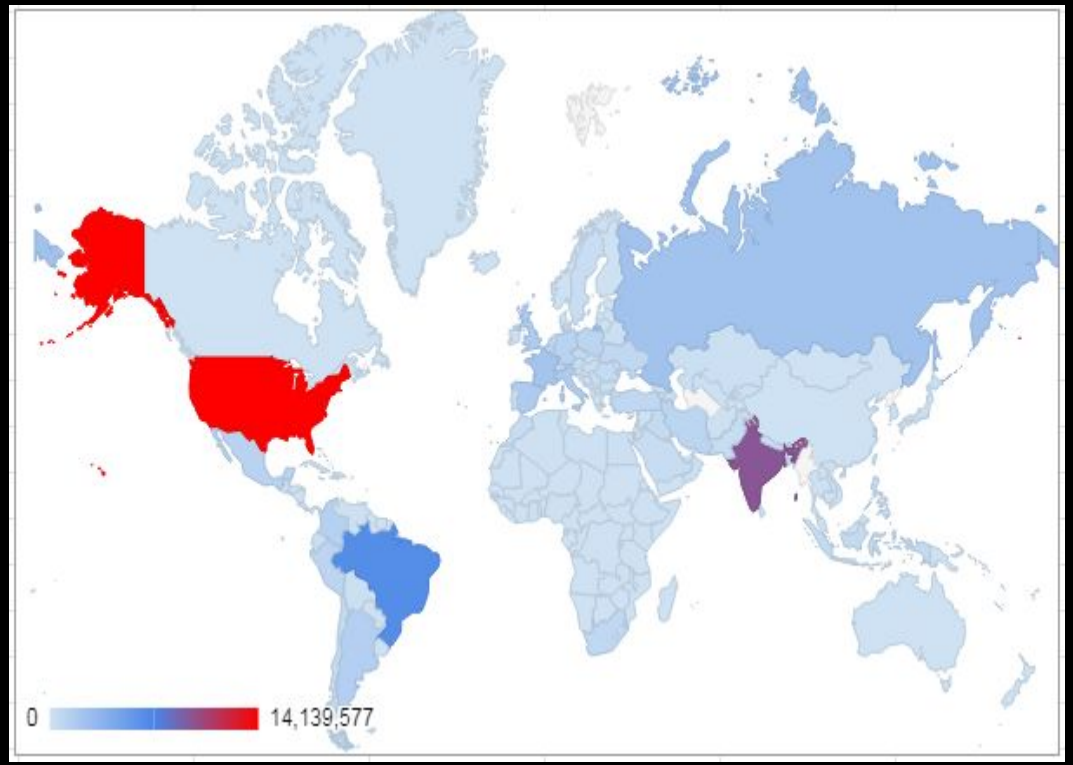
1. Collection of data from various sources
2. Plotting of appropriate graphs to visualize the effects of covid-19
3. Interpreting the graphs by making hypothesis
4. Testing the hypothesis using various statistical methods
5. Concluding the results from the project

COVID 19 in World

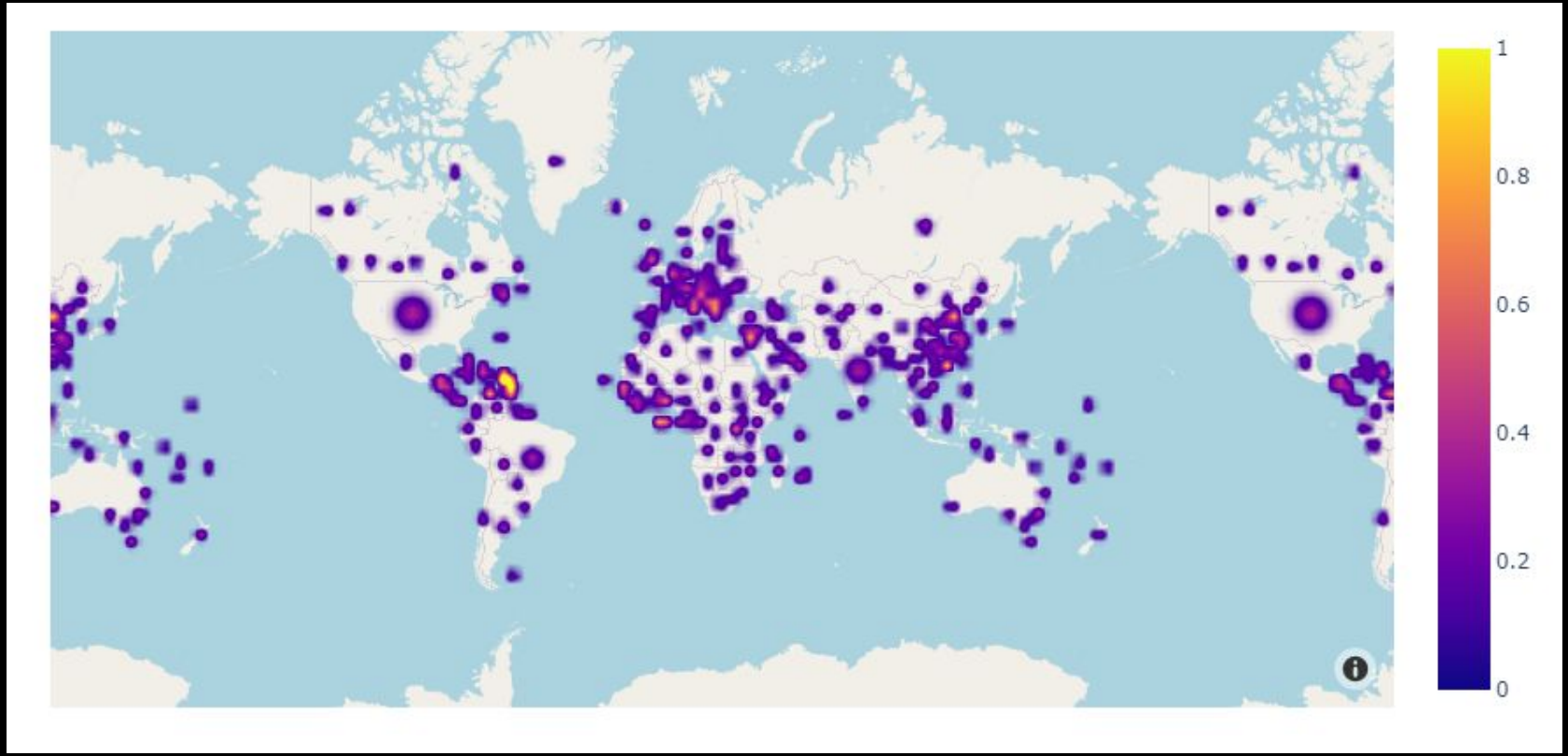
Some plots to show the various trends of the Covid 19 Virus in the world.All graphs were made either using python or excel.

World heat map

From the heat map we can conclude that in the World, US had the highest number of covid 19 cases. Then comes India and Brazil.



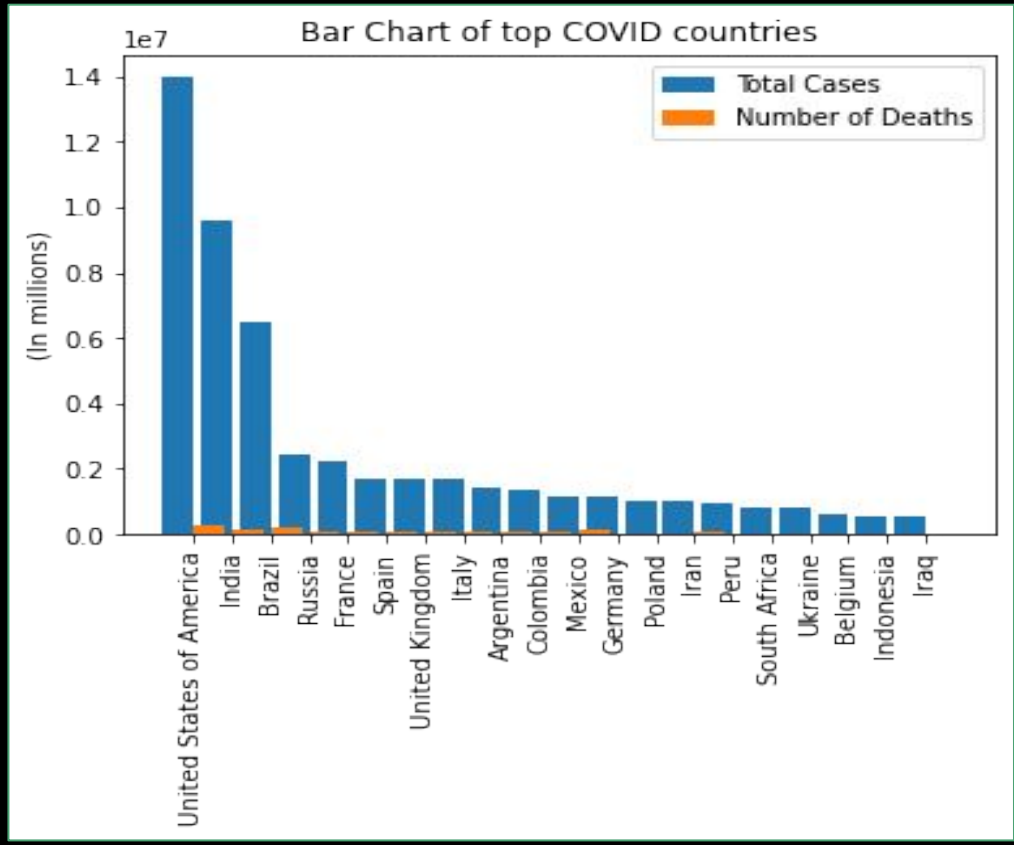
World heat map



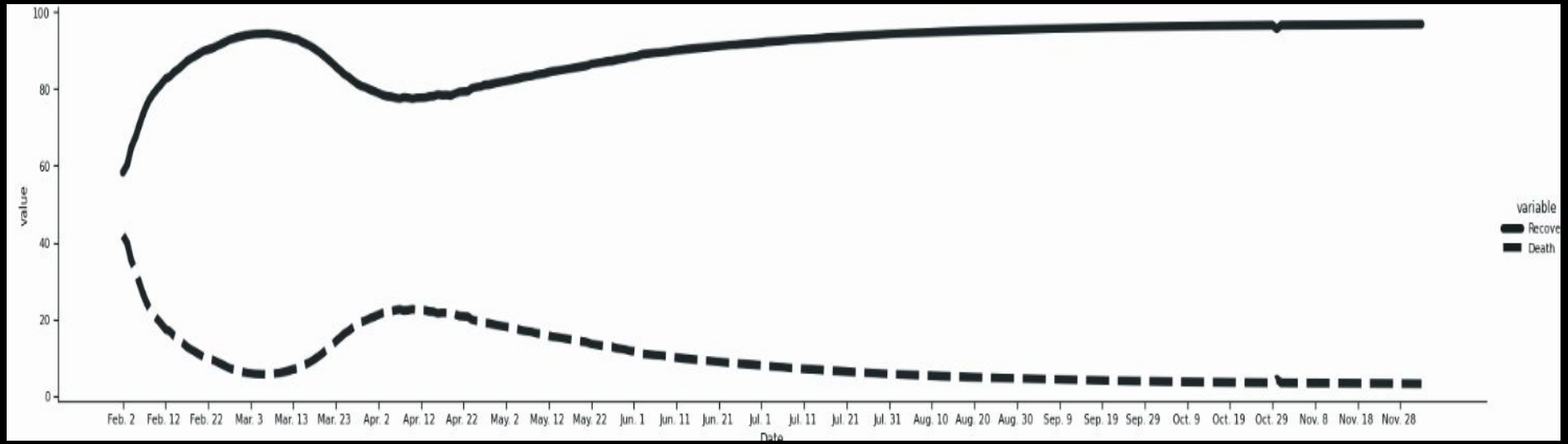
Bar Graph - top 20(cases)

Conclusion :

- USA has the highest no. of cases.
- Death rate is very low for all the countries.
- Excluding the top 3 countries, all the countries have less than 0.5 million cases.



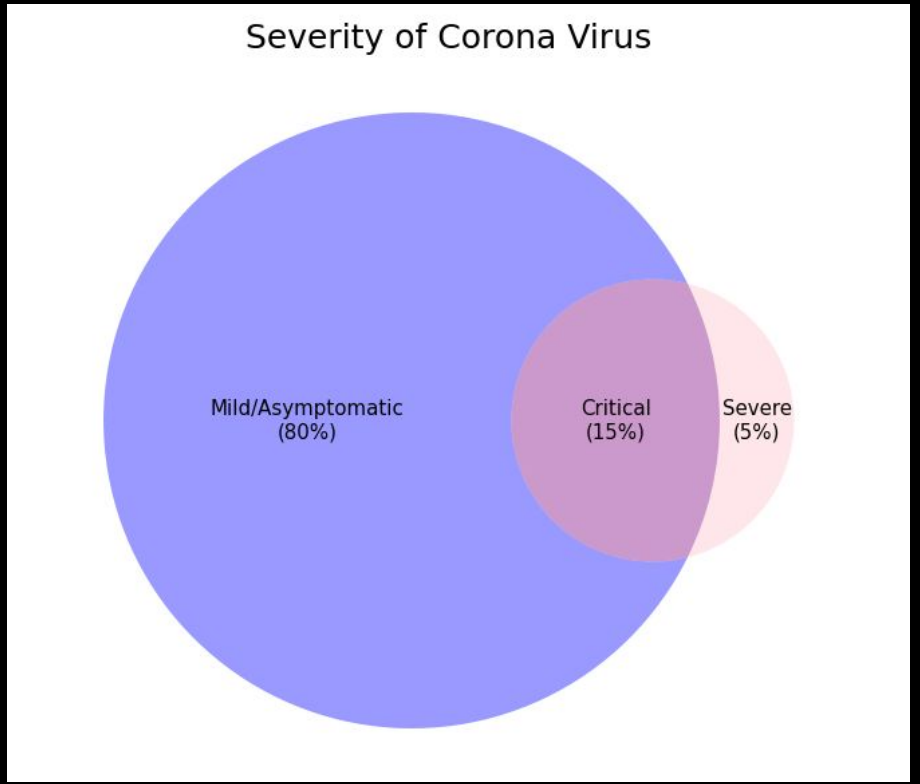
Recovery Rate vs Death Rate : Line graph



From the graph we can see that over time death rate has decreased and recovery rate has increased

Severity of covid 19 : Venn Diagram

According to a generic figure given by WHO, around 80 % of cases were mild/asymptomatic, 15% were critical and needed to be hospitalized and 5% cases were severe where people needed ventilators.

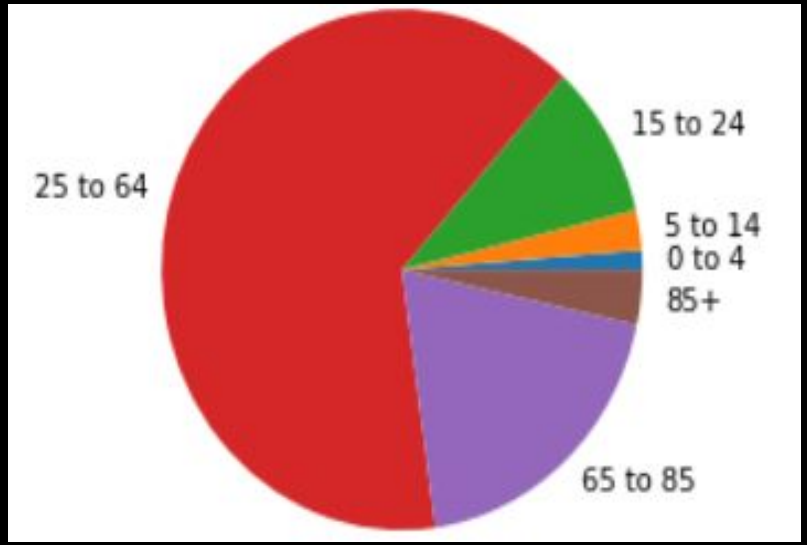


Pie chart age vs covid

CONCLUSIONS:

1. The deadly virus infects people between 25 to 64 the most worldwide.

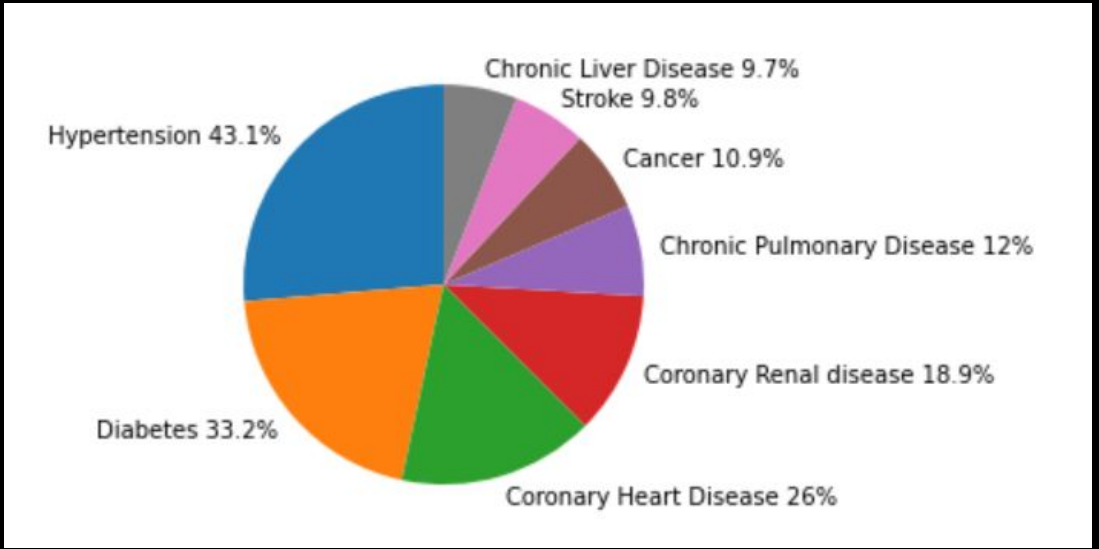
2. Children between 0 to 4 are least affected however preventive measures should always be taken.



pie chart-disease wise

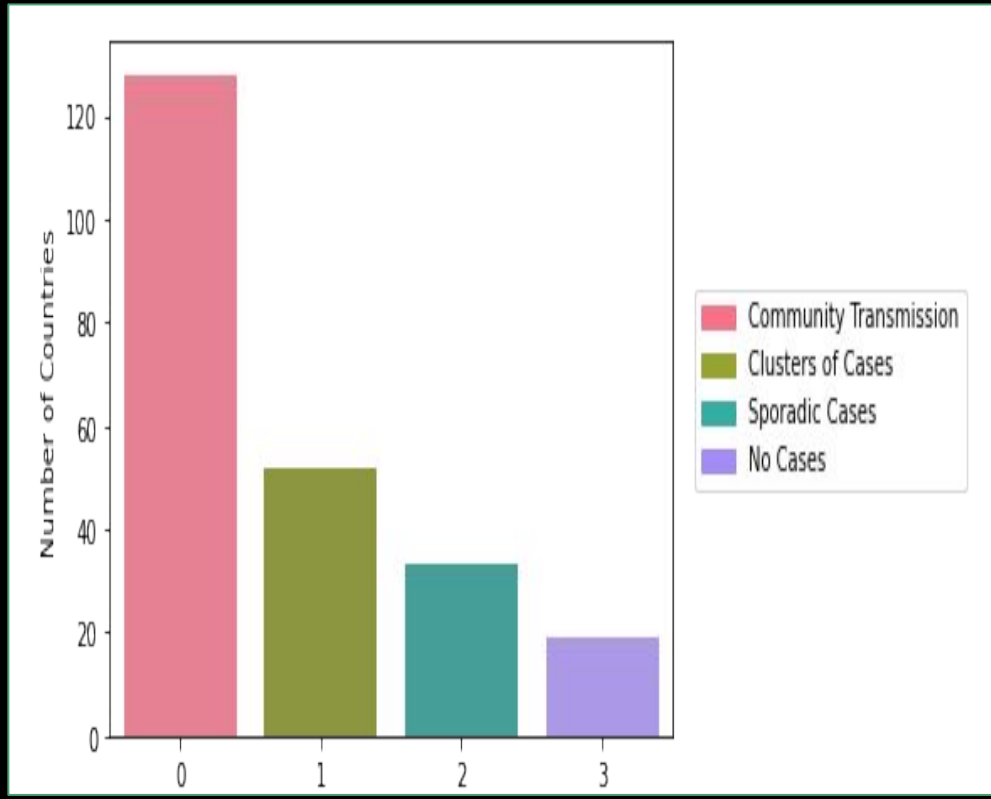
Conclusion:

1. Majority of deaths are caused in those with hypertension.
2. Least number of deaths are in those with chronic liver disease.

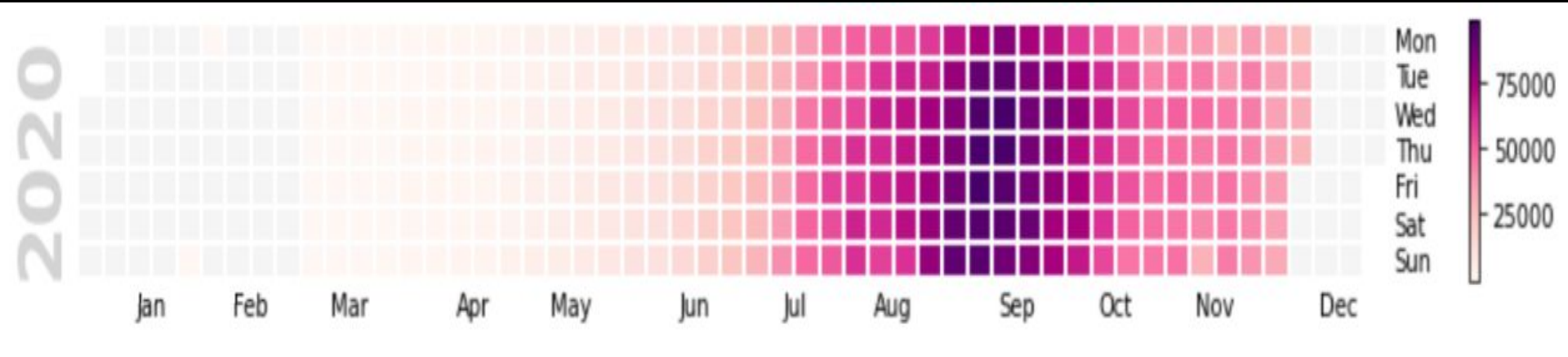


Bar plot for Method of transmission

Conclusion : The most common method of transmission is community transmission, and the least common is sporadic cases.



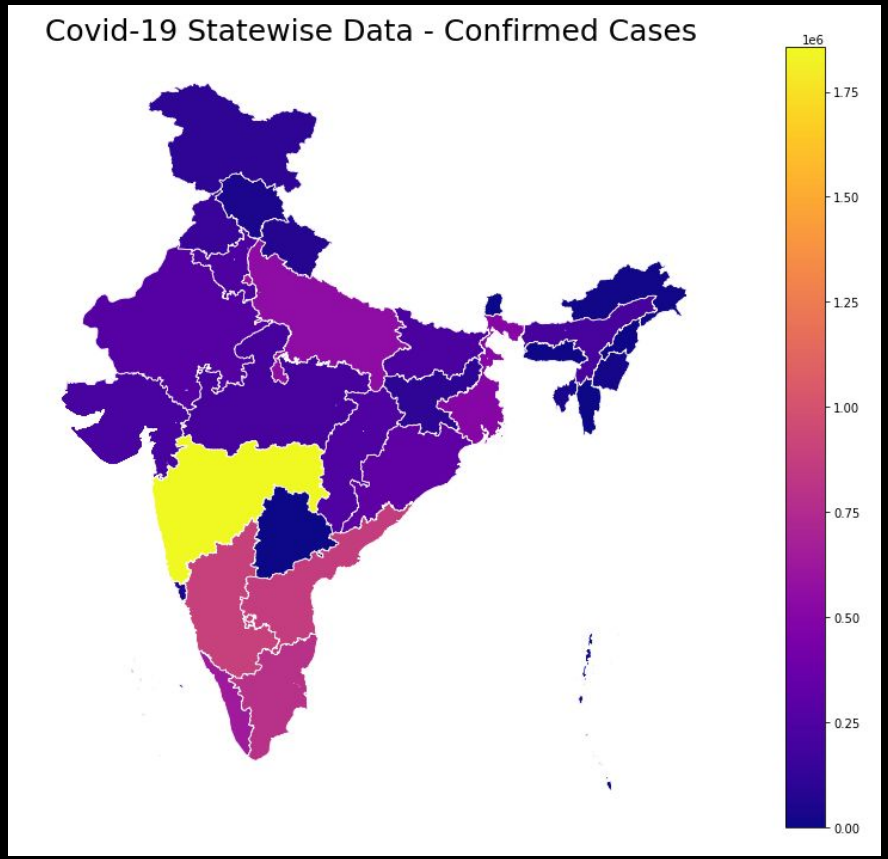
Daily cases India : Calendar plot



We can see from this calendar plot the rise of covid 19 in the year 2020. The intensity of colour is directly proportional to the number of cases. The maximum number of cases were reported in Aug-Sep.

Cases in India State wise : Heat map

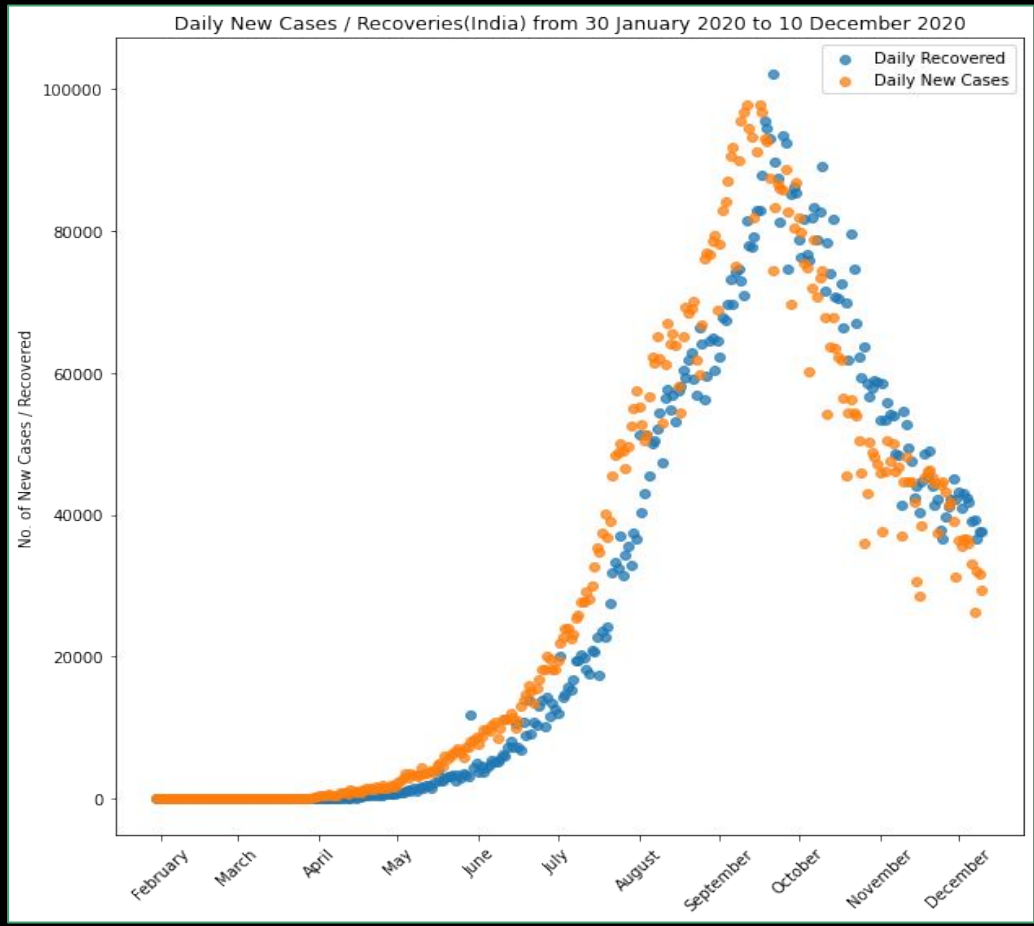
From the heat map we can conclude that in India, Maharashtra had the highest number of covid 19 cases.



Scatter Plot for Daily Cases and Recoveries(India)

Conclusion :

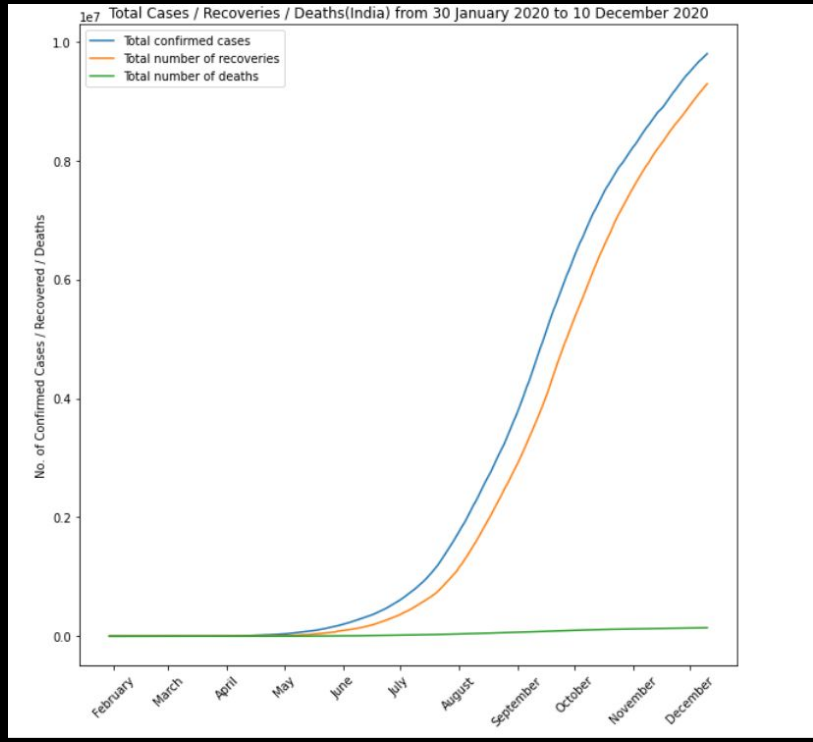
- Before October, most of the days, no. of new cases were greater than no. of daily recovered.
- Since October, most of the days, no. of daily recovered became greater than no. of new cases.



Line Graph for total cases vs recovery vs deaths in india

Conclusion:

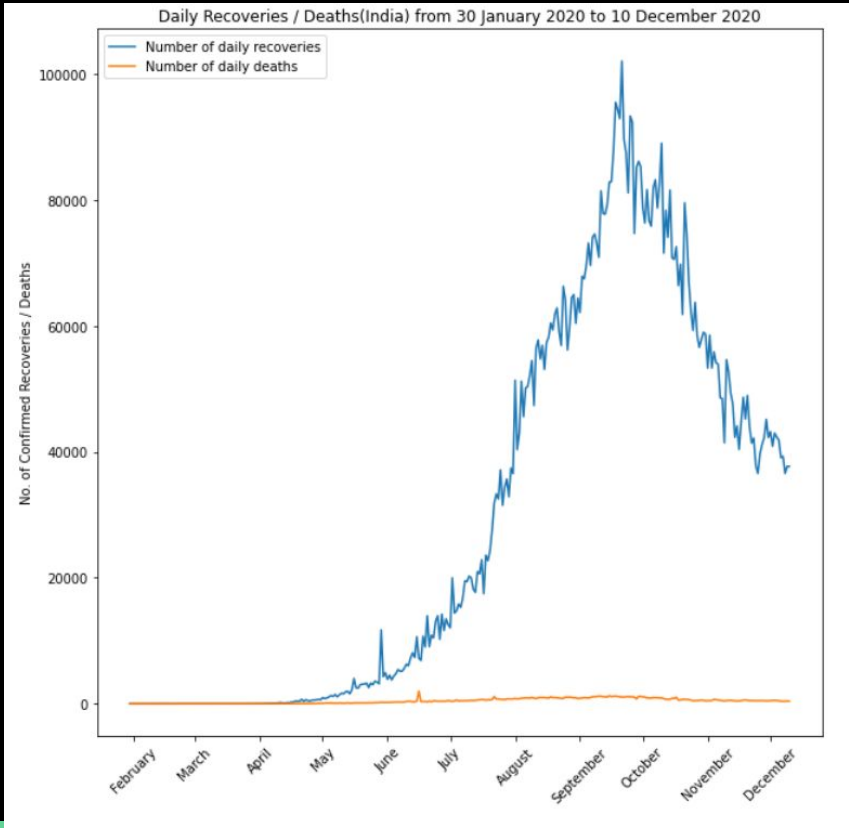
1. The total number of deaths are very small
2. The recovery number follows the same pattern as total number with a small offset.
3. The total number of cases follow normal distribution.



Line-graph of daily recoveries vs deaths

Conclusion:

1. The daily deaths are a small number as compared to number of recoveries.
2. The recovery pattern follows normal distribution.

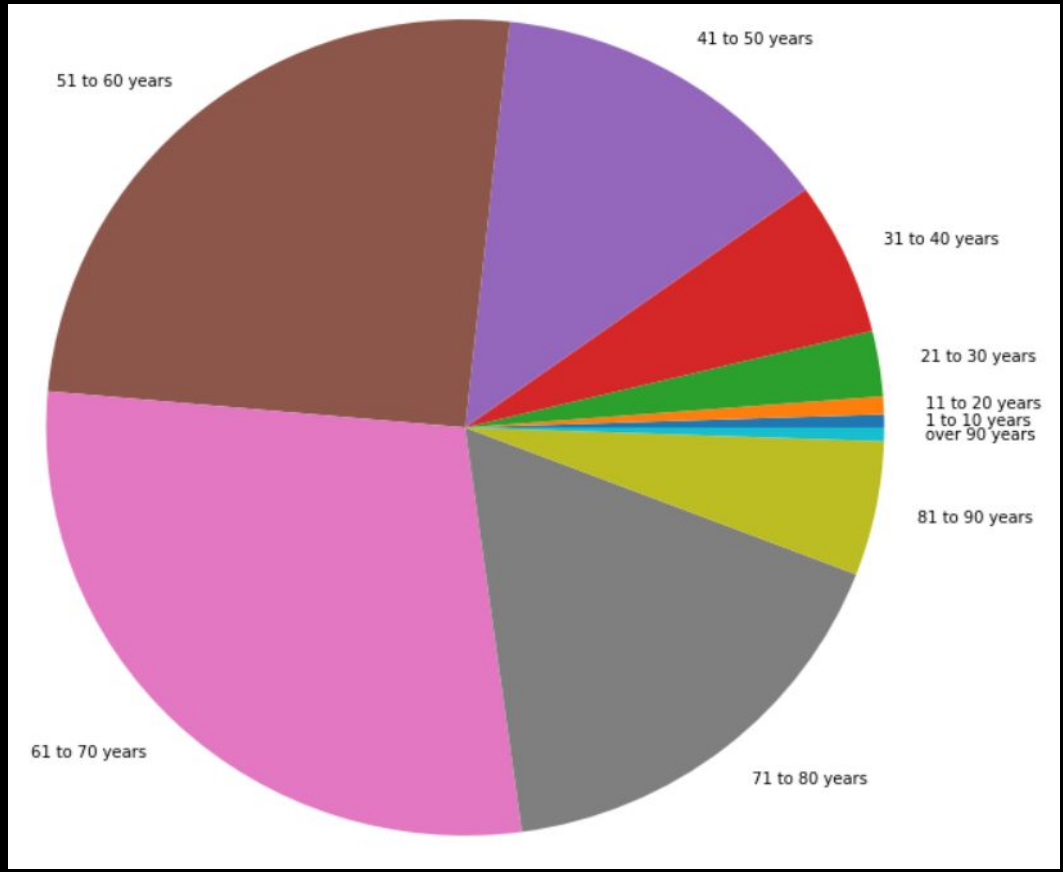


PIE CHART INDIA(AGE WISE)

CONCLUSIONS:

1. Corona virus affects the people of age 61-70 the most. Elder people should take utmost care by staying at homes.

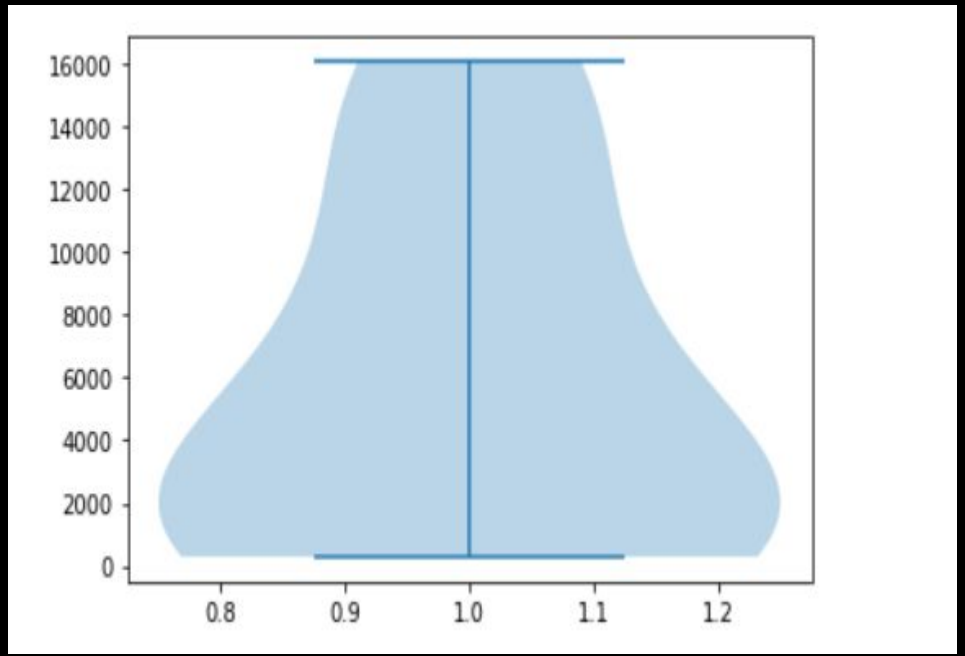
2. The virus least affects small children from 1-10 years however preventive measures should always be taken.



VIOLIN PLOT FOR INDIA(AGE WISE)

Conclusion:

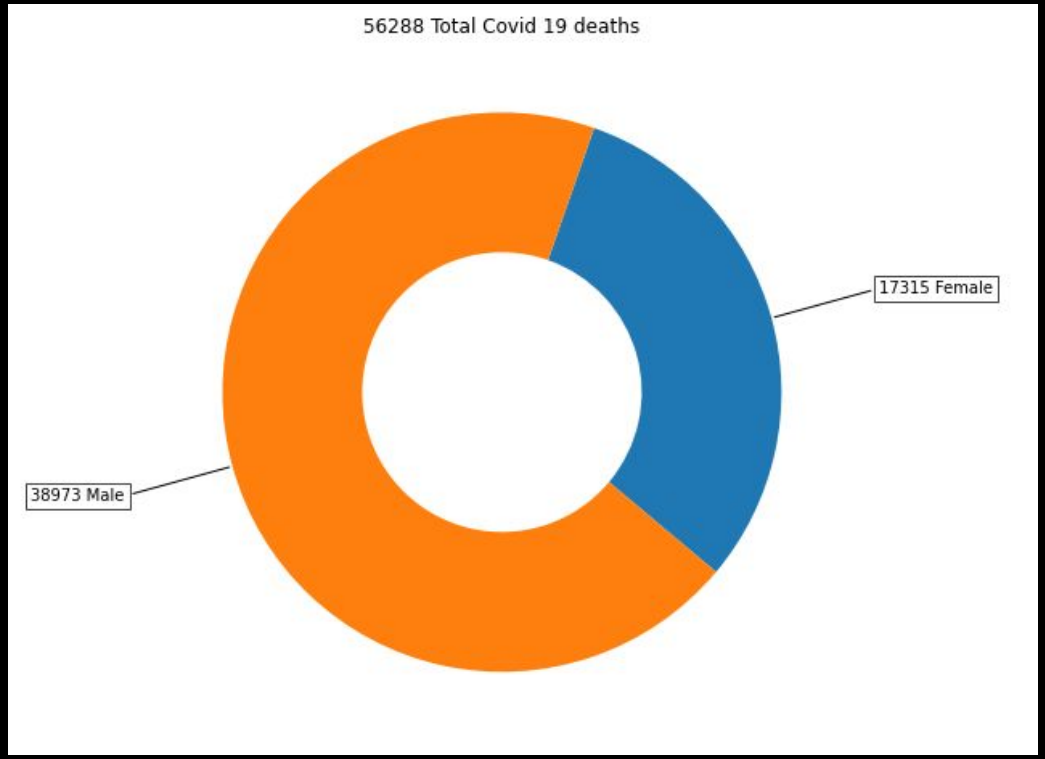
The median number of cases is 2000



Gender vs death rate : Donut chart

In a sample of 56288 people who died due to covid 19, it was found that 38973 were male and 17315 were female.

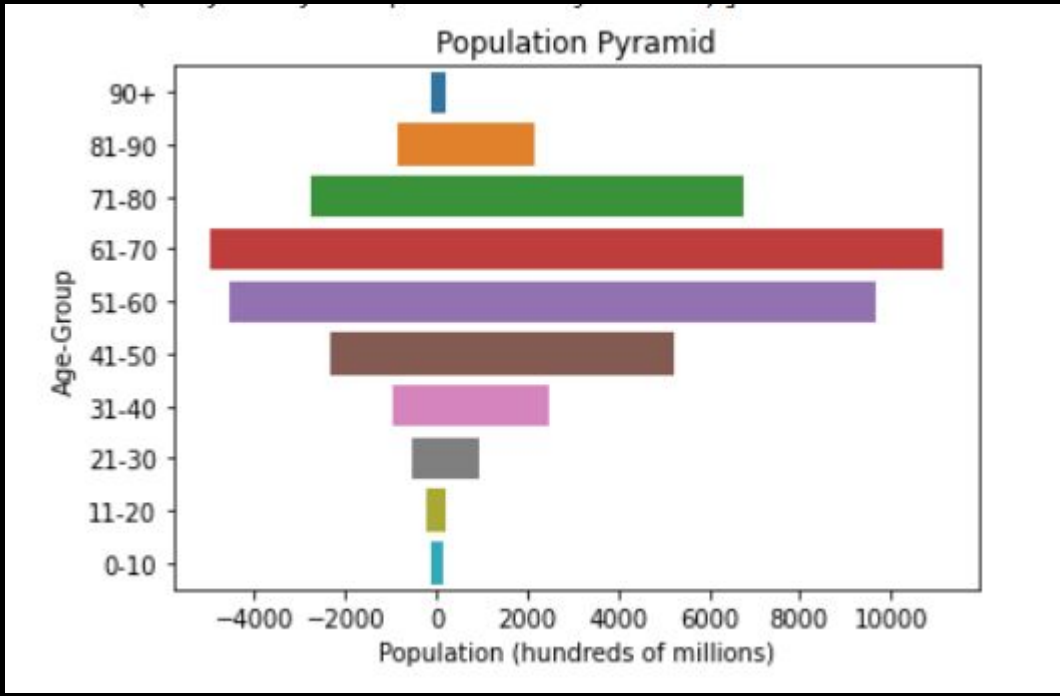
The donut plot clearly shows the difference in death rate of males and females. Death rate of men is almost 2.5% more than women



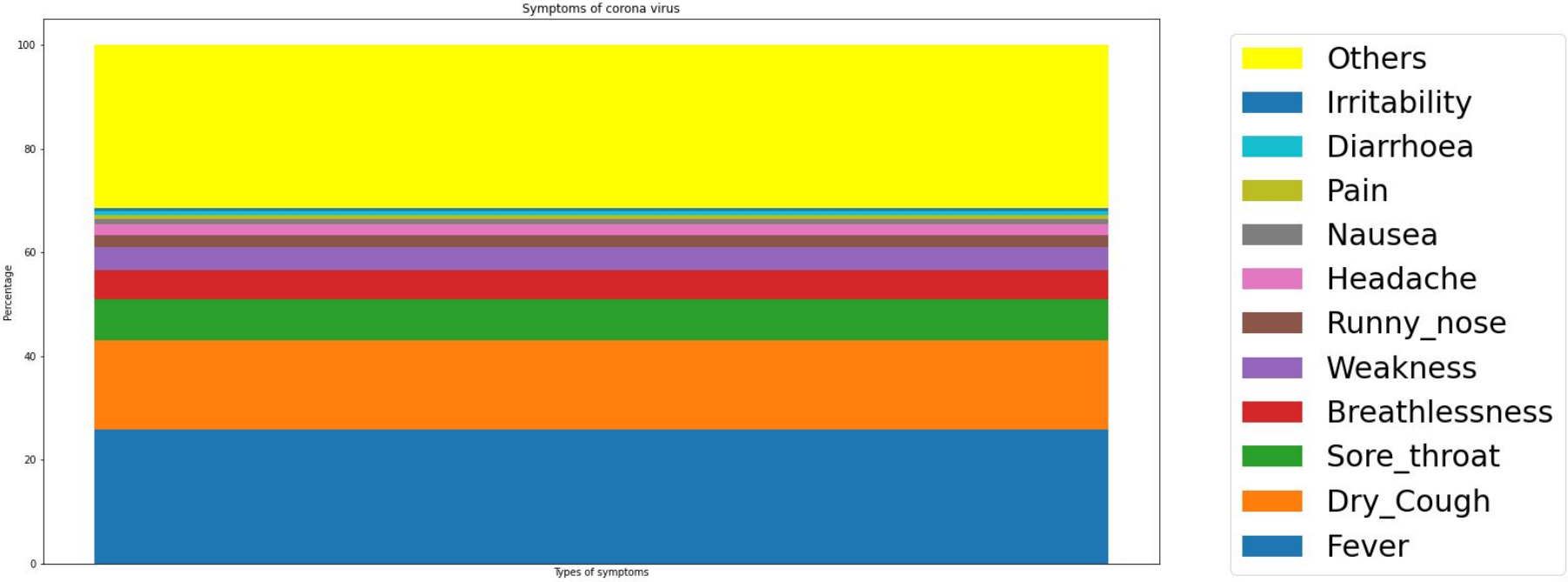
Population pyramid gender vs age

Conclusion:

1. The male deaths are more than female deaths.
2. Most deaths are in the age group 61 - 70.



Symptoms of Covid 19-India : Stacked bar graph

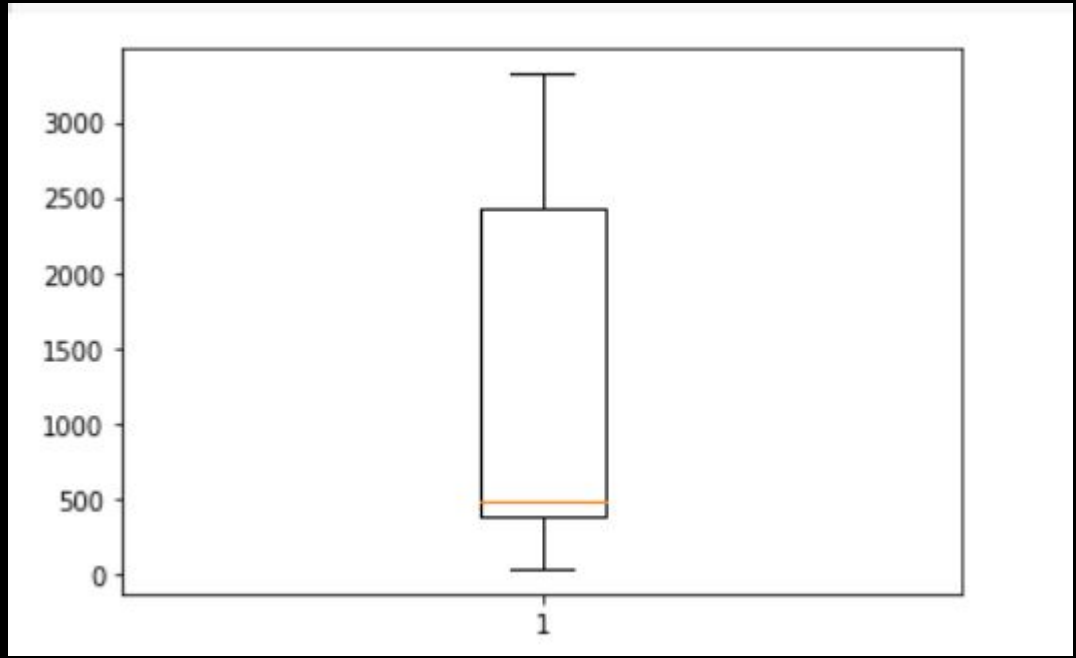


The stacked bar graph shows the percentage of each symptom faced by people.

AGE VS CORONA DEATHS(MUMBAI)

CONCLUSIONS:

1. The median is 490.
2. The first quartile is 300
3. The third quartile is 2500
4. Minimum value is 35
5. Maximum value is 3321
6. There are no outliers



Hypothesis Testing

Hypothesis testing is an act in statistics whereby an analyst tests an assumption regarding a population parameter. The methodology employed by the analyst depends on the nature of the data used and the reason for the analysis.

In hypothesis testing, an analyst tests a statistical sample, with the goal of providing evidence on the plausibility of the null hypothesis.

Statistical analysts test a hypothesis by measuring and examining a random sample of the population being analyzed. All analysts use a random population sample to test two different hypotheses: the null hypothesis and the alternative hypothesis.

The null hypothesis is usually a hypothesis of equality between population parameters; e.g., a null hypothesis may state that the population mean return is equal to zero. The alternative hypothesis is effectively the opposite of a null hypothesis; e.g., the population mean return is not equal to zero. Thus, they are mutually exclusive, and only one can be true. However, one of the two hypotheses will always be true.

Growth of COVID-19 in INDIA

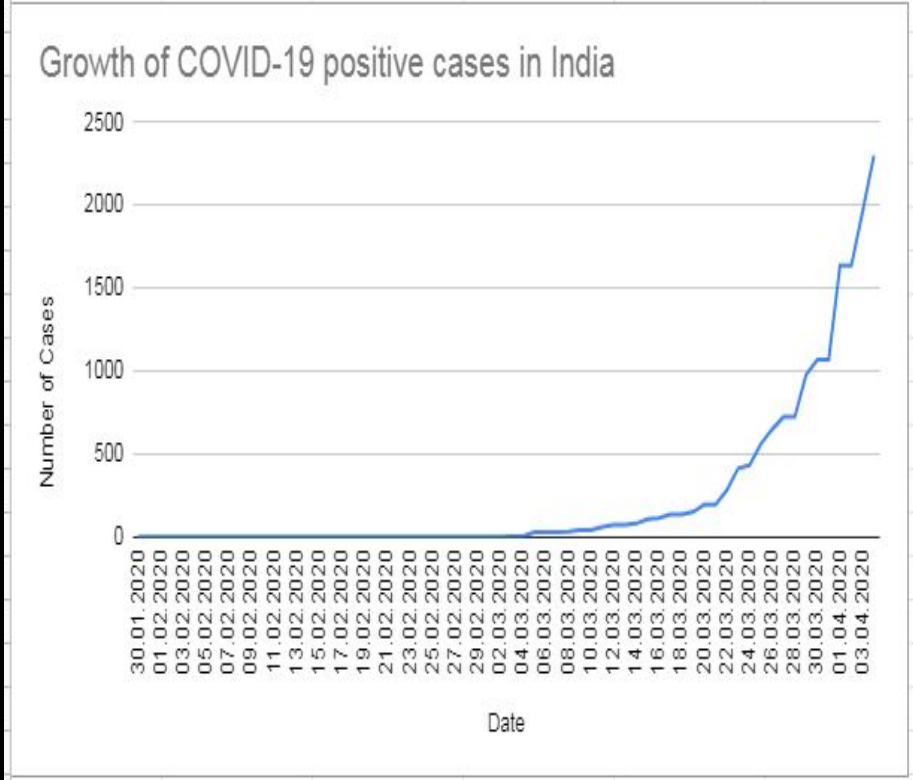
Here, the growth profile of COVID 19 has been analyzed in two segments i.e. one is for India and another is for the World consisting of all affected countries. For each of the segments, the analysis has been done in terms of two factors viz. i) total COVID 19 positive cases and ii) total COVID 19 death cases. The data was collected from the situation reports published by World Health Organization (WHO) in its website (www.who.int) from 30.01.2020 i.e. the day on which the first confirmed case of a COVID 19 positive was reported in India to 04.04.2020.
Firstly, it was statistically analyzed whether COVID 19 positive and death cases are significantly growing in India. Paired t tests (for two sample means) were done to assess whether the COVID 19 positive and death cases are significantly growing. The null hypothesis (H_0) and alternate hypothesis (H_1) were designed as follows:

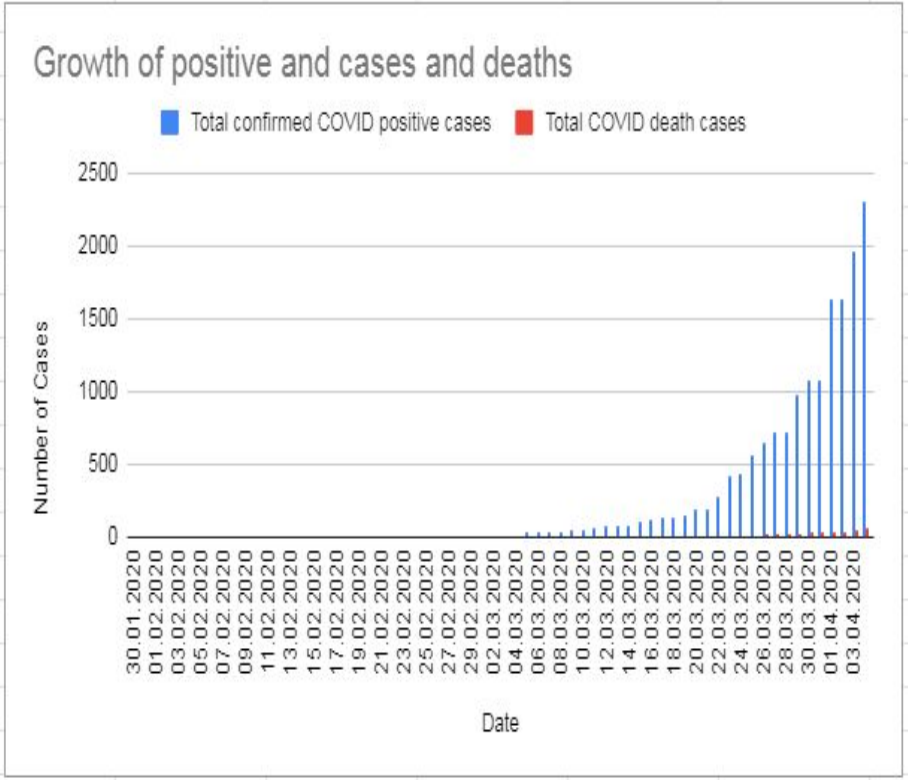
H_0: The COVID cases are not significantly growing week by week in India

H_1: The COVID cases are significantly growing week by week in India

Let's analyze this data in a week wise manner for cases in India and the World. Do the tests with 5% level of significance. The P values obtained in the tests for paired two weeks mean COVID positive and death cases in India are given in the next slide.

P values of t tests- Paired two weeks mean COVID cases in India.		
	P value for COVID positive cases	P value for COVID death cases
1st week vs. 2nd week	2.898 x 10-6	9.981 x 10-2
2nd week vs. 3rd week	1.184x 10-5	7.215 x 10-6
3rd week vs. 4th week	1.389 x 10-4	1.321 x 10-3
4th week vs. 5th week	1.554 x 10-4	2.681 x 10-5


Growth of COVID-19 positive cases in India
Number of Cases
2500
2000
1500
1000
500
0
30.01.2020
01.02.2020
03.02.2020
05.02.2020
07.02.2020
09.02.2020
11.02.2020
13.02.2020
15.02.2020
17.02.2020
19.02.2020
21.02.2020
23.02.2020
25.02.2020
27.02.2020
29.02.2020
02.03.2020
04.03.2020
06.03.2020
08.03.2020
10.03.2020
12.03.2020
14.03.2020
16.03.2020
18.03.2020
20.03.2020
22.03.2020
24.03.2020
26.03.2020
28.03.2020
30.03.2020
01.04.2020
03.04.2020
Date


Growth of positive and cases and deaths
Total confirmed COVID positive cases
Total COVID death cases
Number of Cases
2500
2000
1500
1000
500
0
30.01.2020
01.02.2020
03.02.2020
05.02.2020
07.02.2020
09.02.2020
11.02.2020
13.02.2020
15.02.2020
17.02.2020
19.02.2020
21.02.2020
23.02.2020
25.02.2020
27.02.2020
29.02.2020
02.03.2020
04.03.2020
06.03.2020
08.03.2020
10.03.2020
12.03.2020
14.03.2020
16.03.2020
18.03.2020
20.03.2020
22.03.2020
24.03.2020
26.03.2020
28.03.2020
30.03.2020
01.04.2020
03.04.2020
Date

Growth of COVID-19 in WORLD

Similarly, it was statistically analyzed whether COVID positive and death cases are significantly increasing in the world for a test period of 01.03.2020 to 04.04.2020. The null hypothesis (H0) and alternate hypothesis (H1) were as under:

H_0: The COVID cases are not significantly growing week by week in the World.

H_1: The COVID cases are significantly growing week by week in the World.

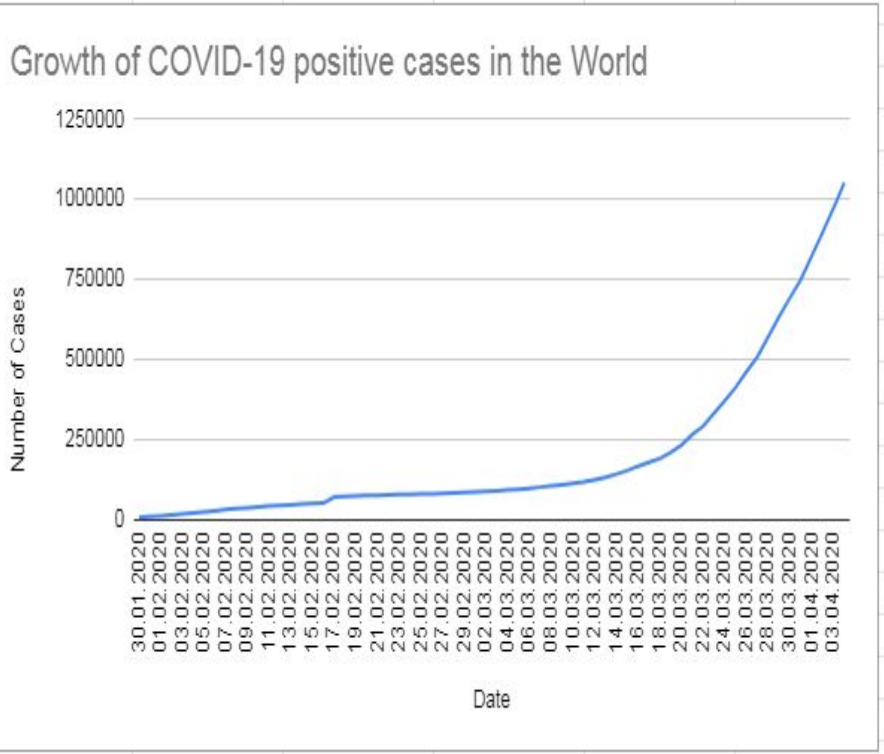
The P values obtained in the tests for paired two weeks mean COVID positive and death cases in the World are presented in the next slide.

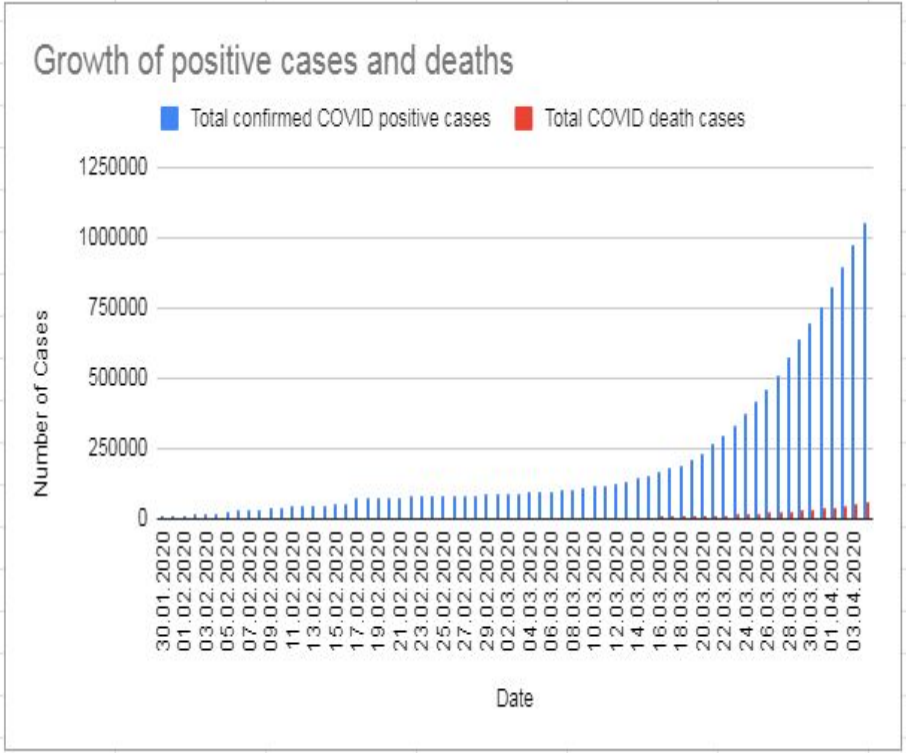
The ***t values*** and the ***Degree of Freedoms(DF)*** of the 2 samples were calculated using the formulas given below:

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\frac{S_X^2}{N_X} + \frac{S_Y^2}{N_Y}}$$

$$DF = \frac{\left(\frac{S_X^2}{N_X} + \frac{S_Y^2}{N_Y}\right)^2}{\frac{\left(\frac{S_X^2}{N_X}\right)^2}{N_X - 1} + \frac{\left(\frac{S_Y^2}{N_Y}\right)^2}{N_Y - 1}}$$

P values of t tests- Paired two weeks mean COVID cases in World.		
	P value for COVID positive cases	P value for COVID death cases
1st week vs. 2nd week	4.988 x 10-5	2.758 x 10-4
2nd week vs. 3rd week	1.046 x 10-4	9.220 x 10-5
3rd week vs. 4th week	3.325 x 10-5	4.272 x 10-5
4th week vs. 5th week	4.025 x 10-7	7.493 x 10-6


Growth of COVID-19 positive cases in the World
1250000
1000000
750000
500000
250000
0
Number of Cases
30.01.2020
01.02.2020
03.02.2020
05.02.2020
07.02.2020
09.02.2020
11.02.2020
13.02.2020
15.02.2020
17.02.2020
19.02.2020
21.02.2020
23.02.2020
25.02.2020
27.02.2020
29.02.2020
02.03.2020
04.03.2020
06.03.2020
08.03.2020
10.03.2020
12.03.2020
14.03.2020
16.03.2020
18.03.2020
20.03.2020
22.03.2020
24.03.2020
26.03.2020
28.03.2020
30.03.2020
01.04.2020
03.04.2020
Date


Growth of positive cases and deaths
Total confirmed COVID positive cases
Total COVID death cases
1250000
1000000
750000
500000
250000
0
Number of Cases
30.01.2020
01.02.2020
03.02.2020
05.02.2020
07.02.2020
09.02.2020
11.02.2020
13.02.2020
15.02.2020
17.02.2020
19.02.2020
21.02.2020
23.02.2020
25.02.2020
27.02.2020
29.02.2020
02.03.2020
04.03.2020
06.03.2020
08.03.2020
10.03.2020
12.03.2020
14.03.2020
16.03.2020
18.03.2020
20.03.2020
22.03.2020
24.03.2020
26.03.2020
28.03.2020
30.03.2020
01.04.2020
03.04.2020
Date

Results:

In case of India, since all the P values are much lesser than the level of significance (i.e. 0.05) for positive cases, it can be concluded that the COVID 19 positive cases are significantly growing week by week during test period (i.e. H_1 accepted). In the same line, it can concluded from the P values of death cases in India that COVID death cases are significantly growing week by week (i.e. H1 accepted) except for 1st to 2nd week wherein H_0 is accepted.

Similarly in case of the World, since all the P values are much lesser than the level of significance (i.e. 0.05) for positive as well as death cases, it can be concluded that the COVID 19 positive and death cases are significantly growing week by week during test period (i.e. H_1 accepted).

Covid 19 Death Rate vs Gender

Is death rate of covid 19 independent of gender?

A survey was conducted and it was found that out **56288** total covid 19 deaths **38973** were male and **17315** were female. We want to prove at a 95% confidence interval that death rate of covid 19 is not independent of gender. In order to do this we have performed a chi square test for independence.

H_0 : Men and women are equally likely to die from covid-19.

H_1 : Death rate is not independent of gender.

The data is represented in the table

Gender	Male	Female
Observed Deaths	38973	17315
Expected Deaths	28144	28144

$$X^2 = \sum \frac{(o-e)^2}{e}$$

CALCULATIONS

χ^2 = 8,333.38

Degree of Freedom = n-1 = 1

Critical value = 3.841

Since χ^2 > **Critical Value**

We can reject $\mathbf{H_0}$.

Therefore, death rate is not independent of gender.

Covid-19 deaths in australia vs germany:

Total deaths in australia=27676(ny)

Total deaths in germany=19314(nx)

Total deaths in australia of age above 60 years=6011(Y)

Total deaths in germany of age above 60 years= 18486(X)

Ho: proportion of people over 60 years who died due to covid in germany is equal to or greater than in australia

H1: proportion of people over 60 years who died due to covid in germany is less than in australia

We will use p test to check the hypothesis.

Covid Vs Diabetes

It is a widely stated hypothesis that the death due to Covid-19 is closely related to a person having diabetes.

In order to analyse this hypothesis we use a sample of 856 people who died due to covid. This group consisted of 64 people who had diabetes and 792 didn't have diabetes. We make the following hypothesis.

H_0: A person with Diabetes is more likely to die than to survive

H_1: A person with Diabetes is equally likely to die and survive

Continued.....

Px cap=X/nx= 0.957129543

Py cap=Y/ny=0.217191791

P cap=(X+Y)/(nx+ny)=0.521324

Z score=157.9812522

P value=0.99997

Alpha =0.05

Since p value>>alpha, We fail to reject Ho.

CONCLUSION: proportion of people over 60 years who died due to covid in germany is equal to or greater than in australia

BIBLIOGRAPHY:

1. https://www.hindustantimes.com/india-news/90-of-those-killed-by-covid-in-india-are-older-than-40-69-are-men/story-glg0Ct4rHQ1YVvZgnckUcM.html
2. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200805-covid-19-sitrep-198.pdf?sfvrsn=f99d1754_2
3. https://timesofindia.indiatimes.com/city/mumbai/94-covid-deaths-in-mumbai-among-middle-aged-patients-senior-citizens/articleshow/77453775.cms
4. https://theprint.in/health/80-covid-patients-in-india-are-asymptomatic-health-ministry-analysis-finds/487761/
5. https://www.downtoearth.org.in/news/health/covid-19-are-80-cases-in-india-really-asymptomatic-70590
6. https://www.worldometers.info/coronavirus/
7. https://www.mohfw.gov.in/
8. https://coronavirus.jhu.edu/
9. https://www.tandfonline.com/doi/full/10.1080/00325481.2020.1786964
10. https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm#Comorbidities
11. https://covid19.who.int/table
12. https://api.covid19india.org/documentation/csv/
13. http://www.who.int
14. https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases

CONTRIBUTIONS:

1.Khushi Yagnik

- Line graph : Recovery rate vs Death rate
- Venn Diagram : Severity of Covid cases
- Heat map: Covid cases in India state wise
- Donut chart : Gender wise cases of a population
- Calendar plot : Daily cases in India
- Stacked Bar graph : Symptoms of covid 19 in India
- Hypothesis : Death rate of covid 19 is not independent of gender.

2.Atharva Swami

- World Heat Map: Latest count of COVID cases and most affected countries
- Growth Hypothesis of COVID cases and deaths in India
- Line and Bar graphs of COVID cases and deaths in India
- Growth Hypothesis of COVID cases and deaths in the World
- Line and Bar graphs of COVID cases and deaths in the World

3. Saloni Singh

- Pie chart for pre-existing condition vs covid
- Line-graph for total cases vs recovery vs deaths in india
- Line-graph of daily recoveries vs deaths
- Daily cases India :Calendar plot
- Population pyramid gender vs age
- box-plot for age and number of deaths

4. Rishika Patel:

- Pie chart of age vs corona deaths (INDIA)
- Pie chart of age vs corona deaths (WORLDWIDE)
- Violin plot for age vs corona deaths(INDIA)
- Box plot for age vs corona deaths(MUMBAI)
- Hypothesis for corona deaths in elder people(above 60 years) of germany vs australia

5. Adarsh Mall :

- Bar Graph : Total cases of top 20 countries
- Bar Graph : Methods of Transmission in various countries
- Scatter Plot : Daily New Cases and Daily Recoveries

6. Dalpat Choudhary :

- Data collection