

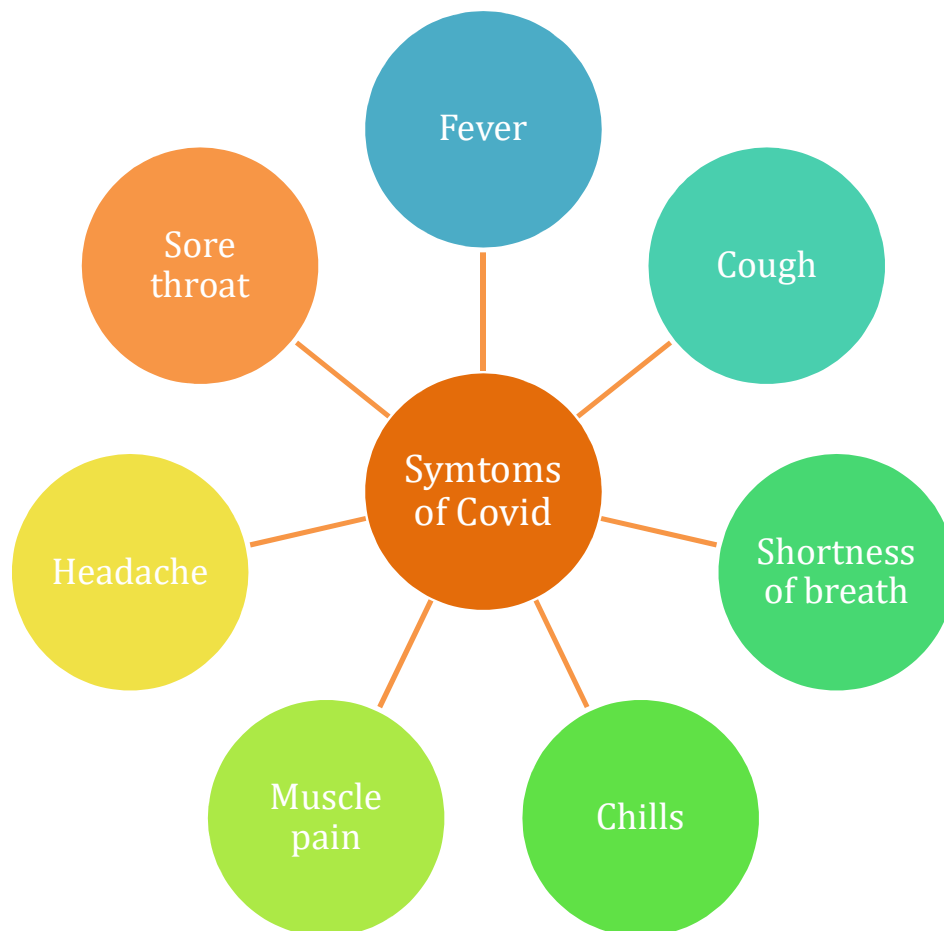
Project 5: Covid-19 India Analysis

Corona viruses are a large family of viruses that may cause respiratory illnesses in humans ranging from common colds to more severe conditions such as Severe Acute Respiratory Syndrome (SARS) and Middle Eastern Respiratory Syndrome (MERS).

COVID-19 can spread from person to person usually through close contact with an infected person or through respiratory droplets that are dispersed into the air when an infected person coughs or sneezes. It may also be possible to get the virus by touching a surface or object contaminated with the virus and then touching your mouth, nose or eyes, but it is not thought to be the main way the virus spreads. Similar to other respiratory illnesses, the symptoms of COVID-19 may include fever, cough, and shortness of breath.

People infected with COVID-19 may experience any range of these symptoms along with aches and pains, nasal congestion, runny nose, sore throat and diarrhea.¹ Symptoms can start to show up anywhere from two to 14 days after exposure to the virus. It may be possible for an infected person who is not yet showing any symptoms to spread the virus.¹ Older persons, and those with pre-existing medical illnesses like heart disease and diabetes, however, seem to be more likely to experience severe respiratory symptoms and complications.

Symptoms of Covid-19:



IMPLEMENTATION CODE:

1: Import all the required libraries and datasets :

```
[8] import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import random
import matplotlib.colors as mcolors

df1 = pd.read_csv('AgeGroupDetails.csv')
df2 = pd.read_csv('HospitalBedsIndia.csv')
df3 = pd.read_csv('ICMRTTestingLabs.csv')
df5 = pd.read_csv('StatewiseTestingDetails.csv')
df4 = pd.read_csv('IndividualDetails.csv')
```

- **Import numpy** : It can be used to perform mathematical operations on arrays such as trigonometric, statistical, and algebraic routines.
- **Import pandas** : it is used for data wrangling and analysis. It is a convenient wrapper around numpy.
- **import seaborn** : it is a visualization library based on matplotlib. It provides a high level interface for drawing attractive and informative statistical graphics.
- **import matplotlib** : it is a plotting library which gives inline plots for quick data analysis.
- **read_csv()** : Data from a data file in the project directory is moved into a pandas dataframe. We can optionally specify the column names.

2: AGE ANALYSIS:

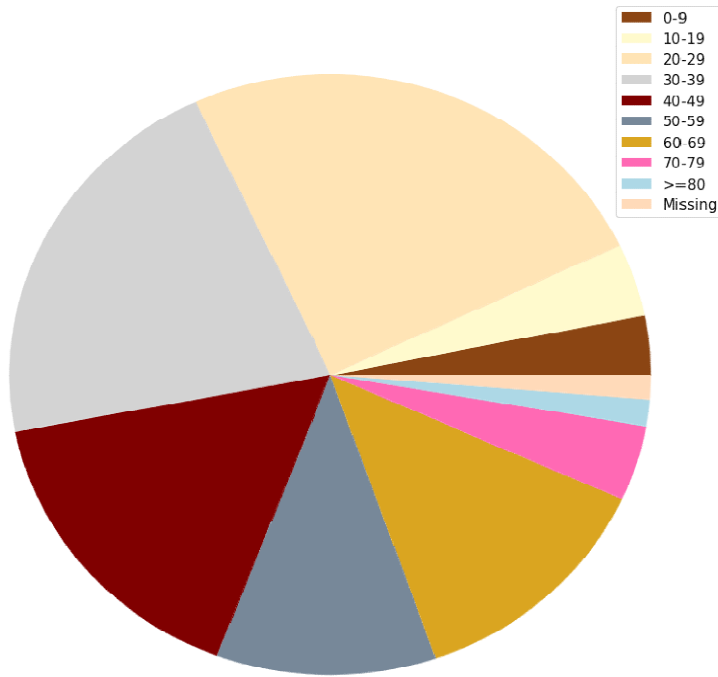
Query - News has Reported that, Covid-19 is more likely to affect elder population compared to the youth, is that true?

```
[11] def plot_pie_charts(x, y, title):
    c = random.choices(list(mcolors.CSS4_COLORS.values()), k = 10)
    plt.figure(figsize=(20,15))
    plt.title(title, size=20)
    plt.pie(y, colors = c)
    plt.legend(x, loc='best', fontsize=15)
    plt.show()

[12] plot_pie_charts(df1['AgeGroup'], df1['TotalCases'], 'Age distribution')
```

Creating a function to plot pie charts and plotting a pie chart of the age distribution of the confirmed cases.

Age distribution



Observation: We can see that most of the cases are of people 20-29 age range with the ranges 30-39, 40-49, 50-59, 60-69 also having a high number of cases.

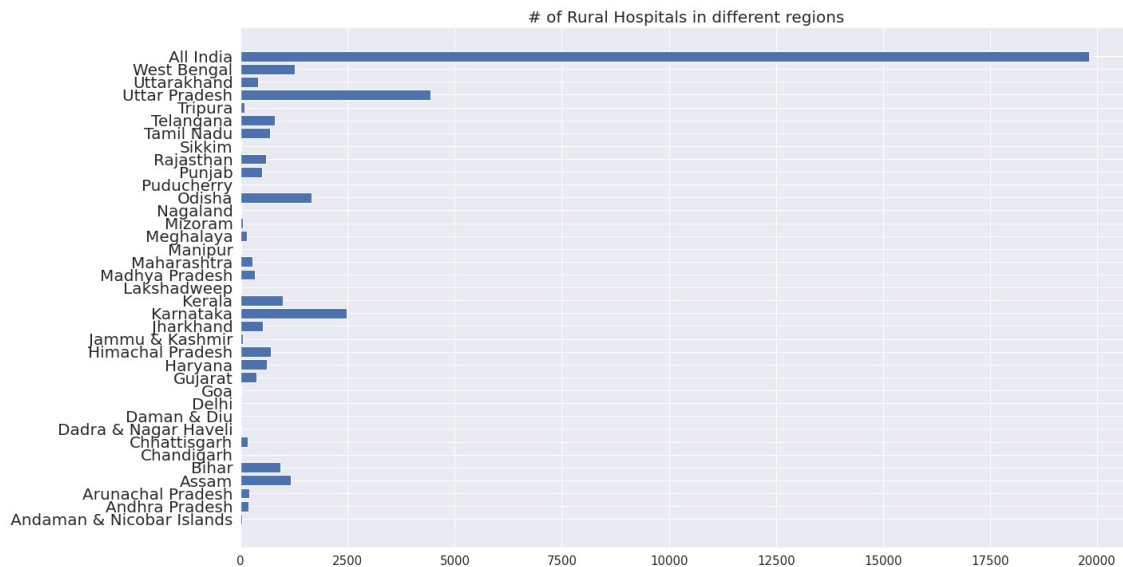
Possible Counter - *Can we give an argument that the numbers are like this due to a higher youth population of India? How can we verify it?*

3. Analysis of number of public health facilities and number of beds:

Query - *The Prime Minister has asked you to plan the migration plans of the labour's as it has become a huge problem, we want them to get the facilities but we have limited resources in cities, how will you handle this situation?*

```
[15] def plot_bar_graphs(x, y, title):  
    plt.figure(figsize=(20, 12))  
    plt.barh(x, y)  
    plt.title(title, size=20)  
    plt.xticks(size=15)  
    plt.yticks(size=20)  
    plt.show()  
    df2 = df2.drop(df2.index[36])
```

```
plot_bar_graphs(df2['State/UT'], df2['NumRuralHospitals_NHP18'], '# of Rural Hospitals in different regions')
```

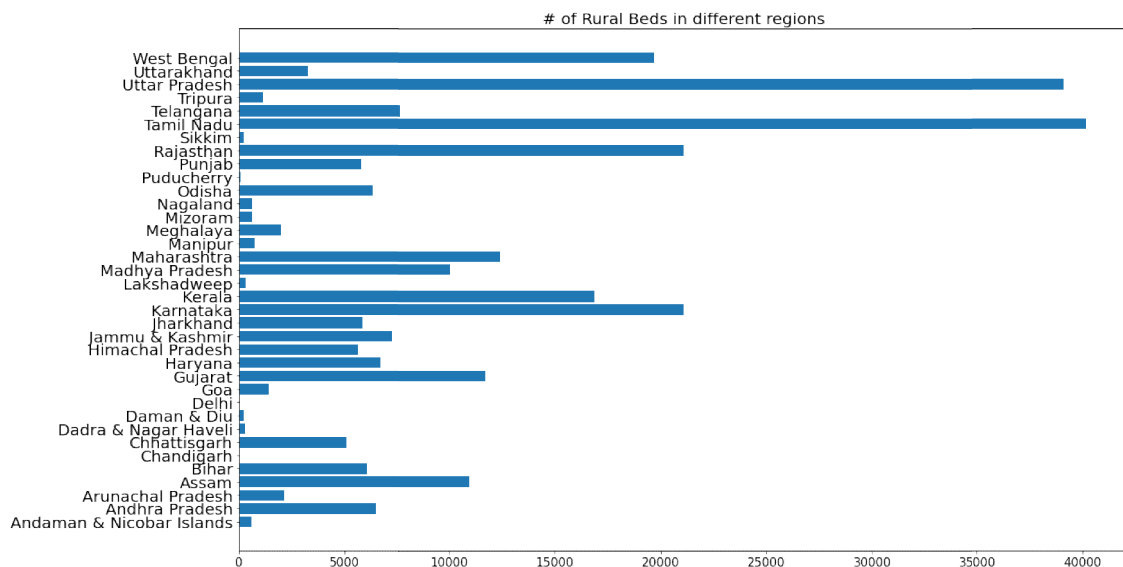


Observation :

- Uttar Pradesh has the most and is the only state with over 4000 rural hospitals.
- Next followed by Karnataka with 2500 rural hospital and stands at second place.

Possible Counter - *We have some of the states which has a much higher number of hospitals and so we can start with moving labour's to Uttar Pradesh, Odisha, Karnataka and in a similar ratio to places with enough medical facilities. But an we be sure of the capacity or hospitals from the numbers?*

```
[17] plot_bar_graphs(df2['State/UT'], df2['NumRuralBeds_NHP18'], '# of Rural Beds in different regions')
```



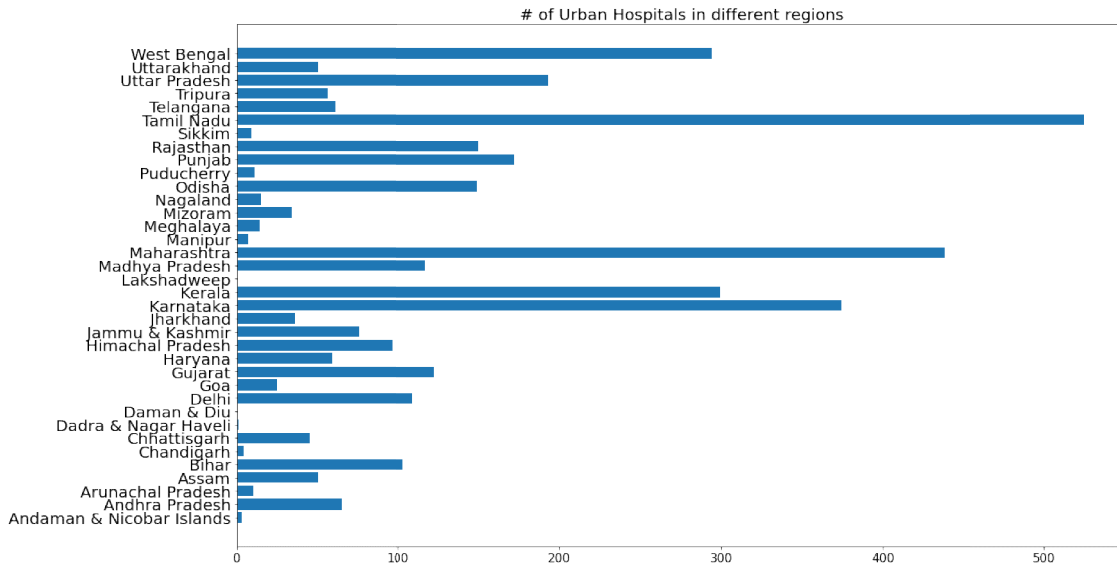
Observation :

- Tamil Nadu is the state with the most Beds in the rural areas with over 40000 beds.
- Uttar Pradesh is a close second.
- Karnataka, Rajasthan and west Bengal are having more than 20000 beds and stands in third position.

Change In Scenario - The Health Ministry was relieved that people will be safe at home and wanted everyone to reach home safely, but at the same time a lot of states with high labor intensive industries didn't wanted the labour's to leave as that will cause a huge problem when the economy restarts (and that is happening today), so couldnt we have done something better?

Possible Counter - *Does Every City needs to evacuate the daily wage workers from the cities?*

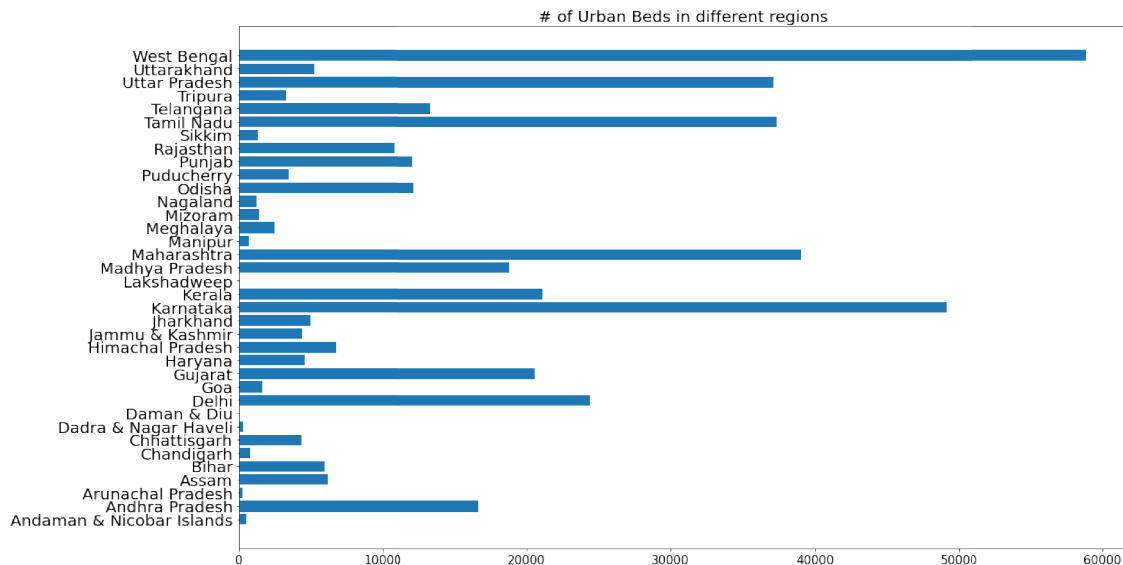
```
[18] plot_bar_graphs(df2['State/UT'], df2['NumUrbanHospitals_NHP18'], '# of Urban Hospitals in different regions')
```



Observation :

- Tamil Nadu has the most urban hospitals with over 500 of them.
- Maharashtra is the only other state with over 400 urban hospitals followed by Karnataka with nearly 400 urban hospitals.

```
[19] plot_bar_graphs(df2['State/UT'], df2['NumUrbanBeds_NHP18'], '# of Urban Beds in different regions')
```



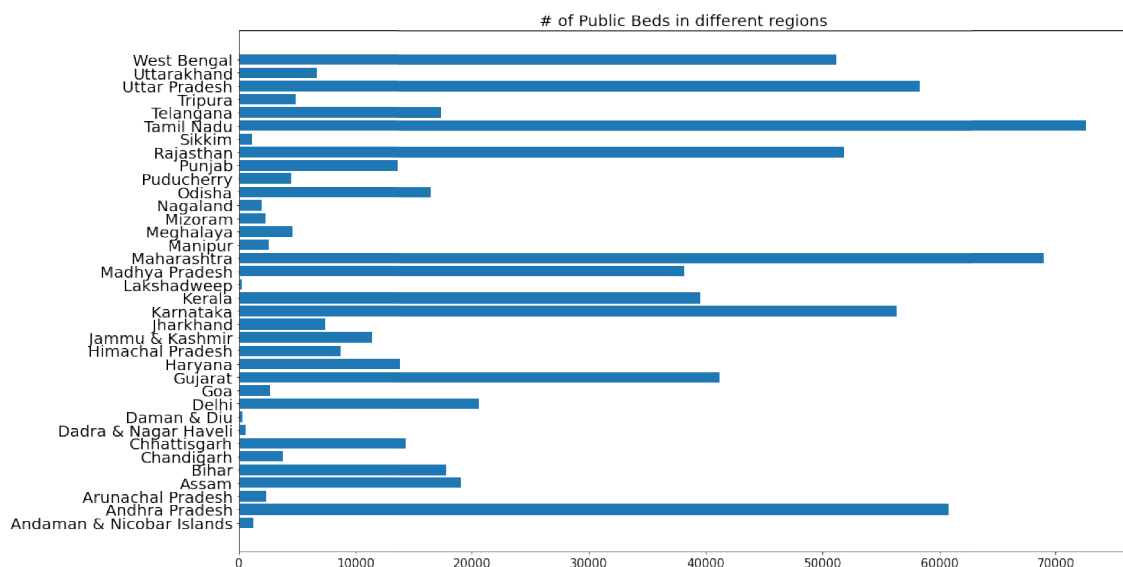
Observation :

- West Bengal has the most beds in Urban areas with almost 60000 beds.
- Karnataka in second place with reaching 50000 beds.
- Thirdly Uttar Pradesh Tamil nadu and Maharashtra with more than 38000 beds

Possible Counter - *Can we give an argument, that it also depends on the population of the state, as regular people will be taking up hospitals as well.*

Query - *Private Hospitals Cost of Treatment is Unaffordable for the Labours, Does all the above queries change there result when we try to focus on public hospitals as hospitals for the poor ?*

```
plot_bar_graphs(df2['State/UT'], df2['NumPublicBeds_HMIS'], '# of Public Beds in different regions')
```



Observation :

- Tamil Nadu has the most public beds with over 70000 of them.
- Maharashtra and Andhra Pradesh are the only other state with over 60000 beds.

Observation - We have some states which have less number of total hospitals but most of them being in public access makes it easier for the lahours to approach like Andhra Pradesh.

Being in a political arena, opposition parties want to do what they can to save there image and so, **some states which had higher positive cases reported that , it is because we test more then others.** (Same arguments are given by DTJ for America's Number of positive cases).

Verify the Information.

How to Handle Lack of Data -Assuming we don't have the test data, the next best thing we can compare is the number of test centers in each cities as they share the testing capacity of the state.

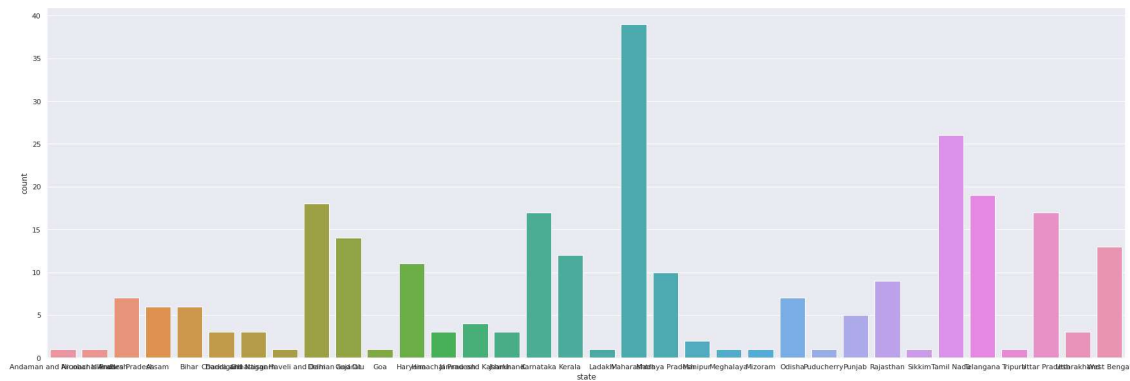
4. Analysis of ICMR testing Labs in the country

```
[23] df3['state'].value_counts()
```

```

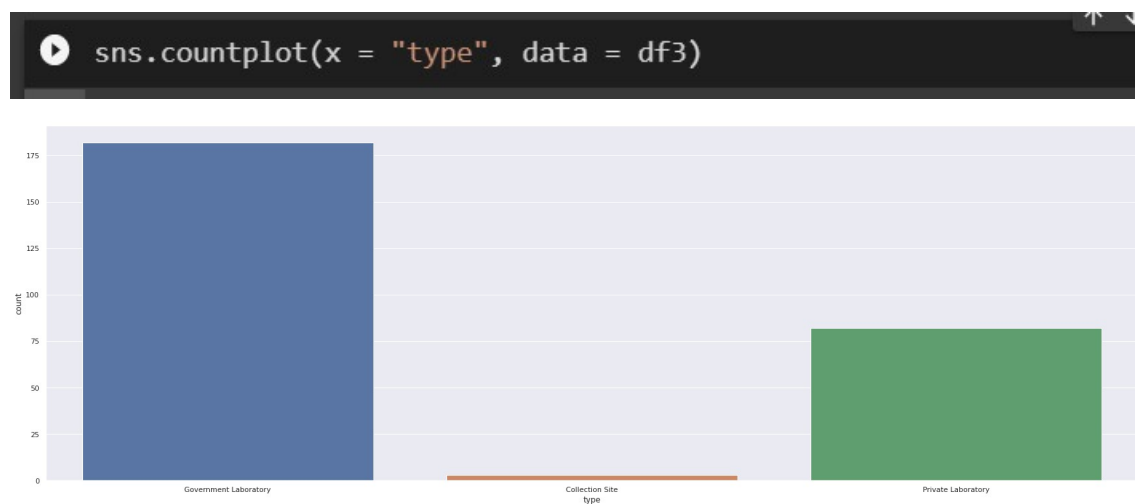
▶ Maharashtra 39
▶ Tamil Nadu 26
▶ Telangana 19
▶ Delhi 18
▶ Karnataka 17
▶ Uttar Pradesh 17
▶ Gujarat 14
▶ West Bengal 13
▶ Kerala 12
▶ Haryana 11
▶ Madhya Pradesh 10
▶ Rajasthan 9
▶ Andhra Pradesh 7
▶ Odisha 7
▶ Assam 6
▶ Bihar 6
▶ Punjab 5
▶ Jammu and Kashmir 4
▶ Jharkhand 3
▶ Uttarakhand 3
▶ Chandigarh 3
▶ Chhattisgarh 3
▶ Himachal Pradesh 3
▶ Manipur 2
▶ Sikkim 1
▶ Tripura 1
▶ Meghalaya 1
▶ Mizoram 1
▶ Arunachal Pradesh 1
▶ Ladakh 1
▶ Puducherry 1
▶ Dadra and Nagar Haveli and Daman and Diu 1
▶ Goa 1
▶ Andaman and Nicobar Islands 1
Name: state, dtype: int64
```

```
[24] sns.set(rc={'figure.figsize':(30,10)})
sns.countplot(x = "state", data = df3)
```



Observation: Maharashtra has the most number of ICMR testing labs with 39 of them.

Tamil nadu and Telangana are 2nd and 3rd with 26 and 19 respectively.



Observation: Out of all the ICMR labs, around 180 are Government Labs and around 80 are Private Labs.

Conclusion Needs More - In practice, we can't ignore facts like the total cases depends on your population example, Goa can never out count Delhi or Mumbai, its entire population is only 7.8 lakh people. We need to consider total population or some percentage value of it to see the positive cases.

5. Analysis of state-wise testing details

Validating Assumptions- When we assume something like We don't have testing data but we can substitute that with Test centers we need to verify if that is a valid assumption and to do that we need to see some information about their correlation as shown.


```
[27] df5_sort = df5.sort_values(by = 'TotalSamples', ascending=False).head()
df5_sort.head()
```

	Date	State	TotalSamples	Negative	Positive
1068	2020-05-20	Tamil Nadu	360068.0	3346311.0	13191.0
1067	2020-05-19	Tamil Nadu	348174.0	334839.0	12448.0
1066	2020-05-18	Tamil Nadu	337841.0	325546.0	11760.0
1065	2020-05-17	Tamil Nadu	326720.0	315019.0	11224.0
1064	2020-05-16	Tamil Nadu	313639.0	302523.0	10585.0

Observation: Clearly, Tamil Nadu has the highest number of samples being tested.

```
[28] df5_sort1 = df5.sort_values(by = 'TotalSamples', ascending=True).head()
df5_sort1.head()
```

	Date	State	TotalSamples	Negative	Positive
765	2020-04-07	Mizoram	58.0	0.0	1.0
764	2020-04-06	Mizoram	58.0	0.0	1.0
804	2020-04-06	Nagaland	60.0	47.0	0.0
806	2020-04-11	Nagaland	70.0	70.0	0.0
805	2020-04-10	Nagaland	70.0	69.0	0.0

Observation: Mizoram and Nagaland have the lowest number of tested samples recorded for a day.

Observation- We can see that towards the extremes the data is synced well as Tamil nade which has second highest testing facility does highest testing and Mizoram and Nagaland which has one of the lowerst test centers does lowest testing.

If you are not sure you can do the median check as well but if you have enough data for testing of all centers then you'll not need to substitute.

Does this disease effect a particular gender more then the others, verify that.

6. Analysis based on Gender, districts, States and number of Active cases :

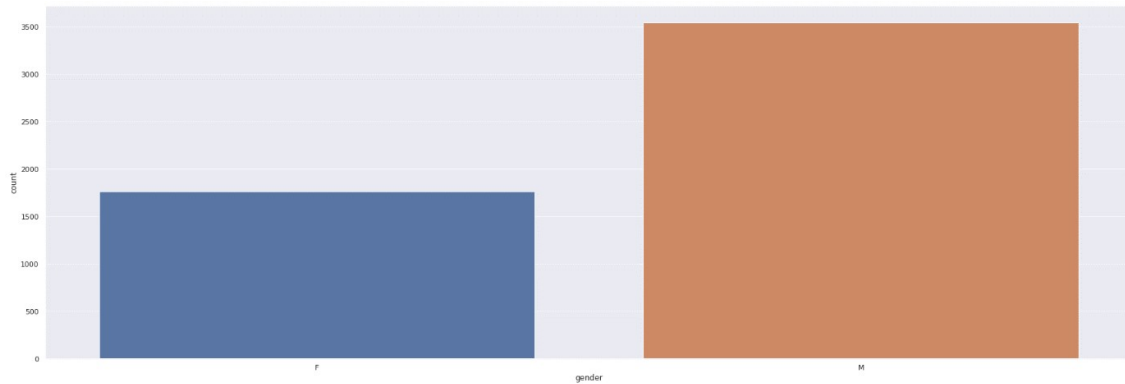
```
[30] df4['gender'].isna().sum()
```

```
22577
```

```
[31] df4['gender'].value_counts()
```

```
M    3547
F    1766
Name: gender, dtype: int64
```

```
[32] sns.countplot(x = "gender", data = df4)
```



Biased/Unreliable Observation: Out of the registered gender entries, patients are 2x as likely to be males than females.

Ask yourself can we really answer that question?

It is clearly visible that we have the gender label missing from 80% of the data, and we are only making assumptions based on the 20% of the labels data so the results we have shared is not a strong one.

Remember this, "Its better to say we can't predict then to do it wrong, +3/-1"

Instructions for Assignments (Comment)

Download the notebook, and state something that you heard in news or you always wanted to know about this pandemic, cross check a statement or discover some pattern in the data. Present that in the comment section, and for people who need help with pandas to get the graph or data link of pandas library is given below you can search for whatever too you need to do your analysis.

LINK TO THE CODE AND DATASET

GOOGLE COLAB:

<https://colab.research.google.com/drive/1Kd3sT3HxjoTFarEbBDSnLRwUa51s6Sxx?usp=sharing>

ON GITHUB:

<https://github.com/Vidyapro/Predictive-Analysis-ML>

Project 5: Covid-19 India Analysis

By: Vidyashree M S

Intern at Utkarshini Edutech
