CORONAVIRUS ANALYSIS IN INDIA

Coronavirus disease (COVID-19) is an infectious disease caused by a newly discovered coronavirus.

- Most people infected with the COVID-19 virus will experience mild to moderate respiratory illness and recover without requiring special treatment.
- Older people, and those with underlying medical problems like cardiovascular disease, diabetes, chronic respiratory disease, and cancer are more likely to develop more serious illness.
- The COVID-19 virus spreads primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, so it is important that people who are sick practice respiratory hygiene and sneezing etiquette.
- The best way to prevent and slow down transmission is to be informed about the COVID-19 virus, the disease it causes and how it spreads. Protect themselves and others from infection by washing your hands or using an alcohol-based rub frequently and not touching your face.
- Protecting oneself and others from infection by washing our hands or using an alcohol-based rub frequently and not touching our face.

To prevent infection and to slow transmission of COVID-19, do the following:

- Washing your hands or using an alcohol-based rub frequently kills any viruses that may be on your hands.
- Maintaining a safe distance from anyone who is coughing or sneezing or otherwise showing signs of illness.
- Avoid touching our face.
- Practicing respiratory hygiene, for example, when coughing or sneezing.
- Staying home if we feel unwell.
- If we have fever, cough and difficulty breathing, seek medical help early.
- Practice physical distancing by covering unnecessary travel and staying away from large groups of people.

Importing Libraries

```
import numpy as np
import pandas as pd
import matplotlib as mp
import seaborn as sns
%matplotlib inline
from matplotlib import pyplot as plt
from mpl_toolkits.mplot3d import *
import plotly.express.offline as px
import plotly.express as px
import plotly.express.colors
import plotly.express.colors as px_c
import plotly.express.colors.qualitative as px_c_q
import plotly.express.colors.sequential as px_c_s
import plotly.express.colors.cyclical as px_c_c
import plotly.express.colors.diverging as px_c_d
import plotly.express.colors.cmocean as px_c_cm
import plotly.express.colors.colorbrewer as px_c_cb
import plotly.express.colors.carto as px_c_ca
from plotly.offline import iplot
from plotly.offline.offline import init_notebook_mode
```

Exploratory Data Analysis and Data Visualization

Importing and Reading the Dataset

```
cvdf_19_india.csv
```

```
data = pd.read_csv('/kaggle/input/covid19-in-india/covid_19_india.csv')
```

Displaying the Dataset

```
data.head()
```

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Death	Confirmed
0	1	30/01/2020	6:00 PM	Kerala	1	0	1	0	1
1	2	31/01/2020	6:00 PM	Kerala	1	0	1	0	1
2	3	01/02/2020	6:00 PM	Kerala	1	0	1	0	1
3	4	02/02/2020	6:00 PM	Kerala	2	0	2	0	2
4	5	03/02/2020	6:00 PM	Kerala	3	0	3	0	3

Displaying the Dataset Columns

```
data.columns
```

```
Index(['Sno', 'Date', 'Time', 'State/UnionTerritory',
       'ConfirmedIndianNational', 'ConfirmedForeignNational', 'Cured',
       'Death', 'Confirmed'],
      dtype='object')
```

Displaying the Dataset Information

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4992 entries, 0 to 4991
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Sno                     4992 non-null   int64
 1   Date                    4992 non-null   object
 2   Time                    4992 non-null   object
 3   State/UnionTerritory    4992 non-null   object
 4   ConfirmedIndianNational 4992 non-null   int64
 5   ConfirmedForeignNational 4992 non-null  int64
 6   Cured                   4992 non-null   int64
 7   Death                   4992 non-null   int64
 8   Confirmed               4992 non-null   int64
dtype: object, int64(6)
memory usage: 351.1 KB
```

Displaying the Dataset Description

```
data.describe()
```

	Sno	Cured	Death	Confirmed
count	4992.000000	4992.000000	4992.000000	4992.000000
mean	2496.500000	1437.109776	48.468950	1517.027244
std	1441.310097	4775.034440	164.174421	4848.263140
min	1.000000	0.000000	0.000000	0.000000
25%	1248.750000	17.750000	0.000000	19.750000
50%	2496.500000	81.000000	1.000000	87.000000
75%	3744.250000	724.000000	20.000000	764.000000
max	4992.000000	34719.000000	1985.000000	35746.000000

Checking presence of some null values in the Dataset

```
data.isnull().sum()
```

```
Sno                       0
Date                      0
Time                      0
State/UnionTerritory      0
ConfirmedIndianNational   0
ConfirmedForeignNational  0
Cured                     0
Death                     0
Confirmed                 0
dtype: int64
```

Displaying the Dataset Columns

```
data.columns
```

Changing the Date Format

```
data['Date'] = pd.to_datetime(data['Date'], dayfirst=True)
data.head()
```

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Death	Confirmed
0	1	2020-01-30	6:00 PM	Kerala	1	0	1	0	1
1	2	2020-01-31	6:00 PM	Kerala	1	0	1	0	1
2	3	2020-02-01	6:00 PM	Kerala	1	0	1	0	1
3	4	2020-02-02	6:00 PM	Kerala	2	0	2	0	2
4	5	2020-02-03	6:00 PM	Kerala	3	0	3	0	3

Sorting values of Confirmed Cases in descending order depending on the State/Union Territory

```
data.sort_values(by=['Confirmed', 'Death', 'Cured'], ascending=[False, False, False])
```

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNational	Cured	Death	Confirmed
4991	4992	2020-12-31	10:00 AM	Telangana	0	0	0	0	0
1	2	2020-01-31	6:00 PM	Kerala	1	0	1	0	1
2	3	2020-02-01	6:00 PM	Kerala	1	0	1	0	1
3	4	2020-02-02	6:00 PM	Kerala	2	0	2	0	2
4	5	2020-02-03	6:00 PM	Kerala	3	0	3	0	3

Total Cases in India

```
data.groupby('Date').sum()[['Confirmed','Death','Cured']].sort_values(by=['Confirmed'], ascending=False)
```

Representation using Bar Chart Plotting

```
data.groupby('Date').sum()[['Confirmed','Death','Cured']].sort_values(by=['Confirmed'], ascending=False).plot(kind='bar',figsize=(20,10))
plt.title('Cases in India')
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.show()
```

Confirmed Cases in the topmost five States - Graphical Representation

```
# Bar Chart
plt.rcParams['figure.figure.max_open_warning'] = 0
data.groupby('State/UnionTerritory').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```

Confirmed Cases in Maharashtra

```
data.groupby('Date').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```

Confirmed Cases in Tamil Nadu

```
data.groupby('Date').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```

Confirmed Cases in Delhi

```
data.groupby('Date').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```


Confirmed Cases in Gujarat

```
data.groupby('Date').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```


Confirmed Cases in Uttar Pradesh

```
data.groupby('Date').sum()[['Confirmed']].\
    sort_values(by=['Confirmed'], ascending=False).\
    head(5).\
    plot(kind='bar',\
         figsize=(20,10),\
         color = 'red',\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases in 5\n', fontsize=20)
plt.show()
```


Scatter Plotting using Data depending on various factors

```
data.sample(n=1000).\
    plot(kind='scatter',\
         x='Confirmed',\
         y='Death',\
         figsize=(20,10),\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases Vs Death\n', fontsize=20)
plt.show()
```


People are recovering at a faster rate

```
data.sample(n=1000).\
    plot(kind='scatter',\
         x='Confirmed',\
         y='Cured',\
         figsize=(20,10),\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases Vs Cured\n', fontsize=20)
plt.show()
```


The death rate is much lower in India

```
data.sample(n=1000).\
    plot(kind='scatter',\
         x='Cured',\
         y='Death',\
         figsize=(20,10),\
         fontsize=20,\
         legend=True)
plt.title('Cured Cases Vs Death\n', fontsize=20)
plt.show()
```


Initially most cases were detected in Foreign Nationals but as days flew by, UK cases in Indian Nationals turned out to be more and more.

```
data.groupby('Date').sum()[['Confirmed', 'ConfirmedIndianNational', 'ConfirmedForeignNational']].\
    plot(kind='line',\
         figsize=(20,10),\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases Vs Nationals\n', fontsize=20)
plt.show()
```


The curve joining is a bit confusing to get the representation. Actually, it provides an overall representation of the whole dataset.

Scatter Plotting using Lines depending on various factors

```
data.sample(n=1000).\
    plot(kind='line',\
         x='Confirmed',\
         y='Death',\
         figsize=(20,10),\
         fontsize=20,\
         legend=True)
plt.title('Confirmed Cases Vs Death\n', fontsize=20)
plt.show()
```


Importing and Reading the Dataset

```
AgeGroupDetails.csv
```

```
!wget --no-check-certificate "https://drive.google.com/uc?export=download&id=1OYr0Ed_LzSRAKN_QwVzoq_ayn7VsbfJ8&confirm=t" -O AgeGroupDetails.csv
```

```
!ls -lrtrh AgeGroupDetails.csv
```

```
-rw-r--r-- 1 root root  1.1K Jun 14 07:41 AgeGroupDetails.csv
```

```
!head AgeGroupDetails.csv
```

```
Sno,AgeGroup,Confirmed,Percentage
1,0-9,1,0.03
2,10-19,136,4.09
3,20-29,1085,32.6
4,30-39,1000,30.04
5,40-49,671,20.16
6,50-59,327,9.82
7,60-69,154,4.63
8,70-79,36,1.08
9,80+,14,0.42
```

```
!wc -l AgeGroupDetails.csv
```

```
10 AgeGroupDetails.csv
```

Representation of the dataset using Pie Chart

```
import matplotlib.colors as mcolors
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt
import numpy as np, numpy.random
import matplotlib.cbook as cbook
import matplotlib.patheffects
%matplotlib inline
```

```
data = pd.read_csv('AgeGroupDetails.csv')
labels = data['AgeGroup']
sizes = data['Confirmed']
# Data to plot
colors = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue', 'white', 'aqua', 'pink', 'orange', 'cyan', 'red']
explode = (0.1, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0)
plt.pie(sizes, explode=explode, labels=labels, colors=colors,
        autopct='%1.1f%%', shadow=True, startangle=140)
plt.axis('equal')
plt.rcParams['figure.figure.max_open_warning'] = 0
plt.rcParams['figure.figure.max_open_warning'] = 0
plt.rcParams['figure.figure.max_open_warning'] = 0
plt.rcParams['figure.figure.max_open_warning'] = 0
plt.title('India - Age Group Wise Distribution\n', fontsize=20)
plt.show()
```


Importing and Reading the Dataset

```
HospitalBeds.csv
```

```
!wget --no-check-certificate "https://drive.google.com/uc?export=download&id=1YTfbGL_LZ3DNNL_QwVzoq_ayn7VsbfJ8&confirm=t" -O HospitalBeds.csv
```

```
!ls -lrtrh HospitalBeds.csv
```

```
-rw-r--r-- 1 root root  1.1K Jun 14 07:41 HospitalBeds.csv
```

```
!head HospitalBeds.csv
```

```
In [2]: !pip install -q -U kaggle
!mkdir ~/.kaggle
!cp kaggle.json ~/.kaggle/
!ls ~/.kaggle
!chmod 600 /root/.kaggle/kaggle.json
!kaggle datasets download -d sudalairajkumar/novel-corona-virus-2019-dataset
!unzip -q novel-corona-virus-2019-dataset.zip -d covid19
!ls covid19
```

```
Dataset URL: https://kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset
covid_19_data.csv  time_series_covid19_confirmed_global.csv
time_series_covid19_deaths_global.csv  time_series_covid19_recovered_global.csv
```

Analysing the various factors present in the Dataset using Visual / Graphical Representation

```
In [3]: !pip install -q -U plotly_express
!pip install -q -U dtale
!pip install -q -U plotnine
!pip install -q -U folium
!pip install -q -U pandas_profiling[notebook]
!pip install -q -U ipywidgets
!jupyter nbextension enable --py --sys-prefix widgetsnbextension
```

```
In [4]: import numpy as np, pandas as pd, matplotlib as mpl, matplotlib.pyplot as plt, seaborn as sns, os
%matplotlib inline
%matplotlib notebook
%matplotlib inline
%matplotlib notebook
%matplotlib inline
%matplotlib notebook
```