AN ANALYSIS ON COVID 19

1.INTRODUCTION:

We area going to an analysis of the covid-19 data available with us for the first and second waves in india to understand the different stages of the coronavirus pandemic during that period.

Reading data

```
In [1]:   # pip install pandas
          #pip install matplotlib
          import pandas as pd
          df=pd.read_csv(r"C:\Users\Dell\OneDrive\Desktop\python\Covid-19 Data Analysis.csv")
          df
```

Out[1]:

	STATES	CONFIRMED CASES	ACTIVE CASES	DEATHS
0	Andaman and Nicobar Islands	4710	99	61
1	Andhra Pradesh	868064	7840	6992
2	Arunachal Pradesh	16282	817	54
3	Assam	212776	3399	981
4	Bihar	234610	5359	1264
5	Chandigarh	17409	1062	277
6	Chhattisgarh	237322	19635	2861
7	Daman and Diu	3332	16	2
8	Delhi	570374	32885	9174
9	Goa	47963	1335	688
10	Gujarat	209780	14970	3989
11	Haryana	234126	18362	2428
12	Himachal Pradesh	40518	8289	645
13	Jammu and Kashmir	110224	4965	1694
14	Jharkhand	109151	2016	964
15	Karnataka	884897	23298	11778
16	Kerala	602982	62025	2244
17	Ladakh	8415	809	117
18	Lakshadweep	0	0	0
19	Madhya Pradesh	206128	14771	3260
20	Maharashtra	1823896	91623	47151
21	Manipur	25045	3198	281
22	Meghalaya	11810	763	111
23	Mizoram	3847	343	5
24	Nagaland	11186	928	64
25	Odisha	318725	4921	1739
26	Puducherry	36968	460	610
27	Punjab	152091	7842	4807
28	Rajasthan	268063	28653	2312
29	Sikkim	4989	248	109
30	Tamil Nadu	781915	10997	11712
31	Telangana	270318	9627	1461
32	Tripura	32726	592	370
33	Uttar Pradesh	543888	24099	7761
34	Uttarakhand	74795	5059	1231
35	West Bengal	483484	24298	8424

Let's us now check if the data has any null Values:

```
In [2]:   df.shape
```

(36, 4)

Let us now check if the data has any null values:

In [3]: `df.isna().sum()`

Out[3]:
```
STATES            0
CONFIRMED CASES   0
ACTIVE CASES      0
DEATHS            0
dtype: int64
```

Let's check the data types:

In [4]: `df.dtypes`

Out[4]:
```
STATES            object
CONFIRMED CASES    int64
ACTIVE CASES       int64
DEATHS             int64
dtype: object
```

We would also like to see the last few values in the dataframe:

In [5]: `df.tail()`

Out[5]:

	STATES	CONFIRMED CASES	ACTIVE CASES	DEATHS
31	Telangana	270318	9627	1461
32	Tripura	32726	592	370
33	Uttar Pradesh	543888	24099	7761
34	Uttarakhand	74795	5059	1231
35	West Bengal	483484	24298	8424

we will check the values for the columns:
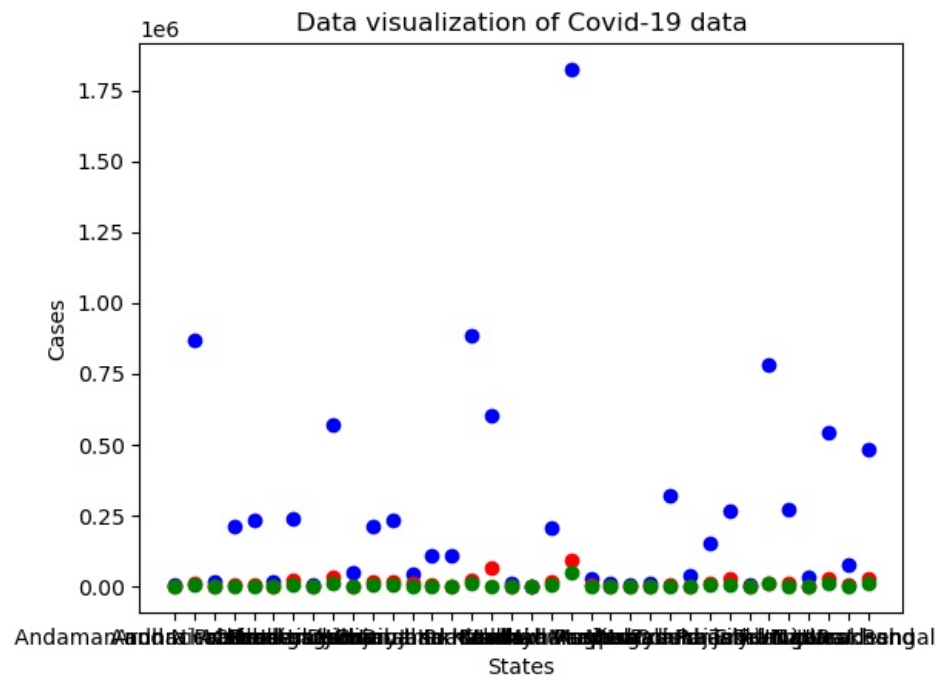
In [6]: `df['ACTIVE CASES'].value_counts`

Out[6]:
```
<bound method IndexOpsMixin.value_counts of 0        99
1       7840
2        817
3       3399
4       5359
5       1062
6      19635
7         16
8      32885
9       1335
10     14970
11     18362
12      8289
13      4965
14      2016
15     23298
16     62025
17       809
18         0
19     14771
20     91623
21      3198
22       763
23       343
24       928
25      4921
26       460
27      7842
28     28653
29       248
30     10997
31      9627
32       592
33     24099
34      5059
35     24298
Name: ACTIVE CASES, dtype: int64>
```

In [8]:
```python
import matplotlib.pyplot as plt
y=df["CONFIRMED CASES"]
x=df["STATES"]
z=df["ACTIVE CASES"]
w=df["DEATHS"]
plt.scatter(x,y,color='blue')
plt.scatter(x,z,color='red')
```

```
plt.scatter(x,w,color='green')
plt.title("Data visualization of Covid-19 data")
plt.xlabel("States")
plt.ylabel("Cases")
```

Out[8]: Text(0, 0.5, 'Cases')

In [ ]:

In [ ]: