In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
df = pd.read_csv('italy-covid-daywise.csv')
```

In [3]:

```
df
```

Out[3]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 248 entries, 0 to 247
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
---  ------      --------------  -----  
 0   date        248 non-null    object 
 1   new_cases   248 non-null    float64
 2   new_deaths  248 non-null    float64
 3   new_tests   135 non-null    float64
dtypes: float64(3), object(1)
memory usage: 7.9+ KB
```

In [ ]:

```

```

In [5]:

```
1 df.describe()
```

Out[5]:

	new_cases	new_deaths	new_tests
count	248.000000	248.000000	135.000000
mean	1094.818548	143.133065	31699.674074
std	1554.508002	227.105538	11622.209757
min	-148.000000	-31.000000	7841.000000
25%	123.000000	3.000000	25259.000000
50%	342.000000	17.000000	29545.000000
75%	1371.750000	175.250000	37711.000000
max	6557.000000	971.000000	95273.000000

In [ ]:

```
1
```

In [6]:

```
1 df[['date','new_cases']]
```

Out[6]:

	date	new_cases
0	2019-12-31	0.0
1	2020-01-01	0.0
2	2020-01-02	0.0
3	2020-01-03	0.0
4	2020-01-04	0.0
...	...	...
243	2020-08-30	1444.0
244	2020-08-31	1365.0
245	2020-09-01	996.0
246	2020-09-02	975.0
247	2020-09-03	1326.0

248 rows × 2 columns

In [ ]:

```
1
```

In [7]:

```
1 covid_copy = df.copy()
```

In [8]:

```
1 covid_copy
```

Out[8]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [9]:

```
1 covid_copy.head()
```

Out[9]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN

In [10]:

```
1 covid_copy.tail()
```

Out[10]:

	date	new_cases	new_deaths	new_tests
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

In [11]:

```
1 covid_copy.at[0,'new_tests']
```

Out[11]:

```
nan
```

In [12]:

```
1 covid_copy['new_tests'][0:100]
```

Out[12]:

```
0    NaN
1    NaN
2    NaN
3    NaN
4    NaN
     ..
95   NaN
96   NaN
97   NaN
98   NaN
99   NaN
Name: new_tests, Length: 100, dtype: float64
```

In [13]:

```
1 covid_copy.new_tests.first_valid_index()
```

Out[13]:

```
111
```

In [14]:

```
1 covid_copy['new_tests'].first_valid_index()
```

Out[14]:

```
111
```

In [15]:

```
1 covid_copy.iloc[111]
```

Out[15]:

```
date          2020-04-20
new_cases         3047.0
new_deaths         433.0
new_tests         7841.0
Name: 111, dtype: object
```

In [16]:

```
1 covid_copy.iloc[110]
```

Out[16]:

```
date          2020-04-19
new_cases         3491.0
new_deaths         480.0
new_tests            NaN
Name: 110, dtype: object
```

In [ ]:

```
1
```

In [17]:

```
1 covid_copy
```

Out[17]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [18]:

```
1 covid_copy.sample(20)
```

Out[18]:

	date	new_cases	new_deaths	new_tests
127	2020-05-06	1075.0	236.0	37771.0
216	2020-08-03	238.0	8.0	13467.0
199	2020-07-17	230.0	20.0	28661.0
60	2020-02-29	238.0	4.0	NaN
108	2020-04-17	3786.0	525.0	NaN
23	2020-01-23	0.0	0.0	NaN
6	2020-01-06	0.0	0.0	NaN
70	2020-03-10	1797.0	98.0	NaN
181	2020-06-29	174.0	22.0	15484.0
172	2020-06-20	-148.0	47.0	29875.0
80	2020-03-20	5322.0	429.0	NaN
163	2020-06-11	202.0	71.0	32991.0
120	2020-04-29	2091.0	382.0	38589.0
63	2020-03-03	347.0	17.0	NaN
125	2020-05-04	1389.0	174.0	22999.0
170	2020-06-18	328.0	43.0	32921.0
240	2020-08-27	1366.0	13.0	57640.0
0	2019-12-31	0.0	0.0	NaN
50	2020-02-19	0.0	0.0	NaN
201	2020-07-19	249.0	14.0	20621.0

In [19]:

```
1 total_new_cases = covid_copy['new_cases'].sum()
```

In [20]:

```
1 total_new_cases
```

Out[20]:

```
271515.0
```

In [21]:

```
1 total_new_death = covid_copy['new_deaths'].sum()
```

In [22]:

```
1 total_new_death
```

Out[22]:

```
35497.0
```

In [23]:

```
1 total_new_test = covid_copy['new_tests'].sum()
```

In [24]:

```
1 total_new_test
```

Out[24]:

```
4279456.0
```

In [ ]:

```
1
```

In [25]:

```
1 death_rate = (total_new_death / total_new_cases)*100
```

In [26]:

```
1 death_rate
```

Out[26]:

```
13.073679170579894
```

In [27]:

```
1 print('Death Rate for Italy Covid 19 is {:.2f} %'.format(death_rate))
```

```
Death Rate for Italy Covid 19 is 13.07 %
```

In [ ]:

```
1
```

```
1 A total of 935310 test were conducted before the tests reports were published
  in public
```

In [28]:

```
1 initial_test = 935310
2 total_test_done = total_new_test + initial_test
```

In [29]:

```
1 total_test_done
```

Out[29]:

5214766.0

In [ ]:

```
1
```

In [30]:

```
1 positivity_rate = (total_new_cases / total_test_done) * 100
```

In [31]:

```
1 positivity_rate
```

Out[31]:

5.206657403227681

In [32]:

```
1 print('Postivity Rate of Overall Test done is {:.2f} %'.format(positivity_rate))
```

```
Postivity Rate of Overall Test done is 5.21 %
```

In [33]:

```
1 covid_copy
```

Out[33]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [34]:

```
1 high_case = covid_copy['new_cases'] > 1000
```

In [35]:

```
1 high_case
```

Out[35]:

```
0      False
1      False
2      False
3      False
4      False
       ...
243     True
244     True
245    False
246    False
247     True
Name: new_cases, Length: 248, dtype: bool
```

In [36]:

```
1 covid_copy[high_case]
```

Out[36]:

	date	new_cases	new_deaths	new_tests
68	2020-03-08	1247.0	36.0	NaN
69	2020-03-09	1492.0	133.0	NaN
70	2020-03-10	1797.0	98.0	NaN
72	2020-03-12	2313.0	196.0	NaN
73	2020-03-13	2651.0	189.0	NaN
...	...	...	...	...
241	2020-08-28	1409.0	5.0	65135.0
242	2020-08-29	1460.0	9.0	64294.0
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
247	2020-09-03	1326.0	6.0	NaN

72 rows × 4 columns

In [37]:

```
1 covid_copy
```

Out[37]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [38]:

```
1 high_death = covid_copy['new_deaths'] > 100
```

In [39]:

```
1 high_death
```

Out[39]:

```
0      False
1      False
2      False
3      False
4      False
       ...
243    False
244    False
245    False
246    False
247    False
Name: new_deaths, Length: 248, dtype: bool
```

In [40]:

```
1 covid_copy[high_death]
```

Out[40]:

	date	new_cases	new_deaths	new_tests
69	2020-03-09	1492.0	133.0	NaN
71	2020-03-11	977.0	167.0	NaN
72	2020-03-12	2313.0	196.0	NaN
73	2020-03-13	2651.0	189.0	NaN
74	2020-03-14	2547.0	252.0	NaN
...	...	...	...	...
144	2020-05-23	652.0	130.0	42579.0
145	2020-05-24	669.0	119.0	34206.0
149	2020-05-28	584.0	117.0	39838.0
152	2020-05-31	416.0	111.0	28948.0
229	2020-08-16	629.0	158.0	22470.0

78 rows × 4 columns

In [ ]:

```
1
```

In [41]:

```
1 covid_copy
```

Out[41]:

	date	new_cases	new_deaths	new_tests
0	2019-12-31	0.0	0.0	NaN
1	2020-01-01	0.0	0.0	NaN
2	2020-01-02	0.0	0.0	NaN
3	2020-01-03	0.0	0.0	NaN
4	2020-01-04	0.0	0.0	NaN
...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0
244	2020-08-31	1365.0	4.0	42583.0
245	2020-09-01	996.0	6.0	54395.0
246	2020-09-02	975.0	8.0	NaN
247	2020-09-03	1326.0	6.0	NaN

248 rows × 4 columns

In [42]:

```
(total_new_cases / total_test_done) * 100
```

Out[42]:

5.206657403227681

In [43]:

```
covid_copy['positivity_rate'] = (covid_copy['new_cases'] / covid_copy['new_test
```

In [44]:

```
covid_copy
```

Out[44]:

	date	new_cases	new_deaths	new_tests	positivity_rate
0	2019-12-31	0.0	0.0	NaN	NaN
1	2020-01-01	0.0	0.0	NaN	NaN
2	2020-01-02	0.0	0.0	NaN	NaN
3	2020-01-03	0.0	0.0	NaN	NaN
4	2020-01-04	0.0	0.0	NaN	NaN
...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2.696999
244	2020-08-31	1365.0	4.0	42583.0	3.205505
245	2020-09-01	996.0	6.0	54395.0	1.831051
246	2020-09-02	975.0	8.0	NaN	NaN
247	2020-09-03	1326.0	6.0	NaN	NaN

248 rows × 5 columns

In [45]:

```
covid_copy.iloc[113]
```

Out[45]:

```
date               2020-04-22
new_cases              2729.0
new_deaths              534.0
new_tests             44248.0
positivity_rate       6.16751
Name: 113, dtype: object
```

In [46]:

```
(total_new_death / total_new_cases)*100
```

Out[46]:

13.073679170579894

In [ ]:

```

```

In [47]:

```
covid_copy['death_rate'] = (covid_copy['new_deaths'] / covid_copy['new_cases'])
```

In [48]:

```
covid_copy
```

Out[48]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate
0	2019-12-31	0.0	0.0	NaN	NaN	NaN
1	2020-01-01	0.0	0.0	NaN	NaN	NaN
2	2020-01-02	0.0	0.0	NaN	NaN	NaN
3	2020-01-03	0.0	0.0	NaN	NaN	NaN
4	2020-01-04	0.0	0.0	NaN	NaN	NaN
...	...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2.696999	0.069252
244	2020-08-31	1365.0	4.0	42583.0	3.205505	0.293040
245	2020-09-01	996.0	6.0	54395.0	1.831051	0.602410
246	2020-09-02	975.0	8.0	NaN	NaN	0.820513
247	2020-09-03	1326.0	6.0	NaN	NaN	0.452489

248 rows × 6 columns

In [ ]:

```

```

In [49]:

```
covid_copy['new_cases'].mean()
```

Out[49]:

1094.8185483870968

In [50]:

```
1 covid_copy['new_deaths'].mean()
```

Out[50]:

```
143.13306451612902
```

In [51]:

```
1 covid_copy['new_tests'].mean()
```

Out[51]:

```
31699.674074074075
```

In [ ]:

```
1
```

In [52]:

```
1 covid_copy.head()
```

Out[52]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate
0	2019-12-31	0.0	0.0	NaN	NaN	NaN
1	2020-01-01	0.0	0.0	NaN	NaN	NaN
2	2020-01-02	0.0	0.0	NaN	NaN	NaN
3	2020-01-03	0.0	0.0	NaN	NaN	NaN
4	2020-01-04	0.0	0.0	NaN	NaN	NaN

In [53]:

```
1 covid_copy['date']
```

Out[53]:

```
0      2019-12-31
1      2020-01-01
2      2020-01-02
3      2020-01-03
4      2020-01-04
          ...
243    2020-08-30
244    2020-08-31
245    2020-09-01
246    2020-09-02
247    2020-09-03
Name: date, Length: 248, dtype: object
```

In [54]:

```
1 covid_copy['date'] = pd.to_datetime(covid_copy['date'])
```

In [55]:

```
1 covid_copy['date']
```

Out[55]:

```
0     2019-12-31
1     2020-01-01
2     2020-01-02
3     2020-01-03
4     2020-01-04
         ...
243   2020-08-30
244   2020-08-31
245   2020-09-01
246   2020-09-02
247   2020-09-03
Name: date, Length: 248, dtype: datetime64[ns]
```

In [56]:

```
1 covid_copy.head()
```

Out[56]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate
0	2019-12-31	0.0	0.0	NaN	NaN	NaN
1	2020-01-01	0.0	0.0	NaN	NaN	NaN
2	2020-01-02	0.0	0.0	NaN	NaN	NaN
3	2020-01-03	0.0	0.0	NaN	NaN	NaN
4	2020-01-04	0.0	0.0	NaN	NaN	NaN

In [57]:

```
1 covid_copy['year'] = pd.DatetimeIndex(covid_copy['date']).year
```

In [58]:

```
1 covid_copy
```

Out[58]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year
0	2019-12-31	0.0	0.0	NaN	NaN	NaN	2019
1	2020-01-01	0.0	0.0	NaN	NaN	NaN	2020
2	2020-01-02	0.0	0.0	NaN	NaN	NaN	2020
3	2020-01-03	0.0	0.0	NaN	NaN	NaN	2020
4	2020-01-04	0.0	0.0	NaN	NaN	NaN	2020
...	...	...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2.696999	0.069252	2020
244	2020-08-31	1365.0	4.0	42583.0	3.205505	0.293040	2020
245	2020-09-01	996.0	6.0	54395.0	1.831051	0.602410	2020
246	2020-09-02	975.0	8.0	NaN	NaN	0.820513	2020
247	2020-09-03	1326.0	6.0	NaN	NaN	0.452489	2020

248 rows × 7 columns

In [59]:

```
1 covid_copy['month'] = pd.DatetimeIndex(covid_copy['date']).month
```

In [60]:

```
1 covid_copy.head()
```

Out[60]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month
0	2019-12-31	0.0	0.0	NaN	NaN	NaN	2019	12
1	2020-01-01	0.0	0.0	NaN	NaN	NaN	2020	1
2	2020-01-02	0.0	0.0	NaN	NaN	NaN	2020	1
3	2020-01-03	0.0	0.0	NaN	NaN	NaN	2020	1
4	2020-01-04	0.0	0.0	NaN	NaN	NaN	2020	1

In [61]:

```
1 covid_copy['day'] = pd.DatetimeIndex(covid_copy['date']).day
```

In [62]:

```
1 covid_copy.head()
```

Out[62]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
0	2019-12-31	0.0	0.0	NaN	NaN	NaN	2019	12	31
1	2020-01-01	0.0	0.0	NaN	NaN	NaN	2020	1	1
2	2020-01-02	0.0	0.0	NaN	NaN	NaN	2020	1	2
3	2020-01-03	0.0	0.0	NaN	NaN	NaN	2020	1	3
4	2020-01-04	0.0	0.0	NaN	NaN	NaN	2020	1	4

In [ ]:

```
1
```

In [63]:

```
1 covid_copy['week_day'] = pd.DatetimeIndex(covid_copy['date']).weekday
```

In [64]:

```
1 covid_copy
```

Out[64]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
0	2019-12-31	0.0	0.0	NaN	NaN	NaN	2019	12	31
1	2020-01-01	0.0	0.0	NaN	NaN	NaN	2020	1	1
2	2020-01-02	0.0	0.0	NaN	NaN	NaN	2020	1	2
3	2020-01-03	0.0	0.0	NaN	NaN	NaN	2020	1	3
4	2020-01-04	0.0	0.0	NaN	NaN	NaN	2020	1	4
...	...	...	...	...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2.696999	0.069252	2020	8	30
244	2020-08-31	1365.0	4.0	42583.0	3.205505	0.293040	2020	8	31
245	2020-09-01	996.0	6.0	54395.0	1.831051	0.602410	2020	9	1
246	2020-09-02	975.0	8.0	NaN	NaN	0.820513	2020	9	2
247	2020-09-03	1326.0	6.0	NaN	NaN	0.452489	2020	9	3

248 rows × 10 columns

0 - Mon 1 - Tues 2 - Wed 3 - Thur 4 - Fri 5 - Sat 6 - Sun

total cases detected on day basis

In [86]:

```
1 covid_copy[covid_copy['week_day'] == 0]['new_cases'].sum()
```

Out[86]:

```
38820.0
```

In [87]:

```
1 covid_copy[covid_copy['week_day'] == 1]['new_cases'].sum()
```

Out[87]:

```
33071.0
```

In [88]:

```
1 covid_copy[covid_copy['week_day'] == 2]['new_cases'].sum()
```

Out[88]:

33383.0

In [89]:

```
1 covid_copy[covid_copy['week_day'] == 3]['new_cases'].sum()
```

Out[89]:

39444.0

In [90]:

```
1 covid_copy[covid_copy['week_day'] == 4]['new_cases'].sum()
```

Out[90]:

41640.0

In [91]:

```
1 covid_copy[covid_copy['week_day'] == 5]['new_cases'].sum()
```

Out[91]:

41503.0

In [92]:

```
1 covid_copy[covid_copy['week_day'] == 6]['new_cases'].sum()
```

Out[92]:

43654.0

deaths per day

In [94]:

```
1 covid_copy[covid_copy['week_day'] == 0]['new_deaths'].sum()
```

Out[94]:

4368.0

In [95]:

```
1 covid_copy[covid_copy['week_day'] == 1]['new_deaths'].sum()
```

Out[95]:

4678.0

In [96]:

```
covid_copy[covid_copy['week_day'] == 2]['new_deaths'].sum()
```

Out[96]:

5211.0

In [97]:

```
covid_copy[covid_copy['week_day'] == 3]['new_deaths'].sum()
```

Out[97]:

5093.0

In [98]:

```
covid_copy[covid_copy['week_day'] == 4]['new_deaths'].sum()
```

Out[98]:

5071.0

In [99]:

```
covid_copy[covid_copy['week_day'] == 5]['new_deaths'].sum()
```

Out[99]:

5495.0

In [100]:

```
covid_copy[covid_copy['week_day'] == 6]['new_deaths'].sum()
```

Out[100]:

5581.0

In [ ]:

```

```

fetching the data forthe month of may

In [65]:

```
may_data = covid_copy[covid_copy.month == 5]
```

In [66]:

```
1 may_data
```

Out[66]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
122	2020-05-01	1872.0	285.0	43732.0	4.280618	15.224359	2020	5	1
123	2020-05-02	1965.0	269.0	31231.0	6.291825	13.689567	2020	5	2
124	2020-05-03	1900.0	474.0	27047.0	7.024809	24.947368	2020	5	3
125	2020-05-04	1389.0	174.0	22999.0	6.039393	12.526998	2020	5	4
126	2020-05-05	1221.0	195.0	32211.0	3.790631	15.970516	2020	5	5
127	2020-05-06	1075.0	236.0	37771.0	2.846099	21.953488	2020	5	6
128	2020-05-07	1444.0	369.0	13665.0	10.567142	25.554017	2020	5	7
129	2020-05-08	1401.0	274.0	45428.0	3.084001	19.557459	2020	5	8
130	2020-05-09	1327.0	243.0	36091.0	3.676817	18.311982	2020	5	9
131	2020-05-10	1083.0	194.0	31384.0	3.450803	17.913204	2020	5	10
132	2020-05-11	802.0	165.0	25823.0	3.105758	20.573566	2020	5	11
133	2020-05-12	744.0	179.0	39620.0	1.877839	24.059140	2020	5	12
134	2020-05-13	1402.0	172.0	37049.0	3.784178	12.268188	2020	5	13
135	2020-05-14	888.0	195.0	41131.0	2.158956	21.959459	2020	5	14
136	2020-05-15	992.0	262.0	39027.0	2.541830	26.411290	2020	5	15
137	2020-05-16	789.0	242.0	40657.0	1.940625	30.671736	2020	5	16
138	2020-05-17	875.0	153.0	33505.0	2.611551	17.485714	2020	5	17
139	2020-05-18	675.0	145.0	26101.0	2.586108	21.481481	2020	5	18
140	2020-05-19	451.0	99.0	40226.0	1.121165	21.951220	2020	5	19
141	2020-05-20	813.0	162.0	38617.0	2.105290	19.926199	2020	5	20
142	2020-05-21	665.0	161.0	40644.0	1.636158	24.210526	2020	5	21
143	2020-05-22	642.0	156.0	42987.0	1.493475	24.299065	2020	5	22

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
144	2020-05-23	652.0	130.0	42579.0	1.531271	19.938650	2020	5	23
145	2020-05-24	669.0	119.0	34206.0	1.955797	17.787743	2020	5	24
146	2020-05-25	531.0	50.0	20676.0	2.568195	9.416196	2020	5	25
147	2020-05-26	300.0	92.0	33944.0	0.883809	30.666667	2020	5	26
148	2020-05-27	397.0	78.0	37299.0	1.064372	19.647355	2020	5	27
149	2020-05-28	584.0	117.0	39838.0	1.465937	20.034247	2020	5	28
150	2020-05-29	593.0	70.0	38233.0	1.551016	11.804384	2020	5	29
151	2020-05-30	516.0	87.0	36051.0	1.431306	16.860465	2020	5	30
152	2020-05-31	416.0	111.0	28948.0	1.437060	26.682692	2020	5	31

In [69]:

```
1 sept_data = covid_copy[covid_copy['month'] == 9]
```

In [70]:

```
1 sept_data
```

Out[70]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
245	2020-09-01	996.0	6.0	54395.0	1.831051	0.602410	2020	9	1
246	2020-09-02	975.0	8.0	NaN	NaN	0.820513	2020	9	2
247	2020-09-03	1326.0	6.0	NaN	NaN	0.452489	2020	9	3

In [71]:

```
1 march_data = covid_copy[covid_copy['month']==3]
```

In [72]:

```
1 march_data
```

Out[72]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day	w
61	2020-03-01	240.0	8.0	NaN	NaN	3.333333	2020	3	1	
62	2020-03-02	561.0	6.0	NaN	NaN	1.069519	2020	3	2	
63	2020-03-03	347.0	17.0	NaN	NaN	4.899135	2020	3	3	
64	2020-03-04	466.0	28.0	NaN	NaN	6.008584	2020	3	4	
65	2020-03-05	587.0	27.0	NaN	NaN	4.599659	2020	3	5	
66	2020-03-06	769.0	41.0	NaN	NaN	5.331599	2020	3	6	
67	2020-03-07	778.0	49.0	NaN	NaN	6.298201	2020	3	7	
68	2020-03-08	1247.0	36.0	NaN	NaN	2.886929	2020	3	8	
69	2020-03-09	1492.0	133.0	NaN	NaN	8.914209	2020	3	9	
70	2020-03-10	1797.0	98.0	NaN	NaN	5.453534	2020	3	10	
71	2020-03-11	977.0	167.0	NaN	NaN	17.093142	2020	3	11	
72	2020-03-12	2313.0	196.0	NaN	NaN	8.473843	2020	3	12	
73	2020-03-13	2651.0	189.0	NaN	NaN	7.129385	2020	3	13	
74	2020-03-14	2547.0	252.0	NaN	NaN	9.893993	2020	3	14	
75	2020-03-15	3497.0	173.0	NaN	NaN	4.947098	2020	3	15	
76	2020-03-16	2823.0	370.0	NaN	NaN	13.106624	2020	3	16	
77	2020-03-17	4000.0	347.0	NaN	NaN	8.675000	2020	3	17	
78	2020-03-18	3526.0	347.0	NaN	NaN	9.841180	2020	3	18	
79	2020-03-19	4207.0	473.0	NaN	NaN	11.243166	2020	3	19	
80	2020-03-20	5322.0	429.0	NaN	NaN	8.060879	2020	3	20	
81	2020-03-21	5986.0	625.0	NaN	NaN	10.441029	2020	3	21	
82	2020-03-22	6557.0	795.0	NaN	NaN	12.124447	2020	3	22	

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day	v
83	2020-03-23	5560.0	649.0	NaN	NaN	11.672662	2020	3	23	
84	2020-03-24	4789.0	601.0	NaN	NaN	12.549593	2020	3	24	
85	2020-03-25	5249.0	743.0	NaN	NaN	14.155077	2020	3	25	
86	2020-03-26	5210.0	685.0	NaN	NaN	13.147793	2020	3	26	
87	2020-03-27	6153.0	660.0	NaN	NaN	10.726475	2020	3	27	
88	2020-03-28	5959.0	971.0	NaN	NaN	16.294680	2020	3	28	
89	2020-03-29	5974.0	887.0	NaN	NaN	14.847673	2020	3	29	
90	2020-03-30	5217.0	758.0	NaN	NaN	14.529423	2020	3	30	
91	2020-03-31	4050.0	810.0	NaN	NaN	20.000000	2020	3	31	

In [73]:

```
1 march_data[['new_cases','new_deaths','new_tests']]
```

Out[73]:

	new_cases	new_deaths	new_tests
61	240.0	8.0	NaN
62	561.0	6.0	NaN
63	347.0	17.0	NaN
64	466.0	28.0	NaN
65	587.0	27.0	NaN
66	769.0	41.0	NaN
67	778.0	49.0	NaN
68	1247.0	36.0	NaN
69	1492.0	133.0	NaN
70	1797.0	98.0	NaN
71	977.0	167.0	NaN
72	2313.0	196.0	NaN
73	2651.0	189.0	NaN
74	2547.0	252.0	NaN
75	3497.0	173.0	NaN
76	2823.0	370.0	NaN
77	4000.0	347.0	NaN
78	3526.0	347.0	NaN
79	4207.0	473.0	NaN
80	5322.0	429.0	NaN
81	5986.0	625.0	NaN
82	6557.0	795.0	NaN
83	5560.0	649.0	NaN
84	4789.0	601.0	NaN
85	5249.0	743.0	NaN
86	5210.0	685.0	NaN
87	6153.0	660.0	NaN
88	5959.0	971.0	NaN
89	5974.0	887.0	NaN
90	5217.0	758.0	NaN
91	4050.0	810.0	NaN

In [74]:

```
1 may_data[['new_cases','new_deaths','new_tests']]
```

Out[74]:

	new_cases	new_deaths	new_tests
122	1872.0	285.0	43732.0
123	1965.0	269.0	31231.0
124	1900.0	474.0	27047.0
125	1389.0	174.0	22999.0
126	1221.0	195.0	32211.0
127	1075.0	236.0	37771.0
128	1444.0	369.0	13665.0
129	1401.0	274.0	45428.0
130	1327.0	243.0	36091.0
131	1083.0	194.0	31384.0
132	802.0	165.0	25823.0
133	744.0	179.0	39620.0
134	1402.0	172.0	37049.0
135	888.0	195.0	41131.0
136	992.0	262.0	39027.0
137	789.0	242.0	40657.0
138	875.0	153.0	33505.0
139	675.0	145.0	26101.0
140	451.0	99.0	40226.0
141	813.0	162.0	38617.0
142	665.0	161.0	40644.0
143	642.0	156.0	42987.0
144	652.0	130.0	42579.0
145	669.0	119.0	34206.0
146	531.0	50.0	20676.0
147	300.0	92.0	33944.0
148	397.0	78.0	37299.0
149	584.0	117.0	39838.0
150	593.0	70.0	38233.0
151	516.0	87.0	36051.0
152	416.0	111.0	28948.0

In [75]:

```
1 may_data['new_cases'].sum()
```

Out[75]:

```
29073.0
```

In [76]:

```
1 may_data['new_tests'].sum()
```

Out[76]:

```
1078720.0
```

In [77]:

```
1 may_data['new_deaths'].sum()
```

Out[77]:

```
5658.0
```

In [78]:

```
1 may_data = may_data[['new_cases','new_deaths','new_tests']]
```

In [80]:

```
1 may_data.sum()
```

Out[80]:

```
new_cases       29073.0
new_deaths       5658.0
new_tests     1078720.0
dtype: float64
```

In [ ]:

```
1
```

In [81]:

```
1 covid_copy
```

Out[81]:

	date	new_cases	new_deaths	new_tests	positivity_rate	death_rate	year	month	day
0	2019-12-31	0.0	0.0	NaN	NaN	NaN	2019	12	31
1	2020-01-01	0.0	0.0	NaN	NaN	NaN	2020	1	1
2	2020-01-02	0.0	0.0	NaN	NaN	NaN	2020	1	2
3	2020-01-03	0.0	0.0	NaN	NaN	NaN	2020	1	3
4	2020-01-04	0.0	0.0	NaN	NaN	NaN	2020	1	4
...	...	...	...	...	...	...	...	...	...
243	2020-08-30	1444.0	1.0	53541.0	2.696999	0.069252	2020	8	30
244	2020-08-31	1365.0	4.0	42583.0	3.205505	0.293040	2020	8	31
245	2020-09-01	996.0	6.0	54395.0	1.831051	0.602410	2020	9	1
246	2020-09-02	975.0	8.0	NaN	NaN	0.820513	2020	9	2
247	2020-09-03	1326.0	6.0	NaN	NaN	0.452489	2020	9	3

248 rows × 10 columns

In [83]:

```
1 covid_month = covid_copy.groupby('month')[['new_cases','new_deaths','new_tests'
```

In [85]:

```
1 covid_month.sum()
```

Out[85]:

month	new_cases	new_deaths	new_tests
1	3.0	0.0	0.0
2	885.0	21.0	0.0
3	100851.0	11570.0	0.0
4	101852.0	16091.0	419591.0
5	29073.0	5658.0	1078720.0
6	7772.0	1404.0	830354.0
7	6722.0	388.0	797692.0
8	21060.0	345.0	1098704.0
9	3297.0	20.0	54395.0
12	0.0	0.0	0.0

In [ ]:

```
1
```

In [101]:

```
1 df1 = pd.read_csv('covid_19_india.csv')
```

In [102]:

```
1 df1.head()
```

Out[102]:

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

In [104]:

```
1 df1
```

Out[104]:

	Sno	Date	Time	State/UnionTerritory	ConfirmedIndianNational	ConfirmedForeignNatic
0	1	2020-01-30	6:00 PM	Kerala	1	
1	2	2020-01-31	6:00 PM	Kerala	1	
2	3	2020-02-01	6:00 PM	Kerala	2	
3	4	2020-02-02	6:00 PM	Kerala	3	
4	5	2020-02-03	6:00 PM	Kerala	3	
...	...	...	...	...	...	
16845	16846	2021-07-07	8:00 AM	Telangana	-	
16846	16847	2021-07-07	8:00 AM	Tripura	-	
16847	16848	2021-07-07	8:00 AM	Uttarakhand	-	
16848	16849	2021-07-07	8:00 AM	Uttar Pradesh	-	
16849	16850	2021-07-07	8:00 AM	West Bengal	-	

16850 rows × 12 columns

In [ ]:

```
1
```