

In [1]:

```
1 import pandas as pd
2 import numpy as np
```

In [3]:

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

In [4]:

```
1 df
```

Out[4]:

	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 [7]:

```
1 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 []:

```
1
```

In [8]:

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

Out[8]:

	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 [9]:

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

Out[9]:

	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 [10]:

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

In [11]:

```
1 covid_copy
```

Out[11]:

	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 [13]:

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

Out[13]:

	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 [14]:

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

Out[14]:

	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 [15]:

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

Out[15]:

nan

In [18]:

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

Out[18]:

```
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 [22]:

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

Out[22]:

111

In [23]:

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

Out[23]:

111

In [20]:

1 covid_copy.iloc[111]

Out[20]:

```

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

```

In [21]:

1 covid_copy.iloc[110]

Out[21]:

```

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

```

In []:

1

In [24]:

1 covid_copy

Out[24]:

	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 [25]:

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

Out[25]:

	date	new_cases	new_deaths	new_tests
51	2020-02-20	0.0	0.0	NaN
17	2020-01-17	0.0	0.0	NaN
10	2020-01-10	0.0	0.0	NaN
76	2020-03-16	2823.0	370.0	NaN
72	2020-03-12	2313.0	196.0	NaN
68	2020-03-08	1247.0	36.0	NaN
212	2020-07-30	288.0	6.0	33396.0
101	2020-04-10	4204.0	612.0	NaN
128	2020-05-07	1444.0	369.0	13665.0
187	2020-07-05	235.0	21.0	21166.0
155	2020-06-03	318.0	55.0	20035.0
44	2020-02-13	0.0	0.0	NaN
202	2020-07-20	218.0	3.0	14121.0
59	2020-02-28	250.0	5.0	NaN
42	2020-02-11	0.0	0.0	NaN
132	2020-05-11	802.0	165.0	25823.0
215	2020-08-02	295.0	5.0	24496.0
65	2020-03-05	587.0	27.0	NaN
139	2020-05-18	675.0	145.0	26101.0
49	2020-02-18	0.0	0.0	NaN

In [26]:

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

In [27]:

```
1 total_new_cases
```

Out[27]:

271515.0

In [28]:

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

In [29]:

```
1 total_new_death
```

Out[29]:

35497.0

In [30]:

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

In [31]:

```
1 total_new_test
```

Out[31]:

4279456.0

In []:

```
1
```

In [32]:

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

In [33]:

```
1 death_rate
```

Out[33]:

13.073679170579894

In [34]:

```
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 [35]:

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

In [36]:

```
1 total_test_done
```

Out[36]:

5214766.0

In []:

```
1
```

In [37]:

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

In [38]:

```
1 positivity_rate
```

Out[38]:

5.206657403227681

In [39]:

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

Postivity Rate of Overall Test done is 5.21 %

In [43]:

```
1 covid_copy
```

Out[43]:

	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 [40]:

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


In [41]:

```
1 high_case
```

Out[41]:

```
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 [42]:

```
1 covid_copy[high_case]
```

Out[42]:

	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 [44]:

```
1 covid_copy
```

Out[44]:

	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 [45]:

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

In [46]:

```
1 high_death
```

Out[46]:

```
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 [47]:

1 covid_copy[high_death]

Out[47]:

	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 [48]:

1 covid_copy

Out[48]:

	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 []:

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

In [49]:

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

In [50]:

```
1 covid_copy
```

Out[50]:

	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 [53]:

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

Out[53]:

```
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 []:

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

In []:

```
1
```

In [54]:

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

In [56]:

```
1 covid_copy
```

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
...
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 []:

```
1
```

In [57]:

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

Out[57]:

1094.8185483870968

In [58]:

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

Out[58]:

143.13306451612902

In [59]:

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

Out[59]:

31699.674074074075

In []:

```
1
```

In [60]:

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

Out[60]:

	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 [61]:

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

Out[61]:

```
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 [62]:

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

In [63]:

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

Out[63]:

```
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 [64]:

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

Out[64]:

	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 [66]:

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

In [67]:

```
1 covid_copy
```

Out[67]:

	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 [68]:

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

In [69]:

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

Out[69]:

	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 [70]:

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

In [71]:

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

Out[71]:

	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 [72]:

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


In [75]:

```
1 covid_copy
```

Out[75]:

	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

In []:

```
1
```