In [1]:
```
from urllib.request import urlretrieve
```

In [2]:
```
italy_covid_url = 'https://gist.githubusercontent.com/aakashns/f6a004fa20c84fec53262f9a8

urlretrieve(italy_covid_url, 'italy-covid-daywise.csv')
```

Out[2]:
```
('italy-covid-daywise.csv', <http.client.HTTPMessage at 0x25532915f10>)
```

In [3]:
```
!pip install pandas --upgrade --quiet
```

```
ERROR: Could not install packages due to an OSError: [WinError 5] Access is denied:
'c:\\programdata\\anaconda3\\lib\\site-packages\\pandas-1.2.4.dist-info\\direct_url.jso
n'
Consider using the `--user` option or check the permissions.
```

In [4]:
```
import pandas as pd
```

In [5]:
```
covid_df = pd.read_csv('italy-covid-daywise.csv')
```

In [6]:
```
type(covid_df)
```

Out[6]:
```
pandas.core.frame.DataFrame
```

In [7]:
```
covid_df
```

Out[7]:

	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 [8]:
```
covid_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 [9]:
```
covid_df.describe()
```

Out[9]:

	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 [10]:
```
# Pandas format is simliar to this
covid_data_dict = {
    'date':       ['2020-08-30', '2020-08-31', '2020-09-01', '2020-09-02', '2020-09-03'],
    'new_cases':  [1444, 1365, 996, 975, 1326],
    'new_deaths': [1, 4, 6, 8, 6],
    'new_tests': [53541, 42583, 54395, None, None]
}
```

In [11]:
```
covid_df.columns
```

Out[11]:
```
Index(['date', 'new_cases', 'new_deaths', 'new_tests'], dtype='object')
```

In [12]:
```
covid_df.shape
```

Out[12]:
```
(248, 4)
```

In [13]:
```
covid_df.columns
```

Out[13]:
```
Index(['date', 'new_cases', 'new_deaths', 'new_tests'], dtype='object')
```

In [14]:
```
covid_df.shape
```

Out[14]:
```
(248, 4)
```

In [15]:
```
covid_data_dict['new_cases']
```

Out[15]:
```
[1444, 1365, 996, 975, 1326]
```

In [16]:
```
covid_df['new_cases']
```

Out[16]:
```
0         0.0
1         0.0
2         0.0
3         0.0
4         0.0
        ...
243    1444.0
244    1365.0
245     996.0
246     975.0
247    1326.0
Name: new_cases, Length: 248, dtype: float64
```

In [17]:
```
type(covid_df['new_cases'])
```

Out[17]:
```
pandas.core.series.Series
```

In [18]:
```
covid_df['new_cases'][246]
```

Out[18]:
```
975.0
```

In [19]:
```
covid_df['new_tests'][240]
```

Out[19]:
```
57640.0
```

In [20]:
```
covid_df.at[246, 'new_cases']
```

Out[20]:
```
975.0
```

In [21]:
```
covid_df.at[240, 'new_cases']
```

Out[21]:
```
1366.0
```

In [22]:
```
cases_df = covid_df[['date', 'new_cases']]
cases_df
```

Out[22]:

	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 [23]:
```
covid_df_copy = covid_df.copy()
```

In [24]:
```
covid_df_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]:
```
covid_df.loc[243]
```

Out[25]:
```
date          2020-08-30
new_cases         1444.0
new_deaths           1.0
new_tests        53541.0
Name: 243, dtype: object
```

In [26]:
```
type(covid_df.loc[243])
```

Out[26]:
```
pandas.core.series.Series
```

In [27]:
```
covid_df.head()
```

Out[27]:

	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 [28]:
```
covid_df.tail()
```

Out[28]:

	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 [29]:
```
covid_df.tail(10)
```

Out[29]:

	date	new_cases	new_deaths	new_tests
238	2020-08-25	953.0	4.0	45798.0
239	2020-08-26	876.0	4.0	58054.0
240	2020-08-27	1366.0	13.0	57640.0
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
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 [30]:
```
covid_df.head(3)
```

Out[30]:

	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

In [31]:
```
covid_df.head(8)
```

Out[31]:

	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
5	2020-01-05	0.0	0.0	NaN
6	2020-01-06	0.0	0.0	NaN
7	2020-01-07	0.0	0.0	NaN

In [32]:
```
covid_df.at[0, 'new_tests']
```

Out[32]:
```
nan
```

In [33]:
```
type(covid_df.at[0, 'new_cases'])
```

Out[33]:
```
numpy.float64
```

In [34]:
```
covid_df.loc[108:113]
```

Out[34]:

	date	new_cases	new_deaths	new_tests
108	2020-04-17	3786.0	525.0	NaN
109	2020-04-18	3493.0	575.0	NaN
110	2020-04-19	3491.0	480.0	NaN
111	2020-04-20	3047.0	433.0	7841.0
112	2020-04-21	2256.0	454.0	28095.0
113	2020-04-22	2729.0	534.0	44248.0

In [35]:
```
covid_df.sample(10)
```

Out[35]:

	date	new_cases	new_deaths	new_tests
114	2020-04-23	3370.0	437.0	37083.0
166	2020-06-14	346.0	78.0	29545.0
110	2020-04-19	3491.0	480.0	NaN
233	2020-08-20	642.0	7.0	49662.0
29	2020-01-29	0.0	0.0	NaN
60	2020-02-29	238.0	4.0	NaN
244	2020-08-31	1365.0	4.0	42583.0
187	2020-07-05	235.0	21.0	21166.0
19	2020-01-19	0.0	0.0	NaN
30	2020-01-30	0.0	0.0	NaN

In [36]:
```
covid_df.sample(5)
```

Out[36]:

	date	new_cases	new_deaths	new_tests
47	2020-02-16	0.0	0.0	NaN
82	2020-03-22	6557.0	795.0	NaN
188	2020-07-06	192.0	7.0	13771.0
169	2020-06-17	210.0	34.0	33957.0
124	2020-05-03	1900.0	474.0	27047.0

In [37]:
```
#Analyse the data from Data frames
total_cases = covid_df.new_cases.sum()
total_deaths = covid_df.new_deaths.sum()
```

In [38]:
```
print('The number of reported cases is {} and the number of reported deaths is {}.'.form
```

```
The number of reported cases is 271515 and the number of reported deaths is 35497.
```

In [39]:
```
total_cases = covid_df['new_cases'].sum()
print('The number of reported cases is {}.'.format(int(total_cases)))
```

```
The number of reported cases is 271515.
```

In [40]:
```
type(total_cases)
```

Out[40]:
```
numpy.float64
```

In [41]:
```
type(total_deaths)
```

Out[41]:
```
numpy.float64
```

In [ ]:

In [ ]:

In [ ]: