```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer
```

```
df=pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Datasets/owid-covid-data.csv')
df
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothe
0	AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	Na
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	Na
2	AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	Na
3	AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	Na
4	AFG	Asia	Afghanistan	2020-02-28	5.0	0.0	Na
...	...	...	...	...	...	...	
232321	ZWE	Africa	Zimbabwe	2022-10-30	257893.0	0.0	0
232322	ZWE	Africa	Zimbabwe	2022-10-31	257893.0	0.0	0
232323	ZWE	Africa	Zimbabwe	2022-11-01	257893.0	0.0	0
232324	ZWE	Africa	Zimbabwe	2022-11-02	257893.0	0.0	0
232325	ZWE	Africa	Zimbabwe	2022-11-03	257893.0	0.0	0

232326 rows × 67 columns

```
df.shape
```

```
(232326, 67)
```

```
df.head(10)
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smooth
0	AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	Na
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	Na

```
df.describe
```

```
<bound method NDFrame.describe of        iso_code continent     location        date  total_cases  new_cases  \
0           AFG      Asia  Afghanistan  2020-02-24          5.0        5.0
1           AFG      Asia  Afghanistan  2020-02-25          5.0        0.0
2           AFG      Asia  Afghanistan  2020-02-26          5.0        0.0
3           AFG      Asia  Afghanistan  2020-02-27          5.0        0.0
4           AFG      Asia  Afghanistan  2020-02-28          5.0        0.0
...         ...       ...          ...         ...          ...        ...
232321      ZWE    Africa     Zimbabwe  2022-10-30     257893.0        0.0
232322      ZWE    Africa     Zimbabwe  2022-10-31     257893.0        0.0
232323      ZWE    Africa     Zimbabwe  2022-11-01     257893.0        0.0
232324      ZWE    Africa     Zimbabwe  2022-11-02     257893.0        0.0
232325      ZWE    Africa     Zimbabwe  2022-11-03     257893.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
0                      NaN           NaN         NaN                  NaN
1                      NaN           NaN         NaN                  NaN
2                      NaN           NaN         NaN                  NaN
3                      NaN           NaN         NaN                  NaN
4                      NaN           NaN         NaN                  NaN
...                    ...           ...         ...                  ...
232321                 0.0        5606.0         0.0                  0.0
232322                 0.0        5606.0         0.0                  0.0
232323                 0.0        5606.0         0.0                  0.0
232324                 0.0        5606.0         0.0                  0.0
232325                 0.0        5606.0         0.0                  0.0

        ...  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0       ...           NaN                  37.746                         0.5
1       ...           NaN                  37.746                         0.5
2       ...           NaN                  37.746                         0.5
3       ...           NaN                  37.746                         0.5
4       ...           NaN                  37.746                         0.5
...     ...           ...                     ...                         ...
232321  ...          30.7                  36.791                         1.7
232322  ...          30.7                  36.791                         1.7
232323  ...          30.7                  36.791                         1.7
232324  ...          30.7                  36.791                         1.7
232325  ...          30.7                  36.791                         1.7

        life_expectancy  human_development_index  population  \
0                 64.83                    0.511  41128772.0
1                 64.83                    0.511  41128772.0
2                 64.83                    0.511  41128772.0
3                 64.83                    0.511  41128772.0
4                 64.83                    0.511  41128772.0
...                 ...                      ...         ...
232321            61.49                    0.571  16320539.0
232322            61.49                    0.571  16320539.0
232323            61.49                    0.571  16320539.0
232324            61.49                    0.571  16320539.0
232325            61.49                    0.571  16320539.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                        NaN                          NaN
1                                        NaN                          NaN
2                                        NaN                          NaN
3                                        NaN                          NaN
4                                        NaN                          NaN
```

```
df.drop(['new_deaths_smoothed','new_cases_per_million','total_cases_per_million','new_cases_smoothed'],axis=1,inplace=True)
```

```
df.shape
```

```
(232326, 63)
```

```
df.rename(columns={'date':'Date','location':'Country','continent':'Continent','iso_code':'ISO_code'},inplace=True)
df
```

	ISO_code	Continent	Country	Date	total_cases	new_cases	total_death
0	AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	Na
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	Na
2	AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	Na
3	AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	Na
4	AFG	Asia	Afghanistan	2020-02-28	5.0	0.0	Na
...	...	...	...	...	...	...	.
232321	ZWE	Africa	Zimbabwe	2022-10-30	257893.0	0.0	5606.
232322	ZWE	Africa	Zimbabwe	2022-10-31	257893.0	0.0	5606.
232323	ZWE	Africa	Zimbabwe	2022-11-01	257893.0	0.0	5606.
232324	ZWE	Africa	Zimbabwe	2022-11-02	257893.0	0.0	5606.
232325	ZWE	Africa	Zimbabwe	2022-11-03	257893.0	0.0	5606.

232326 rows × 63 columns

```
continent_unique=list(df.Continent.unique())
continent_unique
```

```
['Asia', nan, 'Europe', 'Africa', 'North America', 'South America', 'Oceania']
```

```
#simple Imputer
imputer=SimpleImputer(strategy='constant')
df2=pd.DataFrame(imputer.fit_transform(df),columns=df.columns)
```

```
df3 = df2.groupby(['Date','Country',])[['Date','Country','total_cases','total_deaths','total_vaccinations']].sum().reset_index(drop=True)
```

```
df3
```

	Date	Country	total_cases	total_deaths	total_vaccinations
0	2020-01-01	Argentina	missing_value	missing_value	missing_value
1	2020-01-01	Mexico	missing_value	missing_value	missing_value
2	2020-01-02	Argentina	missing_value	missing_value	missing_value
3	2020-01-02	Mexico	missing_value	missing_value	missing_value
4	2020-01-03	Argentina	missing_value	missing_value	missing_value
...	...	...	...	...	...
232321	2022-11-03	Wallis and Futuna	3415.0	7.0	missing_value
232322	2022-11-03	World	631707818.0	6597533.0	missing_value

```
df3['total_cases'].replace({'missing_value':0},inplace=True)
```

```
df3
```

	Date	Country	total_cases	total_deaths	total_vaccinations
0	2020-01-01	Argentina	0.0	missing_value	missing_value
1	2020-01-01	Mexico	0.0	missing_value	missing_value
2	2020-01-02	Argentina	0.0	missing_value	missing_value
3	2020-01-02	Mexico	0.0	missing_value	missing_value
4	2020-01-03	Argentina	0.0	missing_value	missing_value
...	...	...	...	...	...
232321	2022-11-03	Wallis and Futuna	3415.0	7.0	missing_value
232322	2022-11-03	World	631707818.0	6597533.0	missing_value

```
df3['total_deaths'].replace({'missing_value':0},inplace=True)
df3['total_vaccinations'].replace({'missing_value':0},inplace=True)
```

```
df3
```

	Date	Country	total_cases	total_deaths	total_vaccinations
0	2020-01-01	Argentina	0.0	0.0	0.0
1	2020-01-01	Mexico	0.0	0.0	0.0
2	2020-01-02	Argentina	0.0	0.0	0.0
3	2020-01-02	Mexico	0.0	0.0	0.0
4	2020-01-03	Argentina	0.0	0.0	0.0
...	...	...	...	...	...
232321	2022-11-03	Wallis and Futuna	3415.0	7.0	0.0
232322	2022-11-03	World	631707818.0	6597533.0	0.0
232323	2022-11-03	Yemen	11944.0	2158.0	0.0

```
df4=df3[df3['total_deaths']>1000000]
countries=df4['Country'].unique()
len(countries)
```

```
10
```

```
country_deaths_graterthan1000000 = list(df4.Country.unique())
country_deaths_graterthan1000000
```

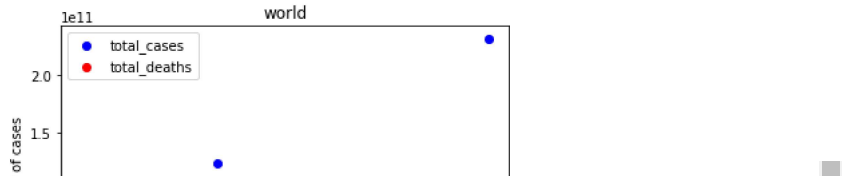
```
['World',
 'High income',
 'Upper middle income',
 'Europe',
 'South America',
 'Asia',
 'Lower middle income',
 'North America',
 'European Union',
 'United States']
```

```
for idx in range(0,len(countries)):
  c = df4[df4['Country']==countries[idx]].reset_index(drop=True)
  plt.scatter(np.arange(0,len(c)),c['total_cases'],color='blue',label='total_cases')
  plt.scatter(np.arange(0,len(c)),c['total_deaths'],color='red',label='total_deaths')
  plt.scatter(np.arange(0,len(c)),c['total_vaccinations'],color='green',label='total_vaccinations')
  plt.title(countries[idx])
```

```
    plt.xlabel("Number of days since first suspect")
    plt.ylabel("Number of cases")
    plt.legend()
    plt.show()
```

```
df5=df4.groupby(['Country'])[['Country','total_cases','total_deaths']].sum().reset_index(drop=True)
```

```
c = df5
plt.scatter(np.arange(0,len(c)),c['total_cases'],color='blue',label='total_cases')
plt.scatter(np.arange(0,len(c)),c['total_deaths'],color='red',label='total_deaths')
plt.title('world')
plt.xlabel('Number of days since first suspect')
plt.ylabel('Number of cases')
plt.legend()
plt.show()
```



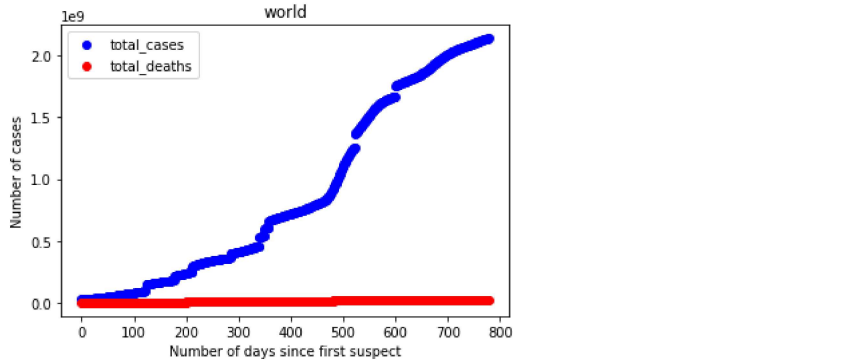

```
date = df4['Date'].unique()
len(date)
```

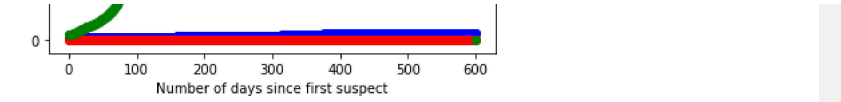
```
779
```

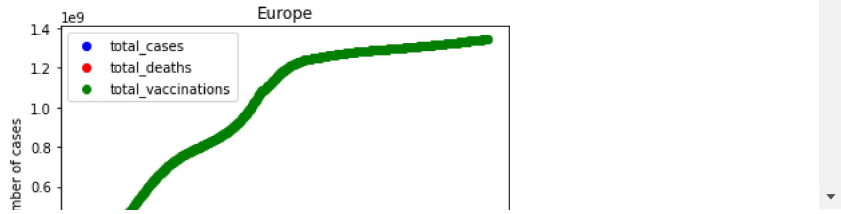
```
df6 = df4.groupby(['Date'])[['Date','total_cases','total_deaths']].sum().reset_index(drop=True)
```

High income

```
c= df6
plt.scatter(np.arange(0,len(c)),c['total_cases'],color='blue',label='total_cases')
plt.scatter(np.arange(0,len(c)),c['total_deaths'],color='red',label='total_deaths')
plt.title('world')
plt.xlabel('Number of days since first suspect')
plt.ylabel('Number of cases')
plt.legend()
plt.show()
```







Colab paid products - Cancel contracts here