**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare

# Covid Dataset

In [1]:
```
# For mathematical operations
import numpy as np

# For handling the data
import pandas as pd

# For plotting a graph
import matplotlib.pyplot as plt

# For labelling the categorical data
from sklearn.preprocessing import LabelEncoder
```

In [2]:
```
# Reading the file from the directory/path
covid_data = pd.read_csv("C:/Users/DHANRAJ008/Downloads/COVID/covid_19_clean_complete.csv")
```

In [3]:
```
# Displaying the data
covid_data
```

Out[3]:

| | Province/State | Country/Region | Lat | Long | Date | Confirmed | Deaths | Recovered | Active | WHO Region |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | Afghanistan | 33.939110 | 67.709953 | 2020-01-22 | 0 | 0 | 0 | 0 | Eastern Mediterranean |
| 1 | NaN | Albania | 41.153300 | 20.168300 | 2020-01-22 | 0 | 0 | 0 | 0 | Europe |
| 2 | NaN | Algeria | 28.033900 | 1.659600 | 2020-01-22 | 0 | 0 | 0 | 0 | Africa |
| 3 | NaN | Andorra | 42.506300 | 1.521800 | 2020-01-22 | 0 | 0 | 0 | 0 | Europe |
| 4 | NaN | Angola | -11.202700 | 17.873900 | 2020-01-22 | 0 | 0 | 0 | 0 | Africa |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 49063 | NaN | Sao Tome and Principe | 0.186400 | 6.613100 | 2020-07-27 | 865 | 14 | 734 | 117 | Africa |
| 49064 | NaN | Yemen | 15.552727 | 48.516388 | 2020-07-27 | 1691 | 483 | 833 | 375 | Eastern Mediterranean |
| 49065 | NaN | Comoros | -11.645500 | 43.333300 | 2020-07-27 | 354 | 7 | 328 | 19 | Africa |
| 49066 | NaN | Tajikistan | 38.861000 | 71.276100 | 2020-07-27 | 7235 | 60 | 6028 | 1147 | Europe |
| 49067 | NaN | Lesotho | -29.610000 | 28.233600 | 2020-07-27 | 505 | 12 | 128 | 365 | Africa |

In [4]:
```
# Checking the correlation of the data
covid_data.corr()
```

Out[4]:

| | Lat | Long | Confirmed | Deaths | Recovered | Active |
|---|---|---|---|---|---|---|
| Lat | 1.000000 | -0.127259 | 0.036665 | 0.070040 | 0.015329 | 0.044392 |
| Long | -0.127259 | 1.000000 | -0.078911 | -0.101340 | -0.052391 | -0.085688 |
| Confirmed | 0.036665 | -0.078911 | 1.000000 | 0.912361 | 0.895506 | 0.950255 |
| Deaths | 0.070040 | -0.101340 | 0.912361 | 1.000000 | 0.763090 | 0.891858 |
| Recovered | 0.015329 | -0.052391 | 0.895506 | 0.763090 | 1.000000 | 0.713088 |
| Active | 0.044392 | -0.085688 | 0.950255 | 0.891858 | 0.713088 | 1.000000 |

In [5]:
```
# Printing first 10 elements of the data set
covid_data.head(10)
```

Out[5]:

| | Province/State | Country/Region | Lat | Long | Date | Confirmed | Deaths | Recovered | Active | WHO Region |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | Afghanistan | 33.93911 | 67.709953 | 2020-01-22 | 0 | 0 | 0 | 0 | Eastern Mediterranean |
| 1 | NaN | Albania | 41.15330 | 20.168300 | 2020-01-22 | 0 | 0 | 0 | 0 | Europe |
| 2 | NaN | Algeria | 28.03390 | 1.659600 | 2020-01-22 | 0 | 0 | 0 | 0 | Africa |
| 3 | NaN | Andorra | 42.50630 | 1.521800 | 2020-01-22 | 0 | 0 | 0 | 0 | Europe |
| 4 | NaN | Angola | -11.20270 | 17.873900 | 2020-01-22 | 0 | 0 | 0 | 0 | Africa |
| 5 | NaN | Antigua and Barbuda | 17.06080 | -61.796400 | 2020-01-22 | 0 | 0 | 0 | 0 | Americas |
| 6 | NaN | Argentina | -38.41610 | -63.616700 | 2020-01-22 | 0 | 0 | 0 | 0 | Americas |
| 7 | NaN | Armenia | 40.06910 | 45.038200 | 2020-01-22 | 0 | 0 | 0 | 0 | Europe |
| 8 | Australian Capital Territory | Australia | -35.47350 | 149.012400 | 2020-01-22 | 0 | 0 | 0 | 0 | Western Pacific |
| 9 | New South Wales | Australia | -33.86880 | 151.209300 | 2020-01-22 | 0 | 0 | 0 | 0 | Western Pacific |

**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare

In [6]:
```
# Printing last 10 elements of the data set
covid_data.tail(10)
```

Out[6]:

| | Province/State | Country/Region | Lat | Long | Date | Confirmed | Deaths | Recovered | Active | WHO Region |
|---|---|---|---|---|---|---|---|---|---|---|
| 49058 | NaN | Malawi | -13.254300 | 34.301500 | 2020-07-27 | 3664 | 99 | 1645 | 1920 | Africa |
| 49059 | Falkland Islands (Malvinas) | United Kingdom | -51.796300 | -59.523600 | 2020-07-27 | 13 | 0 | 13 | 0 | Europe |
| 49060 | Saint Pierre and Miquelon | France | 46.885200 | -56.315900 | 2020-07-27 | 4 | 0 | 1 | 3 | Europe |
| 49061 | NaN | South Sudan | 6.877000 | 31.307000 | 2020-07-27 | 2305 | 46 | 1175 | 1084 | Africa |
| 49062 | NaN | Western Sahara | 24.215500 | -12.885800 | 2020-07-27 | 10 | 1 | 8 | 1 | Africa |
| 49063 | NaN | Sao Tome and Principe | 0.186400 | 6.613100 | 2020-07-27 | 865 | 14 | 734 | 117 | Africa |
| 49064 | NaN | Yemen | 15.552727 | 48.516388 | 2020-07-27 | 1691 | 483 | 833 | 375 | Eastern Mediterranean |
| 49065 | NaN | Comoros | -11.645500 | 43.333300 | 2020-07-27 | 354 | 7 | 328 | 19 | Africa |
| 49066 | NaN | Tajikistan | 38.861000 | 71.276100 | 2020-07-27 | 7235 | 60 | 6028 | 1147 | Europe |
| 49067 | NaN | Lesotho | -29.610000 | 28.233600 | 2020-07-27 | 505 | 12 | 128 | 365 | Africa |

In [7]:
```
# Printing total records of the data set
print(f"Number of records in the dataset is {len(covid_data)}")
```

```
Number of records in the dataset is 49068
```

In [8]:
```
# Printing columns of the data set
print(f"Number of columns in the dataset is {len(covid_data.columns)} \n namely {covid_data.columns}")
```

```
Number of columns in the dataset is 10
 namely Index(['Province/State', 'Country/Region', 'Lat', 'Long', 'Date', 'Confirmed',
       'Deaths', 'Recovered', 'Active', 'WHO Region'],
      dtype='object')
```

In [9]:
```
# Label encoding the data set
lbe = LabelEncoder()
covid_data["WHO Region"] = lbe.fit_transform(covid_data["WHO Region"])
```

In [10]:
```
# Displaying the encoded dataset
covid_data
```

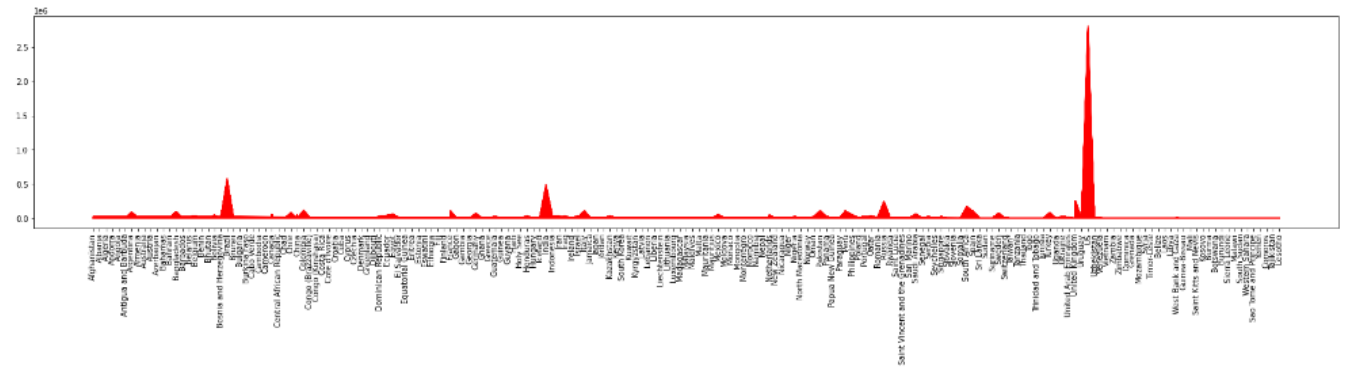
Out[10]:

| | Province/State | Country/Region | Lat | Long | Date | Confirmed | Deaths | Recovered | Active | WHO Region |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | Afghanistan | 33.939110 | 67.709953 | 2020-01-22 | 0 | 0 | 0 | 0 | 2 |
| 1 | NaN | Albania | 41.153300 | 20.168300 | 2020-01-22 | 0 | 0 | 0 | 0 | 3 |
| 2 | NaN | Algeria | 28.033900 | 1.659600 | 2020-01-22 | 0 | 0 | 0 | 0 | 0 |
| 3 | NaN | Andorra | 42.506300 | 1.521800 | 2020-01-22 | 0 | 0 | 0 | 0 | 3 |
| 4 | NaN | Angola | -11.202700 | 17.873900 | 2020-01-22 | 0 | 0 | 0 | 0 | 0 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 49063 | NaN | Sao Tome and Principe | 0.186400 | 6.613100 | 2020-07-27 | 865 | 14 | 734 | 117 | 0 |
| 49064 | NaN | Yemen | 15.552727 | 48.516388 | 2020-07-27 | 1691 | 483 | 833 | 375 | 2 |
| 49065 | NaN | Comoros | -11.645500 | 43.333300 | 2020-07-27 | 354 | 7 | 328 | 19 | 0 |
| 49066 | NaN | Tajikistan | 38.861000 | 71.276100 | 2020-07-27 | 7235 | 60 | 6028 | 1147 | 3 |
| 49067 | NaN | Lesotho | -29.610000 | 28.233600 | 2020-07-27 | 505 | 12 | 128 | 365 | 0 |

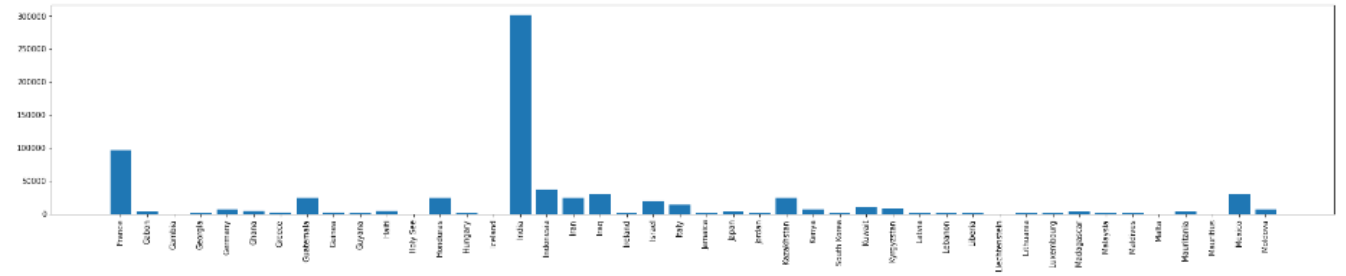
49068 rows × 10 columns

**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare

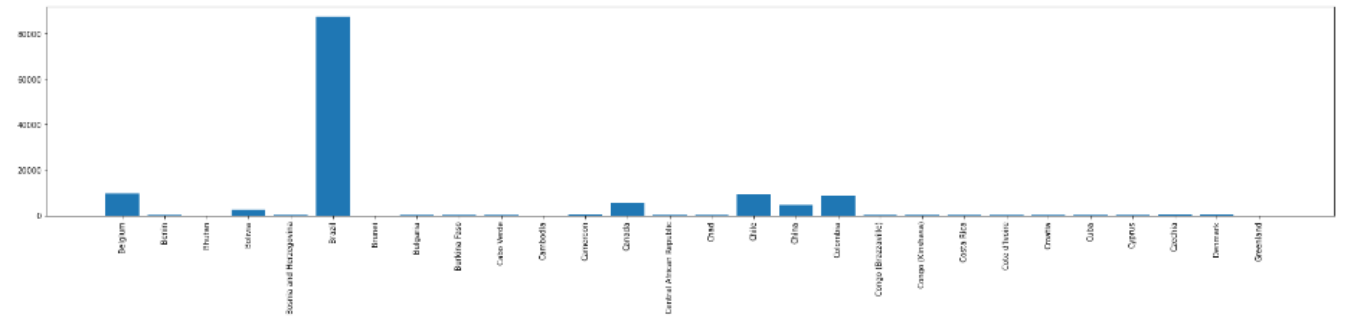
In [11]:
```
# Plotting the graph
fig = plt.figure()
fig.set_size_inches(30,5)
plt.xticks(rotation="vertical")
plt.plot(covid_data["Country/Region"],covid_data["Active"], color ="red")
plt.show()
```



In [12]:
```
# Bar graph of country/region to active cases
fig2 = plt.figure()
fig2.set_size_inches(30,5)
country = covid_data["Country/Region"].iloc[45000:45050]
active = covid_data["Active"].iloc[45000:45050]
plt.xticks(rotation="vertical")
plt.bar(country,active)
plt.show()
```

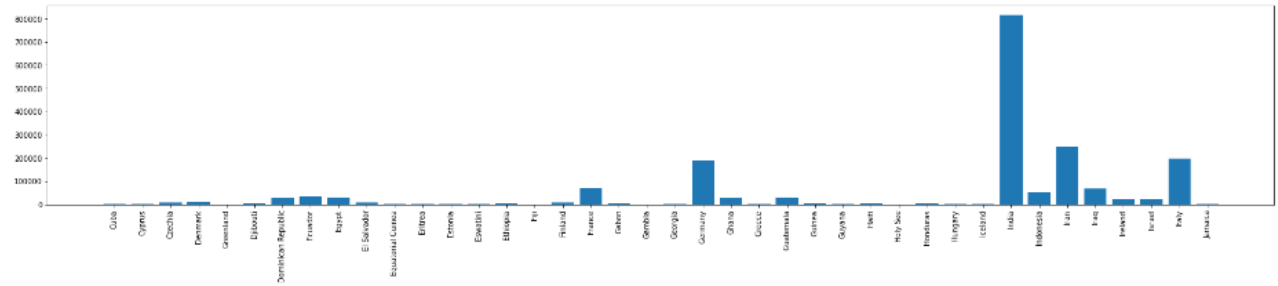


In [13]:
```
# Bar graph of country/region to deaths
fig2 = plt.figure()
fig2.set_size_inches(30,5)
country = covid_data["Country/Region"].iloc[48830:48900]
deaths = covid_data["Deaths"].iloc[48830:48900]
plt.xticks(rotation="vertical")
plt.bar(country,deaths)
plt.show()
```

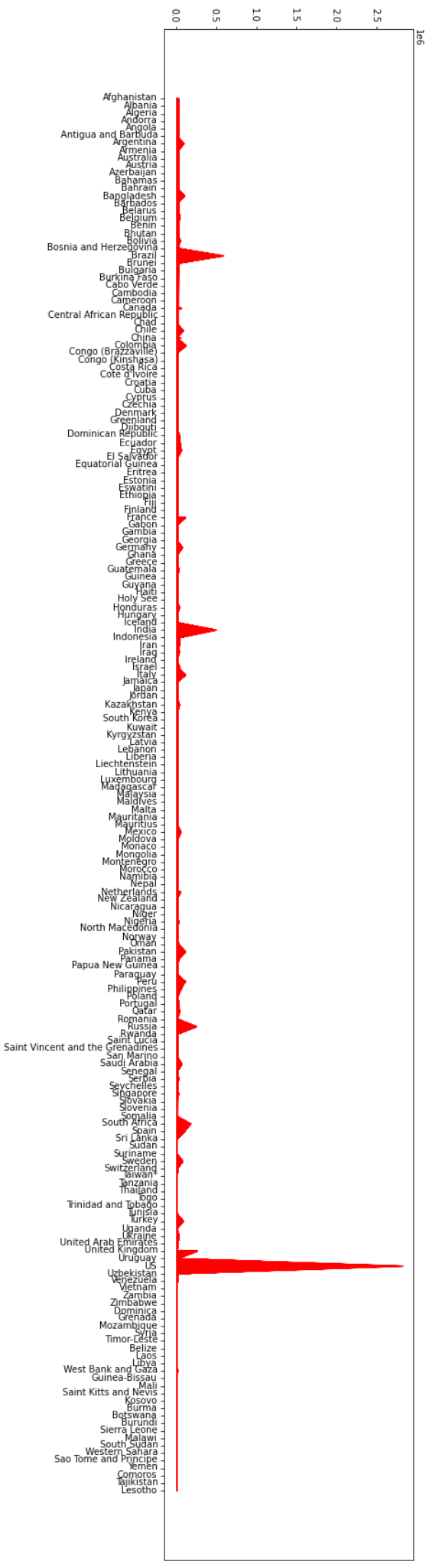


**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare

In [14]:
```
# Bar graph of country/region to recovered cases
fig3 = plt.figure()
fig3.set_size_inches(30,5)
country = covid_data["Country/Region"].iloc[47850:47900]
recovered = covid_data["Recovered"].iloc[47850:47900]
plt.xticks(rotation="vertical")
plt.bar(country,recovered)
plt.show()
```



**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare



**Roll No:** 2193301 **Name:** Dhanraj S Bhanusghare