```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import LabelEncoder
```

```python
df = pd.read_csv("/covid_19_clean_complete.csv")
```

```python
df
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths
0	NaN	Afghanistan	33.939110	67.709953	2020-01-22	0	0
1	NaN	Albania	41.153300	20.168300	2020-01-22	0	0
2	NaN	Algeria	28.033900	1.659600	2020-01-22	0	0
3	NaN	Andorra	42.506300	1.521800	2020-01-22	0	0
4	NaN	Angola	-11.202700	17.873900	2020-01-22	0	0
...	...	...	...	...	...	...	...
49063	NaN	Sao Tome and Principe	0.186400	6.613100	2020-07-27	865	14
49064	NaN	Yemen	15.552727	48.516388	2020-07-27	1691	483

```python
df.shape
```

```
(49068, 10)
```

```python
df.corr()
```

	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

```
df.head()
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	Rec
0	NaN	Afghanistan	33.93911	67.709953	2020-01-22	0	0	
1	NaN	Albania	41.15330	20.168300	2020-01-22	0	0	
2	NaN	Algeria	28.03390	1.659600	2020-01-22	0	0	

```
df.tail(10)
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths
49058	NaN	Malawi	-13.254300	34.301500	2020-07-27	3664	9!
49059	Falkland Islands (Malvinas)	United Kingdom	-51.796300	-59.523600	2020-07-27	13	(
49060	Saint Pierre and Miquelon	France	46.885200	-56.315900	2020-07-27	4	(
49061	NaN	South Sudan	6.877000	31.307000	2020-07-27	2305	4(
49062	NaN	Western Sahara	24.215500	-12.885800	2020-07-27	10	
49063	NaN	Sao Tome and Principe	0.186400	6.613100	2020-07-27	865	1-

```
print(f"Number of rows in the dataset is {len(df)}")
```

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

```
print(f"Number of attributes in the dataset is {len(df.columns)} \n: {df.columns}")
```

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

```
label = LabelEncoder()
df["WHO Region"] = label.fit_transform(df["WHO Region"])
```

```
df
```

|	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths
0	NaN	Afghanistan	33.939110	67.709953	2020-01-22	0	(
1	NaN	Albania	41.153300	20.168300	2020-01-22	0	(
2	NaN	Algeria	28.033900	1.659600	2020-01-22	0	(
3	NaN	Andorra	42.506300	1.521800	2020-01-22	0	(
4	NaN	Angola	-11.202700	17.873900	2020-01-22	0	(
...	...	...	...	...	...	...	..
49063	NaN	Sao Tome and Principe	0.186400	6.613100	2020-07-27	865	1

```
#plot the graph
fig = plt.figure()
fig.set_size_inches(30,10)
plt.xticks(rotation="vertical")
plt.plot(df["Country/Region"], df["Active"], color = "red")
plt.show()
```

1e6

```
#bar graph for active cases
#plot the graph
fig = plt.figure()
fig.set_size_inches(30,10)

plt.xticks(rotation="vertical")
plt.bar(df["Country/Region"].iloc[30000:30050], df["Active"].iloc[30000:30050])
plt.show()
```
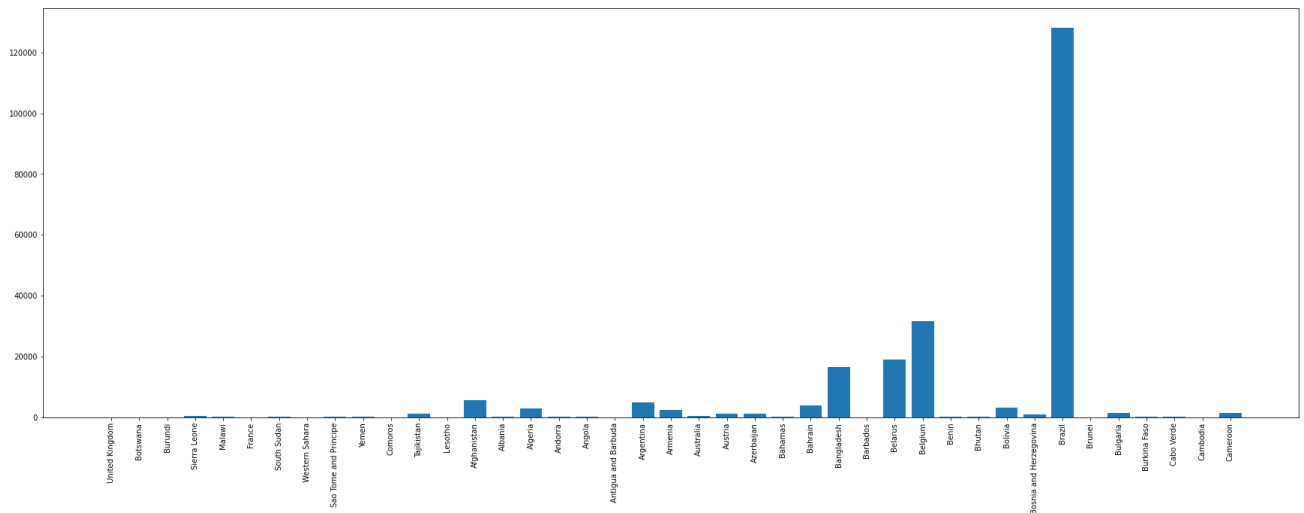


```
#bar graph for deaths
#plot the graph
fig = plt.figure()
fig.set_size_inches(30,5)
plt.xticks(rotation="vertical")
plt.bar(df["Country/Region"].iloc[20000:20050], df["Deaths"].iloc[20000:20050], color = "r
plt.show()
```
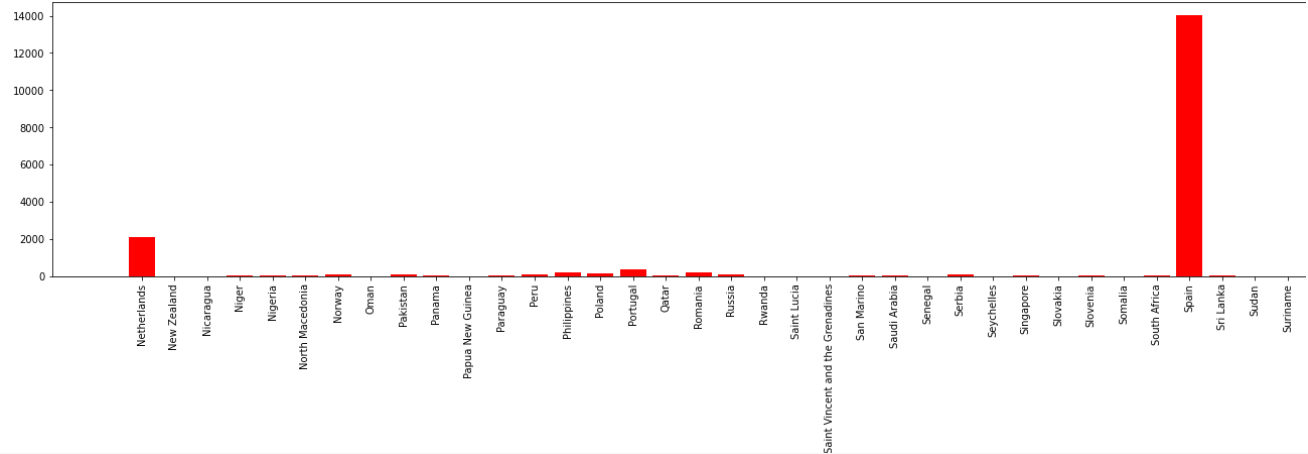
```
#bar graph for recovered cases
fig = plt.figure()
fig.set_size_inches(30,5)
plt.xticks(rotation="vertical")
plt.bar(df["Country/Region"].iloc[15000:15050], df["Recovered"].iloc[15000:15050], color =
plt.show()
```
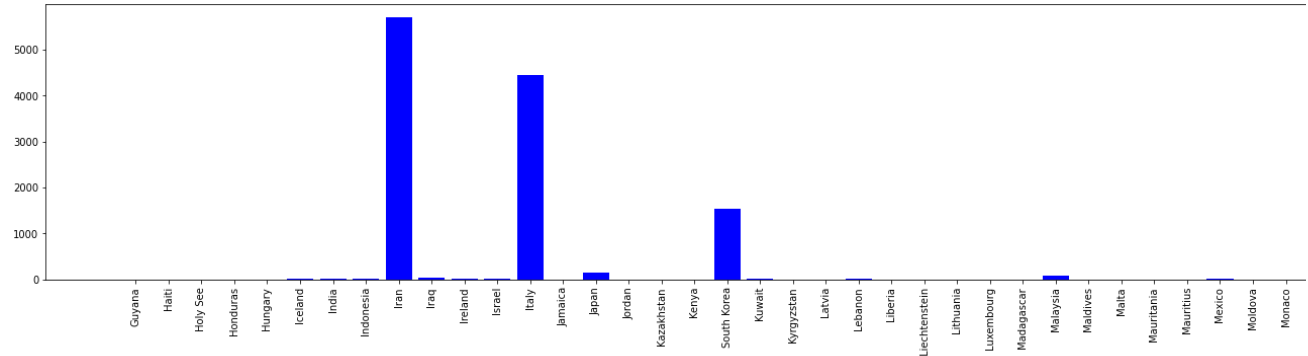
Colab paid products - Cancel contracts here