# Covid Analysis

Link to the Dataset - https://www.kaggle.com/datasets/imdevskp/corona-virus-report (https://www.kaggle.com/datasets/imdevskp/corona-virus-report)

This Dataset includes the Covid Data of seven Months from January 2020 to September 2020

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

In [2]:

```
covid = pd.read_csv('Covid_data.csv')
```

In [3]:

```
covid.head()
```

Out[3]:

| | Date | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered | W |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2020-01-22 | Afghanistan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | M |
| **1** | 2020-01-22 | Albania | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **2** | 2020-01-22 | Algeria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **3** | 2020-01-22 | Andorra | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **4** | 2020-01-22 | Angola | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

In [4]:

```
covid.tail()
```

Out[4]:

| | Date | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered |
|---|---|---|---|---|---|---|---|---|---|
| **35151** | 2020-07-27 | West Bank and Gaza | 10621 | 78 | 3752 | 6791 | 152 | 2 | 0 |
| **35152** | 2020-07-27 | Western Sahara | 10 | 1 | 8 | 1 | 0 | 0 | 0 |
| **35153** | 2020-07-27 | Yemen | 1691 | 483 | 833 | 375 | 10 | 4 | 36 |
| **35154** | 2020-07-27 | Zambia | 4552 | 140 | 2815 | 1597 | 71 | 1 | 465 |
| **35155** | 2020-07-27 | Zimbabwe | 2704 | 36 | 542 | 2126 | 192 | 2 | 24 |

In [5]:

```
covid.describe()
```

Out[5]:

| | Confirmed | Deaths | Recovered | Active | New cases | New deaths |
|---|---|---|---|---|---|---|
| **count** | 3.515600e+04 | 35156.000000 | 3.515600e+04 | 3.515600e+04 | 35156.00000 | 35156.000000 |
| **mean** | 2.356663e+04 | 1234.068239 | 1.104813e+04 | 1.128443e+04 | 469.36375 | 18.603339 |
| **std** | 1.499818e+05 | 7437.238354 | 6.454640e+04 | 8.997149e+04 | 3005.86754 | 115.706351 |
| **min** | 0.000000e+00 | 0.000000 | 0.000000e+00 | -2.000000e+00 | 0.00000 | -1918.000000 |
| **25%** | 1.000000e+00 | 0.000000 | 0.000000e+00 | 0.000000e+00 | 0.00000 | 0.000000 |
| **50%** | 2.500000e+02 | 4.000000 | 3.300000e+01 | 8.500000e+01 | 2.00000 | 0.000000 |
| **75%** | 3.640250e+03 | 78.250000 | 1.286250e+03 | 1.454000e+03 | 75.00000 | 1.000000 |
| **max** | 4.290259e+06 | 148011.000000 | 1.846641e+06 | 2.816444e+06 | 77255.00000 | 3887.000000 |

In [6]:

```
covid.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35156 entries, 0 to 35155
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype 
---  ------          --------------  ----- 
 0   Date            35156 non-null  object
 1   Country/Region  35156 non-null  object
 2   Confirmed       35156 non-null  int64 
 3   Deaths          35156 non-null  int64 
 4   Recovered       35156 non-null  int64 
 5   Active          35156 non-null  int64 
 6   New cases       35156 non-null  int64 
 7   New deaths      35156 non-null  int64 
 8   New recovered   35156 non-null  int64 
 9   WHO Region      35156 non-null  object
dtypes: int64(7), object(3)
memory usage: 2.7+ MB
```

In [7]:

```
covid['Date'] = pd.to_datetime(covid['Date'])
covid.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35156 entries, 0 to 35155
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype         
---  ------          --------------  -----         
 0   Date            35156 non-null  datetime64[ns]
 1   Country/Region  35156 non-null  object        
 2   Confirmed       35156 non-null  int64         
 3   Deaths          35156 non-null  int64         
 4   Recovered       35156 non-null  int64         
 5   Active          35156 non-null  int64         
 6   New cases       35156 non-null  int64         
 7   New deaths      35156 non-null  int64         
 8   New recovered   35156 non-null  int64         
 9   WHO Region      35156 non-null  object        
dtypes: datetime64[ns](1), int64(7), object(2)
memory usage: 2.7+ MB
```

In [8]:

```
covid.isnull().sum()
```

Out[8]:

```
Date              0
Country/Region    0
Confirmed         0
Deaths            0
Recovered         0
Active            0
New cases         0
New deaths        0
New recovered     0
WHO Region        0
dtype: int64
```

In [9]:

```
covid.columns
```

Out[9]:

```
Index(['Date', 'Country/Region', 'Confirmed', 'Deaths', 'Recovered', 'Activ
e',
       'New cases', 'New deaths', 'New recovered', 'WHO Region'],
      dtype='object')
```

In [10]:

```
covid['Date'].min()
```

Out[10]:

```
Timestamp('2020-01-22 00:00:00')
```

In [11]:

```
covid['Date'].max()
```

Out[11]:

```
Timestamp('2020-07-27 00:00:00')
```

In [12]:

```
cvd = covid.copy()
```

In [13]:

```
cvd.head()
```

Out[13]:

| | Date | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered | W |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2020-01-22 | Afghanistan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Me |
| **1** | 2020-01-22 | Albania | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **2** | 2020-01-22 | Algeria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **3** | 2020-01-22 | Andorra | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **4** | 2020-01-22 | Angola | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

In [14]:

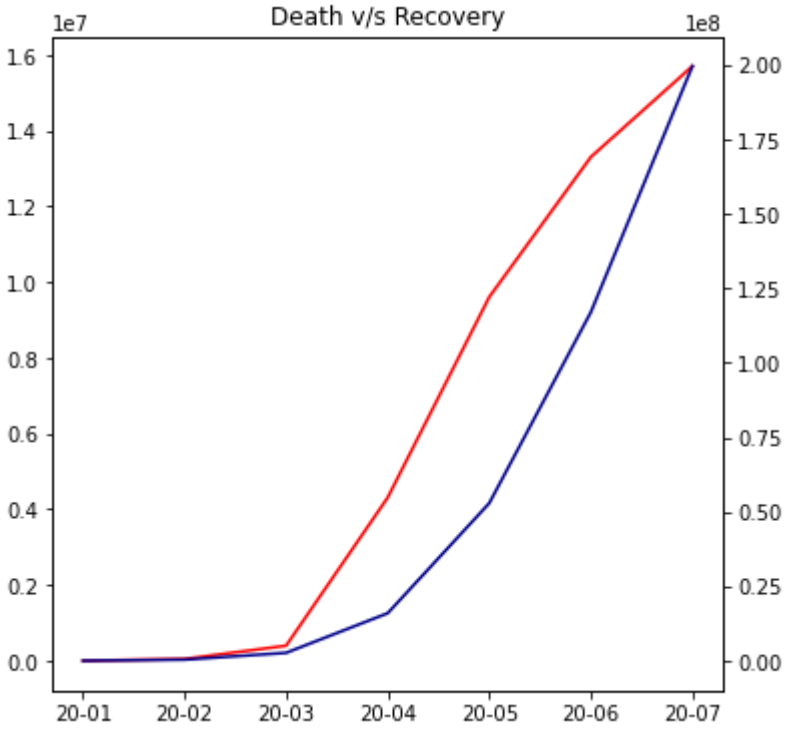
```
cvd['Month Year'] = cvd['Date'].apply(lambda x: x.strftime('%y-%m'))
cvd.head()
```

Out[14]:

| | Date | Country/Region | Confirmed | Deaths | Recovered | Active | New cases | New deaths | New recovered | W |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 2020-01-22 | Afghanistan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Me |
| **1** | 2020-01-22 | Albania | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **2** | 2020-01-22 | Algeria | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **3** | 2020-01-22 | Andorra | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **4** | 2020-01-22 | Angola | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

In [15]:

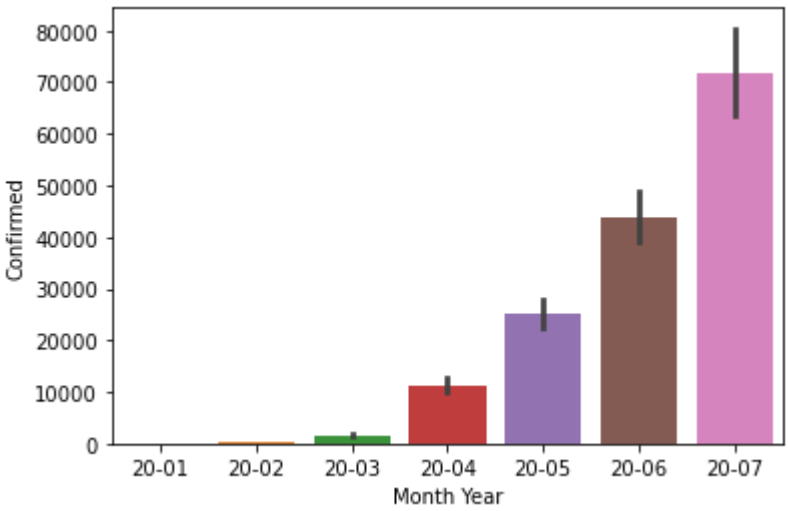
```
Monthly_d = cvd.groupby('Month Year').sum()['Deaths'].reset_index()
Monthly_r = cvd.groupby('Month Year').sum()['Recovered'].reset_index()

fig, ax1 = plt.subplots(figsize=(6, 6))
ax2 = ax1.twinx()

ax1.plot(Monthly_d['Month Year'], Monthly_d['Deaths'], color = 'red')
ax2.plot(Monthly_r['Month Year'], Monthly_r['Recovered'], color = 'darkblue')
plt.title('Death v/s Recovery')

plt.show()
```




We can see that in this graph the recovery rate is less than the death rate for first 7 months

In [16]:

```
#Mean cases - Monthly (sns barplot)
sns.barplot(cvd['Month Year'], cvd['Confirmed'])
plt.show()
```




In this bar chart we can see that the cases we gradually increasing during first three months of Covid but then there was rapid increase in the cases

In [17]:

```
d = cvd.groupby('Country/Region')['Deaths'].max().sort_values(ascending = False).head(10)
pd.DataFrame(d)
```

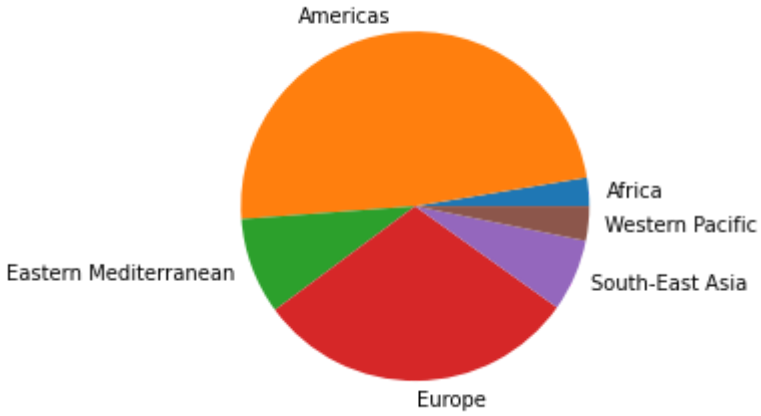
Out[17]:

| Country/Region | Deaths |
|---|---|
| **US** | 148011 |
| **Brazil** | 87618 |
| **United Kingdom** | 45844 |
| **Mexico** | 44022 |
| **Italy** | 35112 |
| **India** | 33408 |
| **France** | 30212 |
| **Spain** | 28752 |
| **Peru** | 18418 |
| **Iran** | 15912 |

The most death occoured in USA during first seven months followes by Brazil and UK

In [18]:

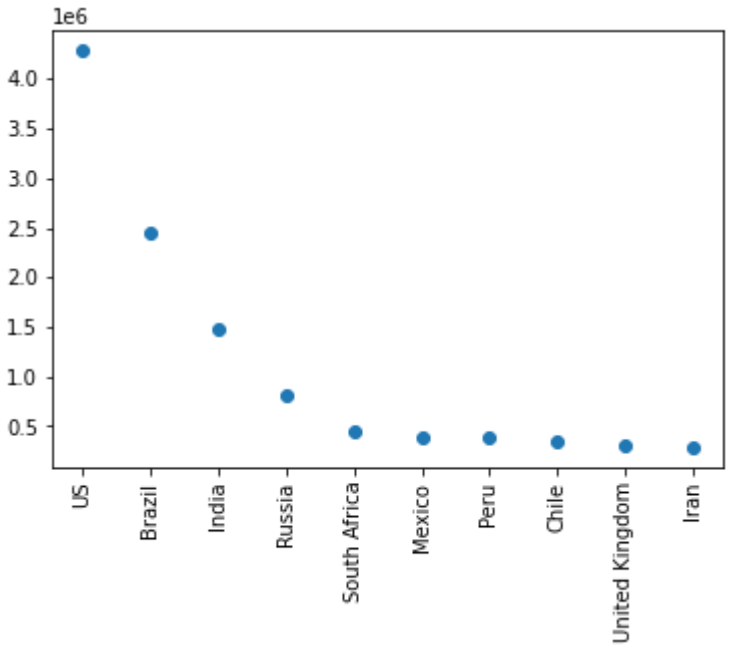
```
a = cvd[['WHO Region', 'Confirmed']]
b = a.groupby(['WHO Region']).sum().reset_index()

plt.pie(b['Confirmed'], labels=b['WHO Region'])
plt.show()
```




America Region was affected the most during first seven months of Covid followed by Europe

In [19]:

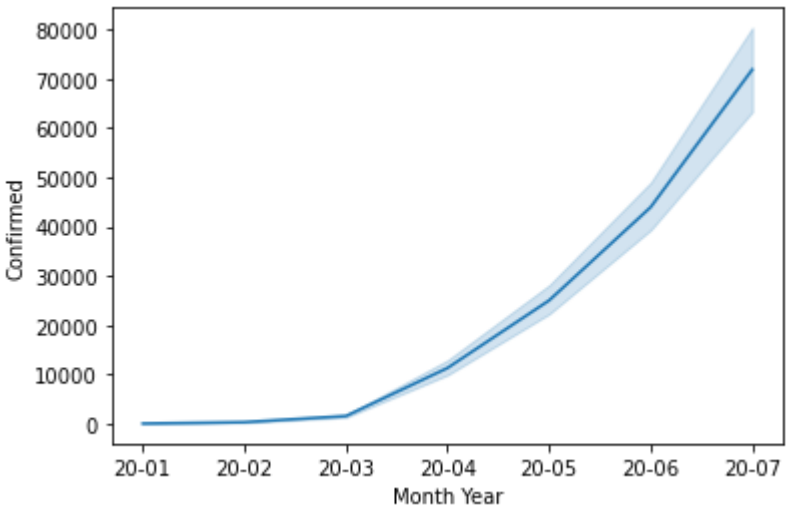
```
a = cvd[['Country/Region', 'Confirmed']]
b = a.groupby('Country/Region')['Confirmed'].max().sort_values(ascending = False).reset_ind

plt.scatter(b['Country/Region'], b['Confirmed'])
plt.xticks(rotation = 'vertical')
plt.show()
```




USA had most covid cases followed by Brazil and India

In [20]:

```
sns.lineplot(x = cvd['Month Year'], y = cvd['Confirmed'])
plt.show()
```




The cases followed the same pattern as death rate. There was gradual increase during first three months but then it there was rapid increase in the cases.

In [21]:

```
a = cvd.groupby('Country/Region')['Recovered'].max().sort_values(ascending = False).head()
pd.DataFrame(a)
```

Out[21]:

| Country/Region | Recovered |
|---|---|
| **Brazil** | 1846641 |
| **US** | 1325804 |
| **India** | 951166 |
| **Russia** | 602249 |
| **Chile** | 319954 |

Brazil had the most recovered patient during first seven months of the pandemic followed by USa and India.

In [ ]:

```

```