```
#COVID ANALYSIS
#SHIVANI TIWARI
# CODE CLAUSE INTERSHIP JUNE-2022

import numpy as np # linear algebra
import pandas as pd # data processing,
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))


import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px


df_country_wise=pd.read_csv('../input/corona-virus-report/country_wise_latest.csv')
df_country_wise.head()
```

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deat / 1 Cas
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.5
1	Albania	4880	144	2745	1991	117	6	63	2.9
2	Algeria	27973	1163	18837	7973	616	8	749	4.1
3	Andorra	907	52	803	52	10	0	0	5.7
4	Angola	950	41	242	667	18	1	0	4.3

```
df_country_wise.shape
```

```
(187, 15)
```

```
df_country_wise.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 187 entries, 0 to 186
Data columns (total 15 columns):
 #   Column                 Non-Null Count  Dtype
---  ------                 --------------  -----
 0   Country/Region         187 non-null    object
 1   Confirmed              187 non-null    int64
 2   Deaths                 187 non-null    int64
 3   Recovered              187 non-null    int64
 4   Active                 187 non-null    int64
```

```
 5   New cases               187 non-null    int64
 6   New deaths              187 non-null    int64
 7   New recovered           187 non-null    int64
 8   Deaths / 100 Cases      187 non-null    float64
 9   Recovered / 100 Cases   187 non-null    float64
 10  Deaths / 100 Recovered  187 non-null    float64
 11  Confirmed last week     187 non-null    int64
 12  1 week change           187 non-null    int64
 13  1 week % increase       187 non-null    float64
 14  WHO Region              187 non-null    object
dtypes: float64(4), int64(9), object(2)
memory usage: 22.0+ KB
```

```
df_country_wise.describe()
```

	Confirmed	Deaths	Recovered	Active	New cases	New d
count	1.870000e+02	187.000000	1.870000e+02	1.870000e+02	187.000000	187.0
mean	8.813094e+04	3497.518717	5.063148e+04	3.400194e+04	1222.957219	28.9
std	3.833187e+05	14100.002482	1.901882e+05	2.133262e+05	5710.374790	120.0
min	1.000000e+01	0.000000	0.000000e+00	0.000000e+00	0.000000	0.0
25%	1.114000e+03	18.500000	6.265000e+02	1.415000e+02	4.000000	0.0
50%	5.059000e+03	108.000000	2.815000e+03	1.600000e+03	49.000000	1.0
75%	4.046050e+04	734.000000	2.260600e+04	9.149000e+03	419.500000	6.0
max	4.290259e+06	148011.000000	1.846641e+06	2.816444e+06	56336.000000	1076.0

```
df_country_wise.isnull().sum()
```

```
Country/Region            0
Confirmed                 0
Deaths                    0
Recovered                 0
Active                    0
New cases                 0
New deaths                0
New recovered             0
Deaths / 100 Cases        0
Recovered / 100 Cases     0
Deaths / 100 Recovered    0
Confirmed last week       0
1 week change             0
1 week % increase         0
WHO Region                0
dtype: int64
```

```
grouped_df_country_wise = df_country_wise[["Country/Region","Confirmed","Deaths","Recovered"]]
grouped_df_country_wise
```

	Country/Region	Confirmed	Deaths	Recovered
0	Afghanistan	36263	1269	25198
1	Albania	4880	144	2745
2	Algeria	27973	1163	18837
3	Andorra	907	52	803
4	Angola	950	41	242
...	...	...	...	...
182	West Bank and Gaza	10621	78	3752
183	Western Sahara	10	1	8
184	Yemen	1691	483	833
185	Zambia	4552	140	2815
186	Zimbabwe	2704	36	542

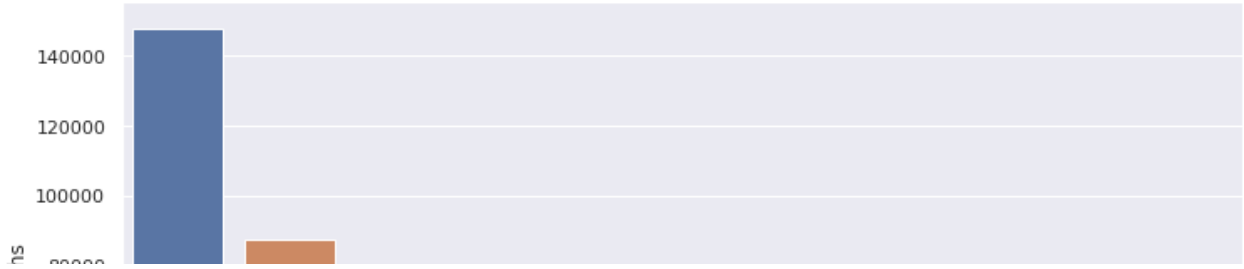
187 rows × 4 columns

```
grouped_df_country_wise = grouped_df_country_wise.sort_values(by="Confirmed",ascending=False)
grouped_df_country_wise = grouped_df_country_wise.reset_index(drop=True)
grouped_df_country_wise
```

	Country/Region	Confirmed	Deaths	Recovered

```
px.bar(grouped_df_country_wise[0:10], x="Country/Region", y="Confirmed",title="Top 10 Countri
```

```
sns.set(rc={'figure.figsize':(12,6)})
sns.barplot(x='Country/Region', y='Deaths', data=grouped_df_country_wise.nlargest(10,'Deaths'
plt.show()
```




```
px.pie(grouped_df_country_wise[0:10],values="Confirmed",names="Country/Region")
```

```
df_covid_19_clean = pd.read_csv('../input/corona-virus-report/covid_19_clean_complete.csv', p
df_covid_19_clean.head()
```

	Province/State	Country/Region	Lat	Long	Date	Confirmed	Deaths	I

```
df_covid_19_clean.shape
```

```
(49068, 10)
```

1	NaN	Albania	41.15330	20.168300	01-22	0	0	

```
df_covid_19_clean.isnull().sum()
```

```
Province/State    34404
Country/Region        0
Lat                   0
Long                  0
Date                  0
Confirmed             0
Deaths                0
Recovered             0
Active                0
WHO Region            0
dtype: int64
```

```
df_covid_19_clean.drop(columns=['Province/State'], inplace=True)
df_covid_19_clean.head()
```

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

```
df_covid_19_clean_new= df_covid_19_clean.groupby(['Date'])[['Confirmed','Deaths']].sum().rese
df_covid_19_clean_new
```

	Date	Confirmed	Deaths
0	2020-01-22	555	17
1	2020-01-23	654	18
2	2020-01-24	941	26
3	2020-01-25	1434	42

```
px.line(df_covid_19_clean_new,x="Date",y="Confirmed",title="World Wide Confirmed Cases ")
```