



Hi !

We are Group C , today we will show you our work on analyzing data on covid-19

Created by:

Mohamed Tarek AbdElhamed

Khalil Hamdy Mohamed

Maram Mohamed Wahaballah

Logina Mohamed Fekry

Supervision of Dr:

Amira Gamal
Yassin



CLEANING DATA

Import **pandas** & **numpy** library

```
[2] import pandas as pd  
import numpy as np  
df=pd.read_excel('/content/covid19.xlsx')
```

```

[3] df['TotalDeaths']=pd.to_numeric(df['TotalDeaths'], errors='coerce')
df['HealthCare (%)'] =100- ((df['TotalDeaths']*100 ) / df['TotalCases'])
df

```

| | Unnamed: 0 | Country,Other | TotalCases | NewCases | TotalDeaths | HealthCare (%) |
|-----|------------|---------------|------------|-----------|-------------|----------------|
| 0 | 0 | USA | 29383324.0 | 12619.0 | 529515.0 | 98.197906 |
| 1 | 1 | India | 11156250.0 | 16927.0 | 157471.0 | 98.588495 |
| 2 | 2 | Brazil | 10647845.0 | 20154.0 | 257562.0 | 97.581088 |
| 3 | 3 | Russia | 4278750.0 | 10535.0 | 87348.0 | 97.958563 |
| 4 | 4 | UK | 4188400.0 | 8692.0 | 123296.0 | 97.056251 |
| ... | ... | ... | ... | ... | ... | ... |
| 224 | 224 | 34442286 | 119818.0 | 822446.0 | 2724.0 | 97.726552 |
| 225 | 225 | 3949024 | 5653.0 | 104626.0 | 64.0 | 98.867858 |
| 226 | 226 | 51331 | 16.0 | 1091.0 | NaN | NaN |
| 227 | 227 | 721 | NaN | 15.0 | NaN | NaN |
| 228 | 228 | 115522246 | 227686.0 | 2564708.0 | 5036.0 | 97.788182 |

229 rows × 6 columns

In this step, we have added a new column called 'HealthCare(%)', It shows the extent of medical development.

✓ [4] df.dropna(inplace=True)

✓ 0s

pd.options.display.max_rows=219

df

| | | | | | | |
|-----|-----|------------------------|-------|------|------|-----------|
| 187 | 187 | Bermuda | 713.0 | 4.0 | 12.0 | 98.316971 |
| 188 | 188 | Diamond Princess | 712.0 | 2.0 | 13.0 | 98.174157 |
| 189 | 189 | Faeroe Islands | 658.0 | 3.0 | 1.0 | 99.848024 |
| 190 | 190 | Mauritius | 619.0 | 11.0 | 10.0 | 98.384491 |
| 191 | 191 | St. Barth | 573.0 | 10.0 | 1.0 | 99.825480 |
| 192 | 192 | Isle of Man | 511.0 | 17.0 | 25.0 | 95.107632 |
| 193 | 193 | Tanzania | 509.0 | 3.0 | 21.0 | 95.874263 |
| 194 | 194 | Cayman Islands | 447.0 | 5.0 | 2.0 | 99.552573 |
| 195 | 195 | Caribbean Netherlands | 429.0 | 2.0 | 4.0 | 99.067599 |
| 196 | 196 | Brunei | 187.0 | 0.0 | 3.0 | 98.395722 |
| 197 | 197 | British Virgin Islands | 153.0 | 2.0 | 1.0 | 99.346405 |
| 198 | 198 | Grenada | 148.0 | 4.0 | 1.0 | 99.324324 |
| 201 | 201 | ... | ... | ... | ... | ... |

✓ 0s completed at 9:50 AM

As we can see here, We represented the least 10 countries in total deaths.

And then we dropped all of rows that have a missing value.

| | | | | | | |
|-----|-----|------------|----------|-----------|--------|------------|
| 218 | 218 | Micronesia | 1.0 | 0.0 | 0.0 | 100.000000 |
| 219 | 219 | Vanuatu | 1.0 | 0.0 | 0.0 | 100.000000 |
| 220 | 220 | China | 89933.0 | 10.0 | 4636.0 | 94.845051 |
| 221 | 221 | 33720845 | 23073.0 | 765344.0 | 1378.0 | 94.027651 |
| 222 | 222 | 25223160 | 72549.0 | 400595.0 | 742.0 | 98.977243 |
| 223 | 223 | 18134879 | 6577.0 | 470591.0 | 128.0 | 98.053824 |
| 224 | 224 | 34442286 | 119818.0 | 822446.0 | 2724.0 | 97.726552 |
| 225 | 225 | 3949024 | 5653.0 | 104626.0 | 64.0 | 98.867858 |
| 228 | 228 | 115522246 | 227686.0 | 2564708.0 | 5036.0 | 97.788182 |

As we can see here, the last 6 rows have invalid data, so we deleted them

```
✓ [6] df.drop([221,222,223,224,225,228], inplace =True)
```

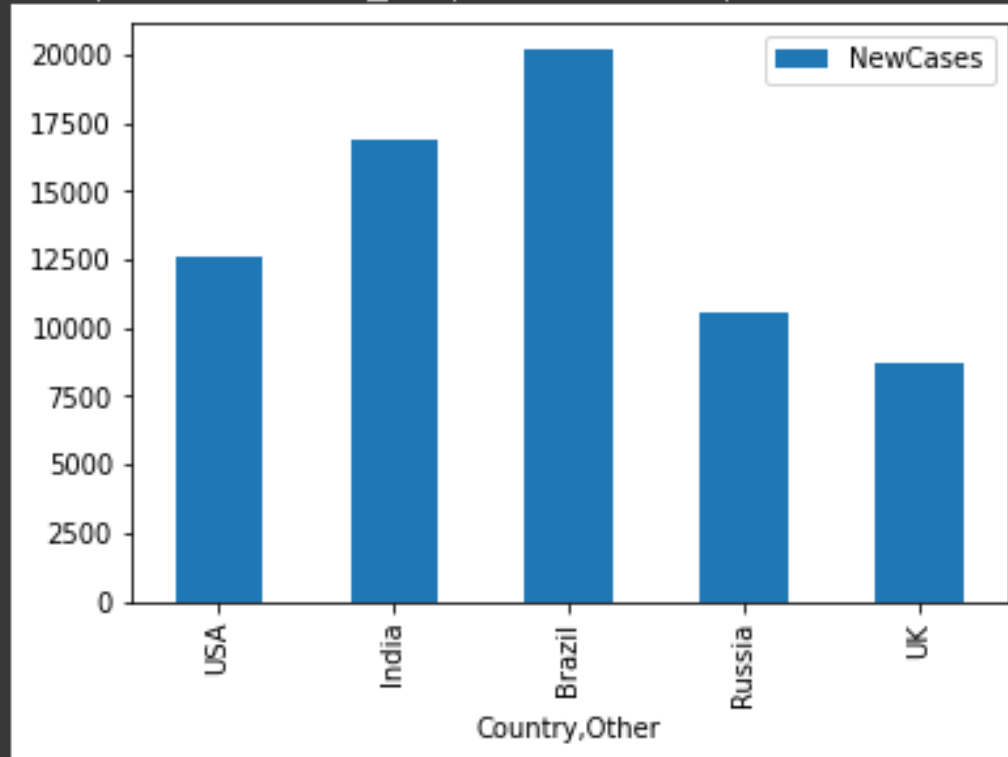


VISUALIZATION DATA


```
✓ [39] df2.plot(kind='bar',y='NewCases',x='Country,Other')
```

Os

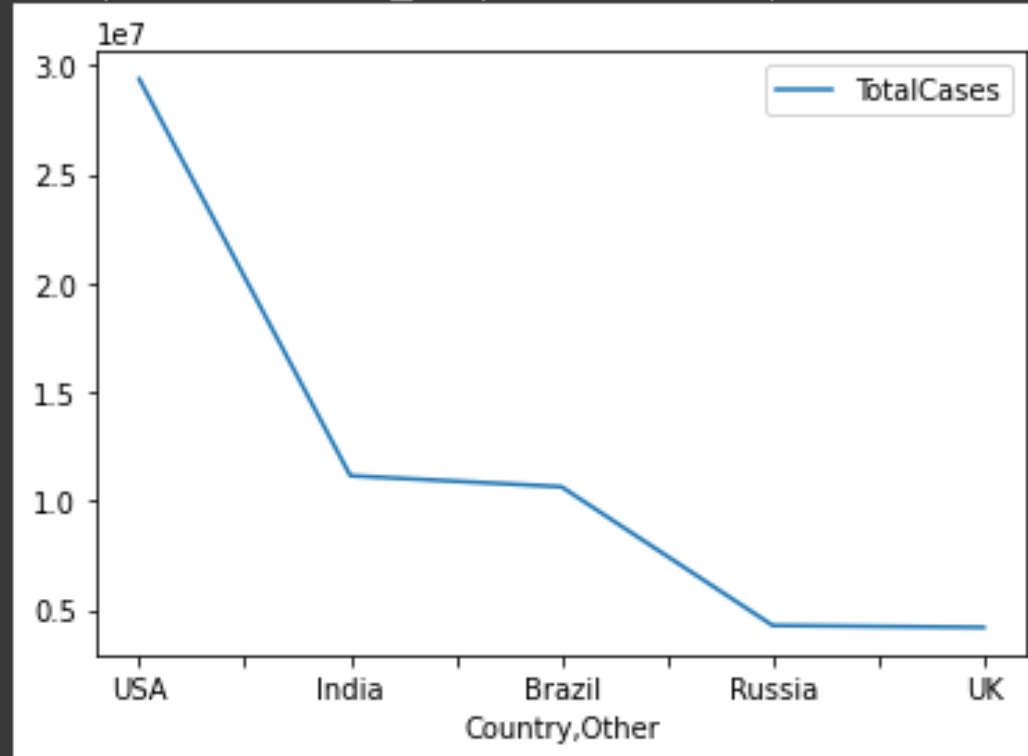
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f7148db4910>
```



As we can see here, We compared the first five rows in terms of new cases

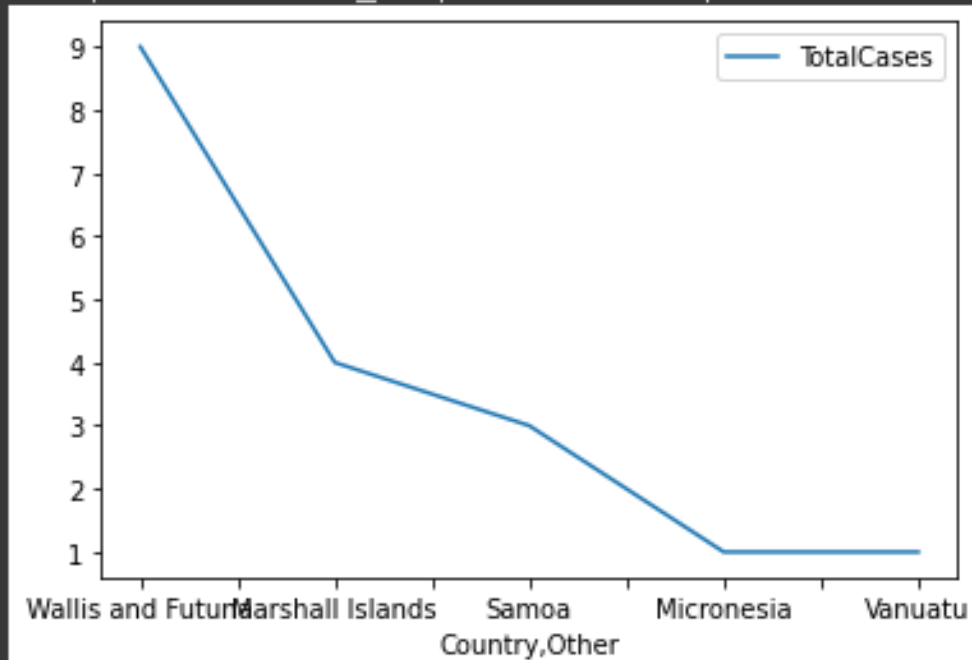
```
df2.plot(kind='line',y='TotalCases',x='Country,Other')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f714ba62950>
```



As we can see here, We represented the data using line plot.

```
✓ [18] df3.plot(kind='line',y='TotalCases',x='Country,Other')  
0s  
<matplotlib.axes._subplots.AxesSubplot at 0x7f714b962390>
```



As we can see here, We represented the data using line plot.

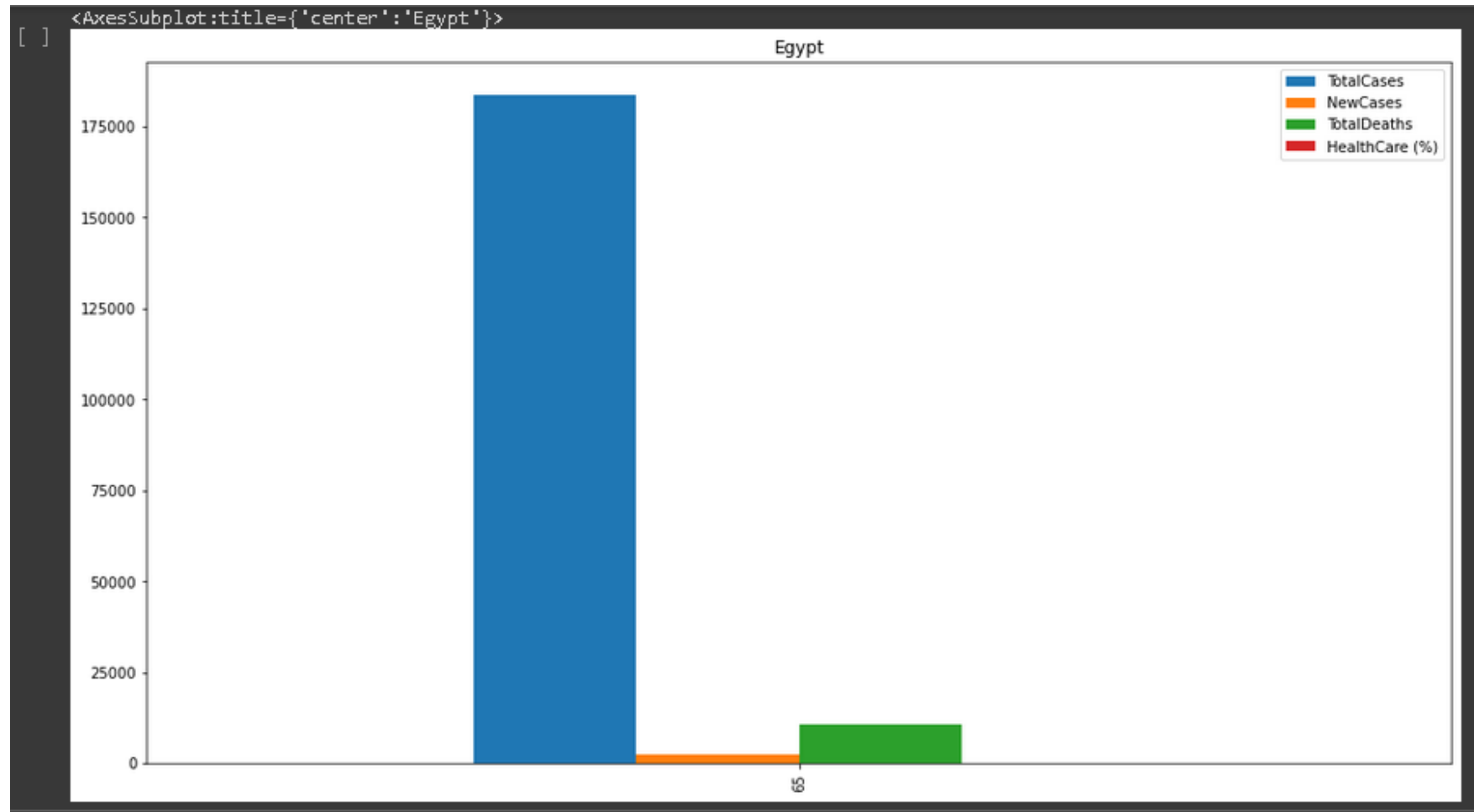
```
[ ] print('The total cases of all countries= ',df['TotalCases'].sum())  
    print('The total new cases of all countries= ',df['NewCases'].sum())  
    print('The total deaths of all countries= ',df['TotalDeaths'].sum())
```

```
The total cases of all countries= 115521719.0  
The total new cases of all countries= 308897.0  
The total deaths of all countries= 2564728.0
```

```
[ ] df_eg=df.groupby('Country,Other')  
    df_eg=df_eg.get_group('Egypt')  
    df_eg.plot(kind='bar',title='Egypt',figsize=(16,9))
```

```
<AxesSubplot:title={'center':'Egypt'}>
```

As we can see here, We represented mathematics operations & bar plot



```
In [16]: df1=df[['Country,Other','TotalDeaths']]
x=df1.sort_values(by='TotalDeaths',ascending=False)
df2= x.tail(10)
df2
```

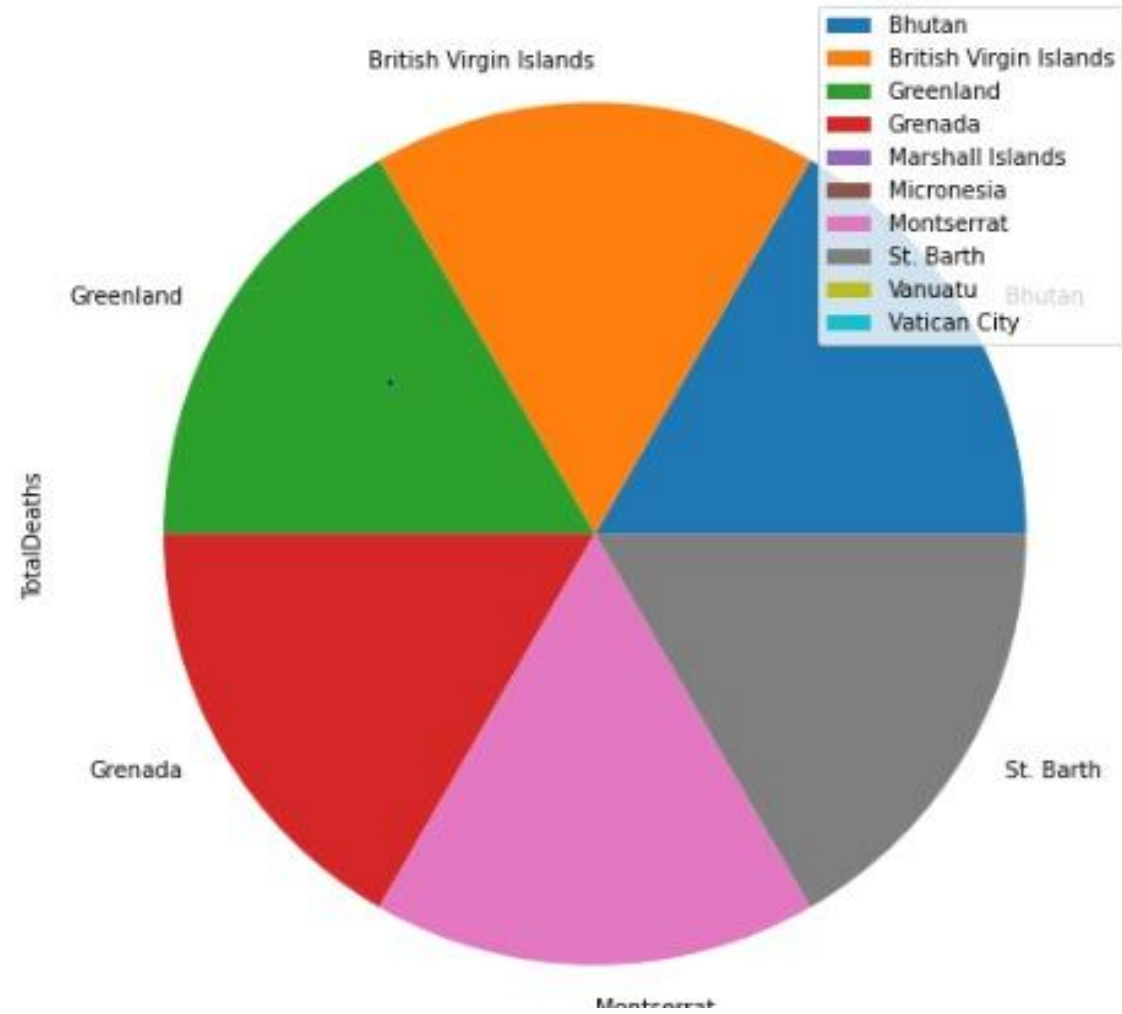
Out[16]:

| | Country,Other | TotalDeaths |
|-----|------------------------|-------------|
| 207 | Greenland | 1.0 |
| 210 | Montserrat | 1.0 |
| 185 | Bhutan | 1.0 |
| 198 | Grenada | 1.0 |
| 197 | British Virgin Islands | 1.0 |
| 191 | St. Barth | 1.0 |
| 208 | Vatican City | 0.0 |
| 216 | Marshall Islands | 0.0 |
| 218 | Micronesia | 0.0 |
| 219 | Vanuatu | 0.0 |

**As we can see here,
We represented the
least 10 countries
in total deaths.**

```
In [19]: df2.groupby(['Country,Other']).sum().plot(kind='pie',y='TotalDeaths',x='Country,Other',title='least 10 countries in total deaths',figsize=(9,9))
```

least 10 countries in total deaths



Thank You

