RAMAIAH INSTITUTE OF TECHNOLOGY, BANGALORE – 560054
(Autonomous Institute, Affiliated to VTU)

Department of Computer Science & Engineering

Internship Report

on

Python Programming

INT411: Intra Institutional Internship

STUDENT NAME : RONIT KUMAR MANJHI

USN : 1MS21EC087

Ramaiah Institute of Technology
(Autonomous Institute, Affiliated to VTU)
MSR Nagar, MSRIT Post, Bangalore-560054

September – October, 2022

RAMAIAH
Institute of Technology

RAMAIAH INSTITUTE OF TECHNOLOGY, BANGALORE – 560054
(Autonomous Institute, Affiliated to VTU)

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that Mr./Ms. _____, a student of Bachelor of Engineering, bearing USN: _____, has successfully completed, 30 Hours: from 29.09.2022 to 15.10.2022 Intra Institutional Internship in Python Programming from the Department of Computer Science & Engineering, M S Ramaiah Institute of Technology, Bangalore.

SL No.	Component	Maximum Marks	Marks Obtained
1	Continuous Evaluation	50	
2	Presentation	20	
3	Report	30	
Total Marks		100	

Signature of the Student with Date **Signature of the Faculty Co-Ordinator**

OVERVIEW OF INTERNSHIP ACTIVITIES

DATE	DAY	NAME OF THE TOPIC COMPLETED
29/09/22	Thusrday	Python-Basics
30/09/22	Friday	Control Statements
01/10/22	Saturday	Data Structures in python
06/10/22	Thursday	Functions, Strings
07/10/22	Friday	Advanced Functions in Python
08/10/22	Saturday	Exception Handling
10/10/22	Monday	Numpy Assessment 1
11/10/22	Tuesday	Pandas
12/10/22	Wednesday	Objects and Classes
13/10/22	Thursday	Assessment 2
14/10/22	Friday	Modules and Packages Project
15/10/22	Saturday	Project Evaluation

TABLE OF CONTENTS

OVERVIEW OF THE PROJECT IN TERMS OF IMPLEMENTATION
PANDAS GOBAL COVID-19 IMPACT ANALYSIS

OBJECTIVE

Upon initial inspection of the data, we can start thinking of some questions about it that we would want to answer.

- Which country is being having the most covid -19 cases outbreak?
- Which are the Top 10 countries on basis of population facing covid-19 outbreak?
- Which countries are having most number of demise?

PLAN OF ACTION

- IMPORTING REQUIRED LIBRARIES

Import in Python helps you to refer to the code, i.e., . functions/objects that are written in another file. **It is also used to import python libraries/packages that are installed using pip(python package manager)**, and you need then to use in your code.

- IMPORTING THE DATASET

When running python programs, we need to use datasets for data analysis. Python has various modules which help us in importing the external data in various file formats to a python program

- DATA AUDIT

You can't make your data work for you until you know what data you're talking about.

We will use HEAD , TAIL and SHAPE to know our Data set.

Then we use INFO to get the information if any column has a NULL value or not.

Now we can do further analysis on our data to answer our questions. Before that, we should see if there are any missing values in our data set. To check if there are any missing values in the entire data set we use the isnull function, then see if there are any values.

Next, we can look at some descriptive statistics of the data frame with the describe method.

This shows some descriptive statistics on the data set. Notice, it only shows the statistics on the numerical columns. From here you can see the following statistics:

- Row count, which aligns to what the shape attribute showed us.
- The mean or average.
- The standard deviation, or how spread out the data is.
- The minimum and maximum value of each column
- The number of items that fall within the first, second, and third percentiles.

EXPLORATORY DATA ANALYSIS

Now we have answered our above listed objectives.

- ## WHAT IS THE OVERALL GLOBAL COVID-19 TREND?

  outbreak of covid-19 resulted in the highest number of covid-19 cases and deaths in the united states. It is comparatively low according to the population.

- ## WHICH ARE THE TOP 10 MOST POPULATED COUNTRIES FACING COVID-19 OUTBREAK?

  we can observe the you entries having most population are facing more outbreak to covid-19 .

- ## WHICH COUNTRIES FACED MOST DEMISE AND WHATS THE TRAND FOR INDIA?

  The most number of demise cases are being observed in united states followed by Brazil even when India has more population then these countries and yet has head to head demise count as of united kingdom and Mexico .

CODES OF MAIN MODULES

#IMPORTING REQUIRED LIBRARIES

```
import pandas as pd
import matplotlib.pyplot as plt
```

#IMPORTING RELATED FILES

```
data = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/transformed_data.csv")
data2 = pd.read_csv("/content/drive/MyDrive/Colab Notebooks/raw_data.csv")
print(data.head())
print("-----------------------------------------------------------------------")
print(data2.head())
```

OUTPUT:

```
  CODE      COUNTRY        DATE    HDI   TC   TD  STI        POP    GDPCAP
0  AFG  Afghanistan  2019-12-31  0.498  0.0  0.0  0.0  17.477233  7.497754
1  AFG  Afghanistan  2020-01-01  0.498  0.0  0.0  0.0  17.477233  7.497754
2  AFG  Afghanistan  2020-01-02  0.498  0.0  0.0  0.0  17.477233  7.497754
3  AFG  Afghanistan  2020-01-03  0.498  0.0  0.0  0.0  17.477233  7.497754
4  AFG  Afghanistan  2020-01-04  0.498  0.0  0.0  0.0  17.477233  7.497754
-----------------------------------------------------------------------
  iso_code     location        date  total_cases  total_deaths  \
0      AFG  Afghanistan  2019-12-31          0.0           0.0
1      AFG  Afghanistan  2020-01-01          0.0           0.0
2      AFG  Afghanistan  2020-01-02          0.0           0.0
3      AFG  Afghanistan  2020-01-03          0.0           0.0
4      AFG  Afghanistan  2020-01-04          0.0           0.0

   stringency_index  population  gdp_per_capita  human_development_index  \
0               0.0    38928341        1803.987                    0.498
1               0.0    38928341        1803.987                    0.498
2               0.0    38928341        1803.987                    0.498
3               0.0    38928341        1803.987                    0.498
4               0.0    38928341        1803.987                    0.498

  Unnamed: 9 Unnamed: 10 Unnamed: 11  Unnamed: 12  Unnamed: 13
0      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
1      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
2      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
3      #NUM!       #NUM!       #NUM!    17.477233  7.497754494
```

OBJECTIVE

Upon initial inspection of the data, we can start thinking of some questions about it that we would want to answer.

- Which country is being having the most covid -19 cases outbreak?
- Which are the Top 10 countries on basis of population facing covid-19 outbreak?
- Which countries are having most number of demise?

```
print(data.shape,data2.shape)
```

OUTPUT: (50418, 9) (50418, 14)

```
print(data.isna().sum())
print("----------------------------------------------------------------------")
print(data2.isna().sum())
```

OUTPUT:

```
CODE          0
COUNTRY       0
DATE          0
HDI        6202
TC            0
TD            0
STI           0
POP           0
GDPCAP        0
dtype: int64
----------------------------------------------------------------------
iso_code                      0
location                      0
date                          0
total_cases                3094
total_deaths              11190
stringency_index           7126
population                    0
gdp_per_capita             5712
human_development_index    6202
Unnamed: 9                    0
Unnamed: 10                   0
Unnamed: 11                   0
Unnamed: 12                   0
Unnamed: 13                   0
dtype: int64
```

```
data["HDI"].fillna(data["HDI"].mean(),inplace = True)
data2["total_cases"].fillna(data2["total_cases"].mean(),inplace = True)
data2["total_deaths"].fillna(data2["total_deaths"].mean(),inplace = True)
data2["stringency_index"].fillna(data2["stringency_index"].mean(),inplace = True)
data2["gdp_per_capita"].fillna(data2["gdp_per_capita"].mean(),inplace = True)
data2["human_development_index"].fillna(data2["human_development_index"].mean(),inplace = True)

print(data.isna().sum())
print("----------------------------------------------------------------------")
print(data2.isna().sum())
```

OUTPUT:

```
index           0
Country Code    0
Country         0
HDI             0
Total Cases     0
Total Deaths    0
Population      0
dtype: int64
--------------------------------------------------------------------------------
iso_code                    0
location                    0
date                        0
total_cases                 0
total_deaths                0
stringency_index            0
population                  0
gdp_per_capita              0
human_development_index     0
Unnamed: 9                  0
Unnamed: 10                 0
Unnamed: 11                 0
Unnamed: 12                 0
Unnamed: 13                 0
dtype: int64
```

```
print(data.describe())
print("--------------------------------------------------------------------------------")
print(data2.describe())
```

OUTPUT:

```
          index        HDI   Total Cases  Total Deaths  Population
count   10.0000  10.000000  1.000000e+01  1.000000e+01   10.000000
mean   134.3000   0.751296  2.135855e+08  7.835122e+06   18.595943
std     62.2826   0.136178  2.361521e+08  7.682455e+06    1.129353
min     27.0000   0.581847  6.094100e+07  1.503642e+06   17.311165
25%     98.7500   0.616490  6.757933e+07  2.563691e+06   17.783344
50%    153.5000   0.766500  7.565768e+07  6.576697e+06   18.354071
75%    177.2500   0.869977  3.411203e+08  7.519042e+06   19.080716
max    200.0000   0.924000  7.474047e+08  2.665928e+07   21.045353
--------------------------------------------------------------------------------
       total_cases   total_deaths  stringency_index    population  \
count  5.041800e+04   50418.000000      50418.000000  5.041800e+04
mean   6.621927e+04    2978.767819         56.162022  4.251601e+07
std    3.919481e+05   12204.916580         25.512844  1.564607e+08
min    0.000000e+00       0.000000          0.000000  8.090000e+02
25%    1.480000e+02      18.000000         41.670000  1.399491e+06
50%    2.057500e+03     200.000000         56.162022  8.278737e+06
75%    2.871075e+04    2978.767819         76.390000  2.913681e+07
max    8.154595e+06  219674.000000        100.000000  1.439324e+09

       gdp_per_capita  human_development_index   Unnamed: 12
count    50418.000000             50418.000000  50418.000000
mean     20818.706240                 0.720139     15.442097
std      19248.613445                 0.150680      2.495039
min        661.240000                 0.000000      6.695799
25%       6253.104000                 0.640000     14.151619
50%      16409.288000                 0.723000     15.929201
75%      27936.896000                 0.825000     17.187513
max     116935.600000                 0.953000     21.087439
```

```
code = list(data["CODE"].unique())
country = list(data["COUNTRY"].unique())
hdi = []
tc = []
td = []
sti = []
population = list(data["POP"].unique())
gdp = []

for i in country:
    hdi.append((data.loc[data["COUNTRY"] == i, "HDI"]).sum()/294)
    tc.append((data2.loc[data2["location"] == i, "total_cases"]).sum())
    td.append((data2.loc[data2["location"] == i, "total_deaths"]).sum())
    sti.append((data.loc[data["COUNTRY"] == i, "STI"]).sum()/294)
    population.append((data2.loc[data2["location"] == i, "population"]).sum()/294)

aggregated_data = pd.DataFrame(list(zip(code, country, hdi, tc, td, population)), columns = ["Country Code", " Country", "HDI","Total Cases", "Total Deaths", "Population"])

print(aggregated_data.head())
```

OUPUT:

	Country Code	Country	HDI	Total Cases	Total Deaths	Population
0	AFG	Afghanistan	0.498000	5.126433e+06	165875.000000	17.477233
1	ALB	Albania	0.600765	1.071951e+06	39992.303457	14.872537
2	DZA	Algeria	0.754000	8.999594e+06	423879.050780	17.596309
3	AND	Andorra	0.659551	6.208916e+05	66446.588559	11.254996
4	AGO	Angola	0.418952	3.040050e+05	35650.142551	17.307957

```
data = aggregated_data.sort_values(by=["Total Cases"], ascending=False)
data.head(10)
```

OUTPUT:

	Country Code	Country	HDI	Total Cases	Total Deaths	Population
200	USA	United States	0.924000	7.474047e+08	2.665928e+07	19.617637
27	BRA	Brazil	0.759000	4.294790e+08	1.457291e+07	19.174732
90	IND	India	0.640000	4.098244e+08	7.464777e+06	21.045353
157	RUS	Russia	0.816000	1.350080e+08	2.390724e+06	18.798668
178	ESP	Spain	0.887969	7.583669e+07	5.704244e+06	17.660427
150	PER	Peru	0.599490	7.547867e+07	3.082592e+06	17.311165
125	MEX	Mexico	0.774000	7.527462e+07	7.537130e+06	18.674802
175	ZAF	South Africa	0.608653	6.501424e+07	1.503642e+06	17.898266
199	GBR	United Kingdom	0.922000	6.159405e+07	7.449150e+06	18.033340
42	COL	Colombia	0.581847	6.094100e+07	1.986773e+06	17.745037

data=data.reset_index()
data
OUTPUT:

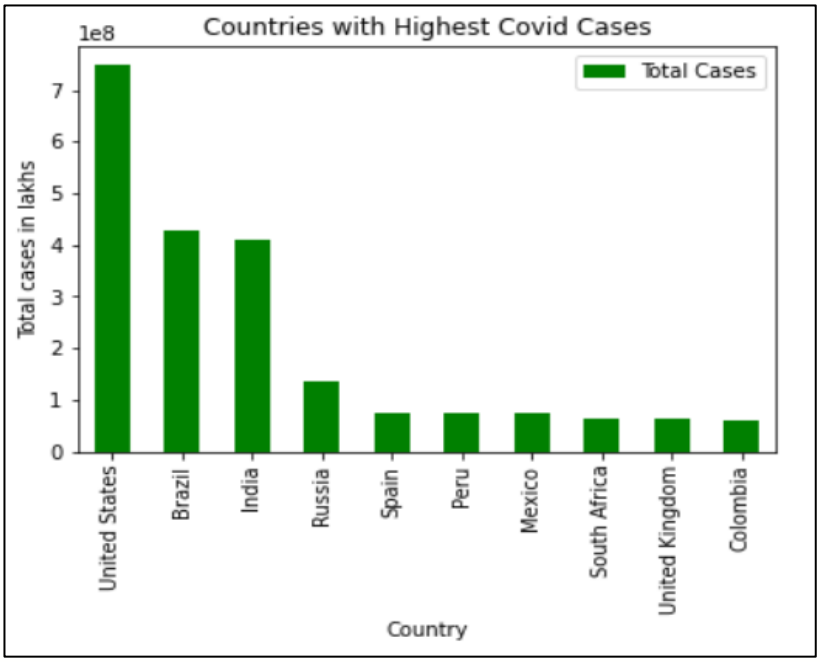
	index	Country Code	Country	HDI	Total Cases	Total Deaths	Population
0	200	USA	United States	0.924000	7.474047e+08	2.665928e+07	19.617637
1	27	BRA	Brazil	0.759000	4.294790e+08	1.457291e+07	19.174732
2	90	IND	India	0.640000	4.098244e+08	7.464777e+06	21.045353
3	157	RUS	Russia	0.816000	1.350080e+08	2.390724e+06	18.798668
4	178	ESP	Spain	0.887969	7.583669e+07	5.704244e+06	17.660427
5	150	PER	Peru	0.599490	7.547867e+07	3.082592e+06	17.311165
6	125	MEX	Mexico	0.774000	7.527462e+07	7.537130e+06	18.674802
7	175	ZAF	South Africa	0.608653	6.501424e+07	1.503642e+06	17.898266
8	199	GBR	United Kingdom	0.922000	6.159405e+07	7.449150e+06	18.033340
9	42	COL	Colombia	0.581847	6.094100e+07	1.986773e+06	17.745037

EXPLORATORY DATA ANALYSIS

1. WHICH ARE THE TOP 10 COUNTRY HAS TOTAL CASES?

```
import matplotlib.pyplot as plt
data.plot(kind="bar",y='Total Cases', x='Country',title="Countries with Highest Covid Cases",color='green')
plt.show()
```

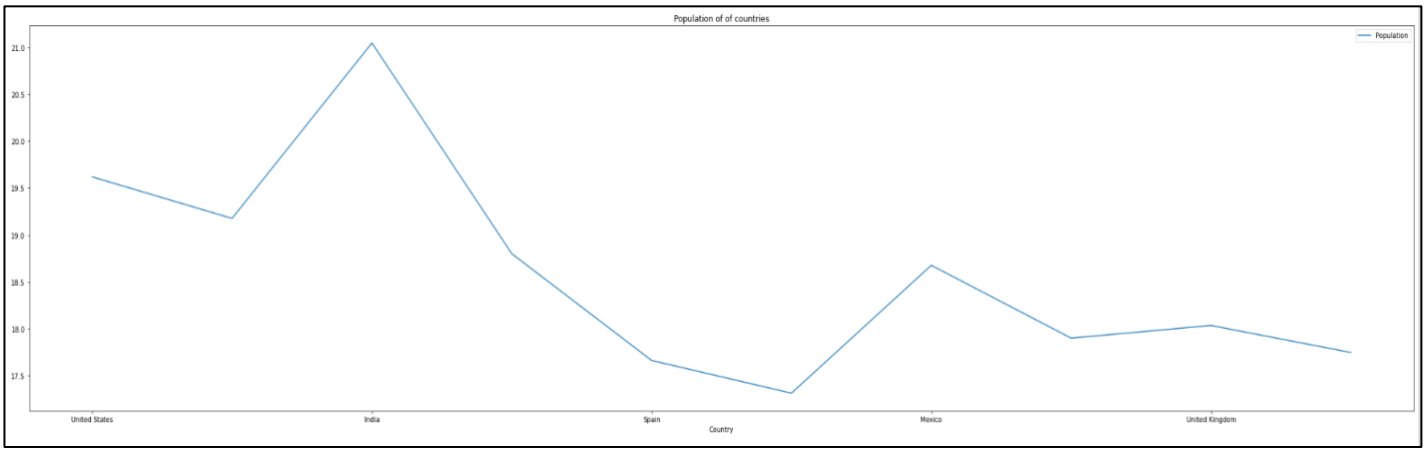
OUTPUT:



2.WHICH ARE THE TOP 10 COUNTRIES BY POPULATION?

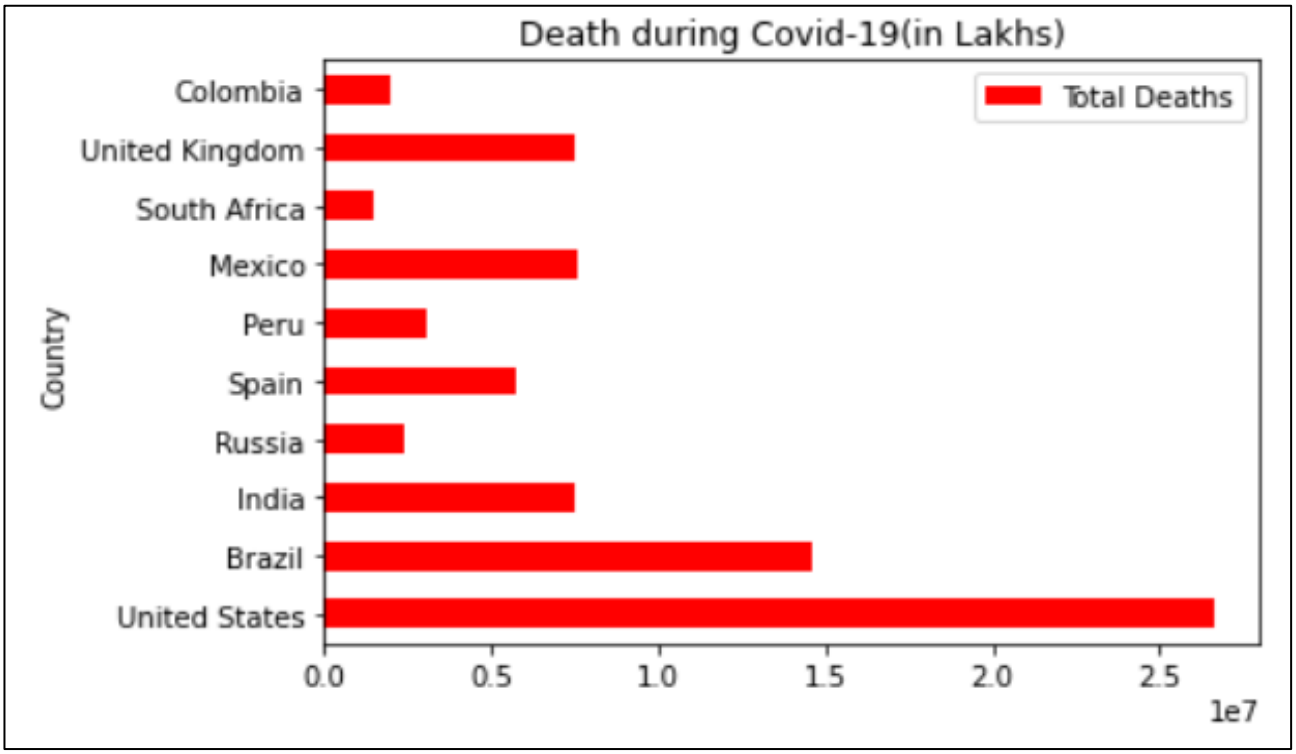
```
data.plot( x='Country', y='Population',title= "Population of of countries")
plt.show()
```

OUTPUT:



3-WHICH ARE COUNTRIES HAVING MOST NUMBER OF DEMISE ?

```
data.head(10).plot( kind="barh",x='Country', y='Total Deaths',title= "Stringency Index during Covid-19",color="red")
```



CONCLUSION

In this task, we studied the spread of covid-19 among the countries and its impact on the global human lifestyle. We saw that the outbreak of covid-19 resulted in the highest number of covid-19 cases and deaths in the united states. It is comparatively low according to the population.

From the first objective we can conclude that majority covid outbreak was in united states followed by Brazil and india respectively.

From the second objective, we can observe the you entries having most population are facing more outbreak to covid-19 .

From the third objective, the most number of demise cases are being observed in united states followed by Brazil even when India has more population then these countries and yet has head to head demise count as of united kingdom and Mexico .