Kingdom of Saudi Arabia
Ministry of Education
University of Jeddah
College of Computer Science and Engineering
Department of Computer Science and Artificial Intelligence

المملكة العربية السعودية
وزارة التعليم
جامعة جدّة
كلية علوم وهندسة الحاسب
قسم علوم الحاسب والذكاء الاصطناعي

Lab 2
CCAI 312 Pattern Recognition
Third Trimester 2023

Student Name: Ruba Khalid Alsulami

Student ID: 2110618

		Max Score	Student Score
PLO S2 / CLO 2 / SO 2	Task 1	2	
PLO C4 / CLO 3 / SO 7	Task 2	2	
Total			

Kingdom of Saudi Arabia
Ministry of Education
University of Jeddah
College of Computer Science and Engineering
Department of Computer Science and Artificial Intelligence

المملكة العربية السعودية
وزارة التعليم
جامعة جدّة
كلية علوم وهندسة الحاسب
قسم علوم الحاسب والذكاء الاصطناعي

Task 1: [PLO S2 / CLO 2 / SO 2] [2 marks]

1. Read covidtotals file and parse lastdate column as date

```
import pandas as pd
covidtotals = pd.read_csv("/kaggle/input/lab2covid/covidtotals.csv",parse_dates=['lastdate'])
```

+ Code + Markdown

2. Set and show the index and size for the COVID data (use iso_code as).

```
covidtotals.set_index("iso_code", inplace=True)
covidtotals.index
```

```
[6]: Index(['AFG', 'ALB', 'DZA', 'AND', 'AGO', 'AIA', 'ATG', 'ARG', 'ARM', 'ABW',
        ...
       'VIR', 'URY', 'UZB', 'VAT', 'VEN', 'VNM', 'ESH', 'YEM', 'ZMB', 'ZWE'],
      dtype='object', name='iso_code', length=210)
```

```
covidtotals.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 210 entries, AFG to ZWE
Data columns (total 11 columns):
 #   Column           Non-Null Count  Dtype
---  ------           --------------  -----
 0   lastdate         210 non-null    datetime64[ns]
 1   location         210 non-null    object
 2   total_cases      210 non-null    int64
 3   total_deaths     210 non-null    int64
 4   total_cases_pm   210 non-null    float64
 5   total_deaths_pm  210 non-null    float64
 6   population       210 non-null    float64
 7   pop_density      198 non-null    float64
 8   median_age       186 non-null    float64
 9   gdp_per_capita   182 non-null    float64
 10  hosp_beds        164 non-null    float64
dtypes: datetime64[ns](1), float64(7), int64(2), object(1)
memory usage: 19.7+ KB
```

3. Show the shape and check to see whether index values are unqiue:

```
[8]: covidtotals.shape
```

```
[8]: (210, 11)
```

```
covidtotals.index.nunique()
```

```
[9]: 210
```

+ Code + Markdown

4. Show a sample of a few rows of the COVID case data:

```
covidtotals.sample(2, random_state=1).T
```

[10]:

iso_code	COG	THA
lastdate	2020-06-01 00:00:00	2020-06-01 00:00:00
location	Congo	Thailand
total_cases	611	3081
total_deaths	20	57
total_cases_pm	110.727	44.14
total_deaths_pm	3.624	0.817
population	5518092.0	69799978.0
pop_density	15.405	135.132
median_age	19.0	40.1
gdp_per_capita	4881.406	16277.671
hosp_beds	NaN	2.1

```
covidtotals.iloc[0:5].T
```

[11]:

iso_code	AFG	ALB	DZA	AND	AGO
lastdate	2020-06-01 00:00:00	2020-06-01 00:00:00	2020-06-01 00:00:00	2020-06-01 00:00:00	2020-06-01 00:00:00
location	Afghanistan	Albania	Algeria	Andorra	Angola
total_cases	15205	1137	9394	764	86
total_deaths	257	33	653	51	4
total_cases_pm	390.589	395.093	214.225	9888.048	2.617
total_deaths_pm	6.602	11.467	14.891	660.066	0.122
population	38928341.0	2877800.0	43851043.0	77265.0	32866268.0
pop_density	54.422	104.871	17.348	163.755	23.89
median_age	18.6	38.0	29.1	NaN	16.8
gdp_per_capita	1803.987	11803.431	13913.839	NaN	5819.495
hosp_beds	0.5	2.89	1.9	NaN	NaN

5. Get the descriptive statistics on the COVID totals and demographic columns

```
covidtotals.describe()
```

[12]:

	total_cases	total_deaths	total_cases_pm	total_deaths_pm	population	pop_density	median_age	gdp_per_capita	hosp
count	2.100000e+02	210.000000	210.000000	210.000000	2.100000e+02	198.000000	186.000000	182.000000	164.(
mean	2.921614e+04	1770.714286	1355.357943	55.659129	3.694276e+07	362.867434	30.627957	19539.154588	3.(
std	1.363978e+05	8705.565857	2625.277497	144.785816	1.425092e+08	1581.438294	9.133152	19862.354091	2.(
min	0.000000e+00	0.000000	0.000000	0.000000	8.090000e+02	0.137000	15.100000	661.240000	0.(
25%	1.757500e+02	4.000000	92.541500	0.884750	1.031042e+06	37.416000	22.250000	4485.329000	1.(
50%	1.242500e+03	25.500000	280.928500	6.154000	6.909866e+06	87.250000	30.250000	13183.081500	2.(
75%	1.011700e+04	241.250000	1801.394750	31.777250	2.615868e+07	214.122000	39.000000	28556.527250	3.(
max	1.790191e+06	104383.000000	19771.348000	1237.551000	1.439324e+09	19347.500000	48.200000	116935.600000	13.(

+ Code + Markdown

Task 2: [PLO S2 / CLO 2 / SO 2] **[2 marks]**

1. Take a closer look at the distribution of values for the cases and deaths columns.

   Hint : Use NumPy's **arange** method to pass a list of floats from 0 to 1.0 to the **quantile** method of the DataFrame.

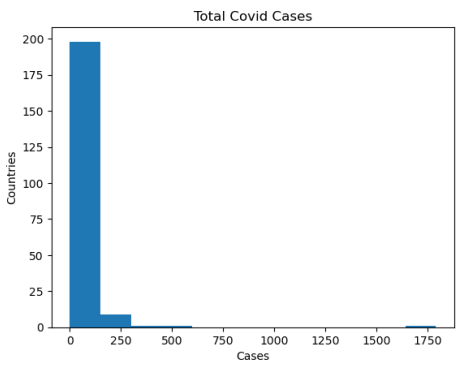
```
total = ['total_cases','total_cases_pm','total_deaths','total_deaths_pm']
covidtotals[total].quantile(np.arange(0.0, 1.1, 0.1))
```

[22]:

	total_cases	total_cases_pm	total_deaths	total_deaths_pm
0.0	0.0	0.0000	0.0	0.0000
0.1	22.9	17.9986	0.0	0.0000
0.2	105.2	56.2910	2.0	0.3752
0.3	302.0	115.4341	6.7	1.7183
0.4	762.0	213.9734	12.0	3.9566
0.5	1242.5	280.9285	25.5	6.1540
0.6	2514.6	543.9562	54.6	12.2452
0.7	6959.8	1071.2442	137.2	25.9459
0.8	16847.2	2206.2982	323.2	49.9658
0.9	46513.1	3765.1363	1616.9	138.9045
1.0	1790191.0	19771.3480	104383.0	1237.5510

2. Draw a histogram of the distribution of total cases.

```
import matplotlib.pyplot as plt

plt.hist(covidtotals['total_cases']/1000, bins=12)

plt.title("Total Covid Cases")

plt.xlabel('Cases')

plt.ylabel("Countries")

plt.show()
```



3. Based on your analysis are the 'total_cases' and ‘total_deaths’ distributions skewed or not? If skewed, to which direction? justify your answer.

Cases and Deaths are skewed to the Right (Positively Skewed)

Mean > Median > Mode