Theory Activity No. 1

Name;- Vivek vilas sanap
PRN:- 202401040091

Roll.no:- cs3-45

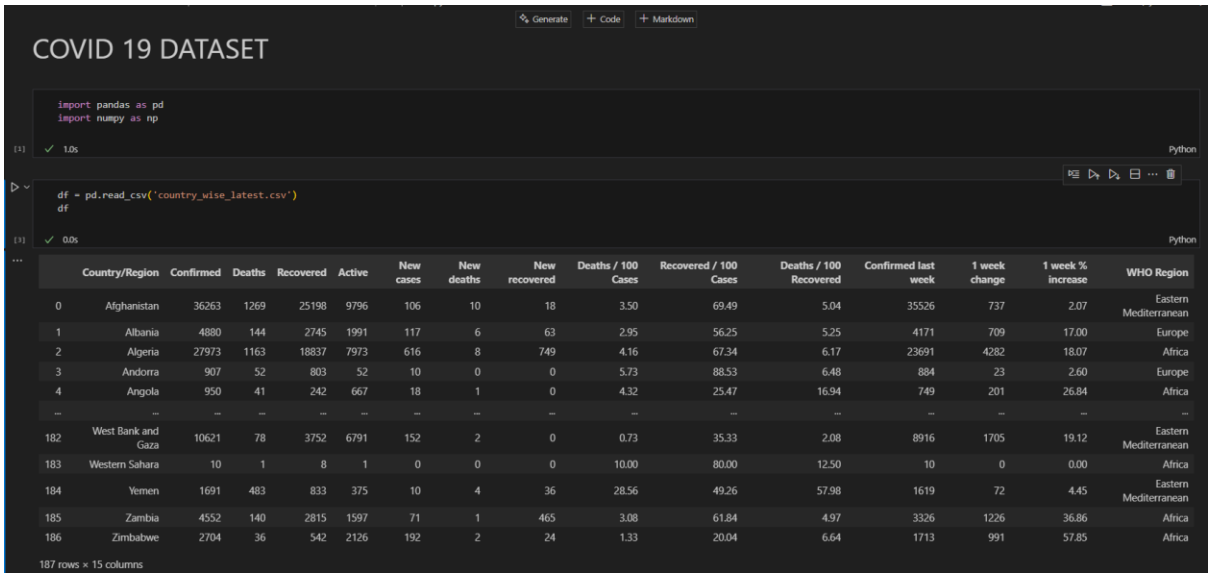
Batch:- c33

Ct_batch:- pe-15

Problem Statement:- Formulate 20 problem statements for a given dataset using Numpy and Pandas

Data_set taken:- COVID-19 Dataset

LINK:- https://www.kaggle.com/datasets/imdevskp/corona-virus-report



COVID 19 DATASET

```
import pandas as pd
import numpy as np
```

[1] ✓ 1.0s Python

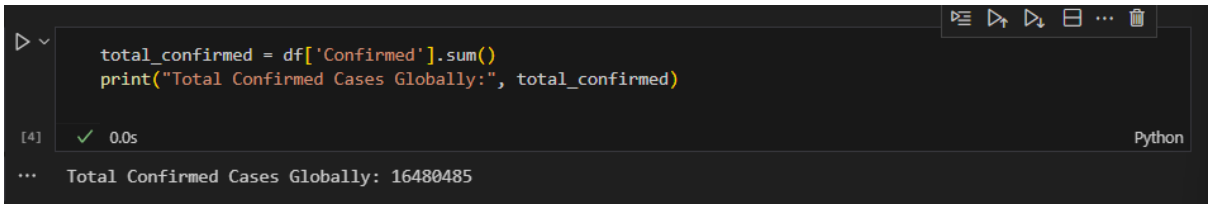
```
df = pd.read_csv('country_wise_latest.csv')
df
```

[3] ✓ 0.0s Python

	Country/Region	Confirmed	Deaths	Recovered	Active	New cases	New deaths	New recovered	Deaths / 100 Cases	Recovered / 100 Cases	Deaths / 100 Recovered	Confirmed last week	1 week change	1 week % increase	WHO Region
0	Afghanistan	36263	1269	25198	9796	106	10	18	3.50	69.49	5.04	35526	737	2.07	Eastern Mediterranean
1	Albania	4880	144	2745	1991	117	6	63	2.95	56.25	5.25	4171	709	17.00	Europe
2	Algeria	27973	1163	18837	7973	616	8	749	4.16	67.34	6.17	23691	4282	18.07	Africa
3	Andorra	907	52	803	52	10	0	0	5.73	88.53	6.48	884	23	2.60	Europe
4	Angola	950	41	242	667	18	1	0	4.32	25.47	16.94	749	201	26.84	Africa
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
182	West Bank and Gaza	10621	78	3752	6791	152	2	0	0.73	35.33	2.08	8916	1705	19.12	Eastern Mediterranean
183	Western Sahara	10	1	8	1	0	0	0	10.00	80.00	12.50	10	0	0.00	Africa
184	Yemen	1691	483	833	375	10	4	36	28.56	49.26	57.98	1619	72	4.45	Eastern Mediterranean
185	Zambia	4552	140	2815	1597	71	1	465	3.08	61.84	4.97	3326	1226	36.86	Africa
186	Zimbabwe	2704	36	542	2126	192	2	24	1.33	20.04	6.64	1713	991	57.85	Africa

187 rows × 15 columns

1. **What is the total number of confirmed COVID-19 cases globally?**



```
total_confirmed = df['Confirmed'].sum()
print("Total Confirmed Cases Globally:", total_confirmed)
```

[4] ✓ 0.0s Python

```
Total Confirmed Cases Globally: 16480485
```

2. **Which country has reported the highest number of deaths?**

```
max_deaths = df['Deaths'].max()
country_max_deaths = df[df['Deaths'] == max_deaths]['Country/Region'].values[0]
print("Country with Most Deaths:", country_max_deaths)
```

[5] ✓ 0.0s Python

```
Country with Most Deaths: US
```

3. **What is the overall global recovery rate (Recovered / Confirmed * 100)?**

```
global_recovery_rate = (df['Recovered'].sum() / df['Confirmed'].sum()) * 100
print("Global Recovery Rate (%):", global_recovery_rate)
```

[6] ✓ 0.0s Python

```
Global Recovery Rate (%): 57.450293483474546
```

4. **What is the average number of active cases per country?**

```
avg_active = df['Active'].mean()
print("Average Active Cases per Country:", avg_active)
```

[7] ✓ 0.0s Python

```
Average Active Cases per Country: 34001.935828877
```

5. **List all countries with more than 10,000 confirmed cases.**

```
countries_10kplus = df[df['Confirmed'] > 10000]['Country/Region'].tolist()
print("Countries with >10,000 cases:", countries_10kplus)
```

[8] ✓ 0.0s Python

```
Countries with >10,000 cases: ['Afghanistan', 'Algeria', 'Argentina', 'Armenia', 'Australia', 'Austria', 'Azerb
```

6. **How many countries reported zero new cases?**

```
zero_new_cases = df[df['New cases'] == 0].shape[0]
print("Countries with Zero New Cases:", zero_new_cases)
```

[9] ✓ 0.0s Python

```
Countries with Zero New Cases: 33
```

7. **What is the median number of deaths per country?**

```
median_deaths = df['Deaths'].median()
print("Median Deaths per Country:", median_deaths)
```

[10] ✓ 0.0s Python

```
Median Deaths per Country: 108.0
```

8. **Which WHO region has the highest total confirmed cases?**

```
region_cases = df.groupby('WHO Region')['Confirmed'].sum()
top_region = region_cases.idxmax()
print("WHO Region with Most Cases:", top_region)
```

[11] ✓ 0.0s Python

```
WHO Region with Most Cases: Americas
```

9. What is the total number of new recovered cases globally?

```
total_new_recovered = df['New recovered'].sum()
print("Total New Recovered Globally:", total_new_recovered)
```

[12] ✓ 0.0s Python

```
Total New Recovered Globally: 174623
```

10.How many countries have a case fatality rate (Deaths / 100 Cases) over 5%?

```
high_fatality = df[df['Deaths / 100 Cases'] > 5].shape[0]
print("Countries with Fatality Rate > 5%:", high_fatality)
```

[13] ✓ 0.0s Python

```
Countries with Fatality Rate > 5%: 29
```

11. Which country had the largest increase in cases in the last week?

```
max_week_change = df['1 week change'].max()
country_max_week_change = df[df['1 week change'] == max_week_change]['Country/Region'].values[0]
print("Country with Largest 1 Week Change:", country_max_week_change)
```

[14] ✓ 0.0s Python

```
Country with Largest 1 Week Change: US
```

12. List all African countries in the dataset.

```
african_countries = df[df['WHO Region'] == 'Africa']['Country/Region'].tolist()
print("African Countries:", african_countries)
```

[15] ✓ 0.0s Python

```
African Countries: ['Algeria', 'Angola', 'Benin', 'Botswana', 'Burkina Faso', 'Burundi', 'Cabo Verde', 'Cameroc
```

13. What is the total number of deaths in the Europe region?

```
europe_deaths = df[df['WHO Region'] == 'Europe']['Deaths'].sum()
print("Total Deaths in Europe:", europe_deaths)
```

[16] ✓ 0.0s Python

```
Total Deaths in Europe: 211144
```

14. Find the country with the highest percentage increase in cases over the last week.

```
max_percent_increase = df['1 week % increase'].max()
country_max_percent_increase = df[df['1 week % increase'] == max_percent_increase]['Country/Region'].values
print("Country with Highest % Increase:", country_max_percent_increase)
```

[17] ✓ 0.0s Python

```
Country with Highest % Increase: Papua New Guinea
```

15.What is the average recovery rate (Recovered / 100 Cases) across all countries?

```
avg_recovery_100 = df['Recovered / 100 Cases'].mean()
print("Average Recovery Rate per 100 Cases:", avg_recovery_100)
```

[18] ✓ 0.0s Python

```
Average Recovery Rate per 100 Cases: 64.82053475935828
```

16. How many countries have more recovered cases than active cases?

```
more_recovered = df[df['Recovered'] > df['Active']].shape[0]
print("Countries with More Recovered than Active:", more_recovered)
```

[20] ✓ 0.0s Python

```
Countries with More Recovered than Active: 140
```

17. Which country has the lowest death rate per 100 recovered cases?

```
min_death_100_recovered = df['Deaths / 100 Recovered'].min()
country_min_death_100_recovered = df[df['Deaths / 100 Recovered'] == min_death_100_recovered]['Country/Regi
print("Country with Lowest Deaths per 100 Recovered:", country_min_death_100_recovered)
```

[21] ✓ 0.0s Python

```
Country with Lowest Deaths per 100 Recovered: Bhutan
```

18. WHAT IS THE SUM OF NEW DEATHS GLOBALLY?

```
total_new_deaths = df['New deaths'].sum()
print("Total New Deaths Globally:", total_new_deaths)
```

[22] ✓ 0.0s Python

```
Total New Deaths Globally: 5415
```

19. What is the average number of confirmed cases last week per country?

```
avg_confirmed_last_week = df['Confirmed last week'].mean()
print("Average Confirmed Last Week per Country:", avg_confirmed_last_week)
```

[23] ✓ 0.0s Python

```
Average Confirmed Last Week per Country: 78682.47593582887
```

20. Which country currently has the highest number of active cases?

```
max_active = df['Active'].max()
country_max_active = df[df['Active'] == max_active]['Country/Region'].values[0]
print("Country with Most Active Cases:", country_max_active)
```

[24] ✓ 0.0s Python

```
Country with Most Active Cases: US
```

Thank you