

ESSENSIAL DATA SCIENCE

Theory Activity No. 1

Name : Swapnil Rajaram Dhokale

Division : ME

Roll No. : ME46

PRN : 202401090030

Subject : EDS

Q.1) Formulate 20 problem statements for a The Blog Authorship Corpus dataset using Numpy and Pandas and Apply Numpy and pandas methods to find the solution for the formulated problem statements.

20 Problem Statements and Solutions on Covid-19 Dataset :

```
import pandas as pd
```

```
import numpy as np
```

```
# Load the dataset
```

```
df = pd.read_csv('/mnt/data/StatewiseTestingDetails(Covid-19).csv')
```

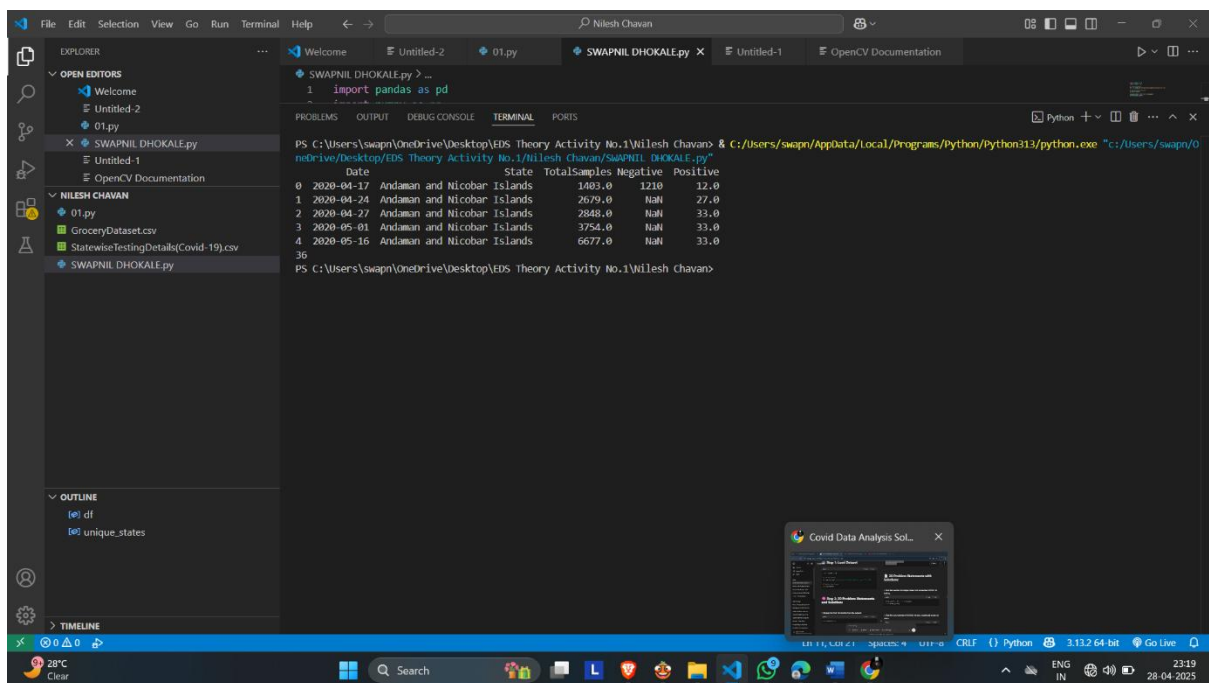
```
# View the first few rows
```

```
print(df.head())
```

1. Find the number of unique states that conducted COVID-19 testing.

```
unique_states = df['State'].nunique()
print(unique_states)
```

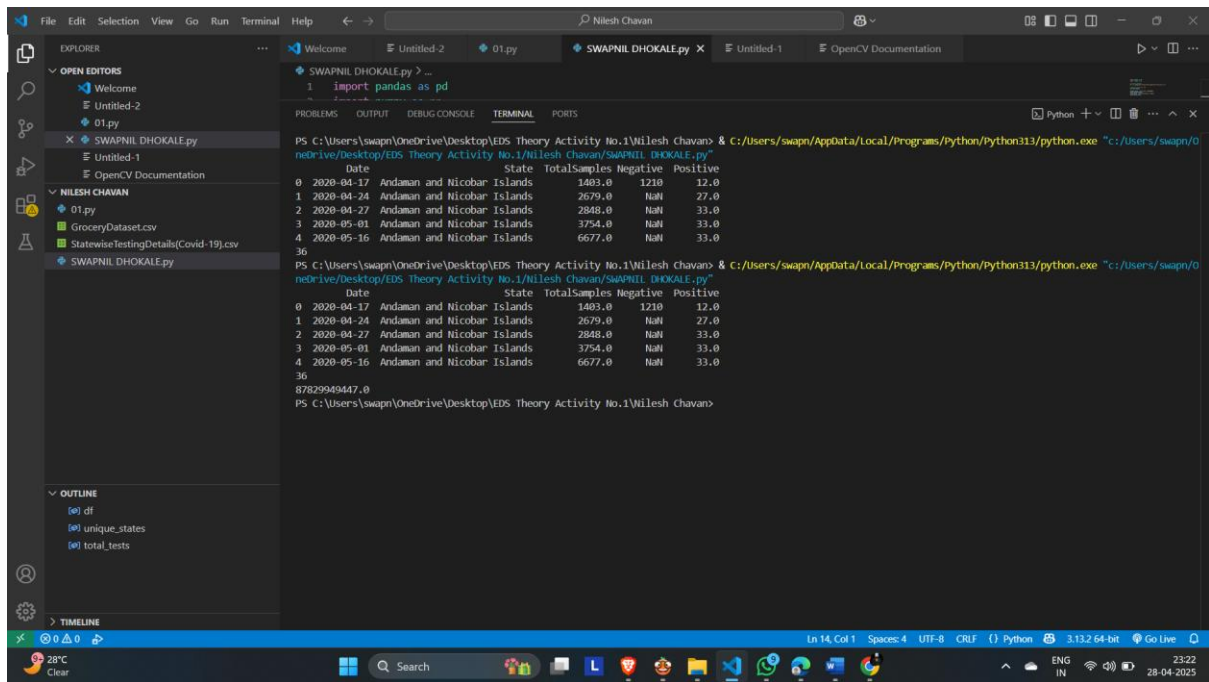
OUTPUT:



2. Find the total number of COVID-19 tests conducted across all states.

```
total_tests = df['TotalSamples'].sum()
print(total_tests)
```

OUTPUT :



3. Identify the state with the highest number of total samples tested.

```

highest_tested_state =
df.groupby('State')['TotalSamples'].max().idxmax()
print(highest_tested_state)

```

OUTPUT:

The screenshot shows a Jupyter Notebook in VS Code with the following content:

```
1 import pandas as pd
```

Date	State	TotalSamples	Negative	Positive
0 2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1 2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2 2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3 2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4 2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0

```
36
```

```
87829949447.0
```

Date	State	TotalSamples	Negative	Positive
0 2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1 2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2 2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3 2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4 2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0

```
36
```

```
87829949447.0
```

```
Uttar Pradesh
```

The notebook also shows an outline of the DataFrame with columns: df, unique_states, total_tests, and highest_tested_state.

4. Find the average number of positive cases across all states.

```
average_positive = df['Positive'].mean()
```

```
print(average_positive)
```

OUTPUT:

```

1 import pandas as pd
2
3 # Load the data
4 df = pd.read_csv('GroceryDataset.csv')
5
6 # Filter the data for Andaman and Nicobar Islands
7 df = df[df['State'] == 'Andaman and Nicobar Islands']
8
9 # Sort the data by Date
10 df = df.sort_values('Date')
11
12 # Display the first 5 rows
13 df.head()

```

Date	State	TotalSamples	Negative	Positive
2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0

5. Calculate the positivity rate (Positive / TotalSamples) for each record.

```
df['Positivity Rate'] = df['Positive'] / df['TotalSamples']
```

```
print(df[['State', 'Date', 'Positivity Rate']])
```

OUTPUT:

```

1 import pandas as pd
2
3 # Load the data
4 df = pd.read_csv('GroceryDataset.csv')
5
6 # Filter the data for Andaman and Nicobar Islands
7 df = df[df['State'] == 'Andaman and Nicobar Islands']
8
9 # Sort the data by Date
10 df = df.sort_values('Date')
11
12 # Display the first 5 rows
13 df.head()

```

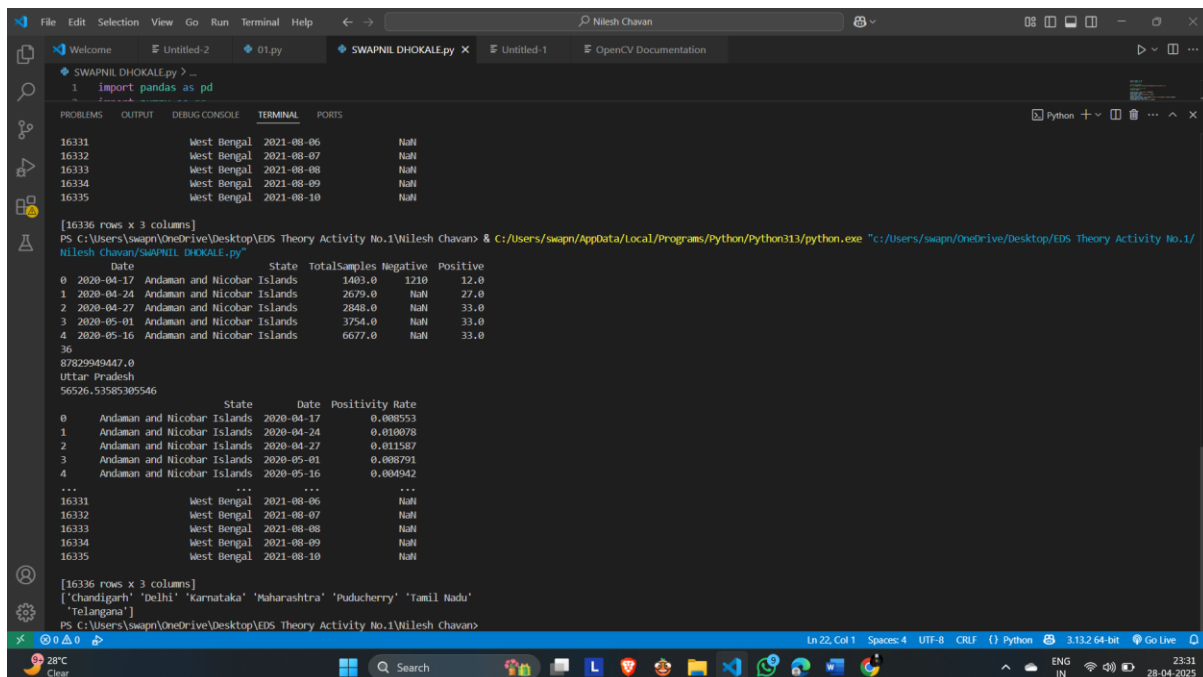
Date	State	TotalSamples	Negative	Positive	Positivity Rate
2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0	0.008553
2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0	0.010078
2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0	0.011587
2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0	0.008791
2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0	0.004942

6. List the states with a positivity rate higher than 10%.

```
high_positivity_states = df[df['Positivity Rate'] >
0.10]['State'].unique()
```

```
print(high_positivity_states)
```

OUTPUT:



```
1 import pandas as pd

[16336 rows x 3 columns]
PS C:\Users\swapi\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan> & C:/Users/swapi/AppData/Local/Programs/Python/Python313/python.exe "C:/Users/swapi/OneDrive/Desktop/EDS Theory Activity No.1/Nilesh Chavan/SWAPNIL DHOKALE.py"

   Date                State  TotalSamples  Negative  Positive
0  2020-04-17  Andaman and Nicobar Islands    1403.0    1210      12.0
1  2020-04-24  Andaman and Nicobar Islands    2679.0     NaN      27.0
2  2020-04-27  Andaman and Nicobar Islands    2848.0     NaN      33.0
3  2020-05-01  Andaman and Nicobar Islands    3754.0     NaN      33.0
4  2020-05-16  Andaman and Nicobar Islands    6677.0     NaN      33.0
36
8782949447.0
Uttar Pradesh
56526.53585305546

   State                Date  Positivity Rate
0  Andaman and Nicobar Islands  2020-04-17    0.008553
1  Andaman and Nicobar Islands  2020-04-24    0.010078
2  Andaman and Nicobar Islands  2020-04-27    0.011587
3  Andaman and Nicobar Islands  2020-05-01    0.008791
4  Andaman and Nicobar Islands  2020-05-16    0.004942
...
...
...
16331  West Bengal  2021-08-06      NaN
16332  West Bengal  2021-08-07      NaN
16333  West Bengal  2021-08-08      NaN
16334  West Bengal  2021-08-09      NaN
16335  West Bengal  2021-08-10      NaN

[16336 rows x 3 columns]
['Chandigarh' 'Delhi' 'Karnataka' 'Maharashtra' 'Puducherry' 'Tamil Nadu'
'Telangana']
PS C:\Users\swapi\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>
```

7. Find the state with the maximum positivity rate on any date.

```
max_positivity_state = df.loc[df['Positivity Rate'].idxmax(),
'State']
```

```
print(max_positivity_state)
```

OUTPUT:

```
# Load the dataset
```

Date	State	TotalSamples	Negative	Positive
2020-04-17	Andaman and Nicobar Islands	1483.0	1210	12.0
2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0

8. Find the total number of entries (records) in the dataset.

```
total_entries = df.shape[0]
```

```
print(total_entries)
```

OUTPUT:

```
# Load the dataset
```

```
total_entries = df.shape[0]
```

```
print(total_entries)
```

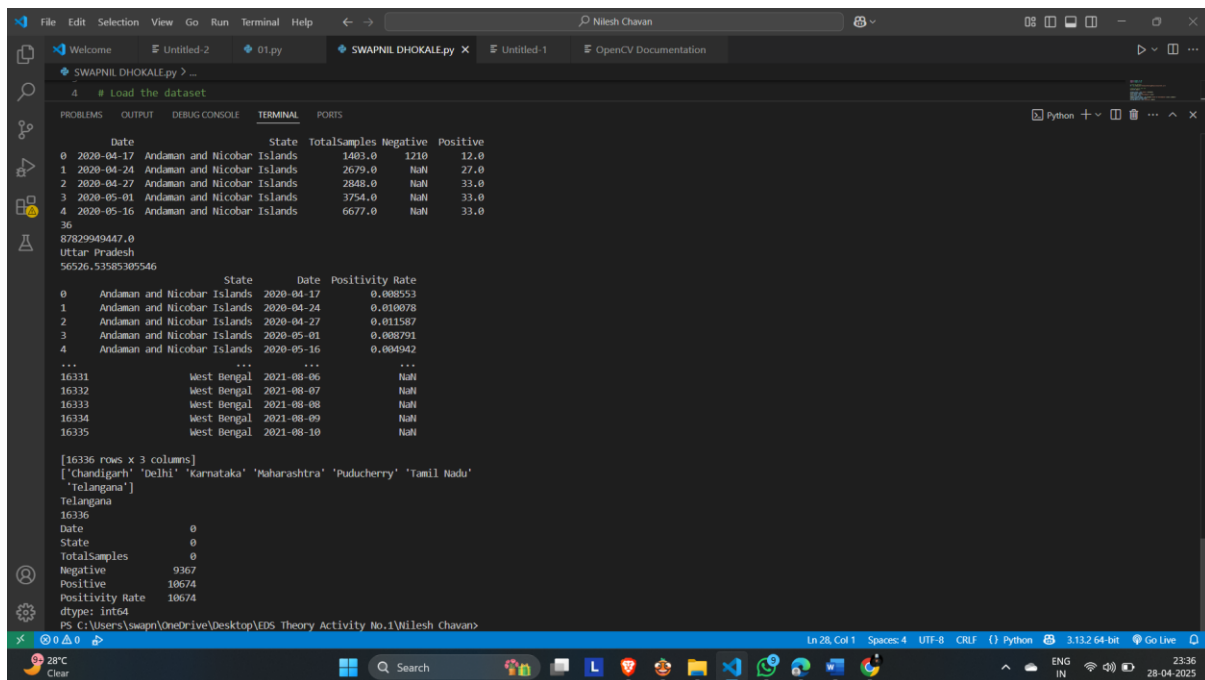
16336

9. Check if there are any missing values in the dataset.

```
missing_values = df.isnull().sum()
```

```
print(missing_values)
```

OUTPUT:



```
File Edit Selection View Go Run Terminal Help Nitesh Chavan
Welcome Untitled-2 01.py SWAPNIL DHOKALE.py X Untitled-1 OpenCV Documentation
4 # Load the dataset

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Date State TotalSamples Negative Positive
0 2020-04-17 Andaman and Nicobar Islands 1403.0 1210 12.0
1 2020-04-24 Andaman and Nicobar Islands 2679.0 NaN 27.0
2 2020-04-27 Andaman and Nicobar Islands 2848.0 NaN 33.0
3 2020-05-01 Andaman and Nicobar Islands 3754.0 NaN 33.0
4 2020-05-16 Andaman and Nicobar Islands 6677.0 NaN 33.0
36
87829949447.0
Uttar Pradesh
56526.53585305546

State Date Positivity Rate
0 Andaman and Nicobar Islands 2020-04-17 0.008953
1 Andaman and Nicobar Islands 2020-04-24 0.010078
2 Andaman and Nicobar Islands 2020-04-27 0.011587
3 Andaman and Nicobar Islands 2020-05-01 0.008791
4 Andaman and Nicobar Islands 2020-05-16 0.004942
...
16331 West Bengal 2021-08-06 NaN
16332 West Bengal 2021-08-07 NaN
16333 West Bengal 2021-08-08 NaN
16334 West Bengal 2021-08-09 NaN
16335 West Bengal 2021-08-10 NaN

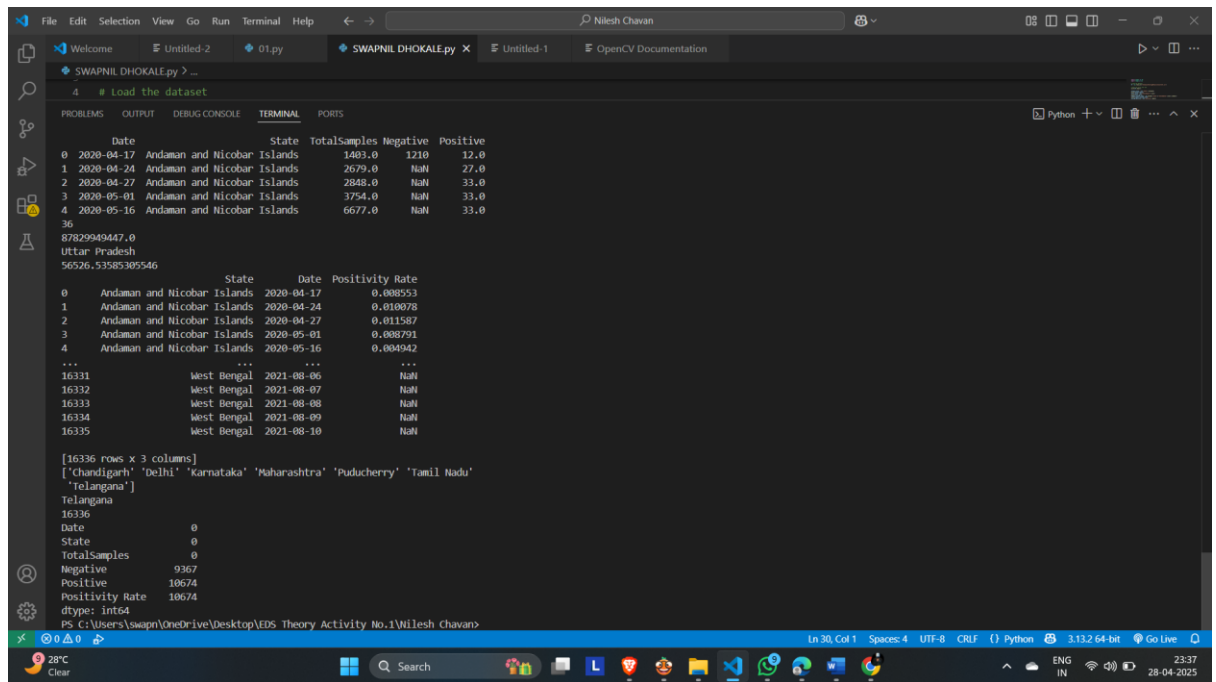
[1636 rows x 3 columns]
['Chandigarh' 'Delhi' 'Karnataka' 'Maharashtra' 'Puducherry' 'Tamil Nadu'
 'Telangana']
16336
Date 0
State 0
TotalSamples 0
Negative 9367
Positive 10674
Positivity Rate 10674
dtype: int64
PS C:\Users\swapn\OneDrive\Desktop\EDS Theory Activity No.1\Nitesh Chavan>
```

10. Fill missing values in 'Positive' and 'TotalSamples' columns with 0.

```
df['Positive'] = df['Positive'].fillna(0)
```

```
df['TotalSamples'] = df['TotalSamples'].fillna(0)
```

OUTPUT:



```
# Load the dataset
df = pd.read_csv('data/covid19_data.csv')

df
```

	Date	State	TotalSamples	Negative	Positive
0	2020-04-17	Andaman and Nicobar Islands	1403.0	1210	12.0
1	2020-04-24	Andaman and Nicobar Islands	2679.0	NaN	27.0
2	2020-04-27	Andaman and Nicobar Islands	2848.0	NaN	33.0
3	2020-05-01	Andaman and Nicobar Islands	3754.0	NaN	33.0
4	2020-05-16	Andaman and Nicobar Islands	6677.0	NaN	33.0
...
16331	2021-08-06	West Bengal	NaN	NaN	NaN
16332	2021-08-07	West Bengal	NaN	NaN	NaN
16333	2021-08-08	West Bengal	NaN	NaN	NaN
16334	2021-08-09	West Bengal	NaN	NaN	NaN
16335	2021-08-10	West Bengal	NaN	NaN	NaN

```
[16336 rows x 5 columns]
['Chandigarh' 'delhi' 'karnataka' 'Maharashtra' 'Puducherry' 'Tamil Nadu'
 'Telangana']
Telangana
16336
Date
0
State
0
TotalSamples
0
Negative
9367
Positive
10674
Positivity Rate
10674
dtype: int64
PS C:\Users\swapi\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>
```

11. Find the first and last date of data collection for COVID-19 tests.

```
first_date = df['Date'].min()
```

```
last_date = df['Date'].max()
```

```
print(first_date, last_date)
```

OUTPUT:

The screenshot shows a Jupyter Notebook with the following content:

```
# Load the dataset
```

	State	Date	Positivity Rate
0	Andaman and Nicobar Islands	2020-04-17	0.008553
1	Andaman and Nicobar Islands	2020-04-24	0.010078
2	Andaman and Nicobar Islands	2020-04-27	0.011587
3	Andaman and Nicobar Islands	2020-05-01	0.008791
4	Andaman and Nicobar Islands	2020-05-16	0.004942
...
16331	West Bengal	2021-08-06	NaN
16332	West Bengal	2021-08-07	NaN
16333	West Bengal	2021-08-08	NaN
16334	West Bengal	2021-08-09	NaN
16335	West Bengal	2021-08-10	NaN

```
[16336 rows x 3 columns]
['Chandigarh' 'Delhi' 'Karnataka' 'Maharashtra' 'Puducherry' 'Tamil Nadu'
 'Telangana']
Telangana
16336
Date
0
State
0
TotalSamples
0
Negative
9367
Positive
10674
Positivity Rate
10674
dtype: int64
2020-04-01 2021-08-10
```

12. Sort the data by 'TotalSamples' in descending order.

```
sorted_df = df.sort_values(by='TotalSamples',
ascending=False)
```

```
print(sorted_df[['State', 'Date', 'TotalSamples']].head(5))
```

OUTPUT:

The screenshot shows the same Jupyter Notebook after sorting the data by 'TotalSamples' in descending order. The output is as follows:

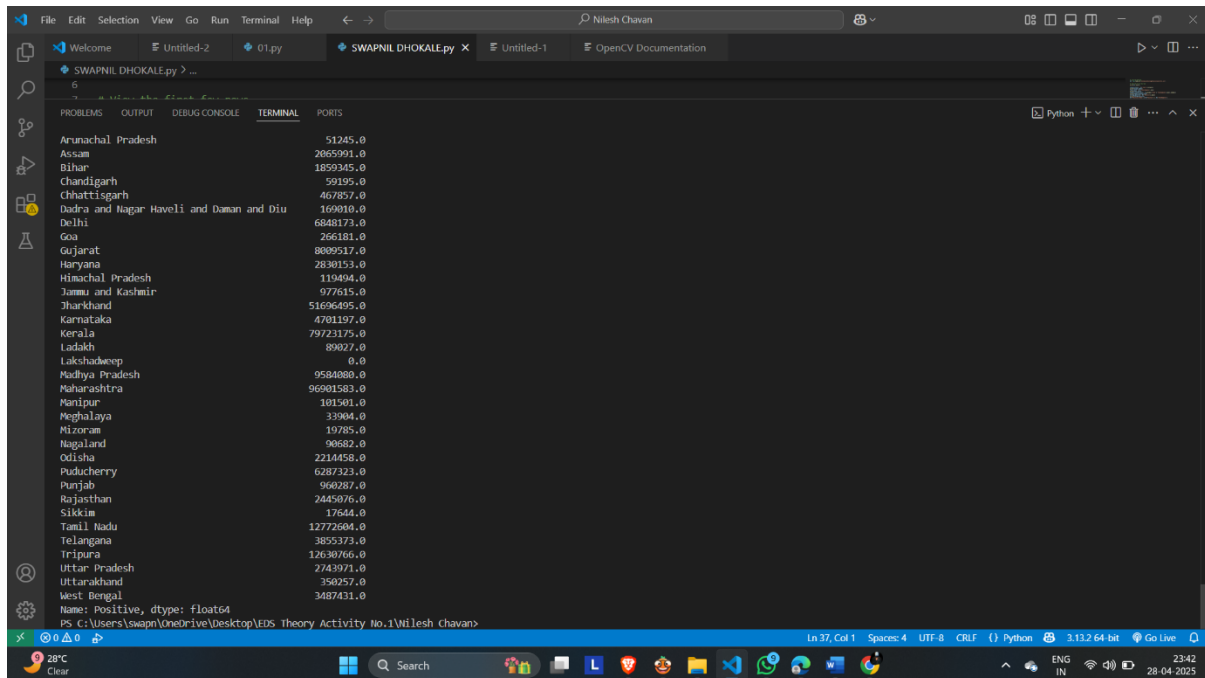
```
[16336 rows x 3 columns]
['Chandigarh' 'Delhi' 'Karnataka' 'Maharashtra' 'Puducherry' 'Tamil Nadu'
 'Telangana']
Telangana
16336
Date
0
State
0
TotalSamples
0
Negative
9367
Positive
10674
Positivity Rate
10674
dtype: int64
2020-04-01 2021-08-10
```

	State	Date	TotalSamples
15251	Uttar Pradesh	2021-08-10	67897856.0
15350	Uttar Pradesh	2021-08-09	67691677.0
15349	Uttar Pradesh	2021-08-08	67476221.0
15348	Uttar Pradesh	2021-08-07	67221784.0
15347	Uttar Pradesh	2021-08-06	66967783.0

13. Group the data by state and calculate total positive cases per state.

```
total_positive_by_state = df.groupby('State')['Positive'].sum()  
print(total_positive_by_state)
```

OUTPUT:



State	Positive
Arunachal Pradesh	51245.0
Assam	2865991.0
Bihar	1859915.0
Chandigarh	59195.0
Chhattisgarh	467857.0
Dadra and Nagar Haveli and Daman and Diu	169010.0
Delhi	6848173.0
Goa	266181.0
Gujarat	8009517.0
Haryana	2830153.0
Himachal Pradesh	119494.0
Jammu and Kashmir	977615.0
Jharkhand	51696495.0
Karnataka	4701197.0
Kerala	79723175.0
Ladakh	89027.0
Lakshadweep	0.0
Madhya Pradesh	9504000.0
Maharashtra	96901583.0
Manipur	101501.0
Meghalaya	33904.0
Mizoram	19785.0
Nagaland	90682.0
Odisha	2214458.0
Puducherry	6287323.0
Punjab	960287.0
Rajasthan	2445076.0
Sikkim	17644.0
Tamil Nadu	12772604.0
Telangana	3855373.0
Tripura	12630766.0
Uttar Pradesh	2743971.0
Uttarakhand	350257.0
West Bengal	3487431.0

14. Find the day when the maximum number of tests was conducted across all states.

```
max_tests_day = df.loc[df['TotalSamples'].idxmax(), 'Date']  
print(max_tests_day)
```

OUTPUT:

```
File Edit Selection View Go Run Terminal Help
Welcome
Untitled-2
01.py
SWAPNIL DHOKALE.py X
Untitled-1
OpenCV Documentation

8 print(df.head())

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Assam 2865991.0
Bihar 1859345.0
Chandigarh 59105.0
Chhattisgarh 467857.0
Dadra and Nagar Haveli and Daman and Diu 169010.0
Delhi 6848173.0
Goa 266181.0
Gujarat 8009517.0
Haryana 2830153.0
Himachal Pradesh 119494.0
Jammu and Kashmir 977615.0
Jharkhand 51696495.0
Karnataka 4701197.0
Kerala 79723175.0
Ladakh 89027.0
Lakshadweep 0.0
Madhya Pradesh 9504800.0
Maharashtra 96001583.0
Manipur 101501.0
Meghalaya 33904.0
Mizoram 19785.0
Nagaland 90682.0
Odisha 2214458.0
Puducherry 6207323.0
Punjab 960287.0
Rajasthan 2445076.0
Sikkim 17644.0
Tamil Nadu 12772604.0
Telangana 3855373.0
Tripura 12630766.0
Uttar Pradesh 2743971.0
Uttarakhand 350257.0
West Bengal 3487431.0
Name: Positive, dtype: float64
2021-08-10
PS C:\Users\swapn\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>
```

15. Get the total tests conducted for a specific state, say Maharashtra.

```
maharashtra_tests = df[df['State'] ==  
'Maharashtra']['TotalSamples'].sum()  
print(maharashtra_tests)
```

OUTPUT:

```
File Edit Selection View Go Run Terminal Help
Welcome
Untitled-2
01.py
SWAPNIL DHOKALE.py X
Untitled-1
OpenCV Documentation

10 unique_states = df['State'].nunique()

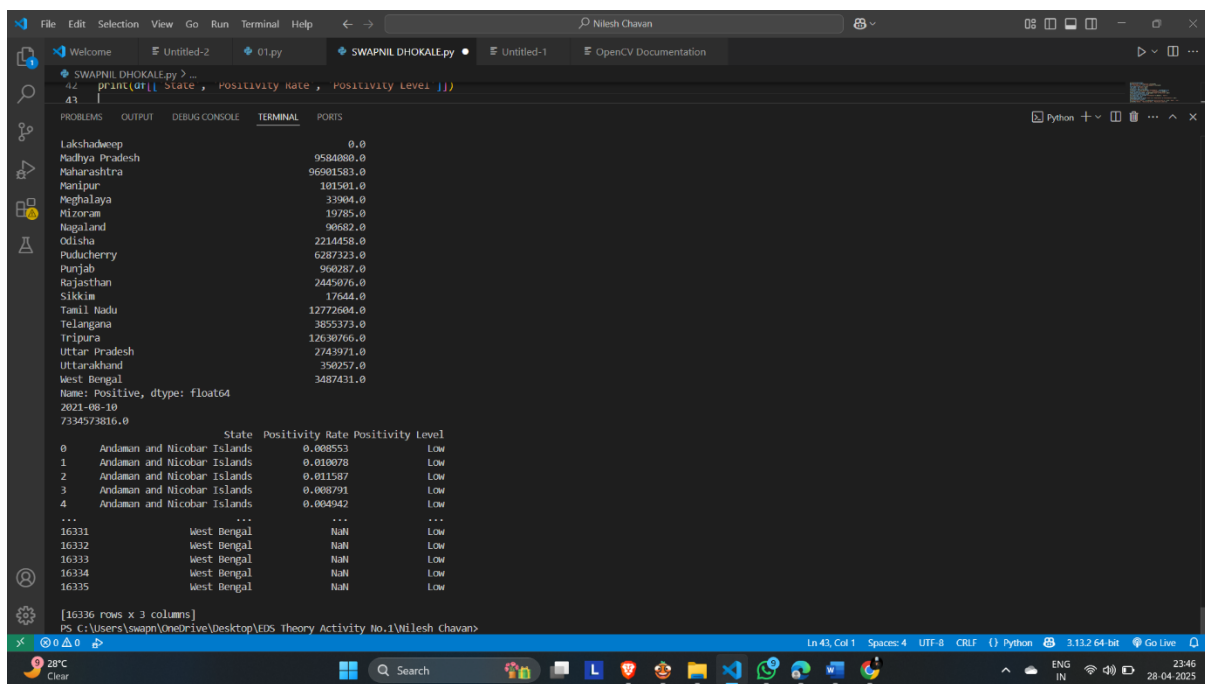
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS
Bihar 1859345.0
Chandigarh 59105.0
Chhattisgarh 467857.0
Dadra and Nagar Haveli and Daman and Diu 169010.0
Delhi 6848173.0
Goa 266181.0
Gujarat 8009517.0
Haryana 2830153.0
Himachal Pradesh 119494.0
Jammu and Kashmir 977615.0
Jharkhand 51696495.0
Karnataka 4701197.0
Kerala 79723175.0
Ladakh 89027.0
Lakshadweep 0.0
Madhya Pradesh 9504800.0
Maharashtra 96001583.0
Manipur 101501.0
Meghalaya 33904.0
Mizoram 19785.0
Nagaland 90682.0
Odisha 2214458.0
Puducherry 6207323.0
Punjab 960287.0
Rajasthan 2445076.0
Sikkim 17644.0
Tamil Nadu 12772604.0
Telangana 3855373.0
Tripura 12630766.0
Uttar Pradesh 2743971.0
Uttarakhand 350257.0
West Bengal 3487431.0
Name: Positive, dtype: float64
2021-08-10
7334573816.0
PS C:\Users\swapn\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>
```

16. Create a new column indicating if the positivity rate is 'High' (>5%) or 'Low'.

```
df['Positivity Level'] = np.where(df['Positivity Rate'] > 0.05, 'High', 'Low')
```

```
print(df[['State', 'Positivity Rate', 'Positivity Level']])
```

OUTPUT:



```
SWARNIL DHOKALE.py
42 print(df[['State', 'Positivity Rate', 'Positivity Level']])
43
```

State	Positivity Rate	Positivity Level
Lakshadweep	0.0	Low
Madhya Pradesh	9584888.0	Low
Maharashtra	96801583.0	Low
Manipur	1015601.0	Low
Meghalaya	33904.0	Low
Mizoram	19785.0	Low
Nagaland	90682.0	Low
Odisha	2214458.0	Low
Puducherry	6287323.0	Low
Punjab	960287.0	Low
Rajasthan	2445876.0	Low
Sikkim	17644.0	Low
Tamil Nadu	12772604.0	Low
Telangana	3855373.0	Low
Tripura	12630766.0	Low
Uttar Pradesh	2743971.0	Low
Uttarakhand	350257.0	Low
West Bengal	3487431.0	Low

```
Name: Positive, dtype: float64
2021-08-10
7334573816.0
```

	State	Positivity Rate	Positivity Level
0	Andaman and Nicobar Islands	0.008553	Low
1	Andaman and Nicobar Islands	0.010078	Low
2	Andaman and Nicobar Islands	0.011587	Low
3	Andaman and Nicobar Islands	0.008791	Low
4	Andaman and Nicobar Islands	0.004942	Low
...
16331	West Bengal	NaN	Low
16332	West Bengal	NaN	Low
16333	West Bengal	NaN	Low
16334	West Bengal	NaN	Low
16335	West Bengal	NaN	Low

```
[16336 rows x 3 columns]
PS C:\Users\swarn\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>
```

17. Find states which never crossed 100 positive cases on any day.

```
states_under_100 = df[df['Positive'] < 100]['State'].unique()
print(states_under_100)
```

OUTPUT:

```

File Edit Selection View Go Run Terminal Help
Nilesh Chavan

SWAPNIL DHOKALE.py x
print(missing_values)

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

Puducherry 6287323.0
Punjab 960287.0
Rajasthan 2445076.0
Sikkim 17600.0
Tamil Nadu 12772604.0
Telangana 3855373.0
Tripura 12630766.0
Uttar Pradesh 2743971.0
Uttarakhand 350257.0
West Bengal 3487431.0

Name: Positive, dtype: float64
2021-08-10
7334573816.0

State Positivity Rate Positivity Level
0 Andaman and Nicobar Islands 0.008553 LOW
1 Andaman and Nicobar Islands 0.010078 LOW
2 Andaman and Nicobar Islands 0.011507 LOW
3 Andaman and Nicobar Islands 0.008791 LOW
4 Andaman and Nicobar Islands 0.004942 LOW
...
16331 West Bengal NaN LOW
16332 West Bengal NaN LOW
16333 West Bengal NaN LOW
16334 West Bengal NaN LOW
16335 West Bengal NaN LOW

[16336 rows x 3 columns]
['Andaman and Nicobar Islands' 'Andhra Pradesh' 'Arunachal Pradesh'
 'Assam' 'Bihar' 'Chandigarh' 'Chhattisgarh'
 'Dadra and Nagar Haveli and Daman and Diu' 'Delhi' 'Goa' 'Gujarat'
 'Haryana' 'Himachal Pradesh' 'Jammu and Kashmir' 'Jharkhand' 'Karnataka'
 'Kerala' 'Ladakh' 'Lakshadweep' 'Madhya Pradesh' 'Maharashtra' 'Manipur'
 'Meghalaya' 'Mizoram' 'Nagaland' 'Odisha' 'Puducherry' 'Punjab'
 'Rajasthan' 'Sikkim' 'Tamil Nadu' 'Telangana' 'Tripura' 'Uttar Pradesh'
 'Uttarakhand' 'West Bengal']
PS C:\Users\swapn\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>

```

18. Find the average total samples tested per state.

```

avg_samples_per_state =
df.groupby('State')['TotalSamples'].mean()
print(avg_samples_per_state)

```

OUTPUT:

```

File Edit Selection View Go Run Terminal Help
Nilesh Chavan

SWAPNIL DHOKALE.py x
print(missing_values)

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

Arunachal Pradesh 3.420978e+05
Assam 6.080424e+06
Bihar 1.511819e+07
Chandigarh 2.082402e+05
Chhattisgarh 3.865413e+06
Dadra and Nagar Haveli and Daman and Diu 3.720157e+04
Delhi 8.815126e+06
Goa 4.097408e+05
Gujarat 9.494690e+06
Haryana 4.180338e+06
Himachal Pradesh 8.752928e+05
Jammu and Kashmir 4.008761e+06
Jharkhand 4.242794e+06
Karnataka 1.379480e+07
Kerala 8.589548e+06
Ladakh 1.433179e+05
Lakshadweep 1.026521e+05
Madhya Pradesh 4.671662e+06
Maharashtra 1.502986e+07
Manipur 4.569285e+05
Meghalaya 3.060135e+05
Mizoram 1.819410e+05
Nagaland 1.071037e+05
Odisha 6.027742e+06
Puducherry 4.907419e+05
Punjab 4.127445e+06
Rajasthan 5.133483e+06
Sikkim 7.214058e+04
Tamil Nadu 1.366841e+07
Telangana 8.109307e+06
Tripura 5.519762e+05
Uttar Pradesh 2.324119e+07
Uttarakhand 1.994440e+06
West Bengal 6.189931e+06

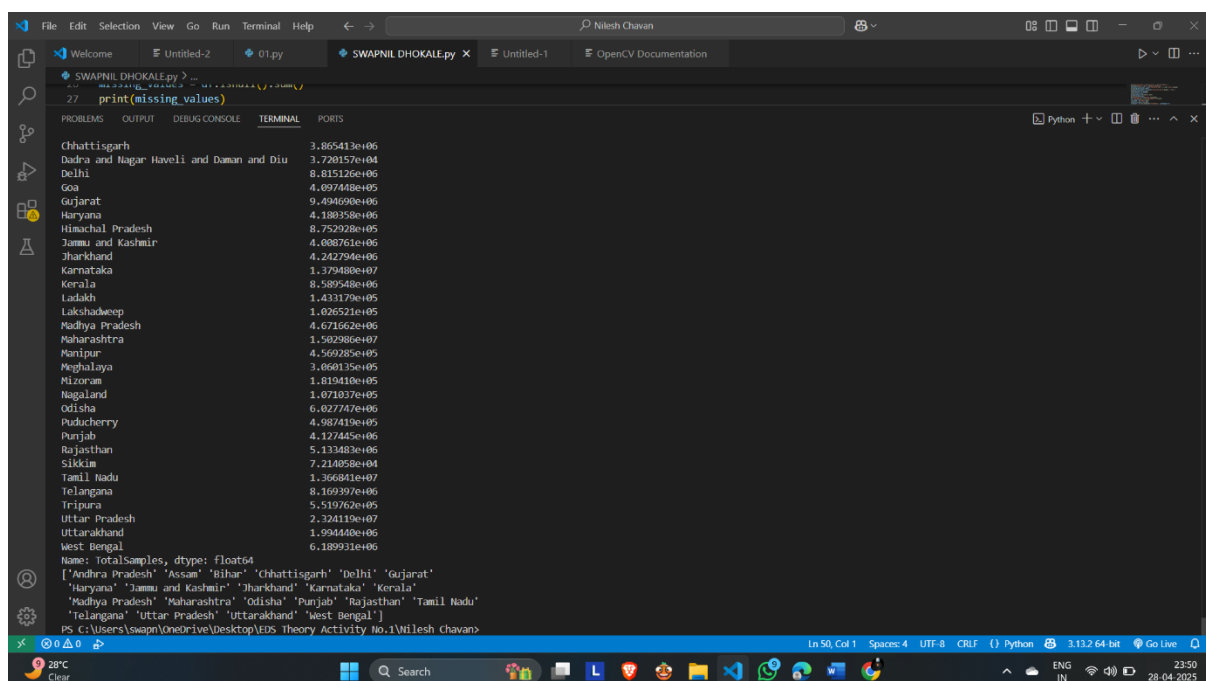
Name: TotalSamples, dtype: float64
PS C:\Users\swapn\OneDrive\Desktop\EDS Theory Activity No.1\Nilesh Chavan>

```

19. Find states where total samples tested are above average.

```
average_tests = df['TotalSamples'].mean()
above_avg_states = df[df['TotalSamples'] >
average_tests]['State'].unique()
print(above_avg_states)
```

OUTPUT:



20. Plot a line graph of total samples tested over time for Delhi.

```
import matplotlib.pyplot as plt

delhi_df = df[df['State'] == 'Delhi'].sort_values('Date')
plt.plot(delhi_df['Date'], delhi_df['TotalSamples'])
plt.title('Total COVID-19 Tests Over Time in Delhi')
```

```
plt.xlabel('Date')
```

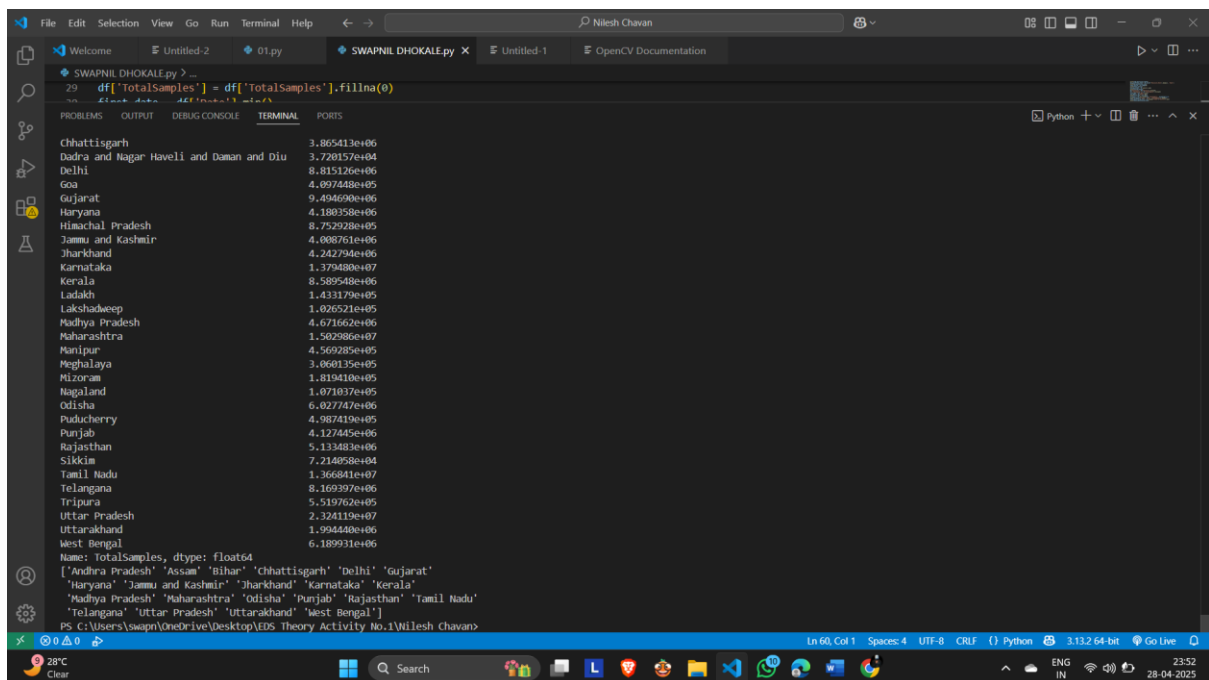
```
plt.ylabel('Total Samples')
```

```
plt.xticks(rotation=45)
```

```
plt.tight_layout()
```

```
plt.show()
```

OUTPUT:



The screenshot shows a VS Code editor with a Python script in a file named 'SWAPNIL DHOKALE.py'. The script contains a single line of code: `df['TotalSamples'] = df['TotalSamples'].fillna(0)`. The terminal output displays a list of Indian states and their corresponding 'TotalSamples' values in scientific notation. The states are listed in descending order of sample count. The output also includes the data type of the 'TotalSamples' column and a list of states where the value was 0.

```
Chhattisgarh 3.865413e+06
Dadra and Nagar Haveli and Daman and Diu 3.728157e+04
Delhi 8.815126e+06
Goa 4.097448e+05
Gujarat 9.494690e+06
Haryana 4.188358e+06
Himachal Pradesh 8.252928e+05
Jammu and Kashmir 4.008761e+06
Jharkhand 4.242794e+06
Karnataka 1.379480e+07
Kerala 8.589548e+06
Ladakh 1.433179e+05
Lakshadweep 1.026521e+05
Madhya Pradesh 4.671662e+06
Maharashtra 1.502886e+07
Manipur 4.569285e+05
Meghalaya 3.060135e+05
Mizoram 1.819418e+05
Nagaland 1.071037e+05
Odisha 6.027747e+06
Puducherry 4.987410e+05
Punjab 4.127445e+06
Rajasthan 5.133483e+06
Sikkim 7.214958e+04
Tamil Nadu 1.366841e+07
Telangana 8.169397e+06
Tripura 5.519762e+05
Uttar Pradesh 2.324119e+07
Uttarakhand 1.994440e+06
West Bengal 6.189931e+06

Name: TotalSamples, dtype: float64
['Andhra Pradesh' 'Assam' 'Bihar' 'Chhattisgarh' 'Delhi' 'Gujarat'
 'Haryana' 'Jammu and Kashmir' 'Jharkhand' 'Karnataka' 'Kerala'
 'Madhya Pradesh' 'Maharashtra' 'Odisha' 'Punjab' 'Rajasthan' 'Tamil Nadu'
 'Telangana' 'Uttar Pradesh' 'Uttarakhand' 'West Bengal']
PS C:\Users\swapn\OneDrive\Desktop\EDS_Theory_Activity_No.1\Nilesh Chavan>
```