Stage2_Task3_Cases

March 14, 2023

0.0.1 Identify 3 counties within a state of your choice with high cases rates.

```
[1]: import pandas as pd
     import numpy as np
     import statistics
```

```
[2]: #I have selected the Albama state for analysis
     selected_state = "AL"
     # reading the confirmed data
     cases = pd.read_csv("../data/covid_confirmed_usafacts.csv")
     cases.head()
```

```
[2]:    countyFIPS          County Name State  StateFIPS  2020-01-22  2020-01-23  \
     0           0  Statewide Unallocated    AL          1           0           0
     1        1001         Autauga County    AL          1           0           0
     2        1003         Baldwin County    AL          1           0           0
     3        1005         Barbour County    AL          1           0           0
     4        1007            Bibb County    AL          1           0           0

        2020-01-24  2020-01-25  2020-01-26  2020-01-27  …  2023-01-07  \
     0           0           0           0           0  …           0
     1           0           0           0           0  …       19205
     2           0           0           0           0  …       68182
     3           0           0           0           0  …        7120
     4           0           0           0           0  …        7808

        2023-01-08  2023-01-09  2023-01-10  2023-01-11  2023-01-12  2023-01-13  \
     0           0           0           0           0           0           0
     1       19205       19205       19205       19205       19318       19318
     2       68182       68182       68182       68182       68518       68518
     3        7120        7120        7120        7120        7188        7188
     4        7808        7808        7808        7808        7855        7855

        2023-01-14  2023-01-15  2023-01-16
     0           0           0           0
     1       19318       19318       19318
     2       68518       68518       68518
     3        7188        7188        7188
```

```
4          7855          7855          7855
```

[5 rows x 1095 columns]

```python
[3]: # using the melt function so that we get the all the dates in one column and␣
     ↪merging will be easy with enrichment data.
     cases_transpose = pd.melt(frame = cases, id_vars=('countyFIPS','County␣
     ↪Name','State','StateFIPS'),var_name=["Date"],value_name='Number of Cases')
     cases_transpose = cases_transpose[cases_transpose['countyFIPS'] != 0]
     cases_transpose.head()
```

```
[3]:    countyFIPS      County Name State  StateFIPS        Date  Number of Cases
     1        1001   Autauga County    AL          1  2020-01-22                0
     2        1003   Baldwin County    AL          1  2020-01-22                0
     3        1005   Barbour County    AL          1  2020-01-22                0
     4        1007      Bibb County    AL          1  2020-01-22                0
     5        1009    Blount County    AL          1  2020-01-22                0
```

```python
[4]: #Dropping the unwanted columns.
     cases_selected_state = cases_transpose[cases_transpose["State"] ==␣
     ↪selected_state]
     cases_selected_state = cases_selected_state.drop(['countyFIPS', 'State',␣
     ↪'StateFIPS'], axis=1)
     cases_selected_state
```

```
[4]:                County Name          Date  Number of Cases
     1          Autauga County    2020-01-22                0
     2          Baldwin County    2020-01-22                0
     3          Barbour County    2020-01-22                0
     4             Bibb County    2020-01-22                0
     5           Blount County    2020-01-22                0
     ...                   ...           ...              ...
     3480433  Tuscaloosa County    2023-01-16            68860
     3480434      Walker County    2023-01-16            23425
     3480435  Washington County    2023-01-16             4309
     3480436      Wilcox County    2023-01-16             3569
     3480437     Winston County    2023-01-16             9200
```

[73097 rows x 3 columns]

```python
[5]: #Number of new cases daily in every county of selected state.
     new_cases_selected_state = cases_selected_state
     new_cases_selected_state['Number of Cases'] = new_cases_selected_state.
     ↪groupby('County Name')['Number of Cases'].diff()
     new_cases_selected_state
```

```
[5]:            County Name        Date   Number of Cases
    1          Autauga County    2020-01-22            NaN
    2          Baldwin County    2020-01-22            NaN
    3          Barbour County    2020-01-22            NaN
    4             Bibb County    2020-01-22            NaN
    5            Blount County    2020-01-22            NaN
    ...                 ...           ...              ...
    3480433  Tuscaloosa County   2023-01-16            0.0
    3480434      Walker County   2023-01-16            0.0
    3480435  Washington County   2023-01-16            0.0
    3480436      Wilcox County   2023-01-16            0.0
    3480437     Winston County   2023-01-16            0.0

    [73097 rows x 3 columns]
```

```
[6]: #Taking the Time range given.
     new_cases_selected_state =␣
      ↪new_cases_selected_state[(new_cases_selected_state["Date"] >= '2022-05-29')␣
      ↪& (new_cases_selected_state["Date"] <= '2023-01-02')]
     new_cases_selected_state
```

```
[6]:            County Name        Date   Number of Cases
    2739595    Autauga County    2022-05-29            0.0
    2739596    Baldwin County    2022-05-29            0.0
    2739597    Barbour County    2022-05-29            0.0
    2739598       Bibb County    2022-05-29            0.0
    2739599      Blount County    2022-05-29            0.0
    ...                 ...           ...              ...
    3435731  Tuscaloosa County   2023-01-02            0.0
    3435732      Walker County   2023-01-02            0.0
    3435733  Washington County   2023-01-02            0.0
    3435734      Wilcox County   2023-01-02            0.0
    3435735     Winston County   2023-01-02            0.0

    [14673 rows x 3 columns]
```

```
[7]: #Summing up the cases in every county and finding the three counties which have␣
      ↪higher number of cases.
     total_cases_selected_state = new_cases_selected_state
     total_cases_selected_state = total_cases_selected_state.groupby('County␣
      ↪Name')['Number of Cases'].sum().reset_index()
     total_cases_selected_state = total_cases_selected_state.sort_values(by =␣
      ↪['Number of Cases'], ascending=False).reset_index(drop=True)
     total_cases_selected_state.head(3)
```

```
[7]:         County Name   Number of Cases
    0   Jefferson County          38529.0
```

```
1    Madison County          19924.0
2     Mobile County          17077.0
```