

Stage 1 Member -Ayodeji Iwayemi

January 28, 2024

1 Name: Ayodeji Iwayemi

1.0.1 Group 1

1.0.2 Individual Member submission

1.0.3 COVID19 Data Analysis

1.0.4 Project Stage - I (Data and Project Understanding)

1.1 - Task 3

1.1.1 Read and display the COVID cases, COVID deaths and Population data

```
[2]: #this cell imports the pandas library and package for use in the notebook
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[3]: #Read the COVID19 cases and save in the 'cases' variable
cases = pd.read_csv('covid_confirmed_usafacts.csv')
```

```
[4]: #Display the COVID19 cases
print('DISPLAY OF COVID19 CASES')
cases
```

DISPLAY OF COVID19 CASES

```
[4]:
```

	countyFIPS	County Name	State	StateFIPS	2020-01-22	\
0	0	Statewide Unallocated	AL	1	0	
1	1001	Autauga County	AL	1	0	
2	1003	Baldwin County	AL	1	0	
3	1005	Barbour County	AL	1	0	
4	1007	Bibb County	AL	1	0	
...	
3188	56037	Sweetwater County	WY	56	0	
3189	56039	Teton County	WY	56	0	
3190	56041	Uinta County	WY	56	0	
3191	56043	Washakie County	WY	56	0	
3192	56045	Weston County	WY	56	0	

	2020-01-23	2020-01-24	2020-01-25	2020-01-26	2020-01-27	...	\
0	0	0	0	0	0	...	
1	0	0	0	0	0	...	
2	0	0	0	0	0	...	
3	0	0	0	0	0	...	
4	0	0	0	0	0	...	
...	
3188	0	0	0	0	0	...	
3189	0	0	0	0	0	...	
3190	0	0	0	0	0	...	
3191	0	0	0	0	0	...	
3192	0	0	0	0	0	...	

	2023-07-14	2023-07-15	2023-07-16	2023-07-17	2023-07-18	2023-07-19	\
0	0	0	0	0	0	0	
1	19913	19913	19913	19913	19913	19913	
2	70521	70521	70521	70521	70521	70521	
3	7582	7582	7582	7582	7582	7582	
4	8149	8149	8149	8149	8149	8149	
...	
3188	12645	12645	12645	12645	12645	12645	
3189	12206	12206	12206	12206	12206	12206	
3190	6468	6468	6468	6468	6468	6468	
3191	2640	2640	2640	2640	2640	2640	
3192	1914	1914	1914	1914	1914	1914	

	2023-07-20	2023-07-21	2023-07-22	2023-07-23
0	0	0	0	0
1	19913	19913	19913	19913
2	70521	70521	70521	70521
3	7582	7582	7582	7582
4	8149	8149	8149	8149
...
3188	12645	12645	12645	12645
3189	12206	12206	12206	12206
3190	6468	6468	6468	6468
3191	2640	2640	2640	2640
3192	1914	1914	1914	1914

[3193 rows x 1269 columns]

```
[5]: #Read the COVID19 cases and save in the 'cases' variable
      deaths = pd.read_csv('covid_deaths_usafacts.csv')
```

```
[6]: print('DISPLAY OF COVID19 DEATHS')
      deaths
```

DISPLAY OF COVID19 DEATHS

```

[6]:      countyFIPS      County Name State  StateFIPS  2020-01-22  \
0          0  Statewide Unallocated    AL          1          0
1         1001      Autauga County    AL          1          0
2         1003      Baldwin County    AL          1          0
3         1005      Barbour County    AL          1          0
4         1007        Bibb County    AL          1          0
...      ...      ...      ...      ...      ...
3188       56037    Sweetwater County    WY          56          0
3189       56039        Teton County    WY          56          0
3190       56041        Uinta County    WY          56          0
3191       56043    Washakie County    WY          56          0
3192       56045      Weston County    WY          56          0

```

```

      2020-01-23  2020-01-24  2020-01-25  2020-01-26  2020-01-27  ...  \
0          0          0          0          0          0  ...
1          0          0          0          0          0  ...
2          0          0          0          0          0  ...
3          0          0          0          0          0  ...
4          0          0          0          0          0  ...
...      ...      ...      ...      ...      ...
3188         0          0          0          0          0  ...
3189         0          0          0          0          0  ...
3190         0          0          0          0          0  ...
3191         0          0          0          0          0  ...
3192         0          0          0          0          0  ...

```

```

      2023-07-14  2023-07-15  2023-07-16  2023-07-17  2023-07-18  2023-07-19  \
0          0          0          0          0          0          0
1         235         235         235         235         235         235
2         731         731         731         731         731         731
3         104         104         104         104         104         104
4         111         111         111         111         111         111
...      ...      ...      ...      ...      ...
3188        142        142        142        142        142        142
3189         16         16         16         16         16         16
3190         43         43         43         43         43         43
3191         51         51         51         51         51         51
3192         24         24         24         24         24         24

```

```

      2023-07-20  2023-07-21  2023-07-22  2023-07-23
0          0          0          0          0
1         235         235         235         235
2         731         731         731         731
3         104         104         104         104
4         111         111         111         111
...      ...      ...      ...
3188        142        142        142        142

```

3189	16	16	16	16
3190	43	43	43	43
3191	51	51	51	51
3192	24	24	24	24

[3193 rows x 1269 columns]

```
[7]: #Read the popoulation and save in the 'population' variable
population = pd.read_csv('covid_county_population_usafacts.csv')
```

```
[8]: print('DISPLAY OF POPULATION')
population
```

DISPLAY OF POPULATION

```
[8]:      countyFIPS      County Name State  population
0           0  Statewide Unallocated    AL           0
1          1001    Autauga County    AL      55869
2          1003    Baldwin County    AL     223234
3          1005    Barbour County    AL      24686
4          1007      Bibb County    AL      22394
...      ...      ...      ...
3190       56037  Sweetwater County    WY      42343
3191       56039      Teton County    WY      23464
3192       56041      Uinta County    WY      20226
3193       56043  Washakie County    WY        7805
3194       56045      Weston County    WY        6927
```

[3195 rows x 4 columns]

1.1.2 Determining the common columns among the three dataframes: cases, deaths, and population

```
[9]: #This section attempts to determine the common columns among the three_
      ↪dataframes for merging purpose
cases_df = set(cases.columns)
deaths_df = set(deaths.columns)
population_df = set(population.columns)
intersec_column = cases_df.intersection(deaths_df).intersection(population_df)
```

```
[10]: #Display the intersecting columns
intersec_column
```

```
[10]: {'County Name', 'State', 'countyFIPS'}
```

1.1.3 Merging the three datasets using the most suitable common column

```
[11]: #Merge the cases and Deaths and name it merged1
merged1 = pd.merge(cases,deaths, how = "inner", left_on =_
↳("countyFIPS","State"), right_on = ("countyFIPS","State"),suffixes =_
↳('cases', 'deaths'))
```

```
[12]: # Display the merged cases and Deaths dataframe
merged1
```

```
[12]:
```

	countyFIPS	County Name	cases	State	StateFIPScases	\
0	0	Statewide Unallocated		AL	1	
1	1001	Autauga County		AL	1	
2	1003	Baldwin County		AL	1	
3	1005	Barbour County		AL	1	
4	1007	Bibb County		AL	1	
...	
3188	56037	Sweetwater County		WY	56	
3189	56039	Teton County		WY	56	
3190	56041	Uinta County		WY	56	
3191	56043	Washakie County		WY	56	
3192	56045	Weston County		WY	56	

	2020-01-22cases	2020-01-23cases	2020-01-24cases	2020-01-25cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	
3188	0	0	0	0	
3189	0	0	0	0	
3190	0	0	0	0	
3191	0	0	0	0	
3192	0	0	0	0	

	2020-01-26cases	2020-01-27cases	...	2023-07-14deaths	\
0	0	0	...	0	
1	0	0	...	235	
2	0	0	...	731	
3	0	0	...	104	
4	0	0	...	111	
...	
3188	0	0	...	142	
3189	0	0	...	16	
3190	0	0	...	43	
3191	0	0	...	51	

3192	0	0	...	24
------	---	---	-----	----

	2023-07-15deaths	2023-07-16deaths	2023-07-17deaths	2023-07-18deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	
4	111	111	111	111	
...	
3188	142	142	142	142	
3189	16	16	16	16	
3190	43	43	43	43	
3191	51	51	51	51	
3192	24	24	24	24	

	2023-07-19deaths	2023-07-20deaths	2023-07-21deaths	2023-07-22deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	
4	111	111	111	111	
...	
3188	142	142	142	142	
3189	16	16	16	16	
3190	43	43	43	43	
3191	51	51	51	51	
3192	24	24	24	24	

	2023-07-23deaths
0	0
1	235
2	731
3	104
4	111
...	...
3188	142
3189	16
3190	43
3191	51
3192	24

[3193 rows x 2536 columns]

```
[13]: #Merge the merged1 with population and name it merged
merged = pd.merge(merged1,population, how = "inner", left_on = "countyFIPS", right_on = ("countyFIPS","State"))
```

```
[14]: # Display the merged cases, deaths and population dataframe
merged
```

```
[14]:
```

	countyFIPS	County Name	cases	State	StateFIPScases	\
0	0	Statewide Unallocated		AL	1	
1	1001	Autauga County		AL	1	
2	1003	Baldwin County		AL	1	
3	1005	Barbour County		AL	1	
4	1007	Bibb County		AL	1	
...	
3188	56037	Sweetwater County		WY	56	
3189	56039	Teton County		WY	56	
3190	56041	Uinta County		WY	56	
3191	56043	Washakie County		WY	56	
3192	56045	Weston County		WY	56	

	2020-01-22cases	2020-01-23cases	2020-01-24cases	2020-01-25cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	
3188	0	0	0	0	
3189	0	0	0	0	
3190	0	0	0	0	
3191	0	0	0	0	
3192	0	0	0	0	

	2020-01-26cases	2020-01-27cases	...	2023-07-16deaths	\
0	0	0	...	0	
1	0	0	...	235	
2	0	0	...	731	
3	0	0	...	104	
4	0	0	...	111	
...	
3188	0	0	...	142	
3189	0	0	...	16	
3190	0	0	...	43	
3191	0	0	...	51	
3192	0	0	...	24	

	2023-07-17deaths	2023-07-18deaths	2023-07-19deaths	2023-07-20deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	

4	111	111	111	111
...
3188	142	142	142	142
3189	16	16	16	16
3190	43	43	43	43
3191	51	51	51	51
3192	24	24	24	24

	2023-07-21deaths	2023-07-22deaths	2023-07-23deaths	\
0	0	0	0	
1	235	235	235	
2	731	731	731	
3	104	104	104	
4	111	111	111	
...	
3188	142	142	142	
3189	16	16	16	
3190	43	43	43	
3191	51	51	51	
3192	24	24	24	

	County Name	population
0	Statewide Unallocated	0
1	Autauga County	55869
2	Baldwin County	223234
3	Barbour County	24686
4	Bibb County	22394
...
3188	Sweetwater County	42343
3189	Teton County	23464
3190	Uinta County	20226
3191	Washakie County	7805
3192	Weston County	6927

[3193 rows x 2538 columns]

1.2 Extracting and Displaying the New York Dataset

```
[15]: # New York Cases extracted from the entire national data of covid-19 cases
casesNY = cases[cases["State"] == "NY"]
```

```
[16]: # Display the New York cases
casesNY
```

```
[16]:   countyFIPS   County Name State StateFIPS 2020-01-22 \
1860         0 Statewide Unallocated   NY         36         0
1861      36001      Albany County   NY         36         0
```


1862	36003	Allegany County	NY	36	0
1863	36005	Bronx County	NY	36	0
1864	36007	Broome County	NY	36	0
...
1918	36115	Washington County	NY	36	0
1919	36117	Wayne County	NY	36	0
1920	36119	Westchester County	NY	36	0
1921	36121	Wyoming County	NY	36	0
1922	36123	Yates County	NY	36	0

	2020-01-23	2020-01-24	2020-01-25	2020-01-26	2020-01-27	...	\
1860	0	0	0	0	0	...	
1861	0	0	0	0	0	...	
1862	0	0	0	0	0	...	
1863	0	0	0	0	0	...	
1864	0	0	0	0	0	...	
...	
1918	0	0	0	0	0	...	
1919	0	0	0	0	0	...	
1920	0	0	0	0	0	...	
1921	0	0	0	0	0	...	
1922	0	0	0	0	0	...	

	2023-07-14	2023-07-15	2023-07-16	2023-07-17	2023-07-18	2023-07-19	\
1860	0	0	0	0	0	0	
1861	80317	80317	80317	80317	80317	80346	
1862	11019	11019	11019	11019	11019	11023	
1863	516506	516506	516506	516506	516506	516827	
1864	60127	60127	60127	60127	60127	60149	
...	
1918	16237	16237	16237	16237	16237	16237	
1919	22585	22585	22585	22585	22585	22592	
1920	340492	340492	340492	340492	340492	340721	
1921	10118	10118	10118	10118	10118	10120	
1922	4543	4543	4543	4543	4543	4544	

	2023-07-20	2023-07-21	2023-07-22	2023-07-23
1860	0	0	0	0
1861	80346	80346	80346	80346
1862	11023	11023	11023	11023
1863	516827	516827	516827	516827
1864	60149	60149	60149	60149
...
1918	16237	16237	16237	16237
1919	22592	22592	22592	22592
1920	340721	340721	340721	340721
1921	10120	10120	10120	10120

1922 4544 4544 4544 4544

[63 rows x 1269 columns]

```
[17]: #New York Deaths
deathsNY = deaths[deaths["State"] == "NY"]
```

```
[18]: # Display the New York deaths
deathsNY
```

```
[18]:
```

	countyFIPS	County Name	State	StateFIPS	2020-01-22	\
1860	0	Statewide Unallocated	NY	36	0	
1861	36001	Albany County	NY	36	0	
1862	36003	Allegany County	NY	36	0	
1863	36005	Bronx County	NY	36	0	
1864	36007	Broome County	NY	36	0	
...	
1918	36115	Washington County	NY	36	0	
1919	36117	Wayne County	NY	36	0	
1920	36119	Westchester County	NY	36	0	
1921	36121	Wyoming County	NY	36	0	
1922	36123	Yates County	NY	36	0	

	2020-01-23	2020-01-24	2020-01-25	2020-01-26	2020-01-27	...	\
1860	0	0	0	0	0	...	
1861	0	0	0	0	0	...	
1862	0	0	0	0	0	...	
1863	0	0	0	0	0	...	
1864	0	0	0	0	0	...	
...	
1918	0	0	0	0	0	...	
1919	0	0	0	0	0	...	
1920	0	0	0	0	0	...	
1921	0	0	0	0	0	...	
1922	0	0	0	0	0	...	

	2023-07-14	2023-07-15	2023-07-16	2023-07-17	2023-07-18	2023-07-19	\
1860	578	578	578	578	578	578	
1861	638	638	638	638	638	638	
1862	166	166	166	166	166	166	
1863	8526	8526	8526	8526	8526	8526	
1864	617	617	617	617	617	617	
...	
1918	140	140	140	140	140	140	
1919	218	218	218	218	218	218	
1920	3031	3031	3031	3031	3031	3031	
1921	98	98	98	98	98	98	

1922	45	45	45	45	45	45
	2023-07-20	2023-07-21	2023-07-22	2023-07-23		
1860	579	579	579	579		
1861	638	638	638	638		
1862	166	166	166	166		
1863	8526	8526	8526	8526		
1864	617	617	617	617		
...		
1918	140	140	140	140		
1919	218	218	218	218		
1920	3032	3032	3032	3032		
1921	98	98	98	98		
1922	45	45	45	45		

[63 rows x 1269 columns]

1.3 Analysing the Last seven days in the USA

```
[19]: #Last seven days of cases
last_seven_days = cases.iloc[:, -7:]
```

```
[20]: # Display the last seven days of cases
last_seven_days
```

```
[20]:      2023-07-17  2023-07-18  2023-07-19  2023-07-20  2023-07-21  2023-07-22  \
0              0              0              0              0              0              0
1          19913          19913          19913          19913          19913          19913
2          70521          70521          70521          70521          70521          70521
3           7582           7582           7582           7582           7582           7582
4           8149           8149           8149           8149           8149           8149
...          ...          ...          ...          ...          ...          ...
3188         12645         12645         12645         12645         12645         12645
3189         12206         12206         12206         12206         12206         12206
3190          6468          6468          6468          6468          6468          6468
3191          2640          2640          2640          2640          2640          2640
3192          1914          1914          1914          1914          1914          1914

      2023-07-23
0              0
1          19913
2          70521
3           7582
4           8149
...          ...
3188         12645
3189         12206
```

```
3190      6468
3191      2640
3192      1914
```

```
[3193 rows x 7 columns]
```

```
[21]: column_sums = last_seven_days.sum()
```

```
[22]: column_sums
```

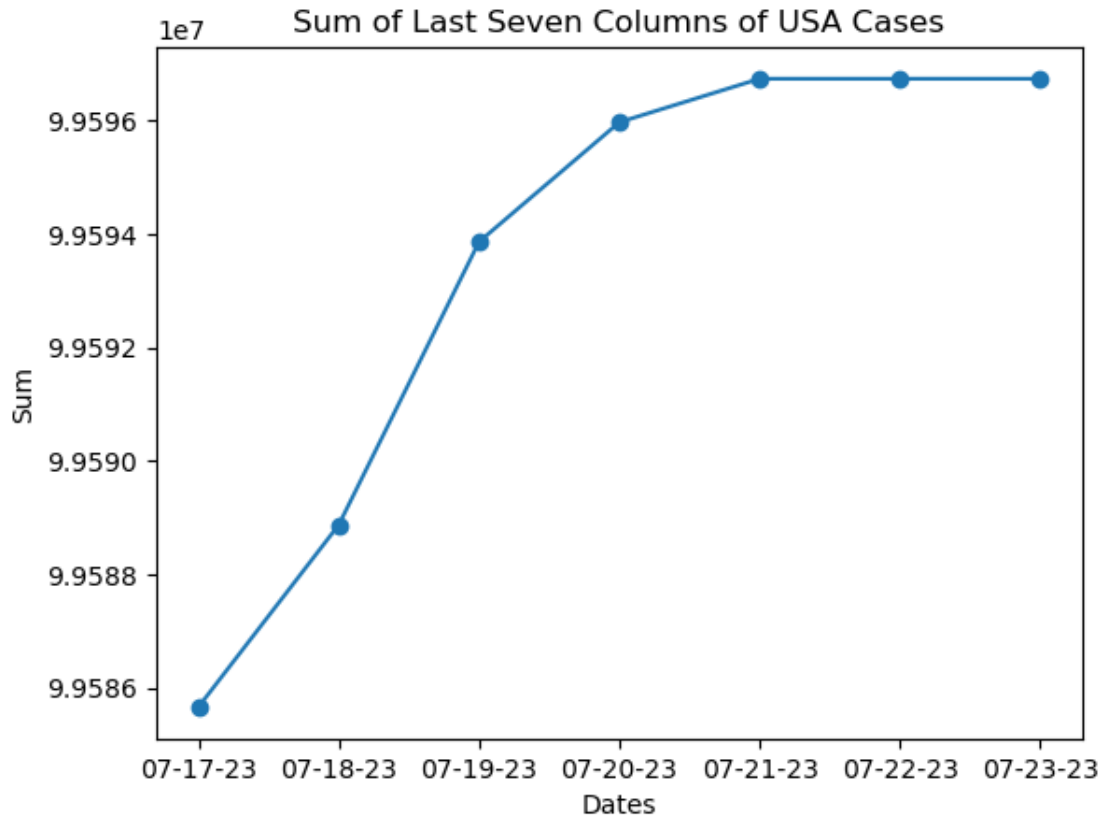
```
[22]: 2023-07-17    99585659
      2023-07-18    99588872
      2023-07-19    99593865
      2023-07-20    99595974
      2023-07-21    99596741
      2023-07-22    99596741
      2023-07-23    99596741
      dtype: int64
```

```
[23]: import pandas as pd
      import matplotlib.pyplot as plt

      x_cases = []
      y_cases = []

      # Using enumerate to iterate over both index and values
      for i, curcol in enumerate(column_sums):
          date = column_sums.index[i]
          x_cases.append(f"{date[-5:]}-23") # Extract the last 5, 6th, 9th, and 10th
          ↪characters
          #x_cases.append(column_sums.index[i][-5,6,9,10:]) # Extract the last 5
          ↪characters from the index
          y_cases.append(curcol)

      plt.plot(x_cases, y_cases, marker='o')
      plt.xlabel('Dates')
      plt.ylabel('Sum')
      plt.title('Sum of Last Seven Columns of USA Cases') # Corrected title
      plt.show()
```



1.3.1 Summary of the Last Seven Days' Infected Cases in the USA (One week Cases)

The COVID 19 cases increased and also remained steady in the last seven days of the dataset under consideration which is that of the USA. Considering the graph and looking at the numbers, the COVID-19 cases are generally increasing over the first few days (that is Monday, July 17, 2023 to Friday, July 21, 2023) but then remain constant from Saturday, July 22, 2023 to Sunday, July 23, 2023. Therefore, during the initial part of the week, the cases increased, but by the end of the week, they remained steady. It can be inferred that the cases increased during the working days of the week due to the increasing work activities but remained constant in the weekend due to the low activities during weekends. ##### Generally this pattern can be caused by: - Workplace and School-Related Transmission - Increased transmission in workplaces or schools during weekdays due to close interactions, commuting, and gatherings. - Reduced workplace and school-related transmission during weekends when people are less likely to be in these settings. - Variations in healthcare system capacity and response during weekdays and weekends. - Strain on healthcare resources during weekdays may result in increased testing and reporting, while weekends may see a more stable capacity. - Changes in people's behavior during weekdays and weekends, affecting exposure and transmission rates. - Compliance with public health measures might vary during the week, influencing the number of new cases.

```
[24]: #Last seven days of Deaths
last_seven_days_deaths = deaths.iloc[:, -7:]
```

```
[25]: last_seven_days_deaths
```

```
[25]:      2023-07-17  2023-07-18  2023-07-19  2023-07-20  2023-07-21  2023-07-22  \
0              0          0          0          0          0          0
1            235         235         235         235         235         235
2            731         731         731         731         731         731
3            104         104         104         104         104         104
4            111         111         111         111         111         111
...          ...          ...          ...          ...          ...          ...
3188          142         142         142         142         142         142
3189           16          16          16          16          16          16
3190           43          43          43          43          43          43
3191           51          51          51          51          51          51
3192           24          24          24          24          24          24

      2023-07-23
0              0
1            235
2            731
3            104
4            111
...          ...
3188          142
3189           16
3190           43
3191           51
3192           24

[3193 rows x 7 columns]
```

```
[26]: column_deaths_sums = last_seven_days_deaths.sum()
```

```
[27]: column_deaths_sums
```

```
[27]: 2023-07-17    1104166
      2023-07-18    1104203
      2023-07-19    1104221
      2023-07-20    1103982
      2023-07-21    1104000
      2023-07-22    1104000
      2023-07-23    1104000
      dtype: int64
```

```
[28]: import pandas as pd
      import matplotlib.pyplot as plt

      x_deaths = []
```

```

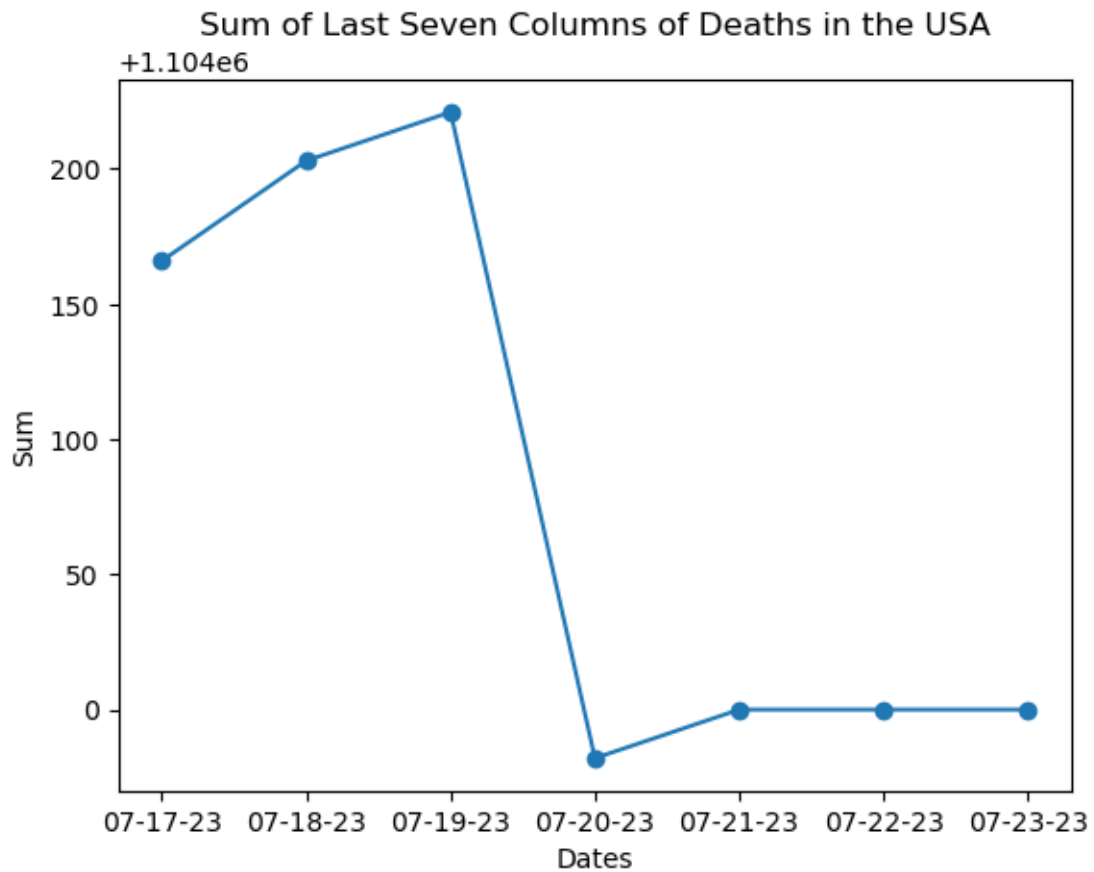
y_deaths = []

# Using enumerate to iterate over both index and values
for i, curcol in enumerate(column_deaths_sums):
    date = column_deaths_sums.index[i]
    x_deaths.append(f"{date[-5:]}-23") # Extract the last 5 characters with
    ↪ the year
    y_deaths.append(curcol)

plt.plot(x_deaths, y_deaths, marker='o')
plt.xlabel('Dates')
plt.ylabel('Sum')
plt.title('Sum of Last Seven Columns of Deaths in the USA') # Corrected title
plt.show()

#plt.plot(column_sums)

```



1.3.2 Summary of the Last Seven Days' Deaths (One week Deaths) in the USA

The COVID 19 deaths increased the first two days of the last seven days. However, the number of deaths reduced sharply in the next day. It remained steady in the last three days of the last seven days. It should be noted that this is with respect to the last seven days of the dataset under consideration in New York state of the USA as shown in the slope of the graph. Looking at the numbers, the COVID-related deaths fluctuate during the week. There is a slight increase from July 17 to July 19, followed by a decrease on July 20, and then the numbers remain steady from July 21 to July 23.

In summary, the data suggests a mixed pattern with some fluctuations during the week, but a general stability in the number of COVID-related deaths towards the end of the provided one-week time frame. It should also be noted that the death cases were steady and constant throughout the weekend starting from Friday, July 21, 2023 to Sunday July 23, 2023. ##### Analysis - Deaths on Mon-Thur: 1104166 (Mon) + 1104203 (Tue) + 1104221 (Wed) + 1103982 (Thu) = 4,416,572 - Deaths on Fri-Sun: 1104000 (Fri) + 1104000 (Sat) + 1104000 (Sun) = 3,312,000 - Difference = (Sum of Deaths on Mon-Thur) - (Sum of Deaths on Fri-Sun) - Difference= 4,416,572 - 3,312,000 ##### Difference = 1,104,572 (Reduced Death cases when compared to the weekdays) - Therefore, the difference between the deaths on Monday through Thursday and the deaths on Friday through Sunday is 1,104,572. ##### Generally this pattern can be caused by: - Healthcare resources, including staff and facilities, may be more strained during weekdays, potentially affecting patient outcomes. - Weekends may see reduced strain on healthcare resources, resulting in better patient care and lower mortality rates. - Fatigue among healthcare professionals during weekdays could impact patient care and outcomes. - Weekends may provide an opportunity for staff to rest, potentially improving patient care and reducing mortality rates. - Reporting Delays - Increased Workplace and School-Related activities during week days and their decrease during weekend

Further analysis, including a detailed examination of local policies, healthcare practices, and reporting mechanisms, would be needed to provide a more definitive understanding of the observed trend.

1.4 Analysing the Last seven days in the NY State of the USA

```
[29]: #Last seven days of cases
last_seven_daysNY = casesNY.iloc[:, -7:]
```

```
[30]: # Display the last seven days of cases
last_seven_daysNY
```

```
[30]:      2023-07-17  2023-07-18  2023-07-19  2023-07-20  2023-07-21  2023-07-22  \
1860              0              0              0              0              0              0
1861          80317          80317          80346          80346          80346          80346
1862          11019          11019          11023          11023          11023          11023
1863         516506         516506         516827         516827         516827         516827
1864          60127          60127          60149          60149          60149          60149
...          ...          ...          ...          ...          ...          ...
1918          16237          16237          16237          16237          16237          16237
1919          22585          22585          22592          22592          22592          22592
1920         340492         340492         340721         340721         340721         340721
```


1921	10118	10118	10120	10120	10120	10120
1922	4543	4543	4544	4544	4544	4544

```

2023-07-23
1860      0
1861    80346
1862    11023
1863   516827
1864    60149
...
1918    16237
1919    22592
1920   340721
1921    10120
1922    4544

```

[63 rows x 7 columns]

```
[31]: column_sumsCasesNY = last_seven_daysNY.sum()
      column_sumsCasesNY
```

```

[31]: 2023-07-17    6703386
      2023-07-18    6703386
      2023-07-19    6706390
      2023-07-20    6706390
      2023-07-21    6706390
      2023-07-22    6706390
      2023-07-23    6706390
      dtype: int64

```

```

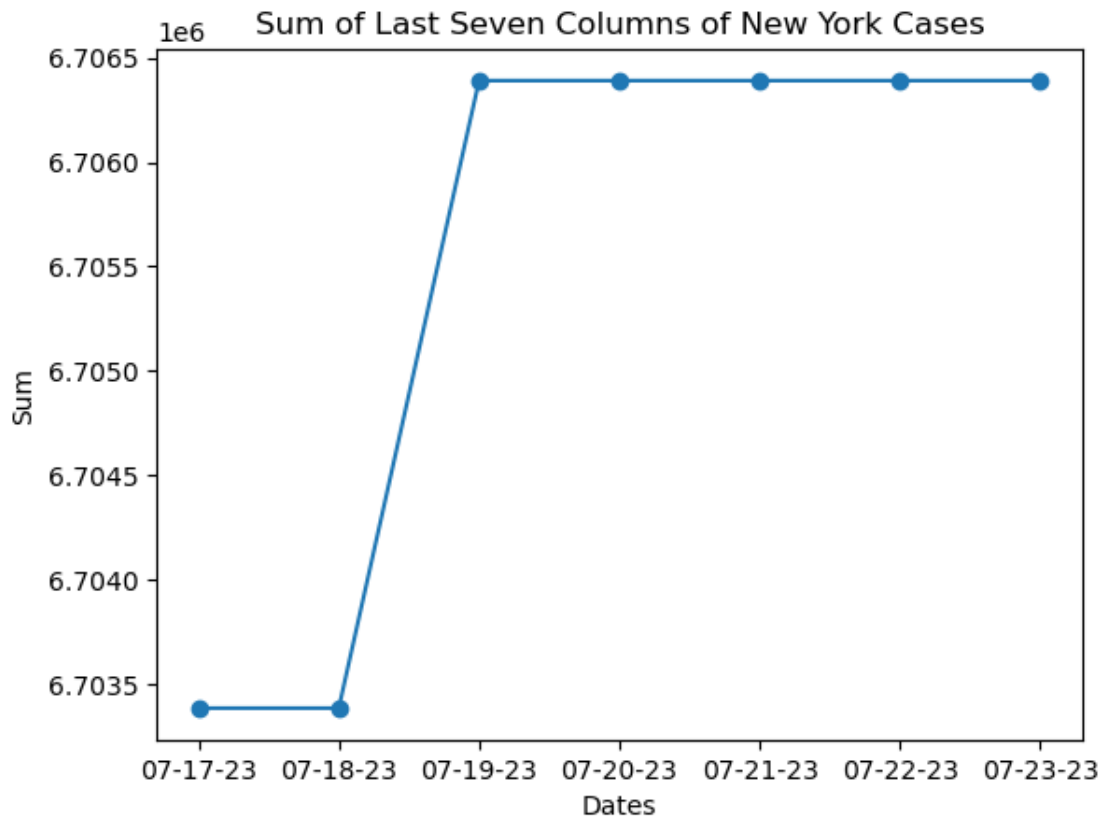
[32]: # Plotting the graph of Last seven days against cases in NY
      x_casesNY = []
      y_casesNY = []

      # Using enumerate to iterate over both index and values
      for i, curcol in enumerate(column_sumsCasesNY):
          date = column_sumsCasesNY.index[i]
          x_casesNY.append(f"{date[-5:]}-23") # Extract the last 5, 6th, 9th, and
          ↪ 10th characters
          #x_cases.append(column_sums.index[i][-5,6,9,10:]) # Extract the last 5
          ↪ characters from the index
          y_casesNY.append(curcol)

      plt.plot(x_casesNY, y_casesNY, marker='o')
      plt.xlabel('Dates')
      plt.ylabel('Sum')
      plt.title('Sum of Last Seven Columns of New York Cases') # Corrected title

```

```
plt.show()
```



1.5 Summary of the trend of the Last Seven Days' Cases (One week Cases) in New York

Looking at the graph, it can be inferred that the COVID-19 cases remained constant from July 17 to July 18, and then increase on July 19. However, from July 19 to July 23, the number of cases appears to be steady, as it remained the same at 6706390.

In summary, for this specific one-week period, the COVID-19 cases increased on July 19 but remained steady from July 19 to July 23. - It should, however, be noted that the cases were constant throughout the weekend without increase or decrease. - It should also be noted that the observation in the entire USA trend of cases is also displayed in the NY State.

```
[33]: #Last seven days of death
last_seven_daysDNY = deathsNY.iloc[:, -7:]
# Display the last seven days of cases
last_seven_daysDNY
```

```
[33]: 2023-07-17 2023-07-18 2023-07-19 2023-07-20 2023-07-21 2023-07-22 \
1860      578      578      578      579      579      579
```

1861	638	638	638	638	638	638
1862	166	166	166	166	166	166
1863	8526	8526	8526	8526	8526	8526
1864	617	617	617	617	617	617
...
1918	140	140	140	140	140	140
1919	218	218	218	218	218	218
1920	3031	3031	3031	3032	3032	3032
1921	98	98	98	98	98	98
1922	45	45	45	45	45	45

2023-07-23

1860	579
1861	638
1862	166
1863	8526
1864	617
...	...
1918	140
1919	218
1920	3032
1921	98
1922	45

[63 rows x 7 columns]

```
[34]: column_sumsDeathsNY = last_seven_daysDNY.sum()
      column_sumsDeathsNY
```

```
[34]: 2023-07-17    77681
      2023-07-18    77681
      2023-07-19    77681
      2023-07-20    77423
      2023-07-21    77423
      2023-07-22    77423
      2023-07-23    77423
      dtype: int64
```

```
[35]: # Plotting the graph of Last seven days against deaths in NY
      x_deathsNY = []
      y_deathsNY = []

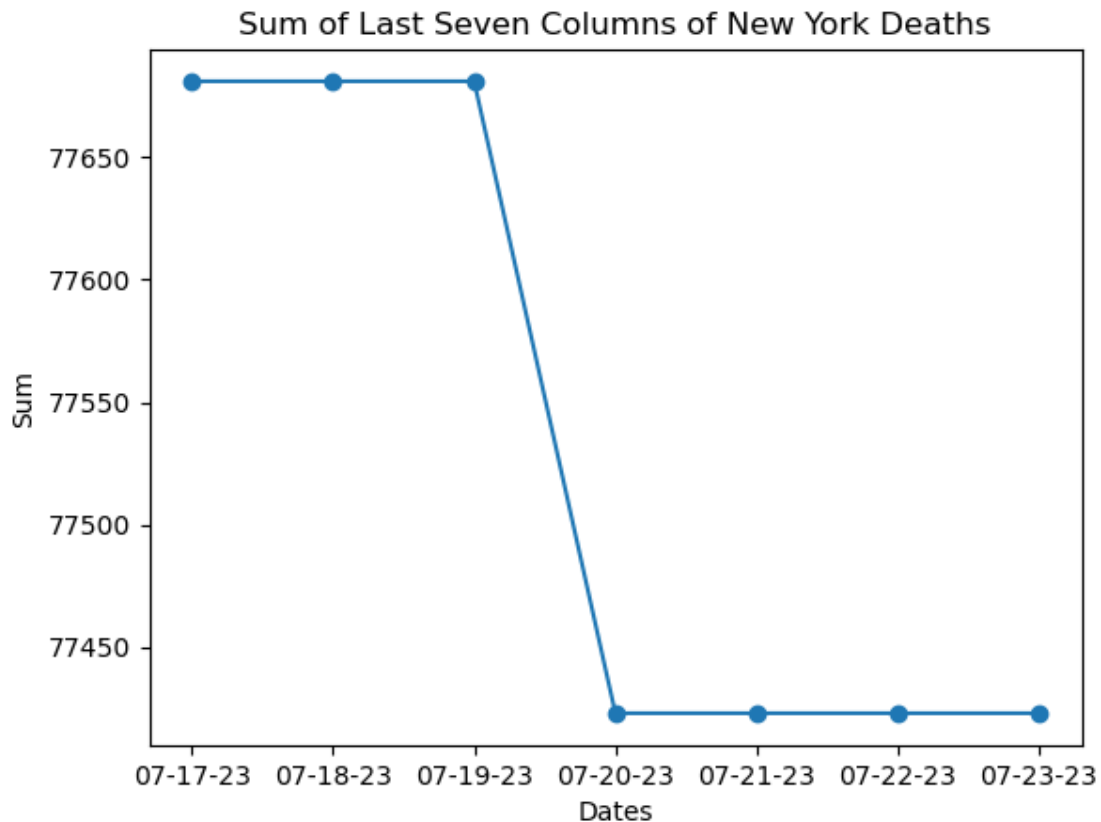
      # Using enumerate to iterate over both index and values
      for i, curcol in enumerate(column_sumsDeathsNY):
          date = column_sumsDeathsNY.index[i]
          x_deathsNY.append(f"{date[-5:]}-23") # Extract the last 5, 6th, 9th, and
          ↪10th characters
```

```

    #x_cases.append(column_sums.index[i][-5,6,9,10:]) # Extract the last 5
    ↪characters from the index
    y_deathsNY.append(curcol)

plt.plot(x_deathsNY, y_deathsNY, marker='o')
plt.xlabel('Dates')
plt.ylabel('Sum')
plt.title('Sum of Last Seven Columns of New York Deaths') # Corrected title
plt.show()

```



1.5.1 Summary of the trend of the Last Seven Days' Deaths (One week Deaths) from Covid19 in New York State

- can be observed that the number of deaths remains constant at 77681 from July 17 to July 19. However, there is a decrease in the number of deaths from July 19 to July 20, and then the number remains constant at 77423 from July 20 to July 23.
- In summary, for this specific one-week period, the COVID-19-related deaths in New York remained constant for the first three days, decreased on the fourth day, and then stayed steady for the remaining days of the week.
- It should also be noted that the observation in the entire USA trend of deaths is also displayed

in the NY State

1.6 Enrichment Dataset - Presidential Election Results Data

1.6.1 Merge enrichment data with super covid19 dataframe

```
[36]: #Display the presidential election results dataframe
presidential = pd.read_csv('president_county_candidate.csv')
presidential
```

```
[36]:
```

	state	county	candidate	party	total_votes	won
0	Delaware	Kent County	Joe Biden	DEM	44552	True
1	Delaware	Kent County	Donald Trump	REP	41009	False
2	Delaware	Kent County	Jo Jorgensen	LIB	1044	False
3	Delaware	Kent County	Howie Hawkins	GRN	420	False
4	Delaware	New Castle County	Joe Biden	DEM	195034	True
...
32172	Arizona	Maricopa County	Write-ins	WRI	1331	False
32173	Arizona	Mohave County	Donald Trump	REP	78535	True
32174	Arizona	Mohave County	Joe Biden	DEM	24831	False
32175	Arizona	Mohave County	Jo Jorgensen	LIB	1302	False
32176	Arizona	Mohave County	Write-ins	WRI	37	False

[32177 rows x 6 columns]

```
[37]: #Display the merged covid19 cases, covid19 deaths, and population dataframe and
      ↪store in variable 'covid_df'
covid_df = merged
covid_df
```

```
[37]:
```

	countyFIPS	County Name	cases	State	StateFIPScases	\
0	0	Statewide Unallocated		AL	1	
1	1001	Autauga County		AL	1	
2	1003	Baldwin County		AL	1	
3	1005	Barbour County		AL	1	
4	1007	Bibb County		AL	1	
...
3188	56037	Sweetwater County		WY	56	
3189	56039	Teton County		WY	56	
3190	56041	Uinta County		WY	56	
3191	56043	Washakie County		WY	56	
3192	56045	Weston County		WY	56	

	2020-01-22cases	2020-01-23cases	2020-01-24cases	2020-01-25cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	

4	0	0	0	0
...
3188	0	0	0	0
3189	0	0	0	0
3190	0	0	0	0
3191	0	0	0	0
3192	0	0	0	0

	2020-01-26cases	2020-01-27cases	...	2023-07-16deaths	\
0	0	0	...	0	
1	0	0	...	235	
2	0	0	...	731	
3	0	0	...	104	
4	0	0	...	111	
...	
3188	0	0	...	142	
3189	0	0	...	16	
3190	0	0	...	43	
3191	0	0	...	51	
3192	0	0	...	24	

	2023-07-17deaths	2023-07-18deaths	2023-07-19deaths	2023-07-20deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	
4	111	111	111	111	
...	
3188	142	142	142	142	
3189	16	16	16	16	
3190	43	43	43	43	
3191	51	51	51	51	
3192	24	24	24	24	

	2023-07-21deaths	2023-07-22deaths	2023-07-23deaths	\
0	0	0	0	
1	235	235	235	
2	731	731	731	
3	104	104	104	
4	111	111	111	
...	
3188	142	142	142	
3189	16	16	16	
3190	43	43	43	
3191	51	51	51	
3192	24	24	24	

	County Name	population
0	Statewide Unallocated	0
1	Autauga County	55869
2	Baldwin County	223234
3	Barbour County	24686
4	Bibb County	22394
...
3188	Sweetwater County	42343
3189	Teton County	23464
3190	Uinta County	20226
3191	Washakie County	7805
3192	Weston County	6927

[3193 rows x 2538 columns]

Observation and Understanding of the covid(cases, deaths, with population) dataframe and the enrichment dataframe for preparation for merging

```
[38]: # Identify the unique data in 'County Namecases' column of the merged Covid19_
      ↪ DataFrame
uniqueCountyNamesCovid = covid_df['County Namecases'].unique()
# Identify the unique data in 'county' column of the presidential DataFrame
uniqueCountyNamesPre = presidential['county'].unique()
print("The Unique data in the 'County Namecases' column of the COVID19_
      ↪ dataframe are:\n", uniqueCountyNamesCovid)
print()
print("The Unique data in the 'county' column of the Presidential dataframe are:
      ↪ \n", uniqueCountyNamesPre)
```

The Unique data in the 'County Namecases' column of the COVID19 dataframe are:
 ['Statewide Unallocated' 'Autauga County' 'Baldwin County' ...
 'Uinta County' 'Washakie County' 'Weston County']

The Unique data in the 'county' column of the Presidential dataframe are:
 ['Kent County' 'New Castle County' 'Sussex County' ... 'La Paz County'
 'Maricopa County' 'Mohave County']

```
[39]: # Display of the covid(cases, deaths, with population) dataframe for_
      ↪ observation and understanding of its columns
covid_df
```

```
[39]:   countyFIPS   County Namecases State StateFIPScases \
0         0 Statewide Unallocated   AL             1
1        1001   Autauga County      AL             1
2        1003   Baldwin County      AL             1
3        1005   Barbour County      AL             1
4        1007    Bibb County        AL             1
...      ...                   ...   ...             ...
```

3188	56037	Sweetwater County	WY	56
3189	56039	Teton County	WY	56
3190	56041	Uinta County	WY	56
3191	56043	Washakie County	WY	56
3192	56045	Weston County	WY	56

	2020-01-22cases	2020-01-23cases	2020-01-24cases	2020-01-25cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	
3188	0	0	0	0	
3189	0	0	0	0	
3190	0	0	0	0	
3191	0	0	0	0	
3192	0	0	0	0	

	2020-01-26cases	2020-01-27cases	...	2023-07-16deaths	\
0	0	0	...	0	
1	0	0	...	235	
2	0	0	...	731	
3	0	0	...	104	
4	0	0	...	111	
...	
3188	0	0	...	142	
3189	0	0	...	16	
3190	0	0	...	43	
3191	0	0	...	51	
3192	0	0	...	24	

	2023-07-17deaths	2023-07-18deaths	2023-07-19deaths	2023-07-20deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	
4	111	111	111	111	
...	
3188	142	142	142	142	
3189	16	16	16	16	
3190	43	43	43	43	
3191	51	51	51	51	
3192	24	24	24	24	

	2023-07-21deaths	2023-07-22deaths	2023-07-23deaths	\
0	0	0	0	

1	235	235	235
2	731	731	731
3	104	104	104
4	111	111	111
...
3188	142	142	142
3189	16	16	16
3190	43	43	43
3191	51	51	51
3192	24	24	24

	County Name	population
0	Statewide Unallocated	0
1	Autauga County	55869
2	Baldwin County	223234
3	Barbour County	24686
4	Bibb County	22394
...
3188	Sweetwater County	42343
3189	Teton County	23464
3190	Uinta County	20226
3191	Washakie County	7805
3192	Weston County	6927

[3193 rows x 2538 columns]

Clean the 'County Namecases' entries of the COVID19 dataframe and renaming the name of the column for uniformity with its equivalent in presidential dataframe

```
[40]: # Clean the 'County Namecases' entries of the COVID19 dataframe by removing
      ↪ whitespaces at the end of the entries
covid_df['County Namecases'] = covid_df['County Namecases'].str.strip()
# Renaming the 'County Namecases' column as 'county' in the covid dataframe to
      ↪ be the same as that of the presidential dataframe
covid_df = covid_df.rename(columns={'County Namecases': 'county'})
covid_df
```

[40]:	countyFIPS	county	State	StateFIPScases	\
0	0	Statewide Unallocated	AL	1	
1	1001	Autauga County	AL	1	
2	1003	Baldwin County	AL	1	
3	1005	Barbour County	AL	1	
4	1007	Bibb County	AL	1	
...	
3188	56037	Sweetwater County	WY	56	
3189	56039	Teton County	WY	56	
3190	56041	Uinta County	WY	56	
3191	56043	Washakie County	WY	56	

3192	56045	Weston County	WY	56	
	2020-01-22cases	2020-01-23cases	2020-01-24cases	2020-01-25cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	
3188	0	0	0	0	
3189	0	0	0	0	
3190	0	0	0	0	
3191	0	0	0	0	
3192	0	0	0	0	
	2020-01-26cases	2020-01-27cases	...	2023-07-16deaths	\
0	0	0	...	0	
1	0	0	...	235	
2	0	0	...	731	
3	0	0	...	104	
4	0	0	...	111	
...	
3188	0	0	...	142	
3189	0	0	...	16	
3190	0	0	...	43	
3191	0	0	...	51	
3192	0	0	...	24	
	2023-07-17deaths	2023-07-18deaths	2023-07-19deaths	2023-07-20deaths	\
0	0	0	0	0	
1	235	235	235	235	
2	731	731	731	731	
3	104	104	104	104	
4	111	111	111	111	
...	
3188	142	142	142	142	
3189	16	16	16	16	
3190	43	43	43	43	
3191	51	51	51	51	
3192	24	24	24	24	
	2023-07-21deaths	2023-07-22deaths	2023-07-23deaths	\	
0	0	0	0		
1	235	235	235		
2	731	731	731		
3	104	104	104		
4	111	111	111		

```

...
3188      142      142      142
3189      16      16      16
3190      43      43      43
3191      51      51      51
3192      24      24      24

```

```

      County Name  population
0  Statewide Unallocated      0
1    Autauga County    55869
2    Baldwin County   223234
3    Barbour County   24686
4    Bibb County     22394
...
3188  Sweetwater County   42343
3189    Teton County    23464
3190    Uinta County    20226
3191  Washakie County    7805
3192    Weston County   6927

```

[3193 rows x 2538 columns]

Replacement of the State Names in the Presidential Dataframe with their two-letter postal abbreviations equivalence in the covid dataframe

```

[41]: # Replacement of the State Names in the Presidential Dataframe with their
      ↪two-letter postal abbreviations
states = {
    'Alaska': 'AK',
    'Alabama': 'AL',
    'Arkansas': 'AR',
    'Arizona': 'AZ',
    'California': 'CA',
    'Colorado': 'CO',
    'Connecticut': 'CT',
    'District of Columbia': 'DC',
    'Delaware': 'DE',
    'Florida': 'FL',
    'Georgia': 'GA',
    'Hawaii': 'HI',
    'Iowa': 'IA',
    'Idaho': 'ID',
    'Illinois': 'IL',
    'Indiana': 'IN',
    'Kansas': 'KS',
    'Kentucky': 'KY',
    'Louisiana': 'LA',
    'Massachusetts': 'MA',

```

```

'Maryland': 'MD',
'Maine': 'ME',
'Michigan': 'MI',
'Minnesota': 'MN',
'Missouri': 'MO',
'Mississippi': 'MS',
'Montana': 'MT',
'North Carolina': 'NC',
'North Dakota': 'ND',
'Nebraska': 'NE',
'New Hampshire': 'NH',
'New Jersey': 'NJ',
'New Mexico': 'NM',
'Nevada': 'NV',
'New York': 'NY',
'Ohio': 'OH',
'Oklahoma': 'OK',
'Oregon': 'OR',
'Pennsylvania': 'PA',
'Rhode Island': 'RI',
'South Carolina': 'SC',
'South Dakota': 'SD',
'Tennessee': 'TN',
'Texas': 'TX',
'Utah': 'UT',
'Virginia': 'VA',
'Vermont': 'VT',
'Washington': 'WA',
'Wisconsin': 'WI',
'West Virginia': 'WV',
'Wyoming': 'WY'
}
presidential['state'] = presidential['state'].replace(states)
presidential

```

```

[41]:

```

	state	county	candidate	party	total_votes	won
0	DE	Kent County	Joe Biden	DEM	44552	True
1	DE	Kent County	Donald Trump	REP	41009	False
2	DE	Kent County	Jo Jorgensen	LIB	1044	False
3	DE	Kent County	Howie Hawkins	GRN	420	False
4	DE	New Castle County	Joe Biden	DEM	195034	True
...
32172	AZ	Maricopa County	Write-ins	WRI	1331	False
32173	AZ	Mohave County	Donald Trump	REP	78535	True
32174	AZ	Mohave County	Joe Biden	DEM	24831	False
32175	AZ	Mohave County	Jo Jorgensen	LIB	1302	False
32176	AZ	Mohave County	Write-ins	WRI	37	False

[32177 rows x 6 columns]

```
[42]: # Display state to confirm the entries have been abbreviated appropriately in
      ↪ the presidential dataframe
uniqueStatesPre = presidential['state'].unique()
uniqueStatesPre
```

```
[42]: array(['DE', 'DC', 'FL', 'GA', 'HI', 'ID', 'IL', 'IN', 'IA', 'KS', 'KY',
        'LA', 'ME', 'MD', 'MA', 'MI', 'MN', 'MS', 'MO', 'MT', 'NE', 'NV',
        'NH', 'NJ', 'NM', 'NY', 'NC', 'ND', 'OH', 'OK', 'OR', 'PA', 'RI',
        'SC', 'SD', 'TN', 'TX', 'UT', 'VT', 'VA', 'WA', 'WV', 'WI', 'WY',
        'AL', 'AK', 'AR', 'CA', 'CO', 'CT', 'AZ'], dtype=object)
```

- Merge the covid_df with the presidential election results dataframe

```
[43]: # the 'state' column in the presidential DataFrame is renamed 'State' for
      ↪ sameness with its equivalent in the covid dataframe
presidential = presidential.rename(columns={'state': 'State'})
presidential
```

```
[43]:
```

	State	county	candidate	party	total_votes	won
0	DE	Kent County	Joe Biden	DEM	44552	True
1	DE	Kent County	Donald Trump	REP	41009	False
2	DE	Kent County	Jo Jorgensen	LIB	1044	False
3	DE	Kent County	Howie Hawkins	GRN	420	False
4	DE	New Castle County	Joe Biden	DEM	195034	True
...
32172	AZ	Maricopa County	Write-ins	WRI	1331	False
32173	AZ	Mohave County	Donald Trump	REP	78535	True
32174	AZ	Mohave County	Joe Biden	DEM	24831	False
32175	AZ	Mohave County	Jo Jorgensen	LIB	1302	False
32176	AZ	Mohave County	Write-ins	WRI	37	False

[32177 rows x 6 columns]

Generate the super covid dataframe: Merging the covid_df with the presidential dataframe

```
[44]: # Generate the super covid dataframe
superCovid = pd.merge(covid_df, presidential, left_on=("county", "State"),
      ↪ right_on = ("county", "State"), how = "inner")
superCovid
```

```
[44]:
```

	countyFIPS	county	State	StateFIPScases	2020-01-22cases	\
0	1001	Autauga County	AL	1	0	
1	1001	Autauga County	AL	1	0	
2	1001	Autauga County	AL	1	0	

3	1001	Autauga County	AL	1	0
4	1003	Baldwin County	AL	1	0
...
19339	56045	Weston County	WY	56	0
19340	56045	Weston County	WY	56	0
19341	56045	Weston County	WY	56	0
19342	56045	Weston County	WY	56	0
19343	56045	Weston County	WY	56	0

	2020-01-23cases	2020-01-24cases	2020-01-25cases	2020-01-26cases	\
0	0	0	0	0	
1	0	0	0	0	
2	0	0	0	0	
3	0	0	0	0	
4	0	0	0	0	
...	
19339	0	0	0	0	
19340	0	0	0	0	
19341	0	0	0	0	
19342	0	0	0	0	
19343	0	0	0	0	

	2020-01-27cases	...	2023-07-20deaths	2023-07-21deaths	\
0	0	...	235	235	
1	0	...	235	235	
2	0	...	235	235	
3	0	...	235	235	
4	0	...	731	731	
...	
19339	0	...	24	24	
19340	0	...	24	24	
19341	0	...	24	24	
19342	0	...	24	24	
19343	0	...	24	24	

	2023-07-22deaths	2023-07-23deaths	County Name	population	\
0	235	235	Autauga County	55869	
1	235	235	Autauga County	55869	
2	235	235	Autauga County	55869	
3	235	235	Autauga County	55869	
4	731	731	Baldwin County	223234	
...	
19339	24	24	Weston County	6927	
19340	24	24	Weston County	6927	
19341	24	24	Weston County	6927	
19342	24	24	Weston County	6927	
19343	24	24	Weston County	6927	

	candidate	party	total_votes	won
0	Donald Trump	REP	19838	True
1	Joe Biden	DEM	7503	False
2	Jo Jorgensen	LIB	350	False
3	Write-ins	WRI	79	False
4	Donald Trump	REP	83544	True
...
19339	Donald Trump	REP	3107	True
19340	Joe Biden	DEM	360	False
19341	Jo Jorgensen	LIB	46	False
19342	Brock Pierce	IND	20	False
19343	Write-ins	WRI	9	False

[19344 rows x 2542 columns]

- Save the supercovid dataframe to a file

```
[45]: superCovid.to_csv('supercovid.csv')
```