

## Stage2\_Task1\_Cases

March 14, 2023

### 0.1 Generate weekly statistics (mean, median, mode) for number of new cases across a specific state.

```
[1]: import pandas as pd
import numpy as np
import statistics
import matplotlib.pyplot as plt
```

```
[2]: #I have selected the Alabama 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]

```
[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

```
[4]: #displaying the data of the selected state.
```

```
cases_selected_state = cases_transpose[cases_transpose["State"] ==
      ↪selected_state]
cases_selected_state.head()
```

```
[4]:
```

	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

```
[5]: #For the selected state Alabama summing the cases per day of all the counties.
```

```
cases_selected_state_daily = cases_selected_state.groupby('Date')['Number of
      ↪Cases'].sum()
cases_selected_state_daily.head()
```

```
[5]: Date
```

2020-01-22	0
2020-01-23	0
2020-01-24	0
2020-01-25	0
2020-01-26	0

Name: Number of Cases, dtype: int64

```
[6]: #Finding out the new cases per day.
```

```
new_cases_selected_state_daily = cases_selected_state_daily.diff().reset_index()
```

```
new_cases_selected_state_daily.head()
```

```
[6]:
```

	Date	Number of Cases
0	2020-01-22	NaN
1	2020-01-23	0.0
2	2020-01-24	0.0
3	2020-01-25	0.0
4	2020-01-26	0.0

```
[7]: #Converting the daily to weekly analysis and finding the mean weekly.
weekly_cases_mean_selected_state = new_cases_selected_state_daily.copy()
weekly_cases_mean_selected_state['Date'] = pd.
    ↳to_datetime(weekly_cases_mean_selected_state['Date']) - pd.to_timedelta(7,
    ↳unit='d')
weekly_cases_mean_selected_state = weekly_cases_mean_selected_state.groupby([pd.
    ↳Grouper(key='Date', freq='W-SUN')])['Number of Cases'].mean()
weekly_cases_mean_selected_state = weekly_cases_mean_selected_state.
    ↳reset_index()
weekly_cases_mean_selected_state.head()
```

```
[7]:
```

	Date	Number of Cases
0	2020-01-19	0.0
1	2020-01-26	0.0
2	2020-02-02	0.0
3	2020-02-09	0.0
4	2020-02-16	0.0

```
[8]: #considering the given range of dates starting from monday. and weekly anlasis
    ↳from monday to sunday.
weekly_cases_mean_selected_state_given_range =
    ↳weekly_cases_mean_selected_state[(weekly_cases_mean_selected_state["Date"]
    ↳>= '2022-05-29') & (weekly_cases_mean_selected_state["Date"] <=
    ↳'2023-01-02')]
weekly_cases_mean_selected_state_given_range =
    ↳weekly_cases_mean_selected_state_given_range.sort_values(by=['Date']).
    ↳reset_index(drop=True)
weekly_cases_mean_selected_state_given_range['Date'] =
    ↳weekly_cases_mean_selected_state_given_range['Date'] + pd.to_timedelta(1,
    ↳unit='d')
weekly_cases_mean_selected_state_given_range
```

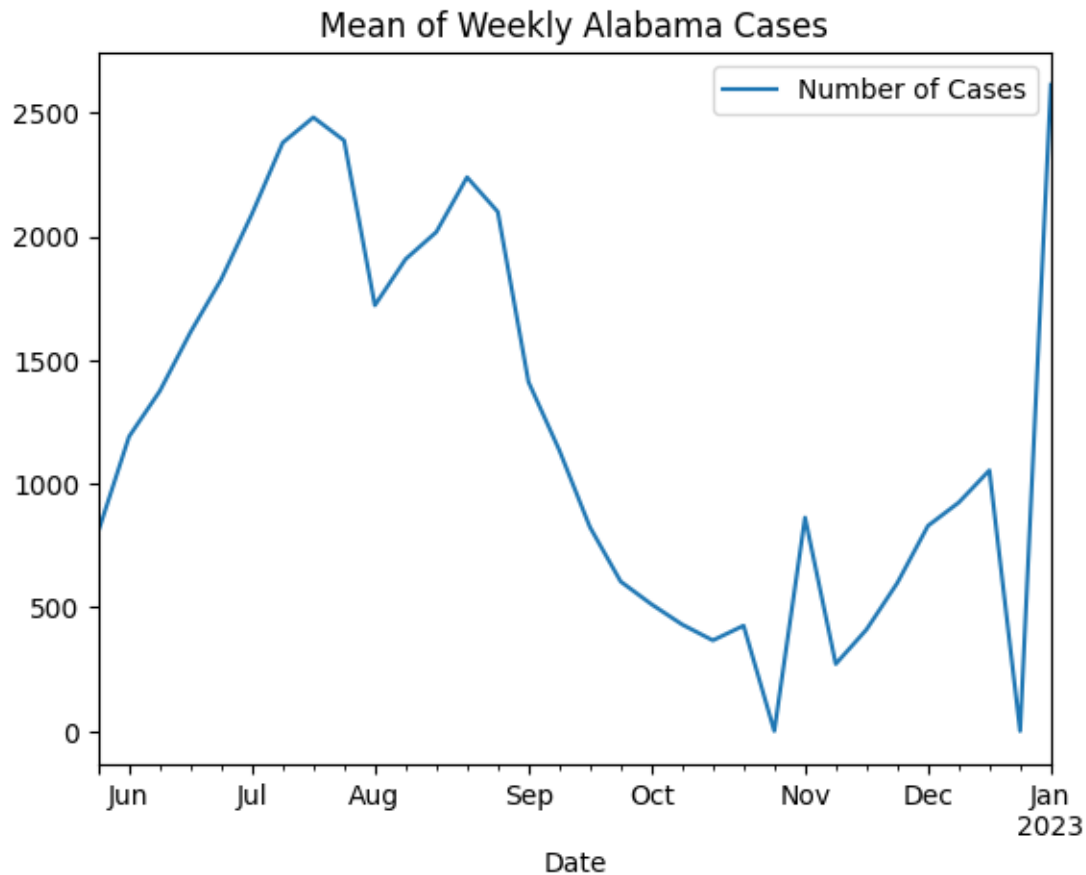
```
[8]:
```

	Date	Number of Cases
0	2022-05-30	806.857143
1	2022-06-06	1190.285714
2	2022-06-13	1374.285714
3	2022-06-20	1612.428571

4	2022-06-27	1826.142857
5	2022-07-04	2090.428571
6	2022-07-11	2378.428571
7	2022-07-18	2480.857143
8	2022-07-25	2387.428571
9	2022-08-01	1721.000000
10	2022-08-08	1907.714286
11	2022-08-15	2017.285714
12	2022-08-22	2238.857143
13	2022-08-29	2099.285714
14	2022-09-05	1411.428571
15	2022-09-12	1136.285714
16	2022-09-19	824.285714
17	2022-09-26	604.428571
18	2022-10-03	512.714286
19	2022-10-10	430.714286
20	2022-10-17	366.571429
21	2022-10-24	426.000000
22	2022-10-31	0.000000
23	2022-11-07	863.142857
24	2022-11-14	271.142857
25	2022-11-21	410.285714
26	2022-11-28	598.000000
27	2022-12-05	829.571429
28	2022-12-12	923.571429
29	2022-12-19	1053.857143
30	2022-12-26	0.000000
31	2023-01-02	2612.857143

```
[9]: #Plotting the mean graph
weekly_cases_mean_selected_state_given_range.plot(x='Date', y='Number of
→Cases', title='Mean of Weekly Alabama Cases')
```

```
[9]: <AxesSubplot: title={'center': 'Mean of Weekly Alabama Cases'}, xlabel='Date'>
```



```
[10]: #Converting the daily to weekly analysis and finding the median weekly.
weekly_cases_median_selected_state = new_cases_selected_state_daily.copy()
weekly_cases_median_selected_state['Date'] = pd.
    ↳to_datetime(weekly_cases_median_selected_state['Date']) - pd.to_timedelta(7,
    ↳unit='d')
weekly_cases_median_selected_state = weekly_cases_median_selected_state.
    ↳groupby([pd.Grouper(key='Date', freq='W-SUN')])['Number of Cases'].median()
weekly_cases_median_selected_state = weekly_cases_median_selected_state.
    ↳reset_index()
weekly_cases_median_selected_state.head()
```

```
[10]:      Date  Number of Cases
0 2020-01-19          0.0
1 2020-01-26          0.0
2 2020-02-02          0.0
3 2020-02-09          0.0
4 2020-02-16          0.0
```

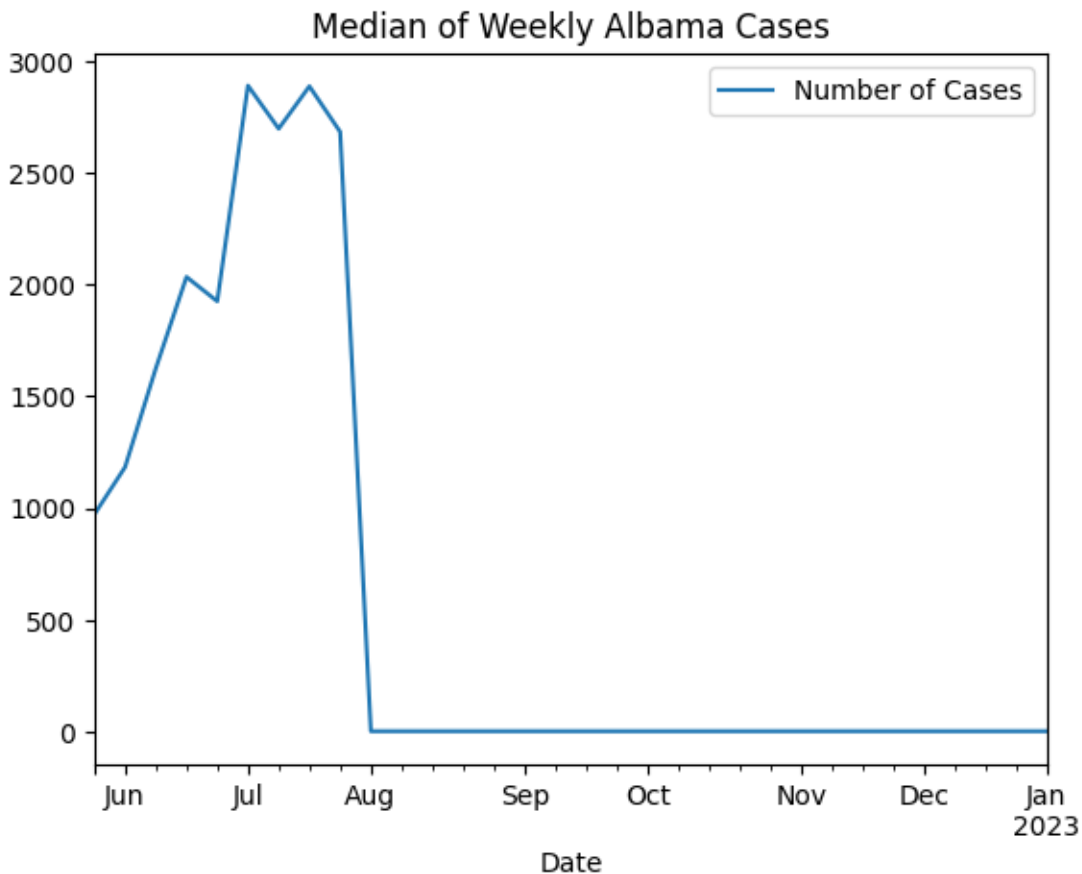
```
[11]: #considering the given range of dates starting from monday. and weekly analysis
      ↪from monday to sunday.
weekly_cases_median_selected_state_given_range =
      ↪weekly_cases_median_selected_state[(weekly_cases_median_selected_state["Date"]
      ↪>= '2022-05-29') & (weekly_cases_median_selected_state["Date"] <=
      ↪'2023-01-02')]
weekly_cases_median_selected_state_given_range =
      ↪weekly_cases_median_selected_state_given_range.sort_values(by=['Date']).
      ↪reset_index(drop=True)
weekly_cases_median_selected_state_given_range['Date'] =
      ↪weekly_cases_median_selected_state_given_range['Date'] + pd.to_timedelta(1,
      ↪unit='d')
weekly_cases_median_selected_state_given_range
```

```
[11]:
```

	Date	Number of Cases
0	2022-05-30	971.0
1	2022-06-06	1183.0
2	2022-06-13	1625.0
3	2022-06-20	2035.0
4	2022-06-27	1926.0
5	2022-07-04	2892.0
6	2022-07-11	2699.0
7	2022-07-18	2889.0
8	2022-07-25	2685.0
9	2022-08-01	0.0
10	2022-08-08	0.0
11	2022-08-15	0.0
12	2022-08-22	0.0
13	2022-08-29	0.0
14	2022-09-05	0.0
15	2022-09-12	0.0
16	2022-09-19	0.0
17	2022-09-26	0.0
18	2022-10-03	0.0
19	2022-10-10	0.0
20	2022-10-17	0.0
21	2022-10-24	0.0
22	2022-10-31	0.0
23	2022-11-07	0.0
24	2022-11-14	0.0
25	2022-11-21	0.0
26	2022-11-28	0.0
27	2022-12-05	0.0
28	2022-12-12	0.0
29	2022-12-19	0.0
30	2022-12-26	0.0
31	2023-01-02	0.0

```
[12]: #Plotting the median graph
weekly_cases_median_selected_state_given_range.plot(x='Date', y='Number of Cases', title = 'Median of Weekly Alabama Cases')
```

```
[12]: <AxesSubplot: title={'center': 'Median of Weekly Alabama Cases'}, xlabel='Date'>
```



```
[13]: #Converting the daily to weekly analysis and finding the mode weekly.
weekly_cases_mode_selected_state = new_cases_selected_state_daily.copy()
weekly_cases_mode_selected_state['Date'] = pd.
    ↳to_datetime(weekly_cases_mode_selected_state['Date']) - pd.to_timedelta(7,
    ↳unit='d')
weekly_cases_mode_selected_state = weekly_cases_mode_selected_state.groupby([pd.
    ↳Grouper(key='Date', freq='W-SUN')])['Number of Cases'].apply(statistics.mode)
weekly_cases_mode_selected_state = weekly_cases_mode_selected_state.
    ↳reset_index()
weekly_cases_mode_selected_state.head()
```

```
[13]:      Date  Number of Cases
0 2020-01-19              0.0
```

1	2020-01-26	0.0
2	2020-02-02	0.0
3	2020-02-09	0.0
4	2020-02-16	0.0

```
[14]: #considering the given range of dates starting from monday. and weekly analysi
↳from monday to sunday.
weekly_cases_mode_selected_state_given_range =
↳weekly_cases_mode_selected_state[(weekly_cases_mode_selected_state["Date"]
↳>= '2022-05-29') & (weekly_cases_mode_selected_state["Date"] <=
↳'2023-01-02')]
weekly_cases_mode_selected_state_given_range =
↳weekly_cases_mode_selected_state_given_range.sort_values(by=['Date']).
↳reset_index(drop=True)
weekly_cases_mode_selected_state_given_range['Date'] =
↳weekly_cases_mode_selected_state_given_range['Date'] + pd.to_timedelta(1,
↳unit='d')
weekly_cases_mode_selected_state_given_range
```

```
[14]:
```

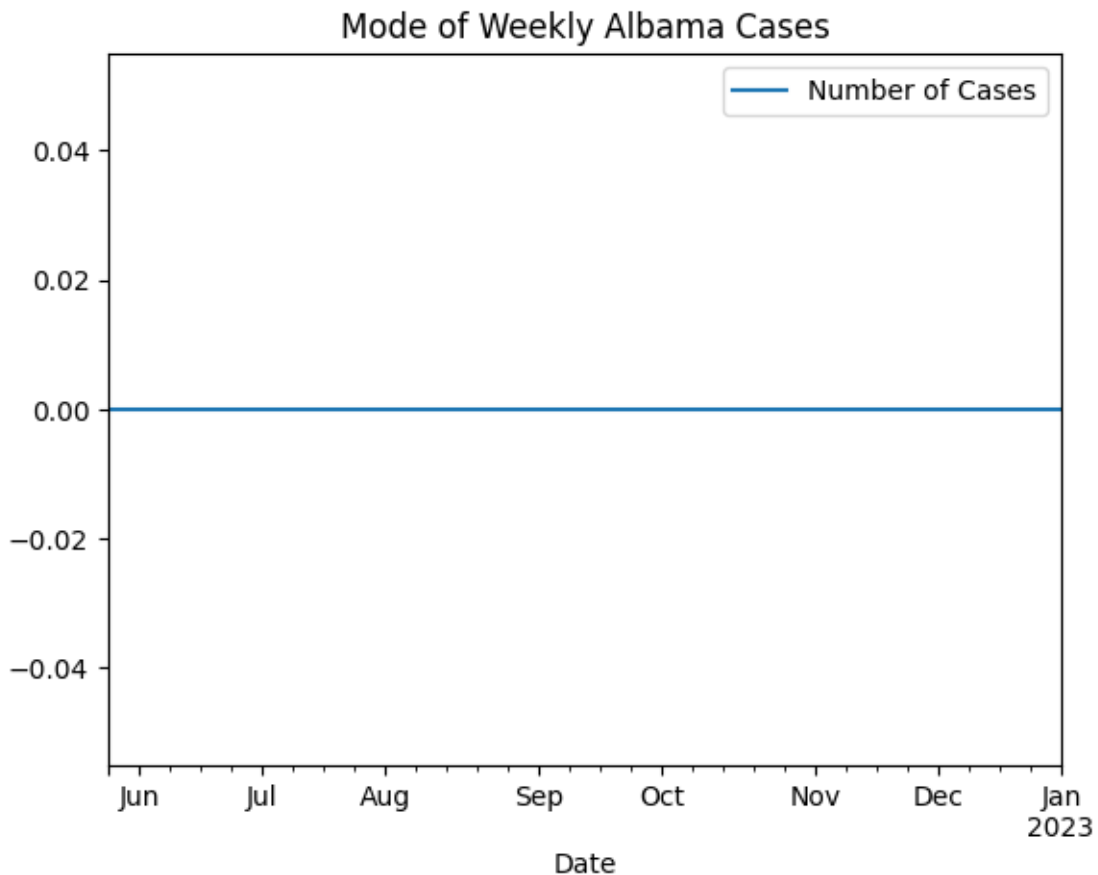
	Date	Number of Cases
0	2022-05-30	0.0
1	2022-06-06	0.0
2	2022-06-13	0.0
3	2022-06-20	0.0
4	2022-06-27	0.0
5	2022-07-04	0.0
6	2022-07-11	0.0
7	2022-07-18	0.0
8	2022-07-25	0.0
9	2022-08-01	0.0
10	2022-08-08	0.0
11	2022-08-15	0.0
12	2022-08-22	0.0
13	2022-08-29	0.0
14	2022-09-05	0.0
15	2022-09-12	0.0
16	2022-09-19	0.0
17	2022-09-26	0.0
18	2022-10-03	0.0
19	2022-10-10	0.0
20	2022-10-17	0.0
21	2022-10-24	0.0
22	2022-10-31	0.0
23	2022-11-07	0.0
24	2022-11-14	0.0
25	2022-11-21	0.0
26	2022-11-28	0.0



27	2022-12-05	0.0
28	2022-12-12	0.0
29	2022-12-19	0.0
30	2022-12-26	0.0
31	2023-01-02	0.0

```
[15]: #Plotting the mode graph
weekly_cases_mode_selected_state_given_range.plot(x='Date', y='Number of
↪Cases', title = 'Mode of Weekly Albama Cases')
```

```
[15]: <AxesSubplot: title={'center': 'Mode of Weekly Albama Cases'}, xlabel='Date'>
```



```
[16]: #plotting the mean,median and mode graphs.
plt.plot(weekly_cases_mean_selected_state_given_range['Date'],
↪weekly_cases_mean_selected_state_given_range['Number of Cases'], label='Mean
↪of Alabama State')
plt.plot(weekly_cases_median_selected_state_given_range['Date'],
↪weekly_cases_median_selected_state_given_range['Number of Cases'],
↪label='Median of Alabama State')
```

```

plt.plot(weekly_cases_mode_selected_state_given_range['Date'],
↪weekly_cases_mode_selected_state_given_range['Number of Cases'], label='Mode_
↪of Alabama State')
plt.title('Mean,Median and Mode of Cases in Albama State')
plt.xlabel('Date')
plt.ylabel('Number of Cases')
plt.legend()
plt.show()

```

