Stage 02 - Member Report

Aman Tej Vidapu, Graduate student

Generate weekly statistics:
I took New York state to do analysis. Mean and median values of mean weekly cases across new york 3835, 1756. Mean and median values of mean weekly deaths across new york 93, 33. (19 million population) - 0.0002. Mode is '0'. I also found mean, median and mode values for the total sum of cases/deaths across New york.

Compare the data against other states and plot trends:
As my task 1 state is New york, I considered taking states which are closely populated as new york or more populated than new york. They are California, Florida, Illinois, Ohio and Texas. So, in order to compare we have to normalise the cases/deaths according to population per each state. Here the selected states population ranges between 12-39 million. So, we are normalising cases/deaths per every 100000 people in the state. Also, we are performing analysis on weekly generated mean cases/deaths across the selected states.

After grouping the data by weeks and normalising. We are calculating mean and median values for all the selected states cases/deaths and making comparisons.
cases:

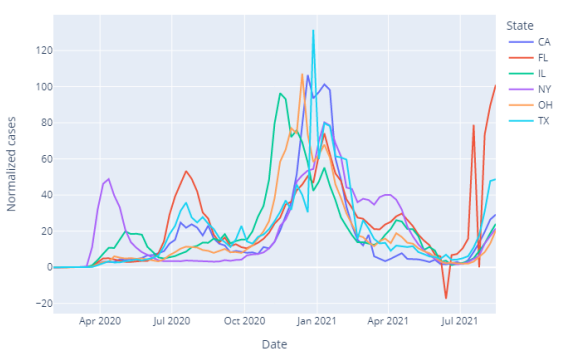
	State	mean	median	Population
0	CA	18.0	8.0	39512223
1	FL	24.0	18.0	21477737
2	IL	20.0	14.0	12671821
3	NY	20.0	9.0	19453561
4	OH	17.0	9.0	11689100
5	TX	20.0	14.0	28995881

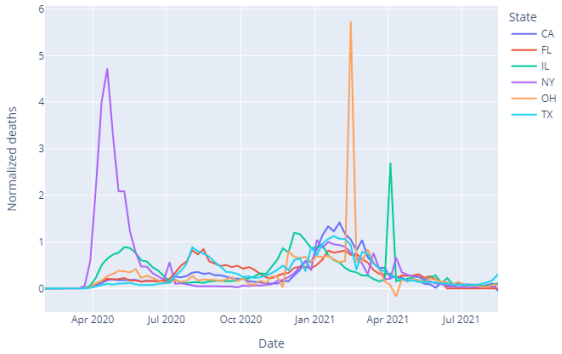
Deaths:

	State	mean	median	Population	rounded_mean	rounded_median
0	CA	0.285385	0.173545	39512223	0.0	0.0
1	FL	0.303608	0.265391	21477737	0.0	0.0
2	IL	0.363149	0.217581	12671821	0.0	0.0
3	NY	0.483782	0.182119	19453561	0.0	0.0
4	OH	0.310952	0.175988	11689100	0.0	0.0
5	TX	0.322809	0.224662	28995881	0.0	0.0

As the data is already normalised, we can not compare correctly on deaths. But, as of cases we can see states with high populations even have moderate mean, median values. Whereas some states with less population have more cases. We can make hypotheses like less medicare, educated people.., other factors for these differences.

Here are the visual plots to understand peaks and dip trends of cases/deaths for selected states.





Ohio has more death rates and Texas has the most number of confirmed cases. We can also see 2-3 peaks for every state, most common around december-january timeline which is a holiday period. New York has one surge with respect to deaths but normalised later. Illinois and Ohio had a surge for deaths during summer and the rest of the states maintained simple trends. At initial hit with covid in New York, there is no medical knowledge about covid. So, the death rate is high. Again with seasonal changes we can observe the peaks in both trends. One more observation that can be made is after a new_cases surge during dec-jan 2020(holiday season) also influenced the death rate by feb-march 2021.

We can observe the case trends similar to the complete United states trend which is present in Team task (satge_2). Simply, these states push trends across the United states.
Death's trend has the first peak match but, gradually trends do not match as exactly because we have different normalisation factors.

Identify five counties within a state of your choice with high cases and death rates:
Top 5 Infected counties of New York in terms of total cases count in that county:

- Bronx County
- Nassau County
- Suffolk County
- Queens County
- Kings County

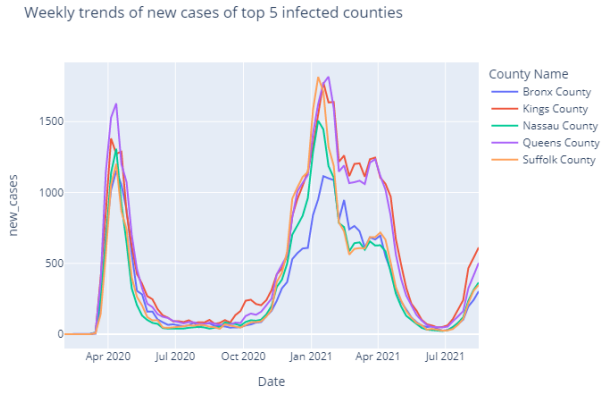
Top 5 Infected counties of New York in terms of total cases count in that county:

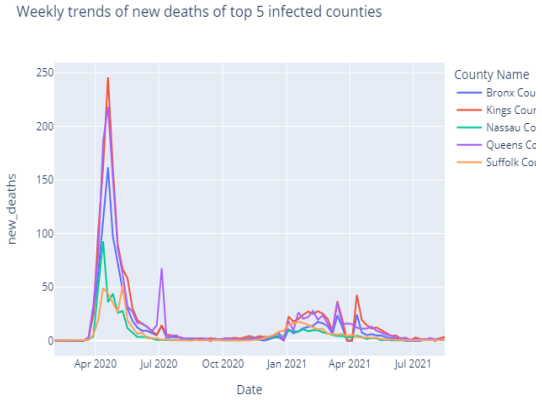
- Bronx County
- New York County
- Suffolk County
- Queens County
- Kings County

Only New York county seems to be having more deaths than the confirmed cases rate. This is an interesting observation that can be more analysed with proper enrichment data.

Trends:
Here are the plots of new_cases, new_deaths per week in these top 5 infected counties. We can understand how one county can influence other counties, as all of them fall into one

state. We can also understand the populations and trends of each county. We are normalising the cases or deaths per 10000 people for conty.





We can see three peaks, the first wave is where covid virus is pretty new and proper knowledge of how to treat it. So, the death rate is high. Later two peaks are in dec-jan 2020, which is an effect of holidays and same lead to hike in death rate. Due to variant in covid, we can see new peak after april 2021.

These county trends match overall New York cases and deaths trends. We can say that these counties are like driving forces for that actual state trend.