

## Data Cleaning Process:

The dataset itself is pretty complete and incisive. However, I noticed that there are some missing data under following heads:

new_case_rate	new_death_rate	new_case_count	new_death_count
---------------	----------------	----------------	-----------------

Therefore, I dropped these missing data so the dataset becomes more complete which is easier for deeper analysis.

## Data Modeling Method:

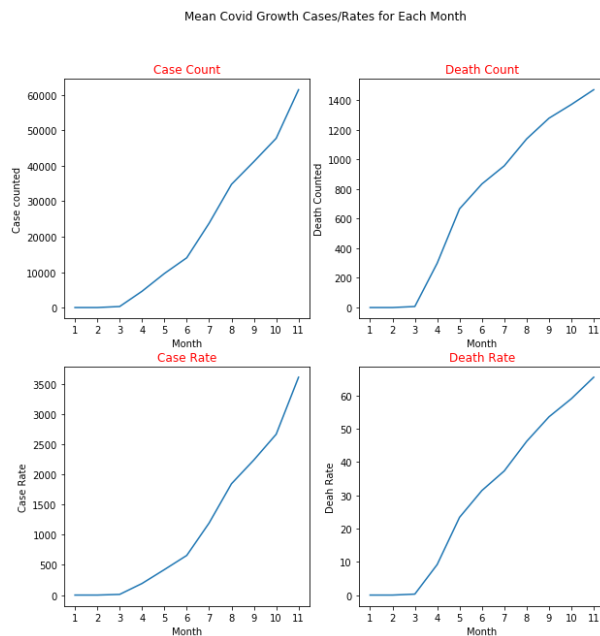
Groupby:

I grouped the dataset by month, and get the mean of following heads for each month from January to November:

case_count	death_count	case_rate	death_rate
------------	-------------	-----------	------------

## Plotting:

Nevertheless, the result is hard to read since it's all numbers, so I used a plotting method to visualize it.



It seems like all four categories are very similar to each other from January to May, with a flat line from January to March, and a steeper line from March to May. But as time goes by, the case count and case rates are still increasing at an increasing rate, but the death count and death rate are starting to decrease at an increasing rate.

## Next Step:

Do mean of new case counted/rate and new death counted/rate

Also need to do city, state, and national level???

**Associate these data with employment and mobility level to better examine how covid impact economics.**