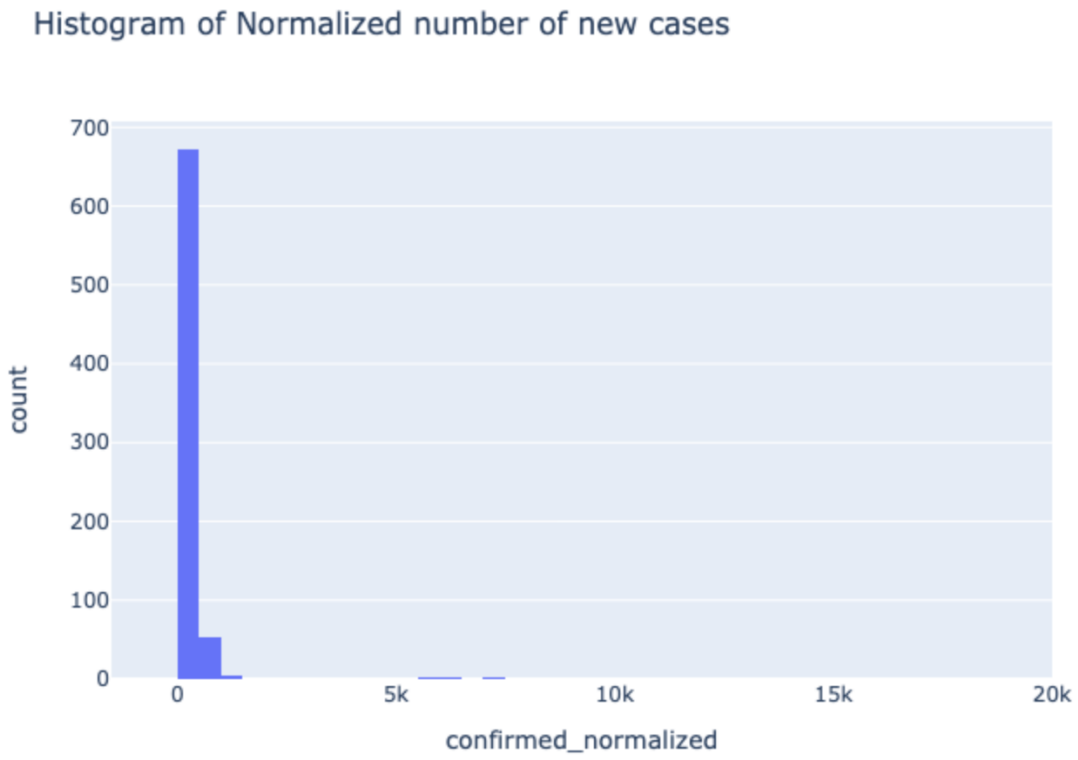
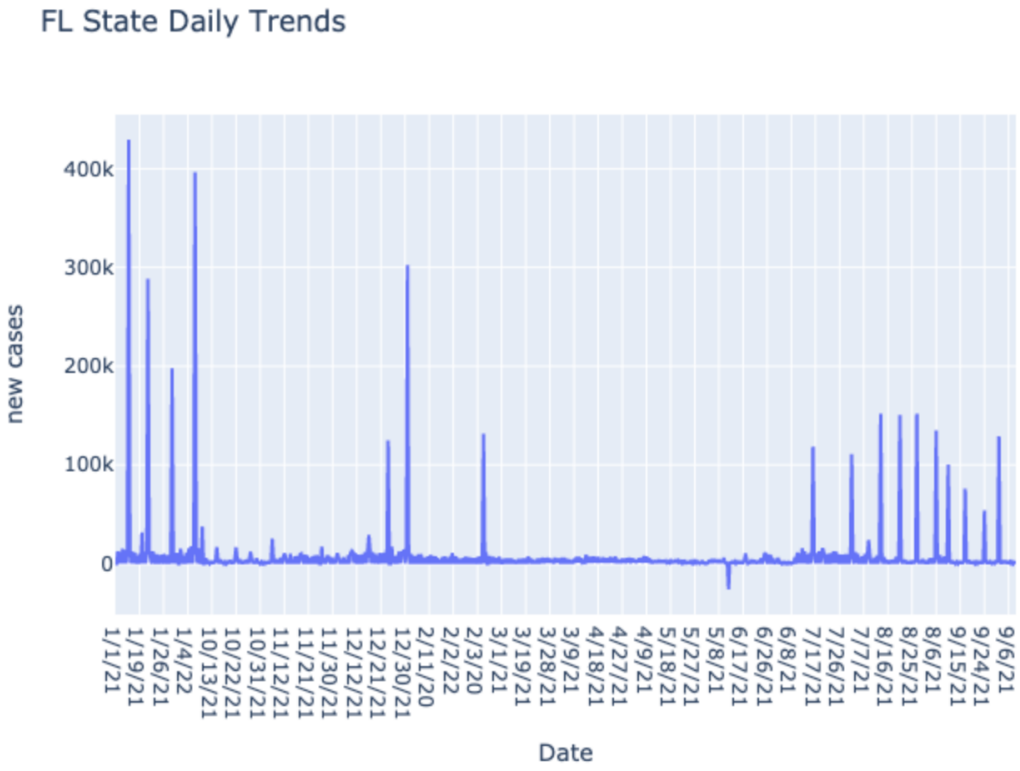
STAGE -3

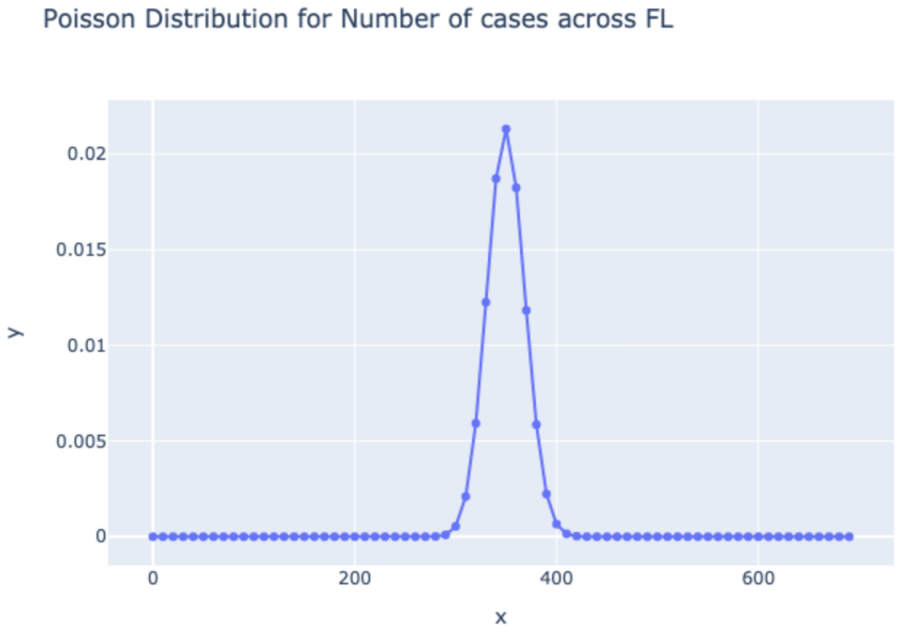
Task 1. I am using Florida as my initial state to compare with other states . I am calculating daily new cases and daily death cases normalized by 1000000. I am using the discrete values then the probability mass function is to be calculated. I am displaying the results in histogram in which we can see the data count is high at the start and later on it decreases.



The distribution I am choosing is Poisson distribution because the data we can see in the graph has positive values and data is discrete. We can observe a covid cases in a certain period and time interval of the day. We know that the Poisson distribution models the probability of certain number of successes within a time interval, So I am using Poisson distribution.



We can observe from the above graph the daily trend of confirmed new cases of state Florida .



Task 2, we compare the distribution and its statistics to 5 other states as Georgia, California, Kentucky, Michigan and Illinois. The statistics are as follows.
I can infer from the diagram that it is using unimodel distribution.

FLORIDA STATE

- MEAN = 349.700
- VARIANCE = 2094522.28
- SKEWNESS = 9.332
- KURTOSIS = 101.541

GEORGIA STATE

- MEAN = 227.73493975903614
- VARIANCE = 199073.803
- SKEWNESS = 8.012
- KURTOSIS = 97.588

CALIFORNIA STATE

- MEAN = 275.445
- VARIANCE = 419471.536
- SKEWNESS =7.487
- KURTOSIS = 73.866

KENTUCKY STATE

- MEAN = 359.348
- VARIANCE = 472831.980
- SKEWNESS = 5.469
- KURTOSIS = 45.836

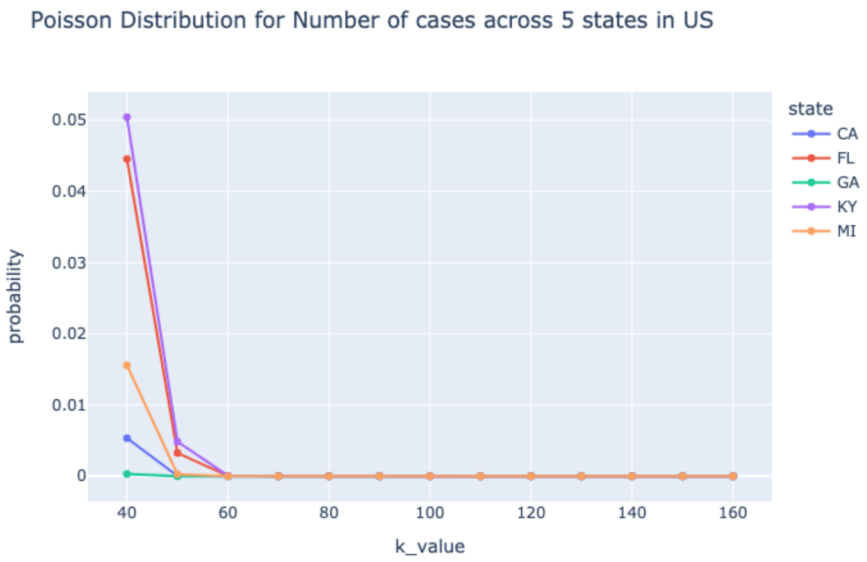
MICHIGAN STATE

- MEAN = 303.382
- VARIANCE = 518015.250
- SKEWNESS = 6.128
- KURTOSIS = 55.370

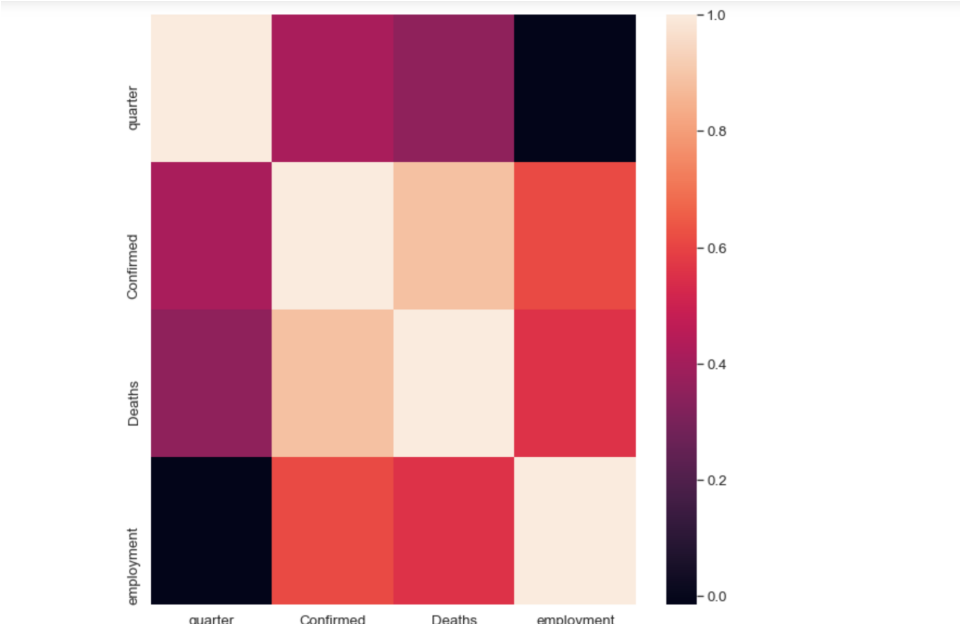
ILLINOIS STATE

- MEAN = 313.879
- VARIANCE = 215980.931
- SKEWNESS = 3.449
- KURTOSIS = 15.524

Task 3.



- The distribution is positive valued in all the states CA,FL,GA,KY,MI
- We can see the state KY is having positive values and GA has the lowest probablity of covid cases.
- **Task 4: Perform corelation between Enrichment data valiables and COVID-19 cases to observe any patterns.**



Based on above correlation of Covid new/death cases and employment data for year 2020 across all 4 quarters, we can form below hypothesis:
We can infer from above heat map as follows:

- The employment rate decreased from the 2020 1st quarter to 4th quarter as the number of COVID cases increased in Florida county.
- The correlation is very evident that the number of confirmed and deaths cases impacted the employment numbers across Florida state.
- The enrichment data has quarterly data of employment rate changes. We can clearly see the negative rate of employment in the 4th quarter of year 2020.
- The first quarter of 2020 has shown a good employment rate as COVID only started impacting from the end of Quarter 1.