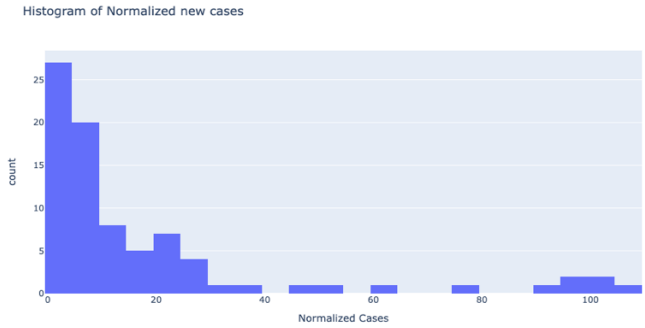
MEMBER TASK

Lahari Chilakuri, Graduate Student

Fit a distribution to the number of COVID-19 new cases:

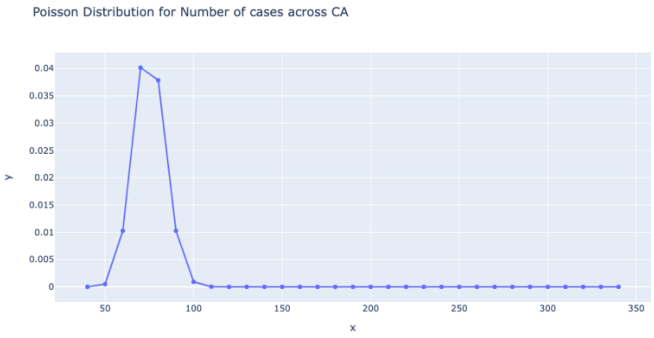
In this task, I have used the data frame which was generated in stage-II. The data contains the information about number of cases and deaths which were normalized by the population.
Here, I have found the mean value for the normalized cases and later calculated the Probability Mass Function for the individual values and plotted a histogram for normalized new cases.

Type of distribution:



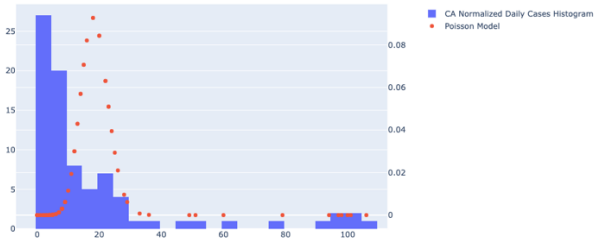
From the histogram, I thought that the Poisson Distribution would better fit because here we are measuring for the number of new cases per week.

Poisson's Distribution formula:
For a random discrete variable X that follows the Poisson distribution, and λ is the average rate of value, then the probability of x is given by:
$f(x) = P(X=x) = (e^{-\lambda} \lambda x)/x!$
Where,
x = 0, 1, 2, 3...
e is the Euler's number (e = 2.718)
λ is an average rate of the expected value and λ = variance, also λ>0



Process of modelling Poisson Distribution:

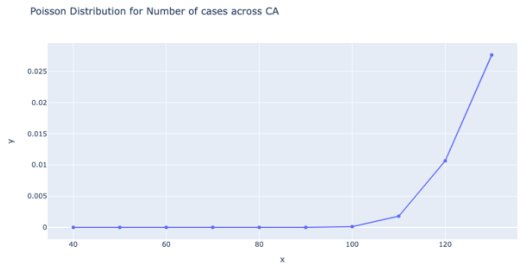
First, I have calculated the mean of normalized number of new cases with 100000 population. The mean value is 18.59 which is according to California state data. To find the probability mass function, pmf.poisson function have been used for values of k to find the probability of normalized number of cases and deaths. The below plot shows the poisson model over the histogram of normalized number of cases.



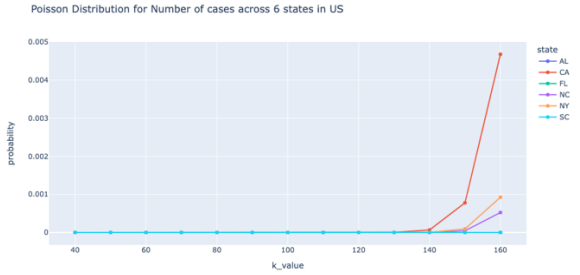
Statistics:

The statistics for the data have been calculated and the results are as shown below:
Measure of center: 18.59036144578313
Measure of spread: 706.7081986482515
Skewness: 2.250870942147231
Kurtosis: 4.222739586165505

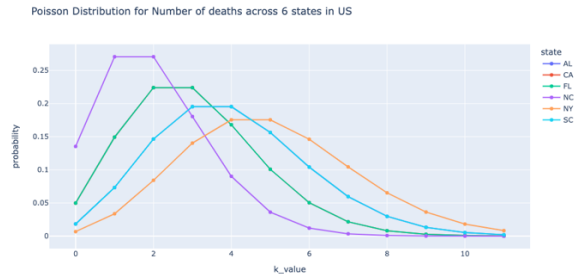
The probability mass function has been calculated at different points from min to max range to observe the probability at different points.
The Poisson distribution of California state is as below



I have chosen North Carolina, South Carolina, Alabama, New York and Florida as five other states. The Poison Distribution for these states for normalized number of cases is shown as below:



The Poison Distribution for these states for normalized number of deaths is shown as below:



Correlation between Enrichment data variables and COVID-19 cases:

The enrichment data which I have used is Hospital beds dataset. In this task, I have made the correlation between the large covid dataset and the enrichment dataset. The normalized number of new cases have been correlated with three different columns of the enrichment dataset. All the three results in the negative correlation.

Formulate hypothesis between Enrichment data and number of cases:

- Counties with small population may have limited resources to the hospitals which leads to maximum utilization to the hospital beds.
- The counties which have the large number of hospitals eventually results the fewer confirmed covid cases.
- The number of confirmed covid cases are fewer in the areas which contains a greater number of hospitals when compared to the areas which contains a smaller number of hospitals

- As the hospitals are categorized as adult and pediatric, we can analyze the ratio of adults and the children who are confirmed to covid.