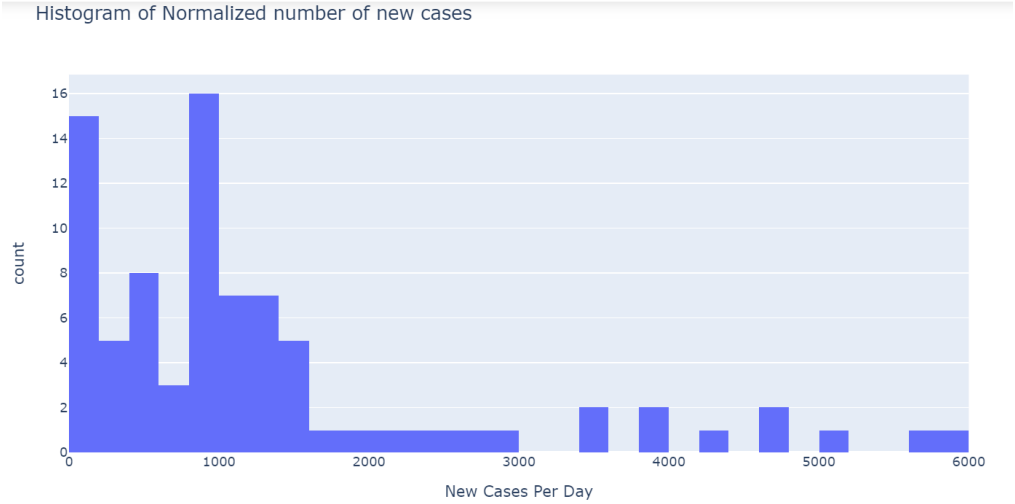
Varsha Veeramaneni, Graduate Student

Use the state data to fit a distribution to the number of COVID-19 new cases:

Type of distribution:

For this problem, I used a Poisson distribution. This is because we can predict the probability if we know the number of times the event occurs times, so for the data, I assume that the Poisson distribution is appropriate and modify the data to get the passion distribution. Here I chose Virginia state to filter the data.

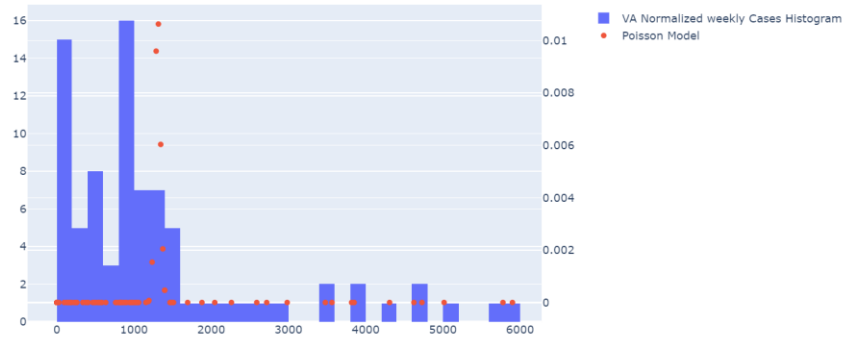
I used 50 bins to create a histogram of our daily cases and calculated the pmf values.



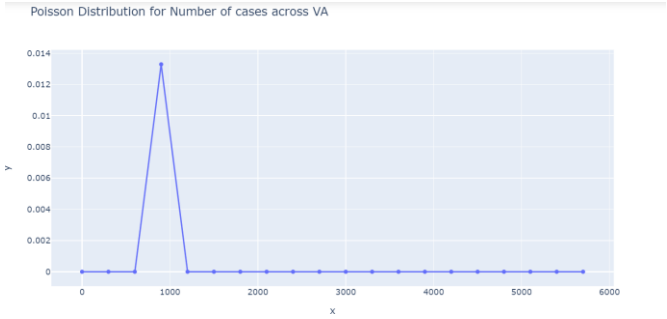
Calculating PMF values:

	Date	New Cases Per Day	New Deaths Per Day	VA PMF
0	2020-01-26	0.0	0.0	0.000000e+00
1	2020-02-02	0.0	0.0	0.000000e+00
2	2020-02-09	0.0	0.0	0.000000e+00
3	2020-02-16	0.0	0.0	0.000000e+00
4	2020-02-23	0.0	0.0	0.000000e+00
...	...	...	...	...
78	2021-07-25	543.0	3.0	4.795512e-127
79	2021-08-01	869.0	5.0	1.016314e-38
80	2021-08-08	1469.0	4.0	6.370421e-07
81	2021-08-15	1880.0	6.0	8.233170e-51
82	2021-08-22	5901.0	19.0	0.000000e+00

Calculate the pmf values using the poission distribution and plot:



Calculating Poisson distribution for Number of cases across VA:



Compare the data against other states:

For Virginia State:

Mean: 1306.578313253012

Variance: 1898394.7102556564

Skewness: 1.7478233451315404

Kurtosis: 2.5908119970838843

For all 6 states:

Mean: 21.225172246807766

Variance: 601.1644322282428

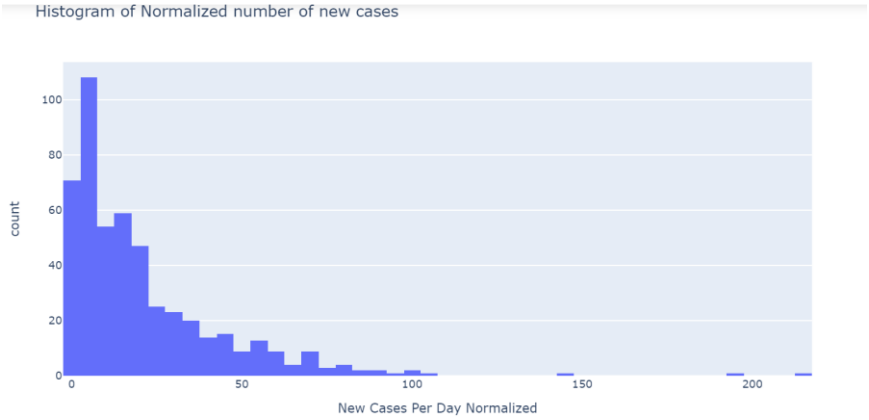
Skewness: 2.7678376099391735

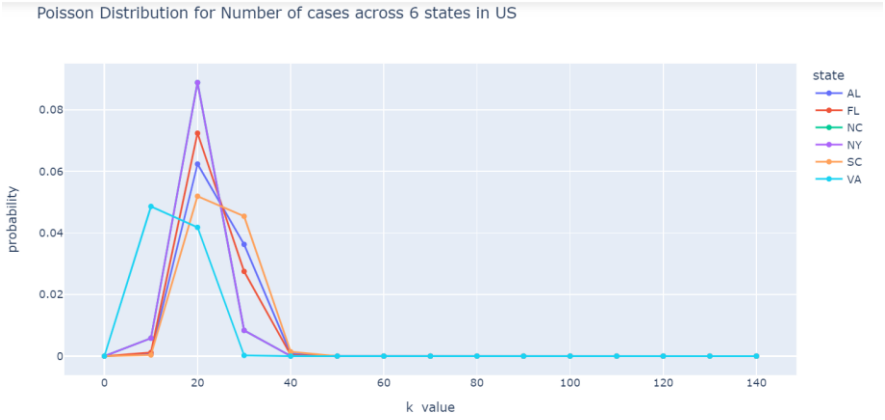
Kurtosis: 13.48950255933966

While comaring virgina state statistics with all other states, I observe that there is a huge difference in Mean,Variance and kurtosis.

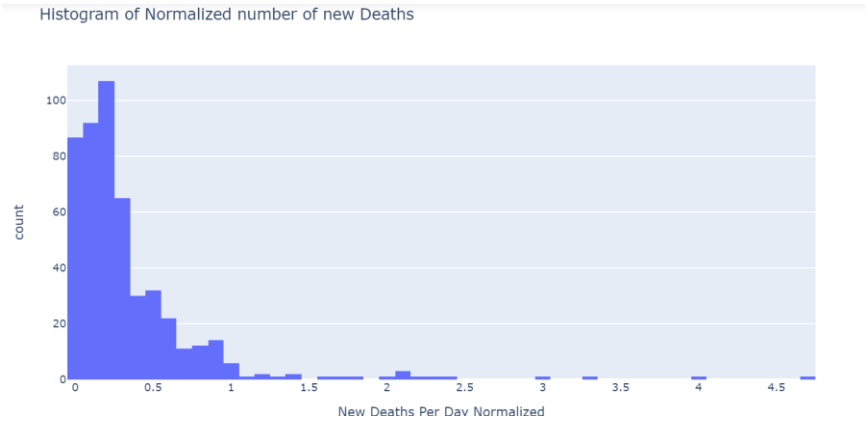
Model a poission distribution of new **COVID-19 cases and deaths:**

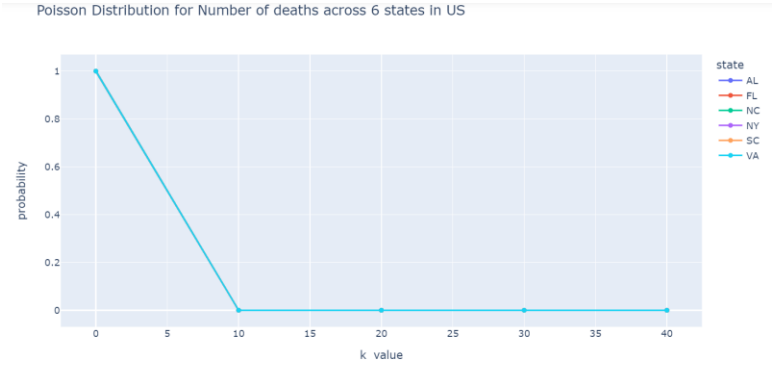
For cases:





For Deaths:





When comparing both Poisson distribution of cases and deaths for all the six states, in the poisson distribution for cases, New York state has highest probability when compared with all other states and followed by Florida. Where as in the poisson distribution for deaths, Virginia has highest probability when compared with all others states.

Perform corelation between Enrichment data valiables and COVID-19 cases :

Here I Have merged the enrichment data variables with the covid data set obtained after selecting six states in above task and calculated the correlation on variables form the enrichment data sets with cases and deaths.

```
merged_data['New Cases Per Day Normalized'].corr(merged_data['total_votes'])
```

```
-0.30062347646940485
```

```
merged_data['New Deaths Per Day Normalized'].corr(merged_data['total_votes'])
```

```
-0.9818956567965462
```

```
merged_data['New Cases Per Day Normalized'].corr(merged_data['won'])
```

```
-0.4821807526750586
```

```
merged_data['New Deaths Per Day Normalized'].corr(merged_data['won'])
```

```
-0.35297186491534205
```

From the above correlation, in all the six states, we observe that the correlation between total votes , i.e, the number of total votes in the sense that there was negative corralation with both cases and deaths. This shows that Voting doesn't increase the covid cases and deaths.

Formulate hypothesis between Enrichment data and number of cases to be compared against states:

As I don't have many enrichment variables for comparison, I compared only cases, deaths and total votes.

1. Are the voting percentages doesn't impact rise in Covid cases?
2. Are the voting percentages doesn't impact rise in Deaths ?
3. Is the rise in covid cases impact more on deaths