Project Stage - II (Data Modeling and Hypothesis Testing)

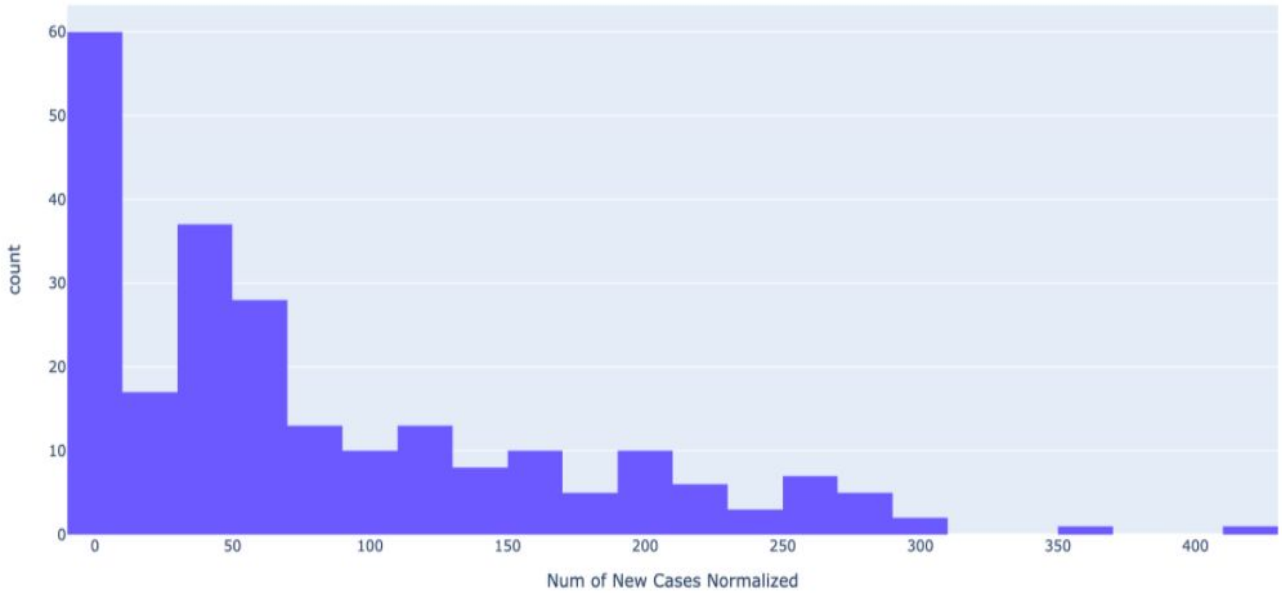
Amulya Yadagani

Distribution and statistics of New cases

Statistics of New Cases

1. The **Mean** of "Number of New Cases Normalized" is 82 and we can say that the average "Number of New Cases Normalized" in this Covid dataset for California State is 82.
2. The **Variance** of "Number of New Cases Normalized" is 7393.9 and this value tells us how large the spread of data("Number of New Cases Normalized") from mean for California state
3. The **Skewness** of "Number of New Cases Normalized" is 1.19, here the positive skew tells that the new cases is asymmetric and left-skewed with tail on the right.
4. The **Kurtosis** of "Number of New Cases Normalized" is 0.86, the positive Kurtosis indicate skewness

Plotting histogram on "Number of New Cases Normalized" variable has resulted in the following plot -



From the observations on histograms -

1. The variable “Num of New Cases” is discrete.

2. From the histogram, we can observe that the plot is left-skewed with a tail extending to the right.

3. The histogram is plotted based on the variable “Number of New Cases Normalized” (arrivals) per day

Therefore, We can say that “Number of New Cases” has a Poisson Distribution for California State.

Calculating and Plotting Poisson Distribution -

By taking the K values for Poisson distribution from 0-430 range, and
Lambda value as the mean of “Num of New Cases Normalized” we can find pmf of Poisson distribution. Where Lambda = 82.2

Plotting the Poisson distribution with K on the x-axis and poisson prob P(x = K) on y-axis -