**TravNur**

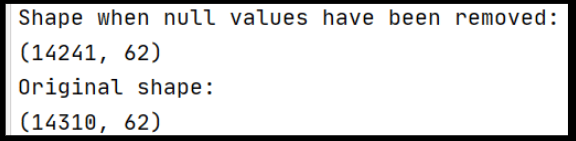
**Improving Healthcare through Data Mining**

Comp 541 Data Mining | Project 4

by Verab Chitchyan, Gabriella de Asis, Emmanoel Dermkrdichyan, Carson Logston, Ahn (Steven) Nguyen

**Null Values**

In our reduced dataset, we found that 69 rows had null values. Instead of filling in the null values with zeros or with the mean value, we decided to remove those 69 rows. Removing the rows would ensure our remaining data was complete and since it was only 69 rows, our dataset was not greatly diminished. We also did not want to fill those rows in with information that might not reflect the actual hospital reporting for that date.



**Figure 1: 69 rows removed**

**Outliers:**

To remove outliers, we used the z-score and removed outliers more than 3 points from the standard deviation.

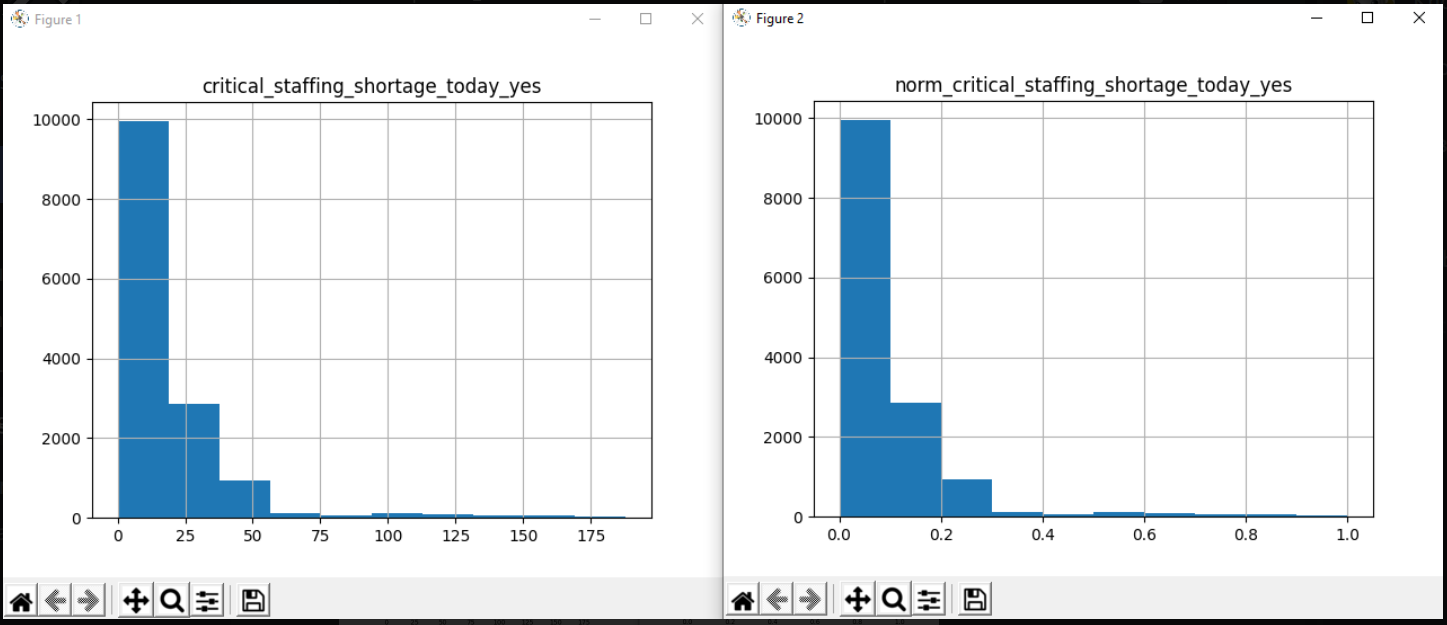
|  |  |
| --- | --- |
| **Figure 2: Dataset containing outliers** | **Figure 3: Most outliers removed** |

**Discretization:**

Our dataset is finite, so we did not need to perform any discretization.

**Data Normalization:**

We normalized our dataset and found that while there was some improvement, it still maintained the same shape and was still positively skewed.



**Figure 4: Non-normalized (left) and normalized (right) dataset**