An Outlier is a data-item/object that deviates significantly from the rest of the (so-called normal)objects. They can be caused by measurement or execution errors. The analysis for outlier detection is referred to as outlier mining. There are many ways to detect the outliers, and the removal process is the data frame same as removing a data item from the panda’s data frame.

Outlier detection using Z Score:

Z score = (x -mean) / std. deviation

If the z score of a data point is more than 3, it indicates that the data point is quite different from the other data points. Such a data point can be an outlier.

In the code we have directly calculated the outliers without calculating the individual z-scores.

​

​

​

​

​

Conclusion:

From the data we conclude that some places are pretty expensive and hence have fallen in the category of outliers.