

Title: Estimates of Variability II

Introduction

In this assessment There will be described Estimates of variability like defining quartiles and what it will give us. There will be chosen retail_19.csv as a dataset and Stock feature for calculations there.

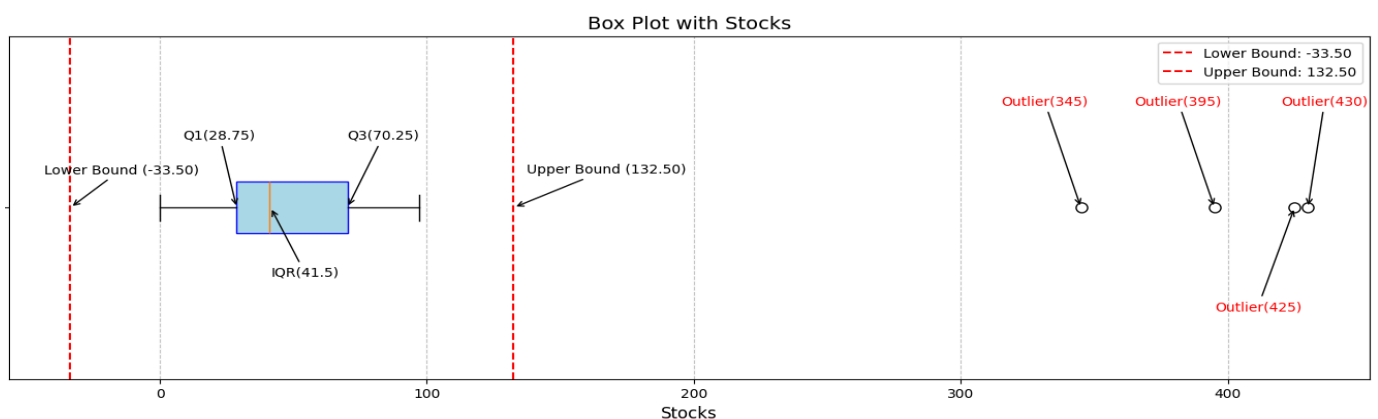
Methods

Before drawing Boxplot itself there should be made some formulas and it starts from sorting our Stock feature. Then, find the range value by subtracting max value of Stock from min value. The Interquartile range, by finding the first quartile and third quartile and subtracting it from the first one we find the Interquartile range.

Findings

By the formula we find that the range in stock feature is 430. Taking a note our first row is 0 and last is 430 the range is not trusting value. So, let's continue, to find the first quartile, multiply 0.25 by the length of rows and to find the third quartile 0.75 by the length of rows. Q1 is 28.75 and Q3 is 70.25, so, the Interquartile range is 41.5. Now, let's define upper and lower bounds to identify the borders. lower bounds are defined by subtracting 1.5 from Q1 and multiplying by IQR and upper is almost the same but adding 1.5 to Q3 and the results -33.50 and 132.50 respectively.

Box plot:



So, in this box plot everything is clear, even outliers. We can see that we have four outliers. Would be better if we trimmed these outliers. And it is clear that most of our data is near to each other looking at upper and lower bounds and boxplot itself with its whiskers.

Conclusion

Summarizing the calculations that we have made during the assessment, it is clear that the upper bound line is so close to the boxplot whisker it means that there is an abyss between last normal value and first outlier.