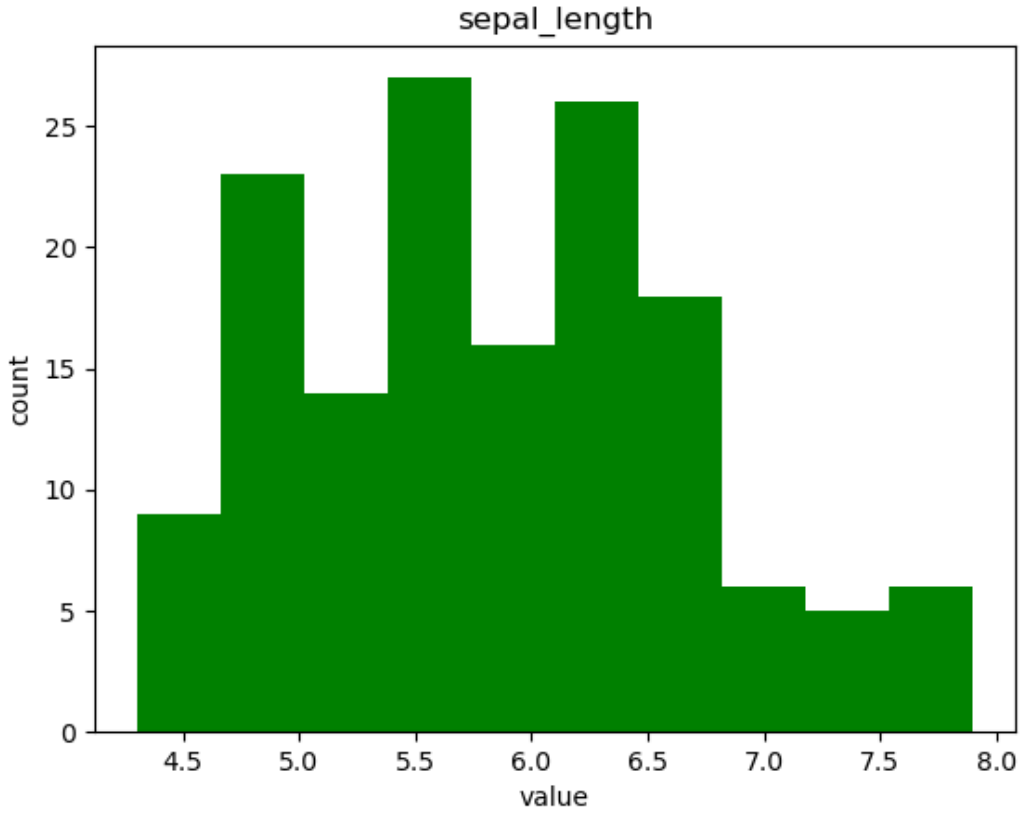
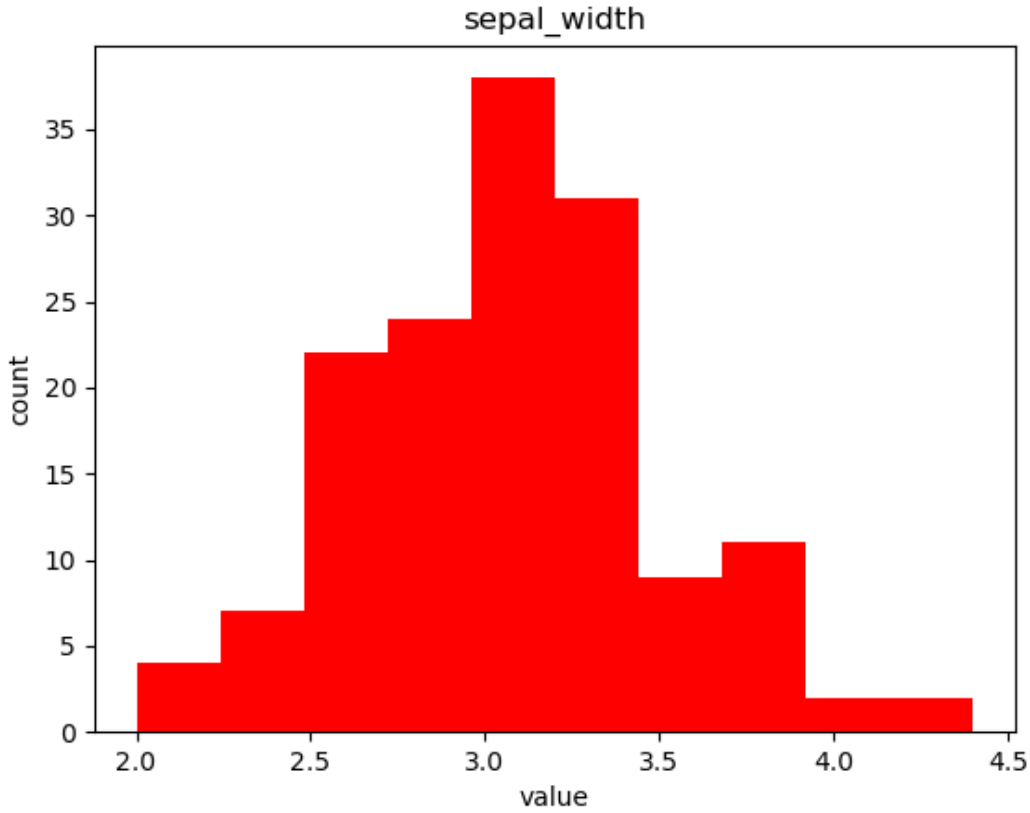
k-Nearest Neighbors (kNN) for anomaly detection

1.Load the data

This scrpit uses two characteristics from the famous Iris dataset. The two characteristic data are distributed as follows:

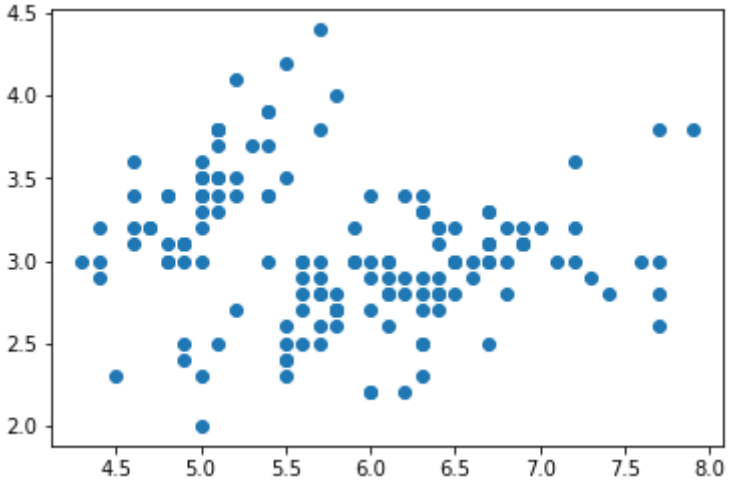




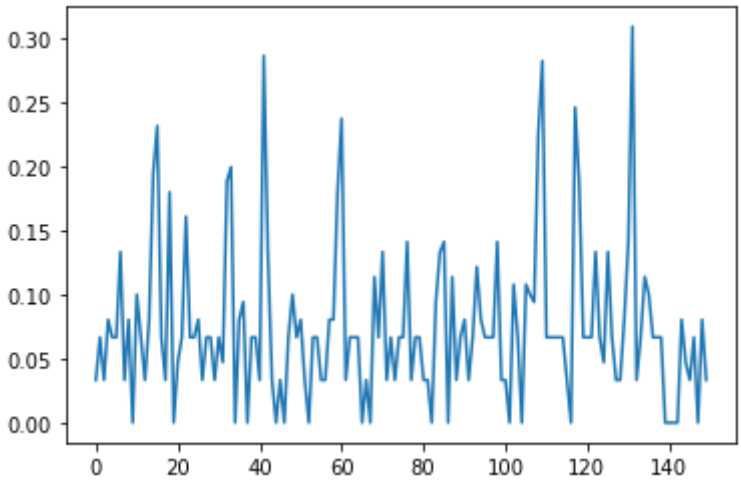
Checking to see if there is an NAn value:

sepal_length doesn't have Nan value.
sepal_width doesn't have Nan value.

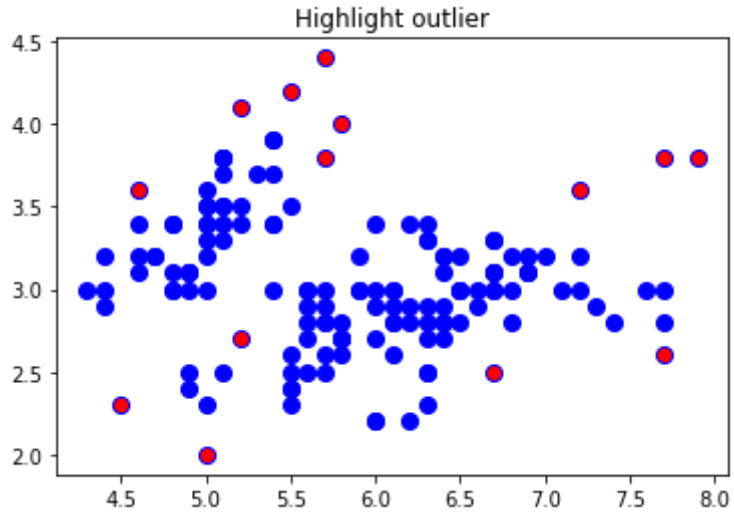
Display distribution of input data:



Show distances and indexes of k-neaighbors from model outputs:



plot outlies



Conclusion:

This example demonstrated how to implement KNN for identifying anomalies in a dataset.I think this is a good clustering method to find outliers. But in the real world, how to deal with large data sets is a question worth considering.