Fundamental Exam Data Science

Score (45)

1. You are now a data scientist in a company, your first task is to detect the outlier data by calculate the IQR (Interquartile Range). In statistics, an outlier is a data point that differs significantly from other observations.

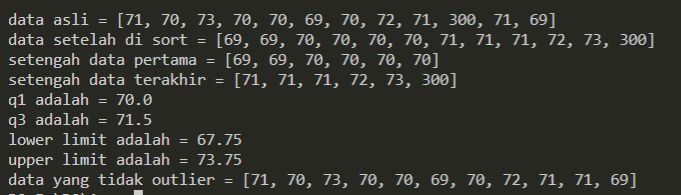
For example I have data = [71, 70, 73, 70, 70, 69, 70, 72, 71, 300, 71, 69].

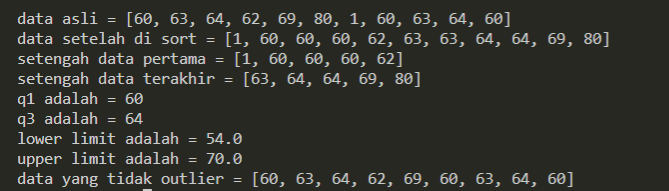
To identify the outlier data , read the instruction below.

1. Sort the data , data = [69, 69, 70, 70, 70, 70, 71, 71, 71, 72, 73, 300]
2. Find the median = 70 + 71 / 2 = 70.5
3. After finding the median, you have to find the Q1 and Q3 data. Because the example data length is 12, so the Q1 is equal to the median of first six data. And Q3 is equal to the median of the last six data.
4. Q1 = 70 + 70 / 2 = 70
5. Q3 = 71 + 72 / 2 = 71.5
6. IQR = Q3 – Q1 = 1.5
7. Find the lower limit and the upper limit, lower limit = Q1 – 1.5×IQR, upper limit = Q3 + 1.5×IQR
8. if a data point is below the lower limit or the data point is above the upper limit, then the data point is oulier

remove\_outlier([71, 70, 73, 70, 70, 69, 70, 72, 71, 300, 71, 69]) => [71, 70, 73, 70, 70, 69, 70, 72, 71, 71, 69]

the answer must be the picture below,



 Another example input and output

Score(20)

1. Return the number (count) of vowels in the given string. We will consider a, e, i, o, and u as vowels for this case. The input string will only consist of lower case letters and/or spaces.

countVowel(‘budi pergi ke pasar’) 🡺 7

countVowel(‘purwadhika’) 🡺 4

Score(30)

1. Given a two-dimensional list of integers, return the flattened version of the list with all the integers in the sorted (ascending) order.

given([[3, 2, 1], [4, 6, 5], [], [9, 7, 8]]) 🡺 [1, 2, 3, 4, 5, 6, 7, 8, 9]

given([[3,4,2,1] , [1,2,3] , [5,4,3,1]]) 🡺 [1,1,1,2,2,3,3,3,4,4,5]

Score(30)

1. Given string of words, you have to calculate the count of every word in the sentences

countWords(‘jangan jangan kamu adalah aku’)

should return

