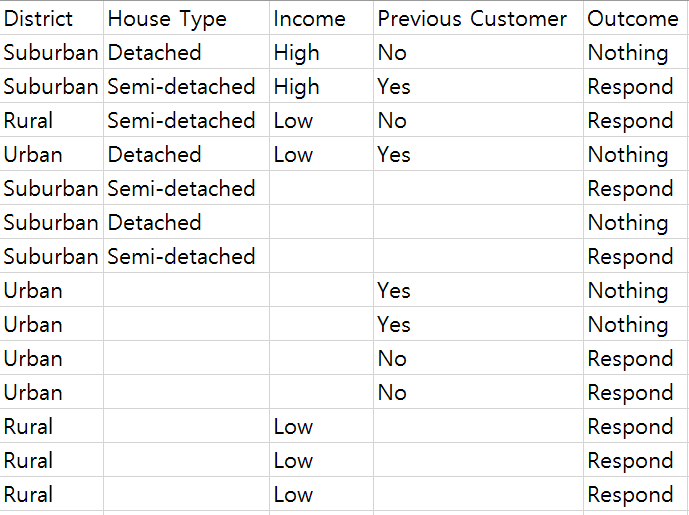
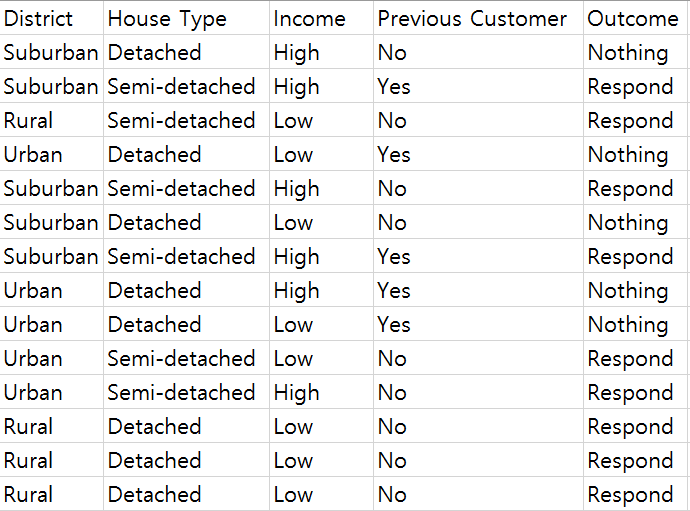
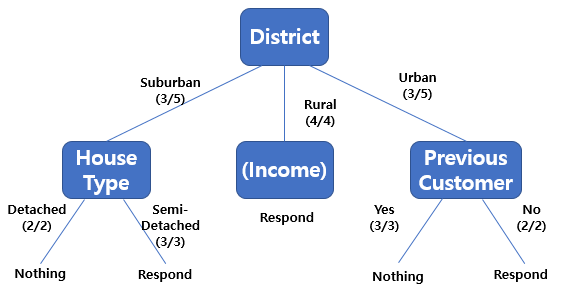
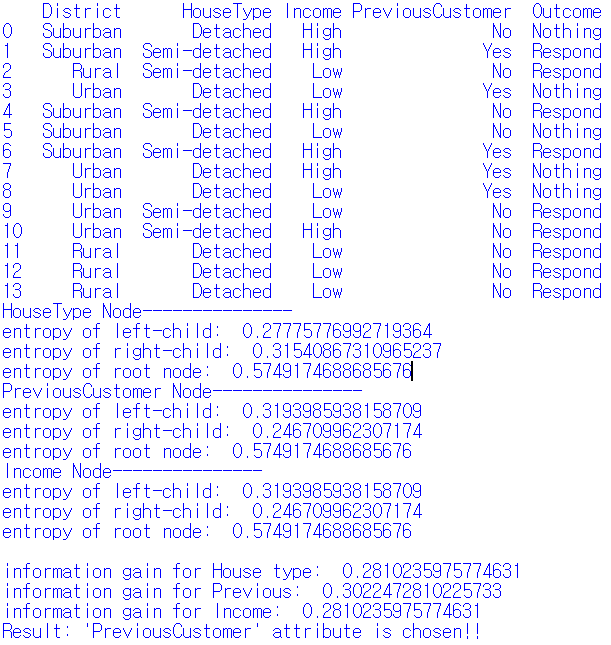
**Data Science homework\_lab4 201533661 이승수**

**Problem 1: Decision Tree**

At below chart, left one is given datas and right one is which I filled and which I used at problem1.

**<given dataset> <optionally filled dataset>**

**<Discision tree of given dataset(left)>**

Following the optionally filled dataset,

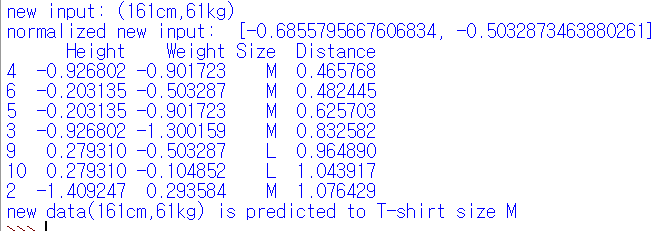
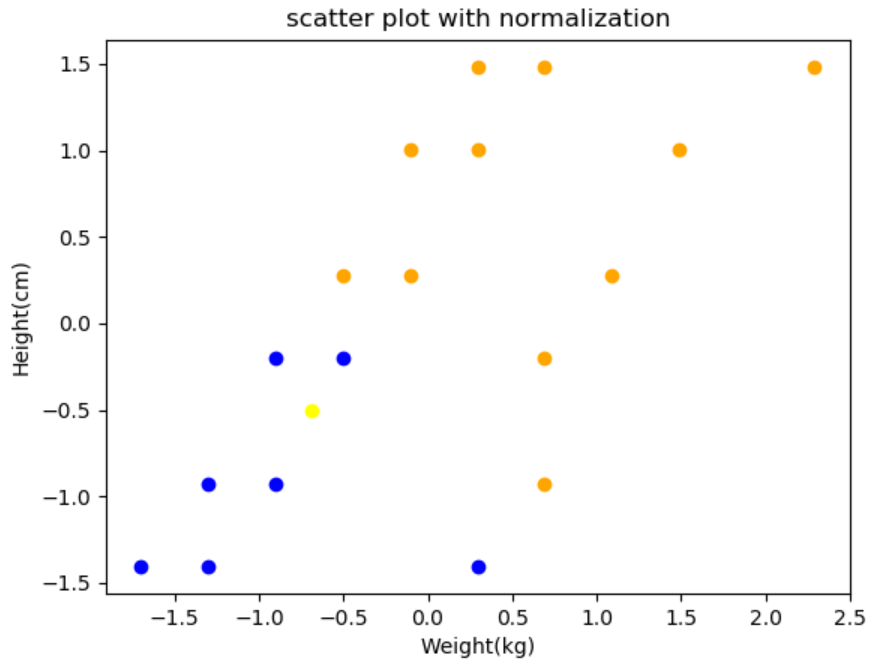
Entropy of each child nodes are as on left.

Calculating information gain of each attributes with my dataset,

‘**PreviousCustomer**’ attribute is chosen with information gain 0.30224… which is biggest.

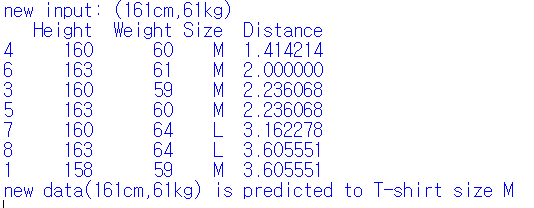
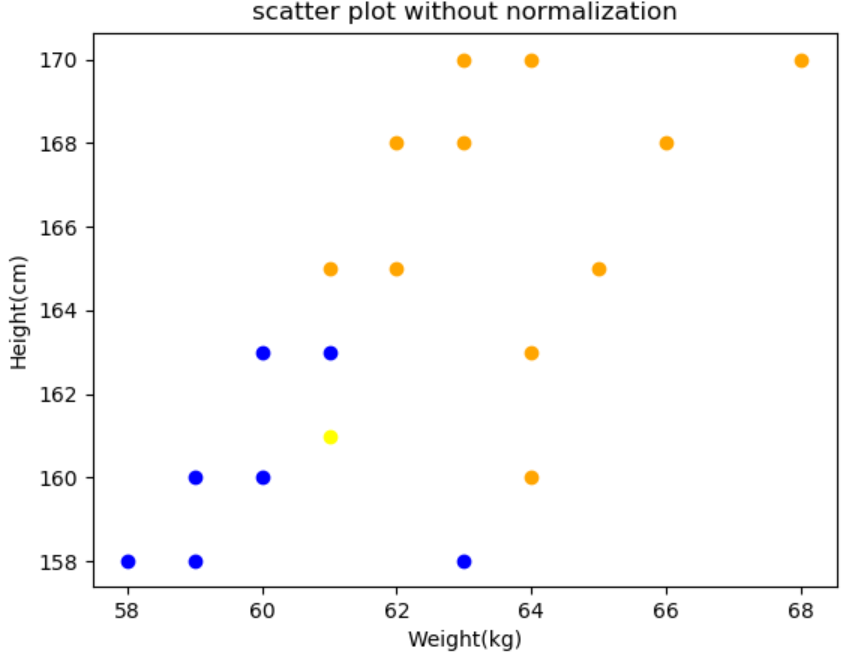
**Problem 2: K-nearest Neighbors Algorithm**

Below scatter plot and output is result with hyper parameter K is 7 with normalization.



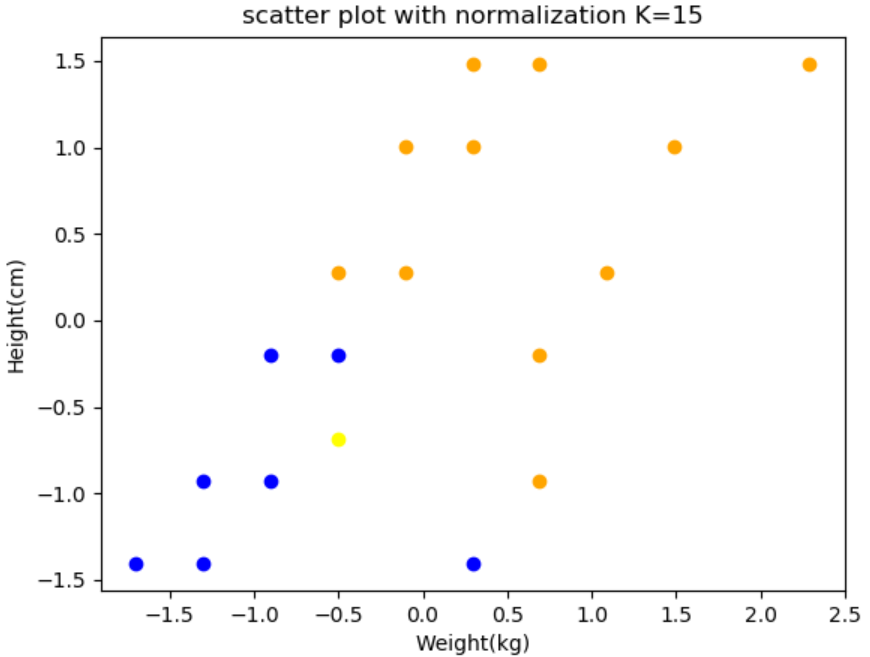
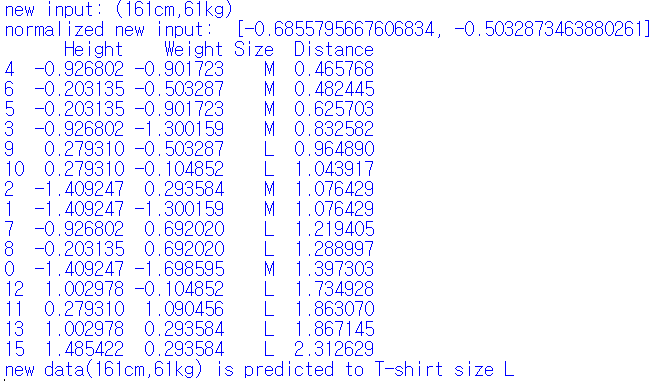
Below scatter plot and output is result with hyper parameter K is 7 without normalization.

Result is not changed.



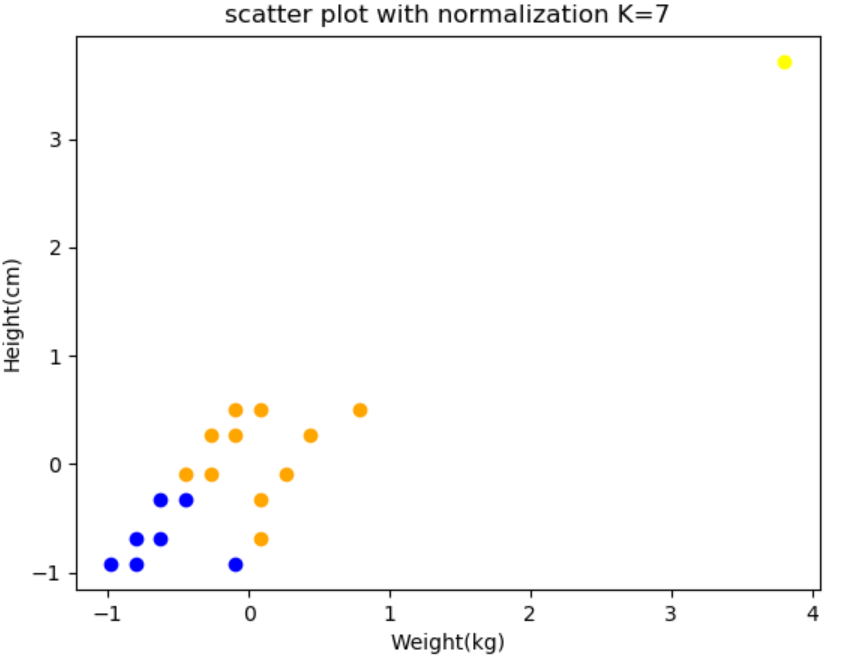
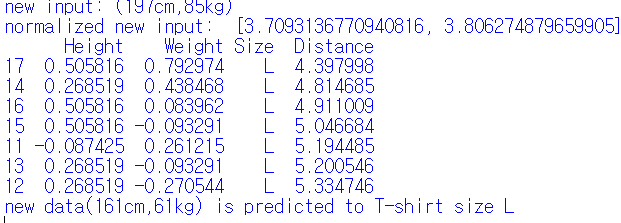
Below scatter plot and output is result with hyper parameter K is 15 with normalization.

Result is changed because of hyper parameter K size becomes bigger.

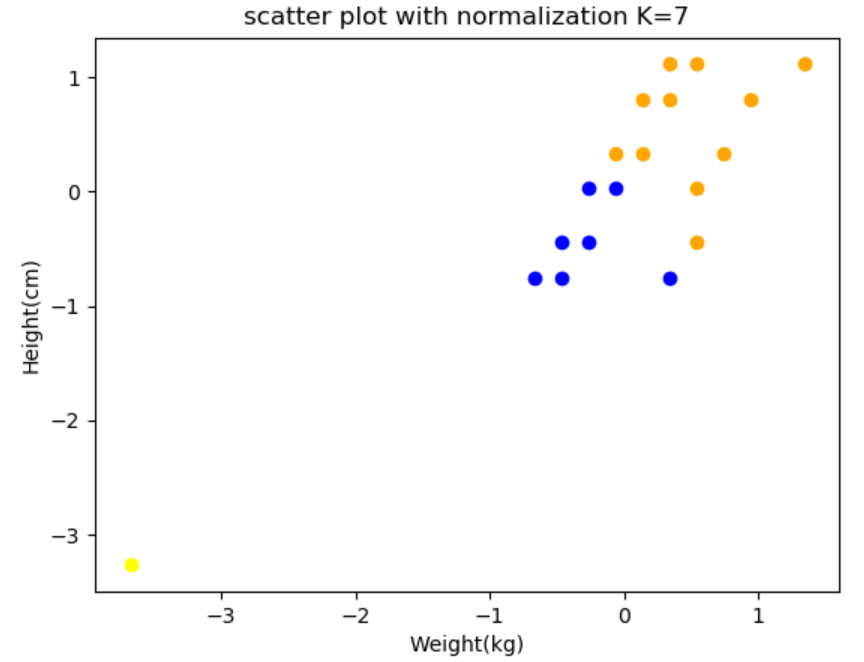
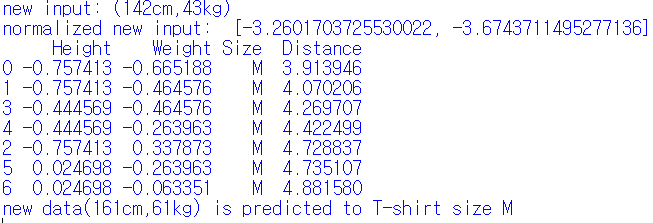
Below scatter plot and output is result with hyper parameter K is 7 with normalization.

New input comes larger than max values.(197cm, 85kg), output becomes changed.

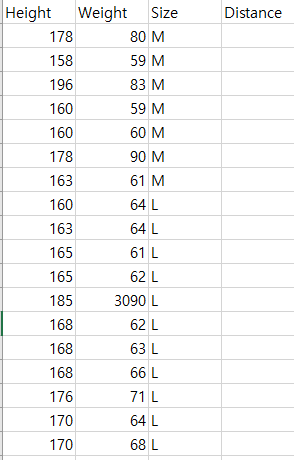
 

Below scatter plot and output is result with hyper parameter K is 7 with normalization.

New input comes smallerr than minimum values.(142cm, 43kg), but output is not changed.

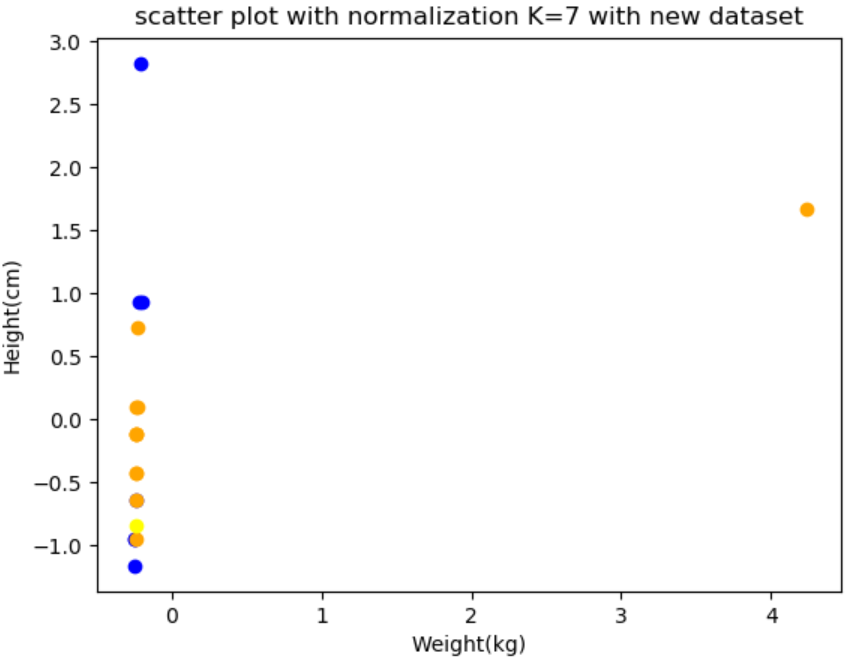
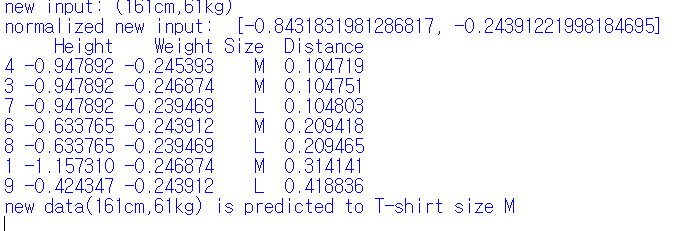
 

Below cases use new datasets on Sheet2.



Below scatter plot and output is result with hyper parameter K is 7 with normalization.

Scatter plot becomes different but result not changed.

Below scatter plot and output is result with hyper parameter K is 7 with normalization.

Result is same as normalized one.

