Problem Statement 1:Graphical user interface, application

Description automatically generated

Here I have imported pandas, numpy and matplotlib.pyplot and then I have read the dataset.

Then I have imported labelencoder for normalize labels.

Graphical user interface, application

Description automatically generated

Here I have described the dataset.

Table

Description automatically generated with low confidence

Here I have given command called head() for viewing the top five rows of the dataset.

Graphical user interface

Description automatically generated with medium confidence

Table

Description automatically generated

Table

Description automatically generated

Here I have viewed the dataset individually by using the above commands.

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Graphical user interface, chart

Description automatically generated

Here I have used scatterplot for width and height and then used scatterplot for mass and color\_score by using the matplotlib.pyplot function. The scatterplot are used to observe relationships between variables. And then plotted the data for height and width.

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

 Precision:

Precision, often referred to as Positive Predicted Value (PPV), is the fraction of samples that the classifier model predicted to be the positive class to the total of number samples that were predicted to be in the positive class. It summarizes how precise the model is out of those predicted as positive; how many of them actually are positive.

## Recall/Sensitivity:

Recall, often referred to as Sensitivity or True Positive Rate (TPR), is the fraction/ratio of samples that the classifier model predicted to be the positive class to the samples that actually belongs to the positive class. It basically summarizes how well the positive class was predicted by the classifier.

## F-Score:

F-Score, often referred to as F-Measure, is a harmonic mean of precision and recall.

Graphical user interface, chart, scatter chart

Description automatically generated

Here I have build a classifier using KNN and then figure out the best value of k with highest r\_score.