CURNEU ASSESSMENT –SD03Q02 HARSITHA.N

1832022

TASK-1-FRUITS DATASET

PROBLEM STATEMENT:

A dataset labelled based on fruit height, width, mass and colour score is given in fruits.xlsx. A classifier based on k Nearest Neighbour (KNN) algorithm is to be crafted for classification.

• Generate scatter plots for various combination of parameters and do the feature engineering meaning thereby which parameters of best suited to build the classifier.

• Split the data into test and training split.

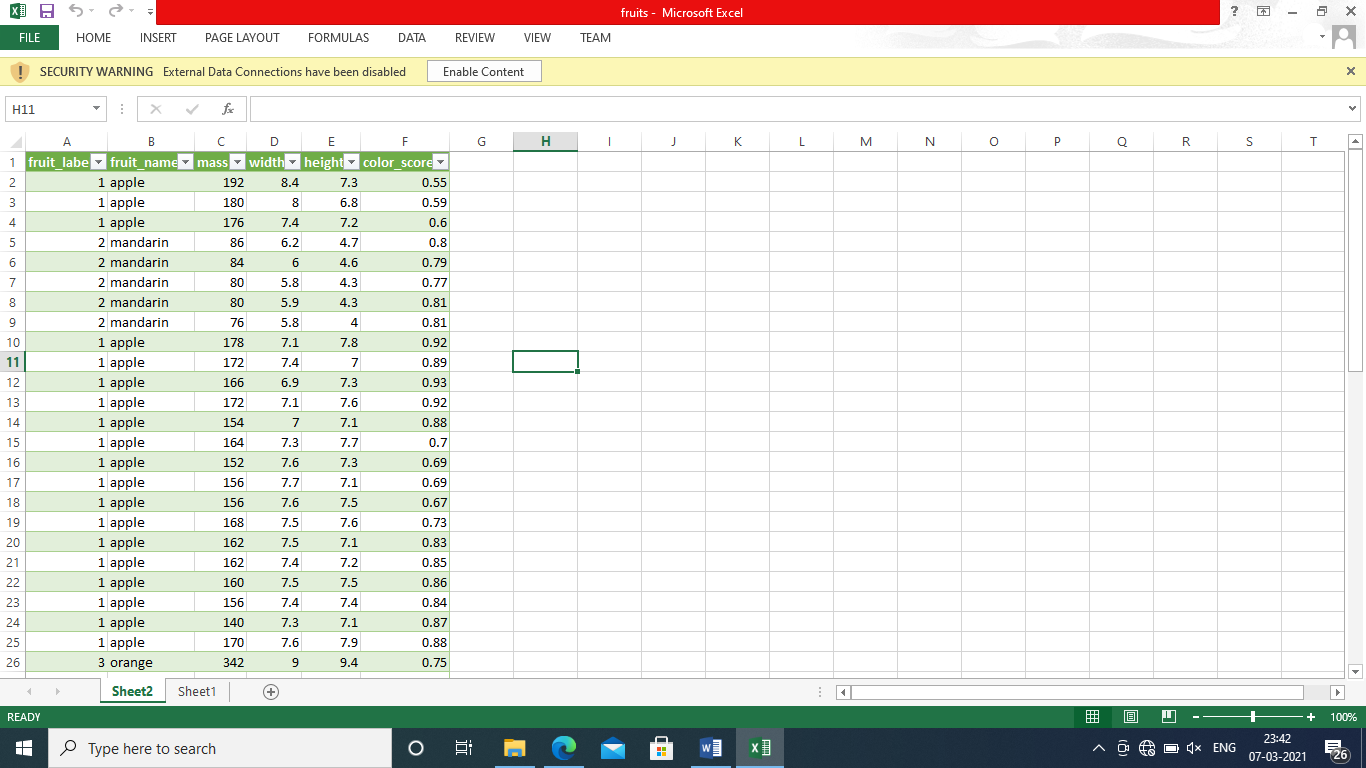
• Building a classifier using KNN from scratch.

• Figure out the best value of k with highest r\_score.

• Run at least three test cases on the parameter and assess the fruit using the classifier.

• Only use python

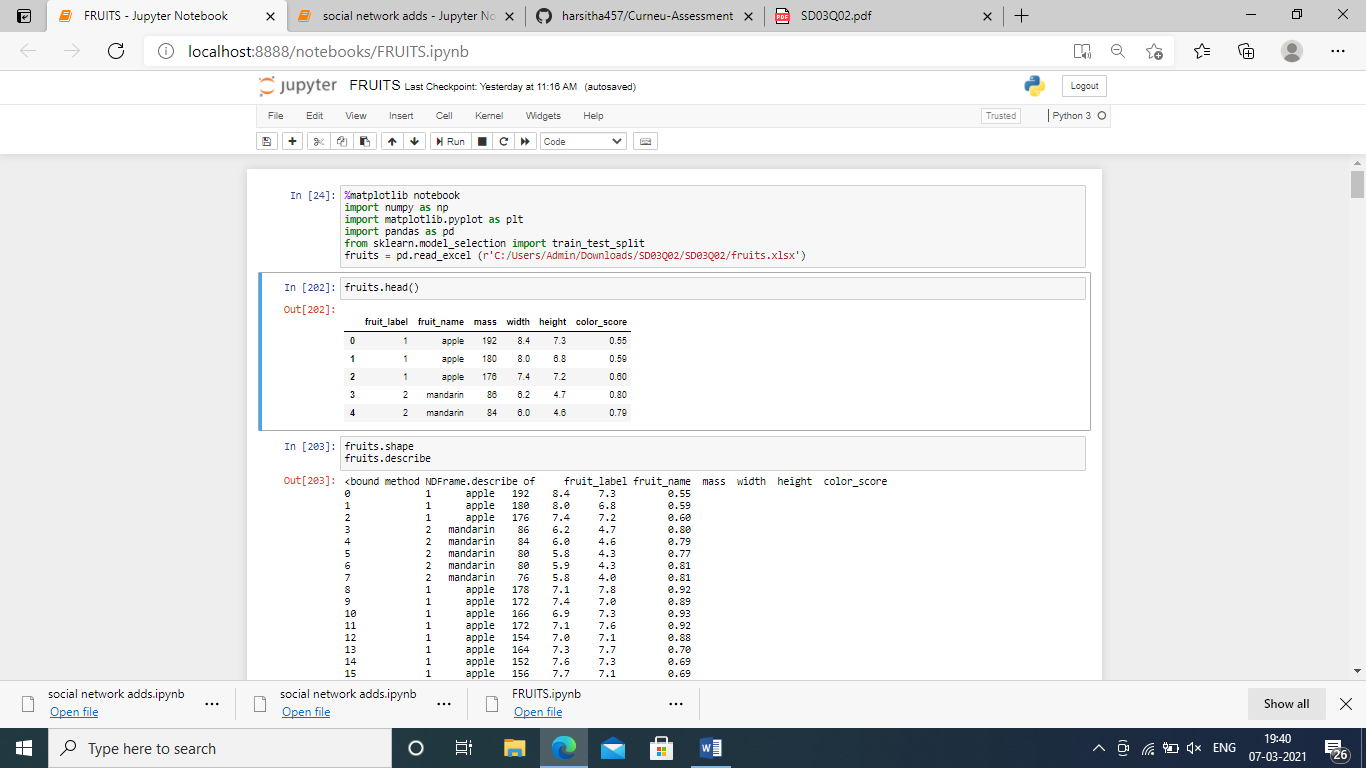
DATASET SAMPLE:



INTERPRETATION:

UNDERSTANDING THE DATASET

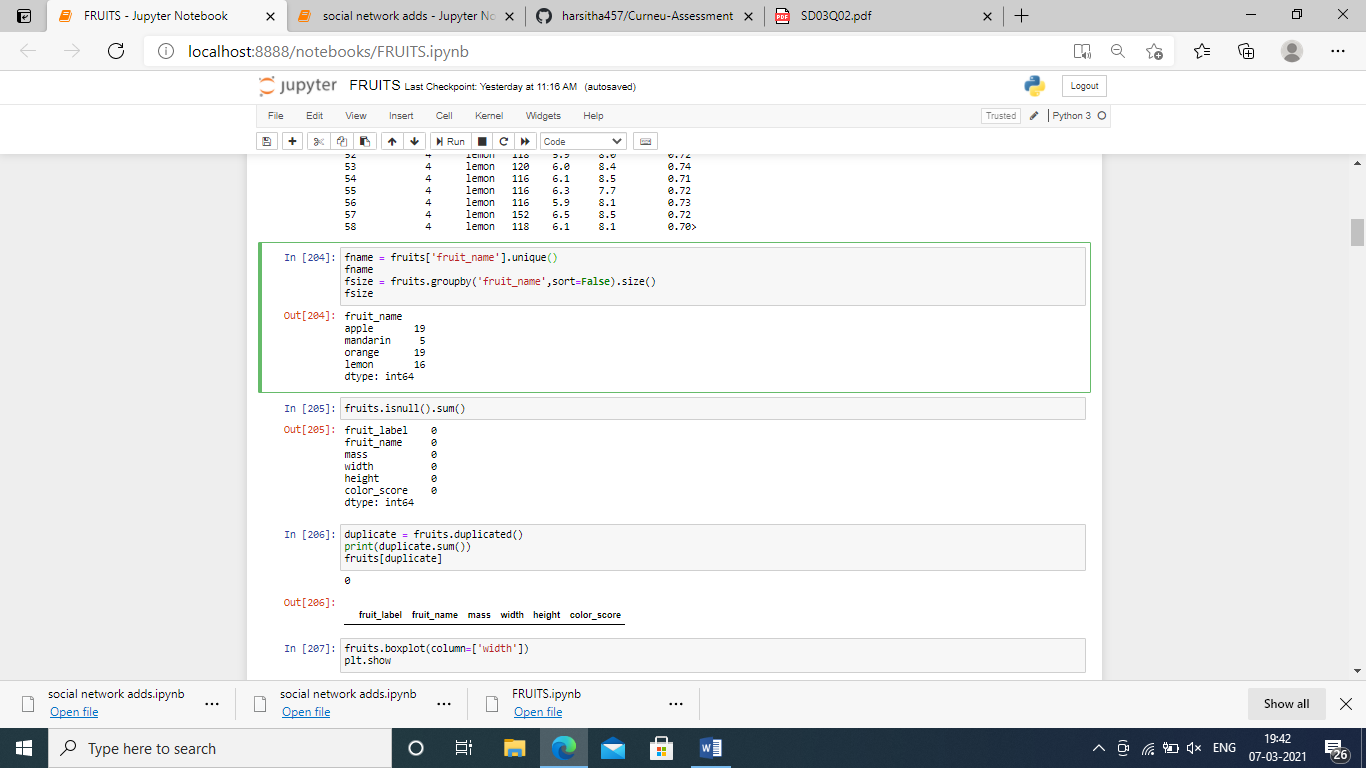
Fruits datatset: The above dataset consists of 6 attributes namely fruit\_label,fruit\_name,mass,width,height,color\_score



PREPROCESSING:

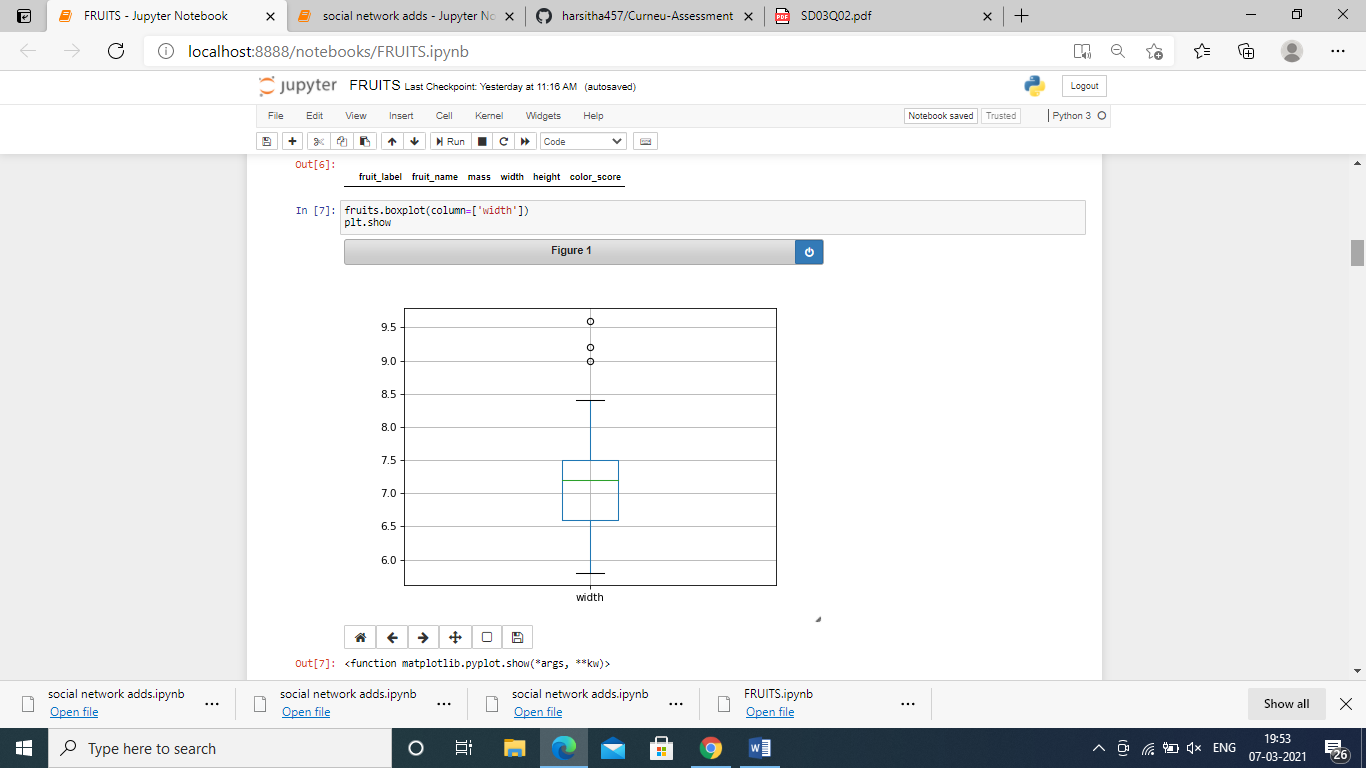
The dataset does not have null value .

The dataset does not any duplicates.

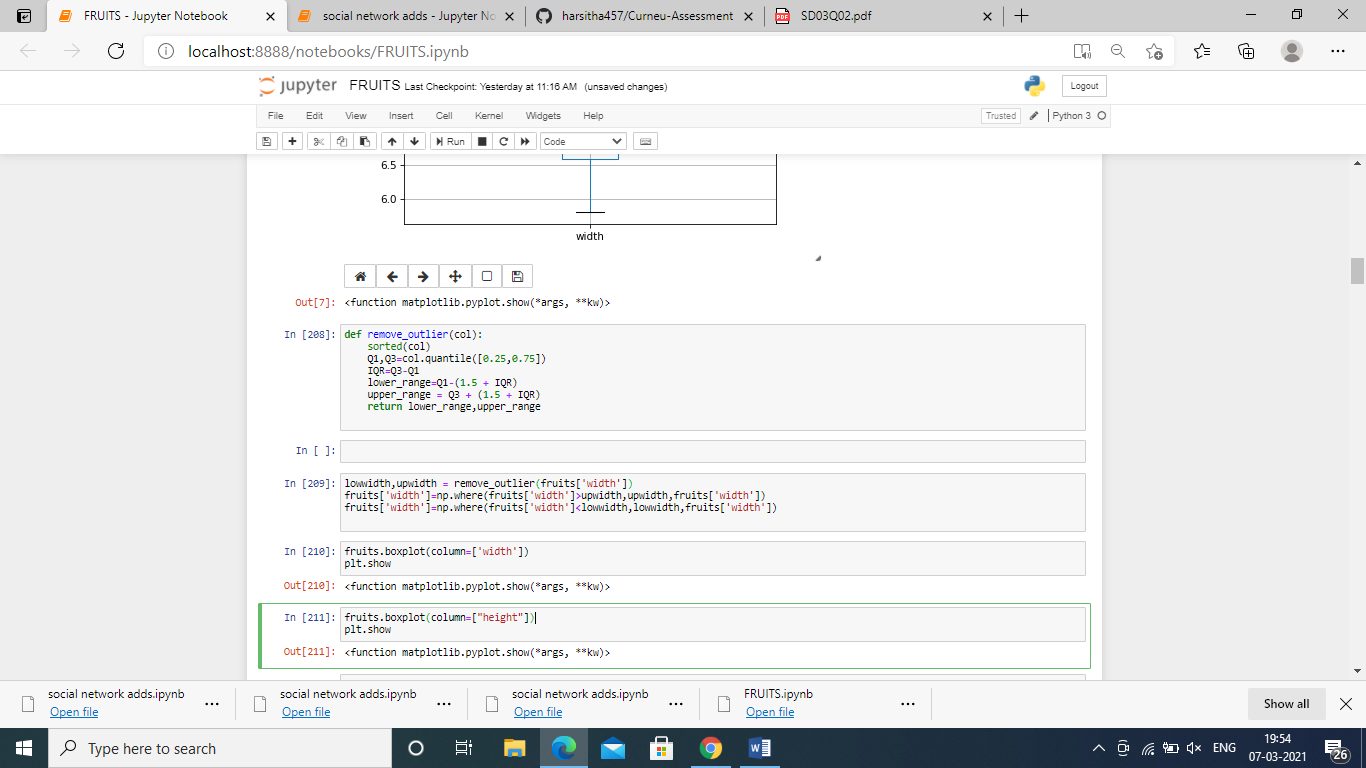


The attribute width ,mass and height had 3 outliers

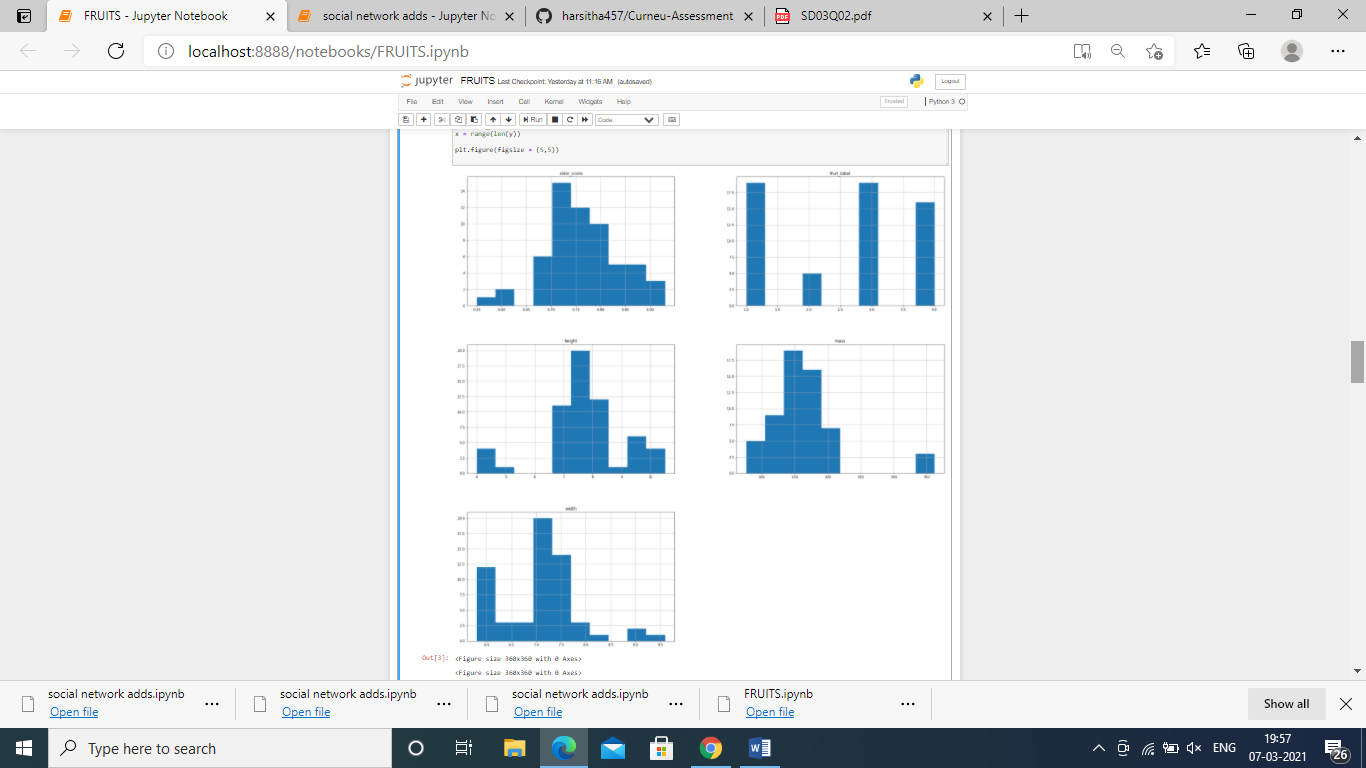
The boxplot for width showing the outliers(3)



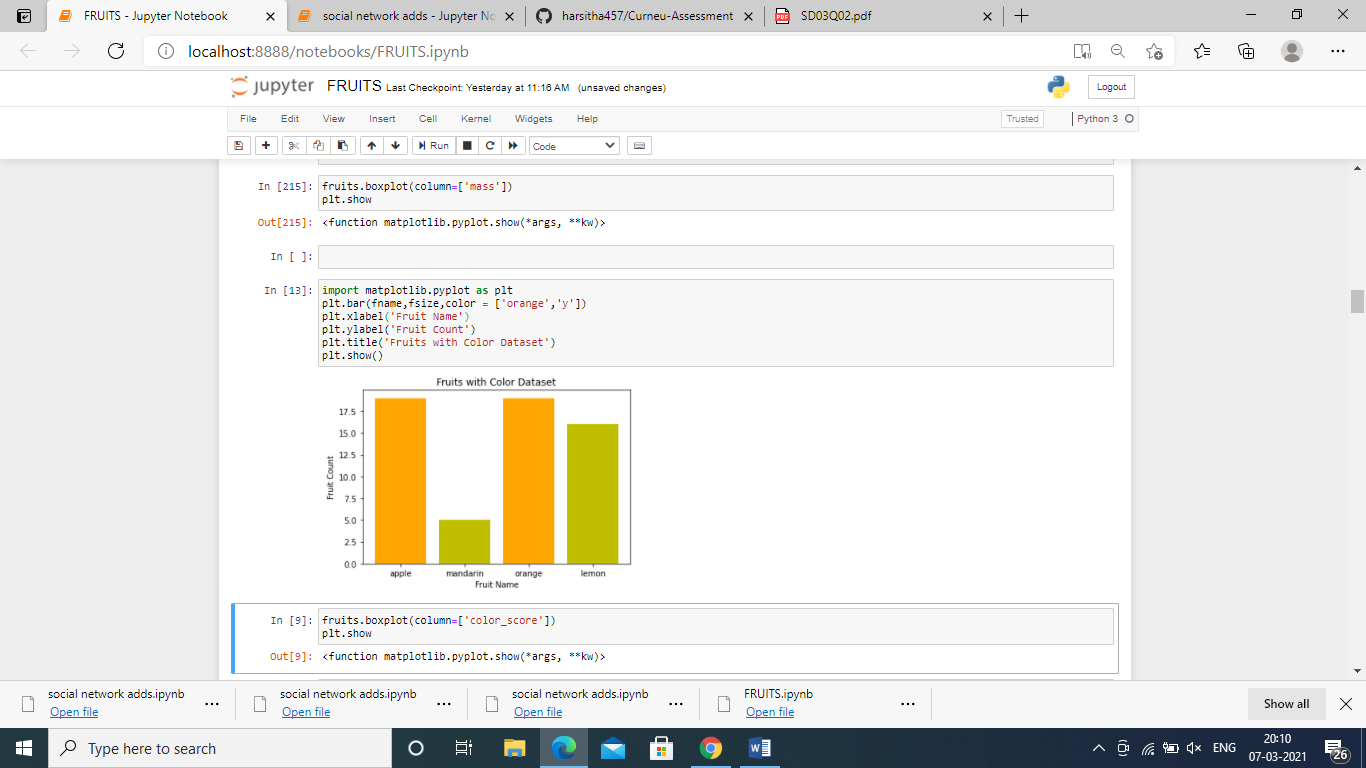
The outliers are removed using the following code



The frequency distribution of each attribute is plotted



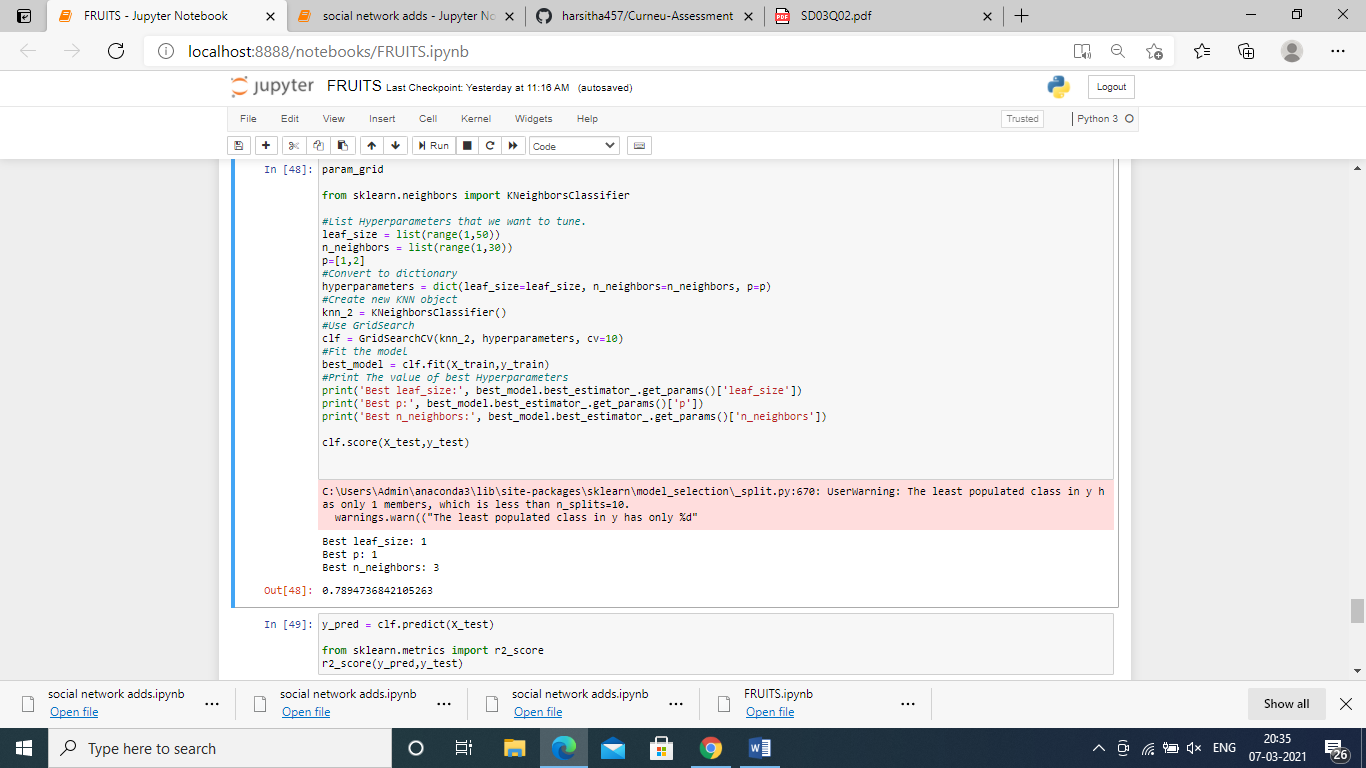
The no of fruits are plotted using a barplot



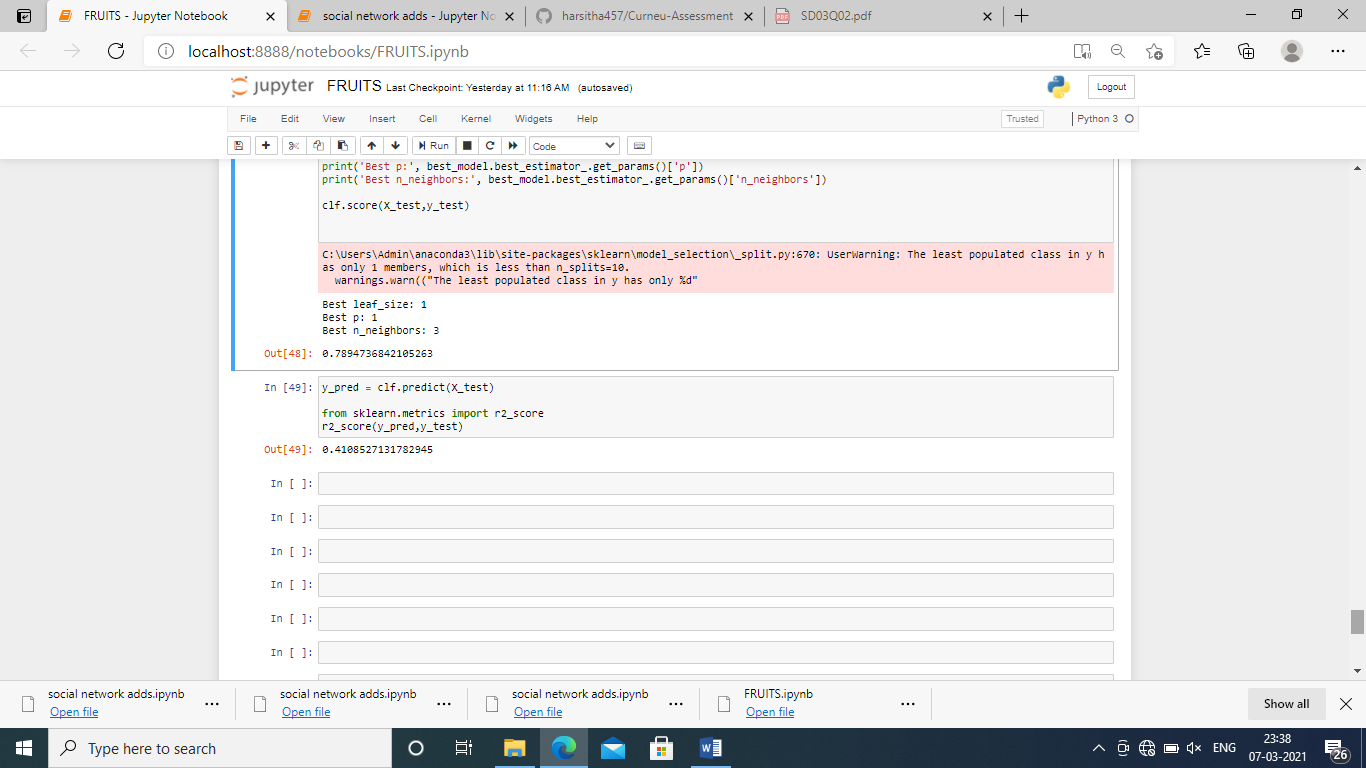
IMPLEMENTATION OF KNN CLASSIFIER

After preprocessing knn model is implemented and hypertunning is done.

The accuracy was found to be 0.68 before hypertuning and 0.78 after hypertuning.



The model accuracy was found to be 0.78 and the r2\_score is found to be 0.41



CONCLUSION:

The dataset is tested and trained using the knn classifier and the accuracy was found to be 0.68 by hypertuning the model the accuracy was about 0.78 with no of neighbors being 3.Before hypertuning the k value considered was 7. The r2\_score is 0.41. To conclude the implemented knn classifier model classifies the data well.