Steps we will be following here:

Import the data

Clean the data : Removing duplicates , remove incomplete / irrelevant data

Split the data into Training / Test Steps ( eg., 80% Training and 20% for Testing)

Create a Model ( Decision Trees )

Traing the Model

Make Predictions

Evaluate and Improve

Libraries and tools we will be using:

Numpy

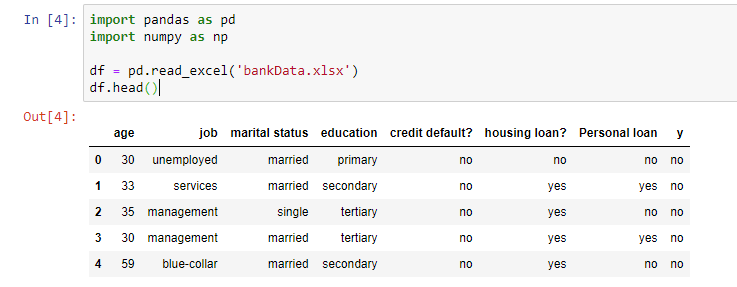
Pandas

MatPlotLib

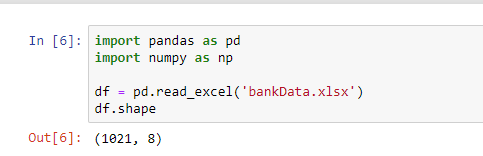
Scikit-Learn

Importing the data set and cleaning the data

1. Peek view of the data set:

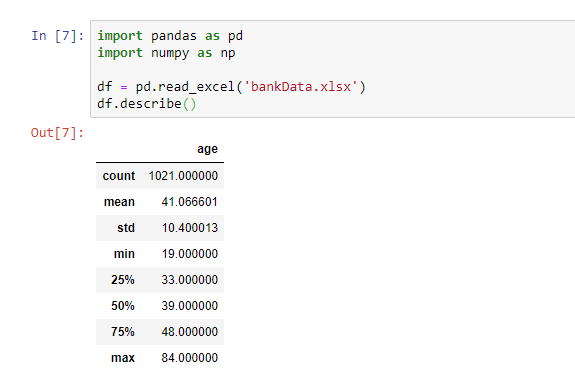


1. Getting shape of the dataset



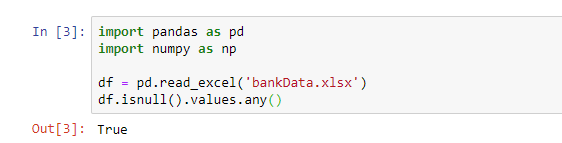
1. Describing the dataset

Our data has only one column that has numeric value

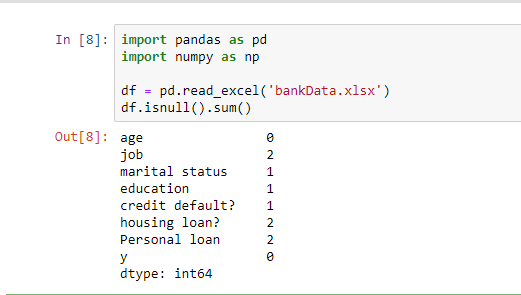


Clean the data : Removing duplicates , remove incomplete / irrelevant data

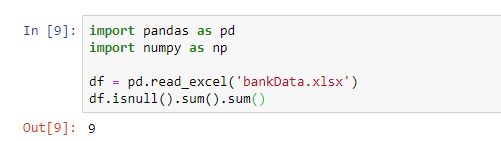
1. Is the dataset contains any missing values ?



1. Finding the null values in the dataset



1. Finding the total count of null values in the data set ( so there were 9 in total )

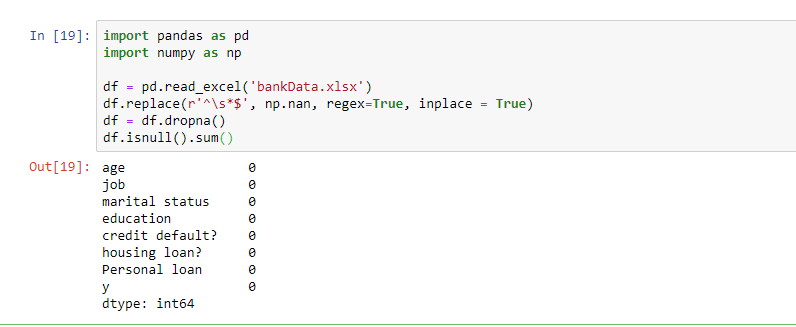


1. All of the missing values datatype were of text so for now we can delete them

First fill the blanks with NaN

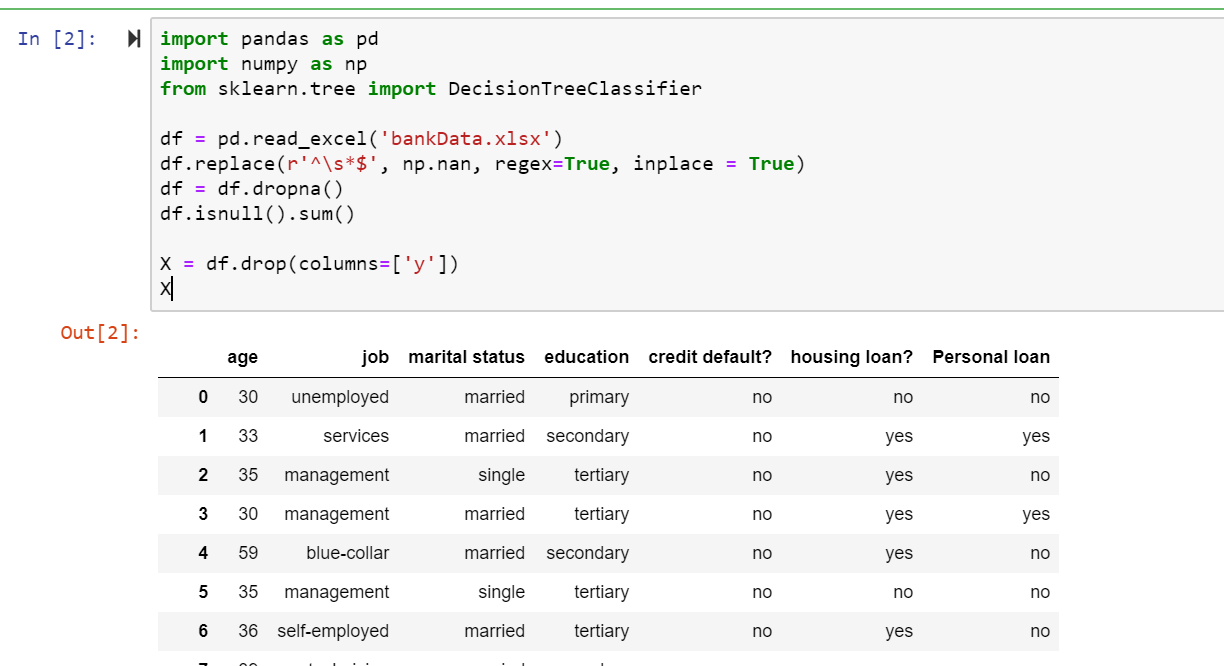


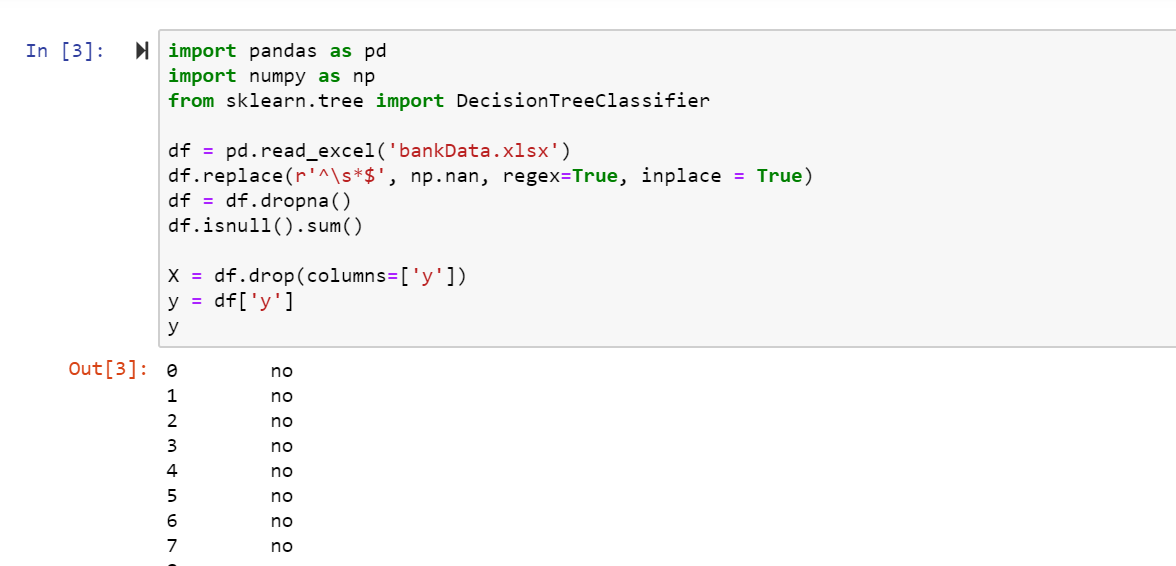
Then drop them all



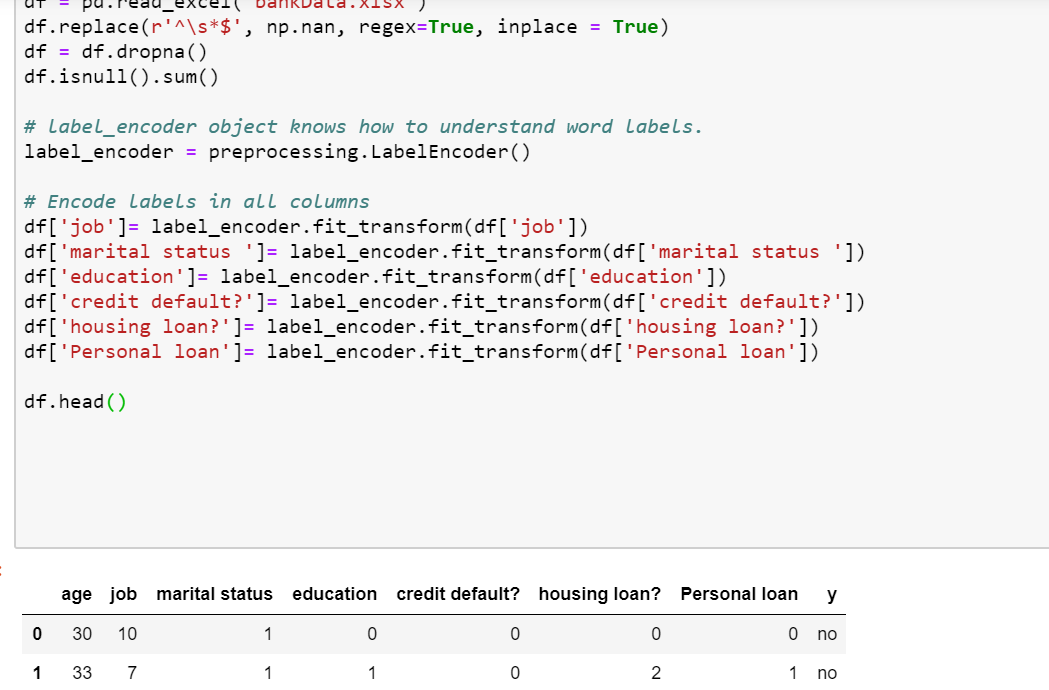
Training the model

1. Learning and predicting ( drop the last column and save the dataset into variable X and variable y will have only last column into it)

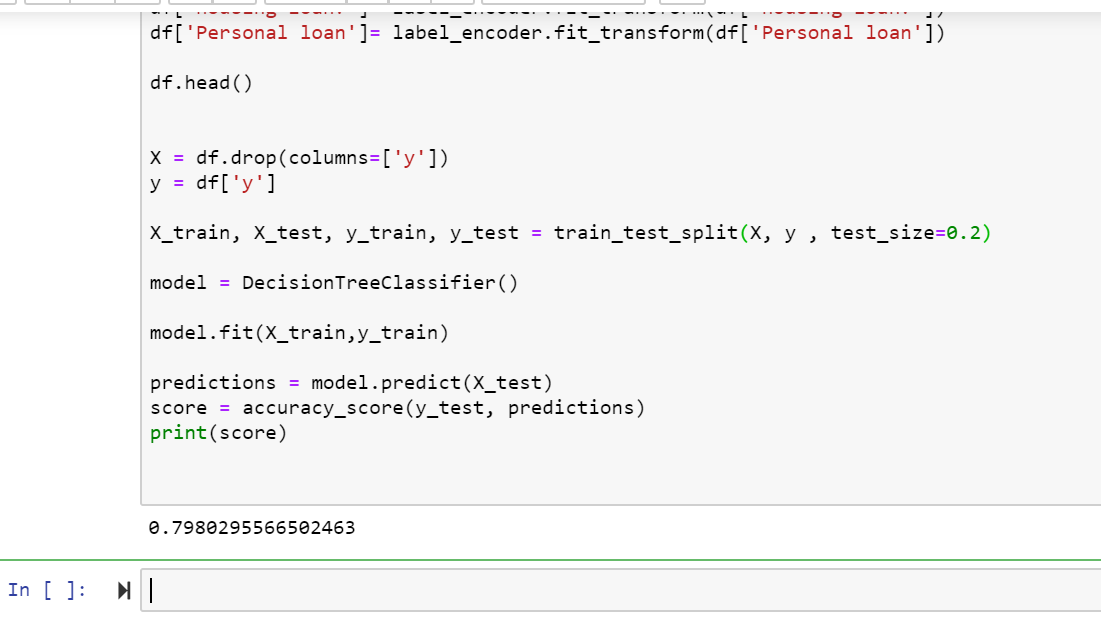




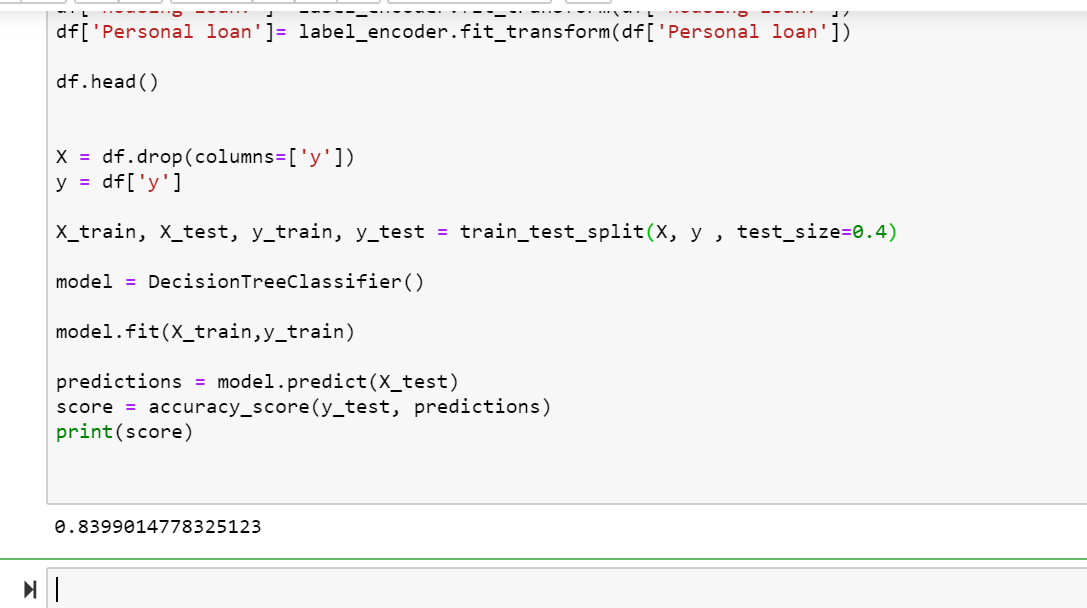
1. Normalize the data using LableEncoder



1. Where we set 80% to train and 20% to test the accuracy is given below



1. Where we set 60% to train and 40% to test the accuracy is given below ( increasing the test size )



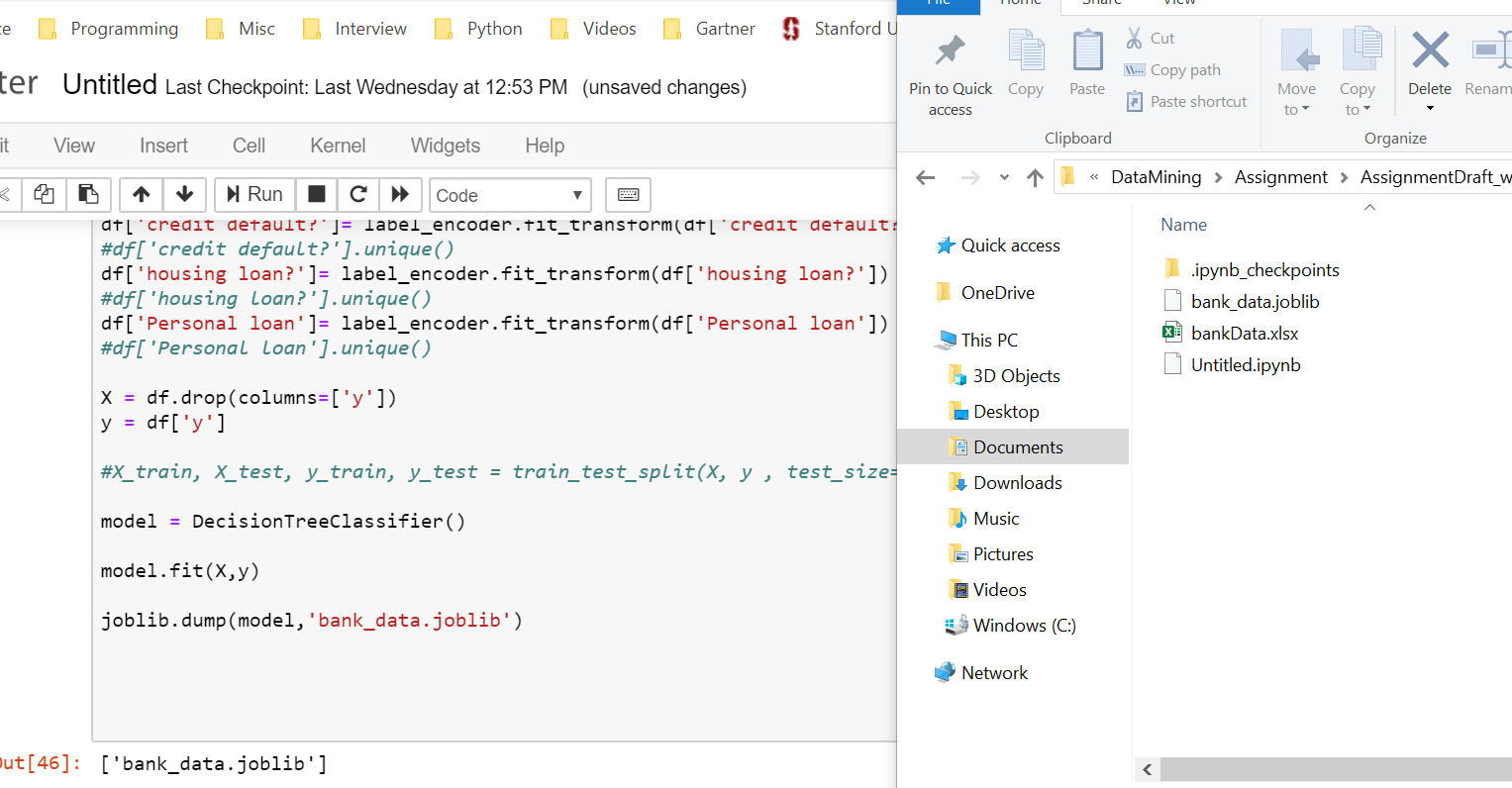
As you see accuracy score grows up

PREDICTION

Lets predict with some data now



NOW LETS SAVE THE MODEL TO HARD DISK



NOW REMOVE ALL CODE AND JUST USE THE SAVED MODEL



VISUALIZING the model as Decision Tree ( save in dot file )



Its gives a DOT File you can open the DOT file in VS code and see the decision tree

