**Python and Deep learning**

**Lab 1 assignment**

**Submitted by Josh Oguntimehin**

**Objective**

**Application/software used; Python**

**The objective is to use python programming language to create a dictionary, find the longest substring without repeating character, create a multiple regression and find the RMSE and R2 ,Handling of Null value,by applying three classification algorithms;-Naïve Baye’s, SVM,KNN on a chosen data. And applying k-means on a dataset.**

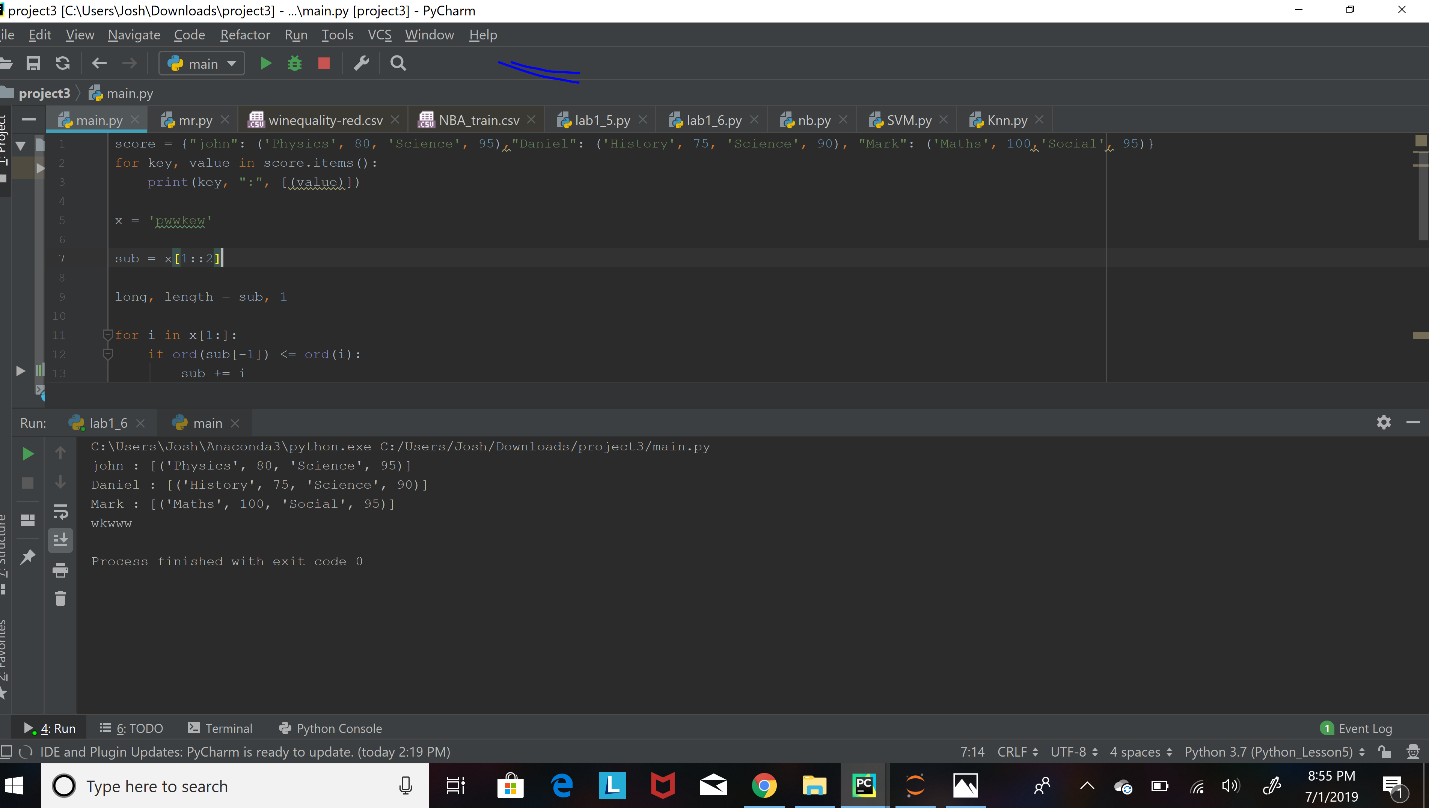
**Approaches/Methods**

1.Creating a dictionary with keys as names and values as list of (subjects, marks) in sorted order and

2.using python to find the longest substrings without repeating characters along with the length as a

tuple

score = {"john": ('Physics', 80, 'Science', 95),"Daniel": ('History', 75, 'Science', 90), "Mark": ('Maths', 100,'Social', 95)}  
for key, value in score.items():  
 print(key, ":", [(value)])  
  
x = 'pwwkew'  
  
sub = x[1::2]  
  
long, length = sub, 1  
  
for i in x[1:]:  
 if ord(sub[-1]) <= ord(i):  
 sub += i



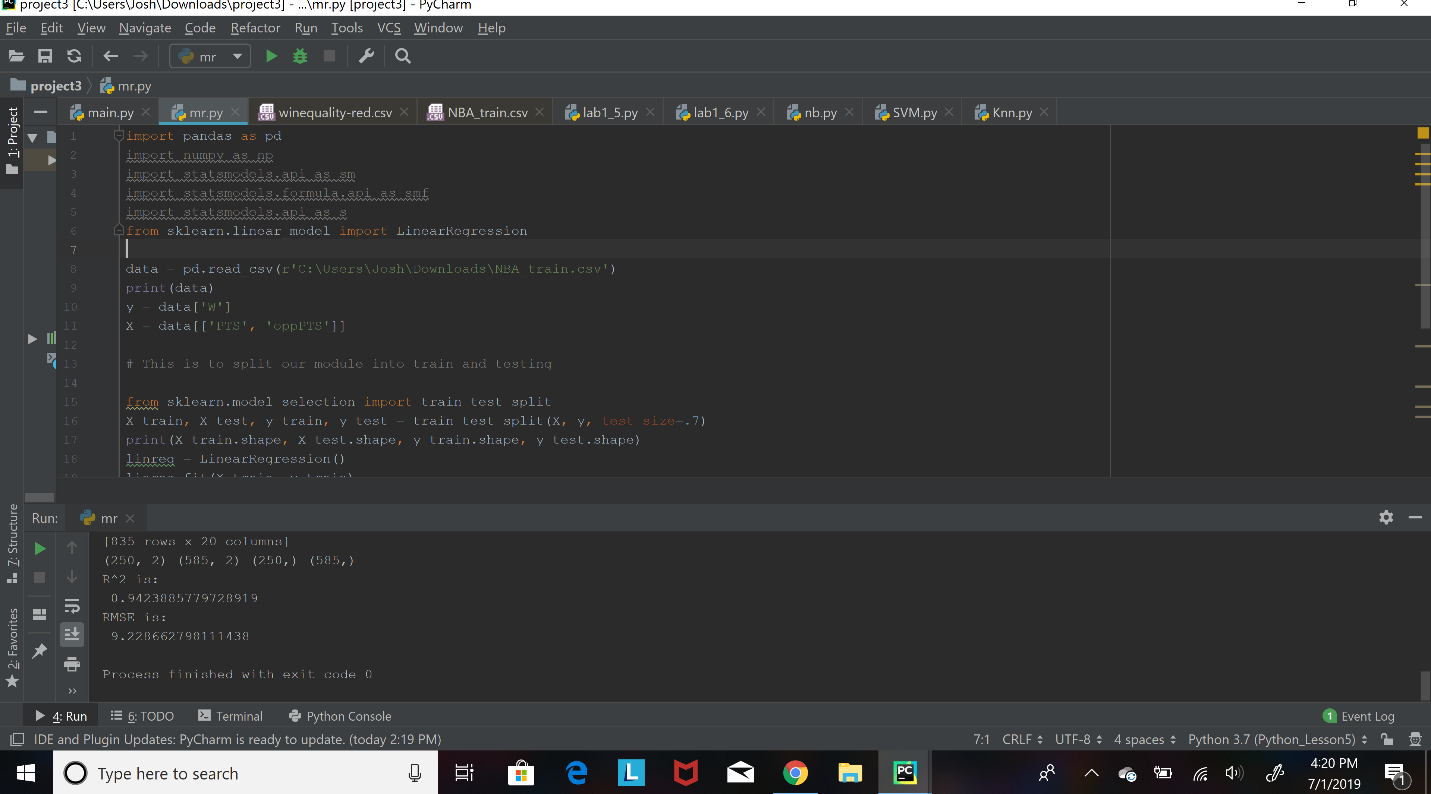
4. l create multiple regression by choosing a dataset of my choice and

evaluate the model using RMSE and R2 and also report if you saw

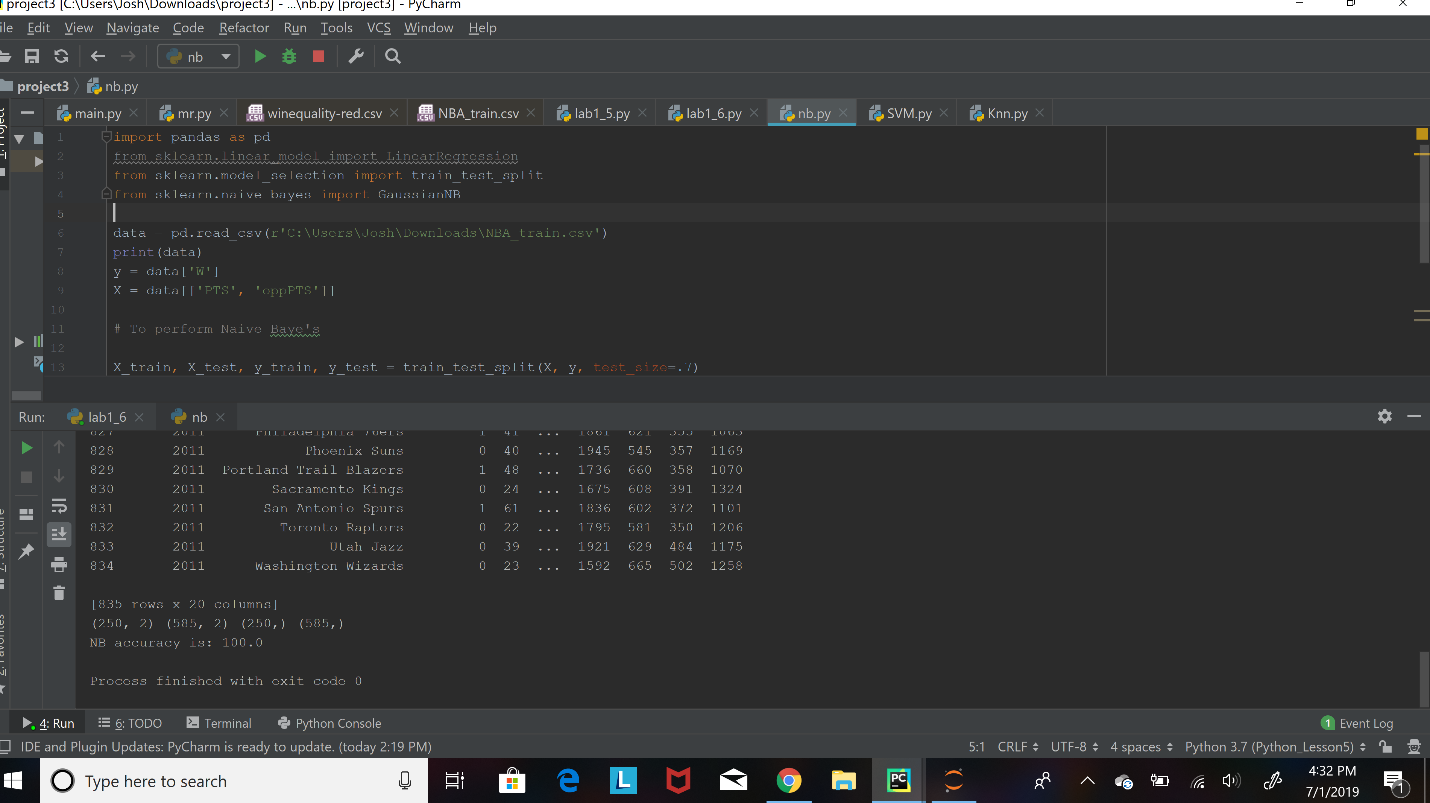
any improvement before and after the EDA

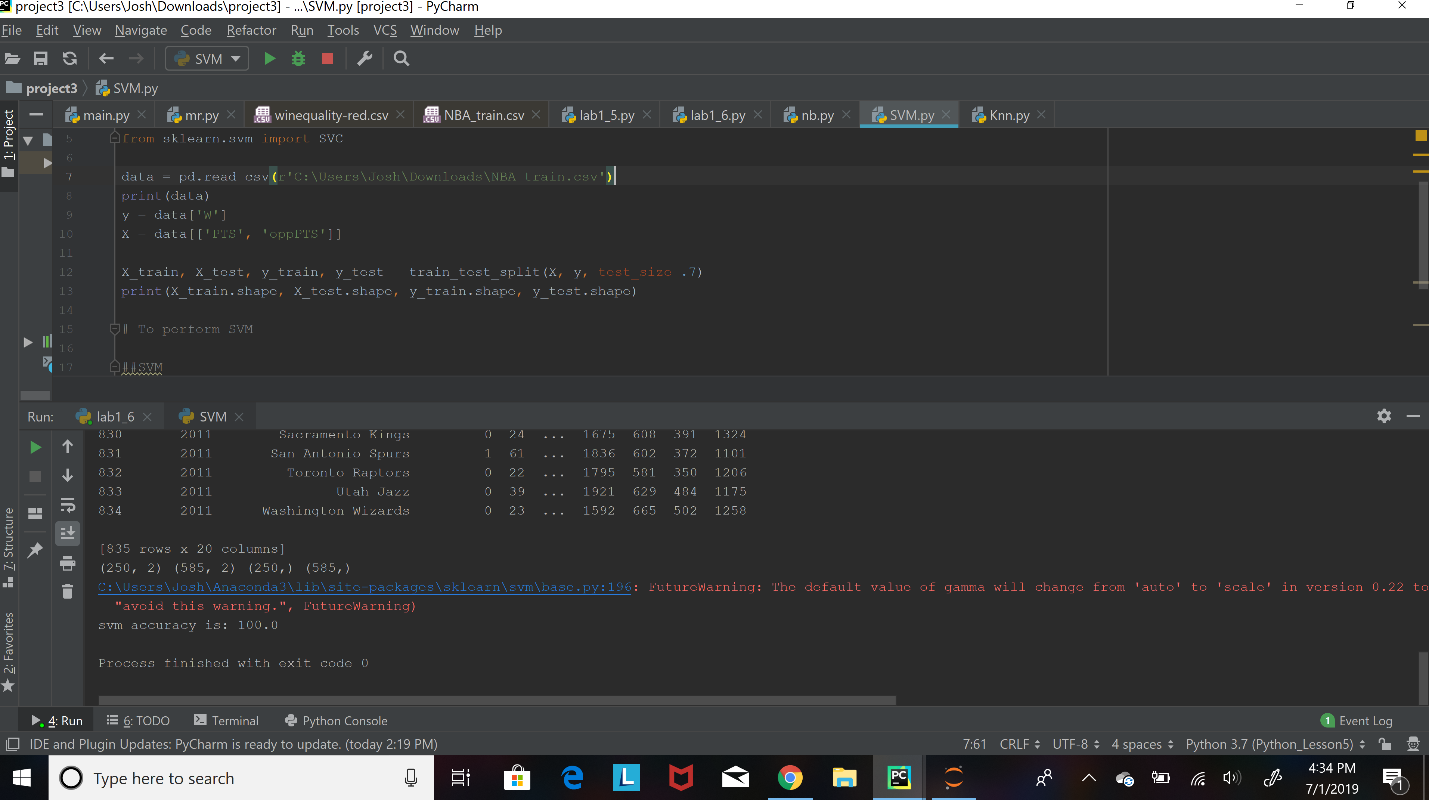
R2;- it tell us how well our selected independent variable(s) explain the variability in your dependent variable(s) and r-squared value is a measure of how close the data are to the fitted regression line

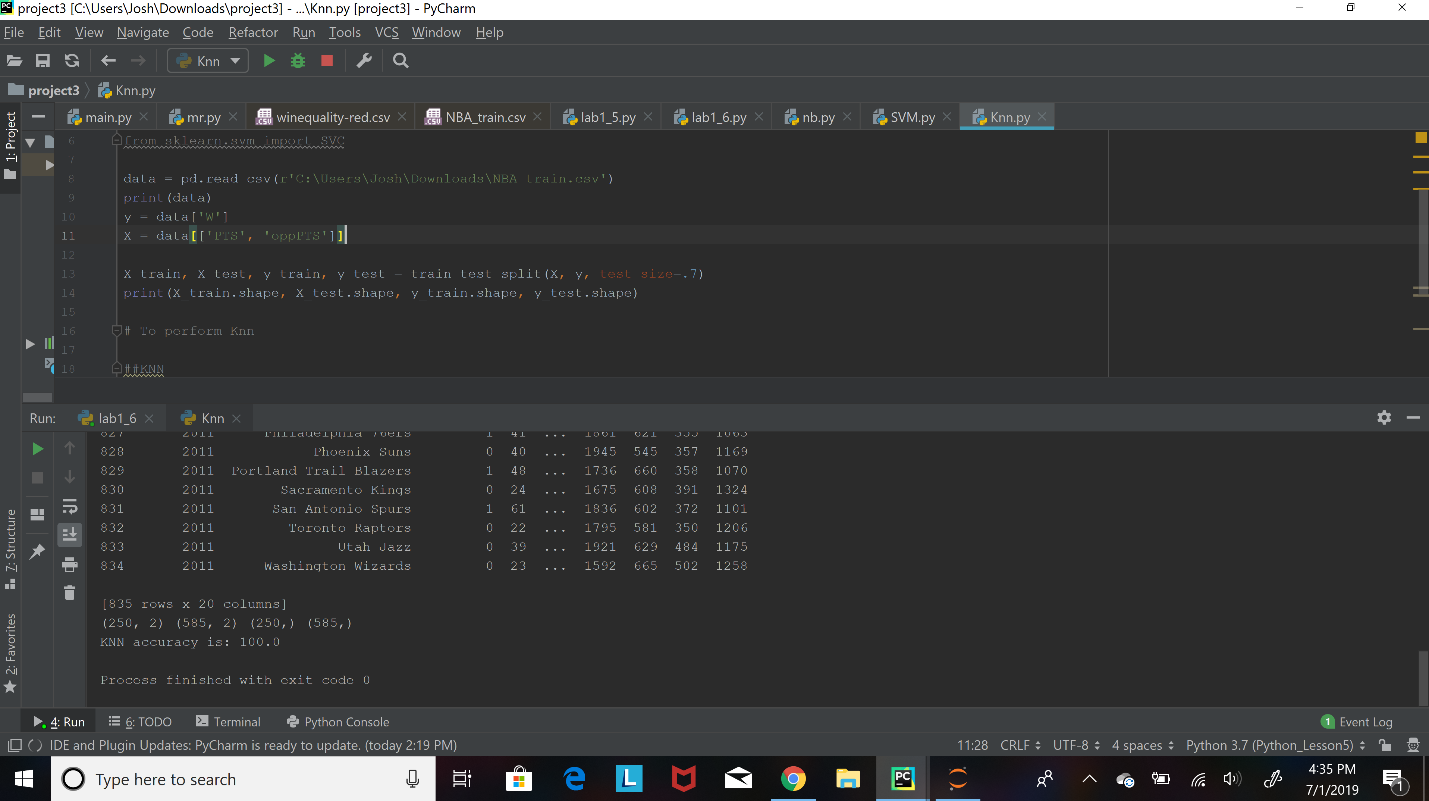
RMSE ;- it represent the standard deviation of difference between the predicted value and observe value.



5. a. i perform exploratory data analysis on the data set (like Handling null values, removing the features not correlated to the target class, encoding the categorical features, ...)



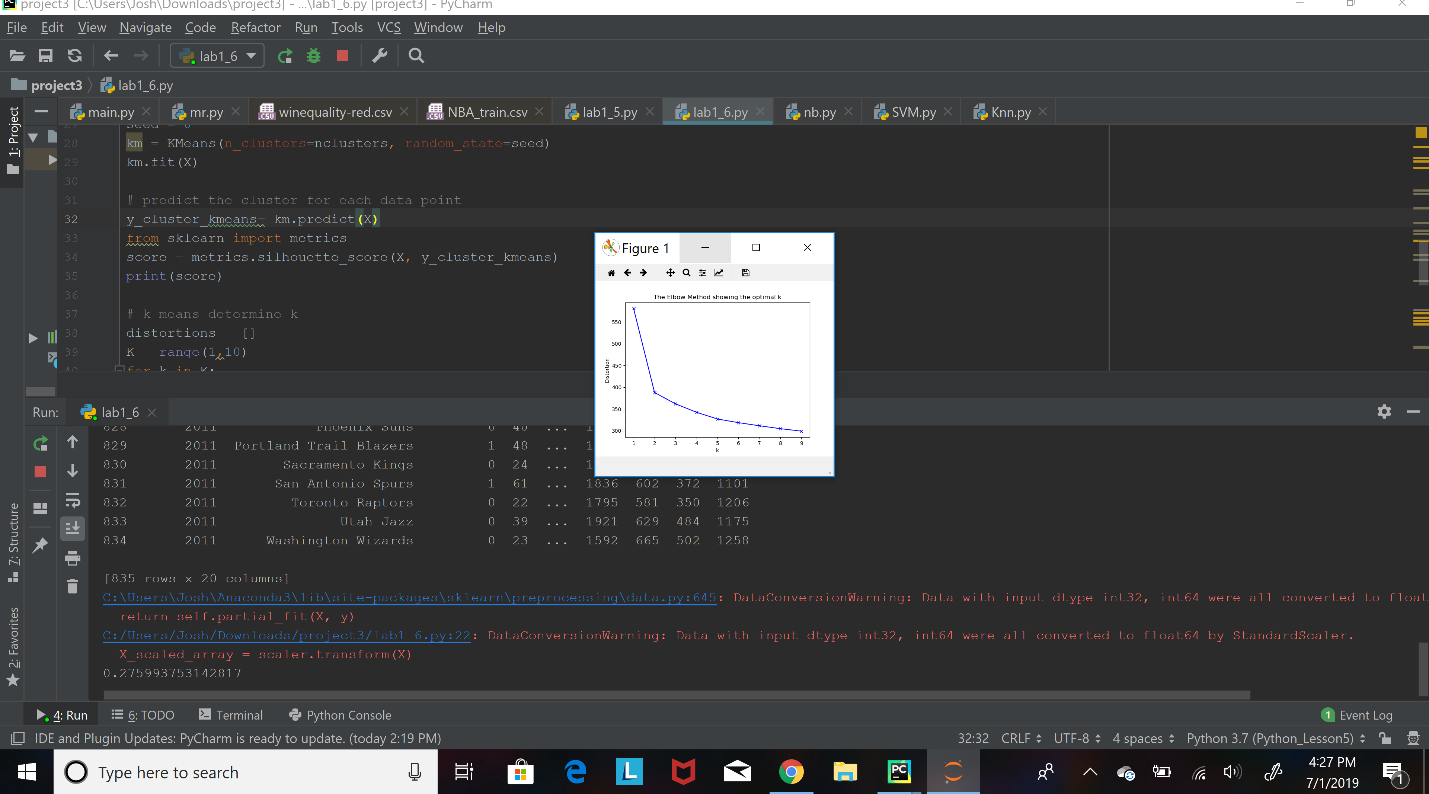




5b, I apply the three classification algorithms Naïve Baye’s, SVM and KNN on the chosen data set and which classifier gives better result, since I got the same score which 100.00 for the three classification algorithm that show all the three give better result.

6a, Choose any dataset of your choice. I apply K-means on the dataset and visualize the clusters using matplotlib or seaborn.

1. We use the elbow method to know the number of cluster



6b, l evaluate with silhouette score which is shown on the screenshot above.

**Reference;-**

**W3resources.**