**Data Mining**

**Assignment 2**

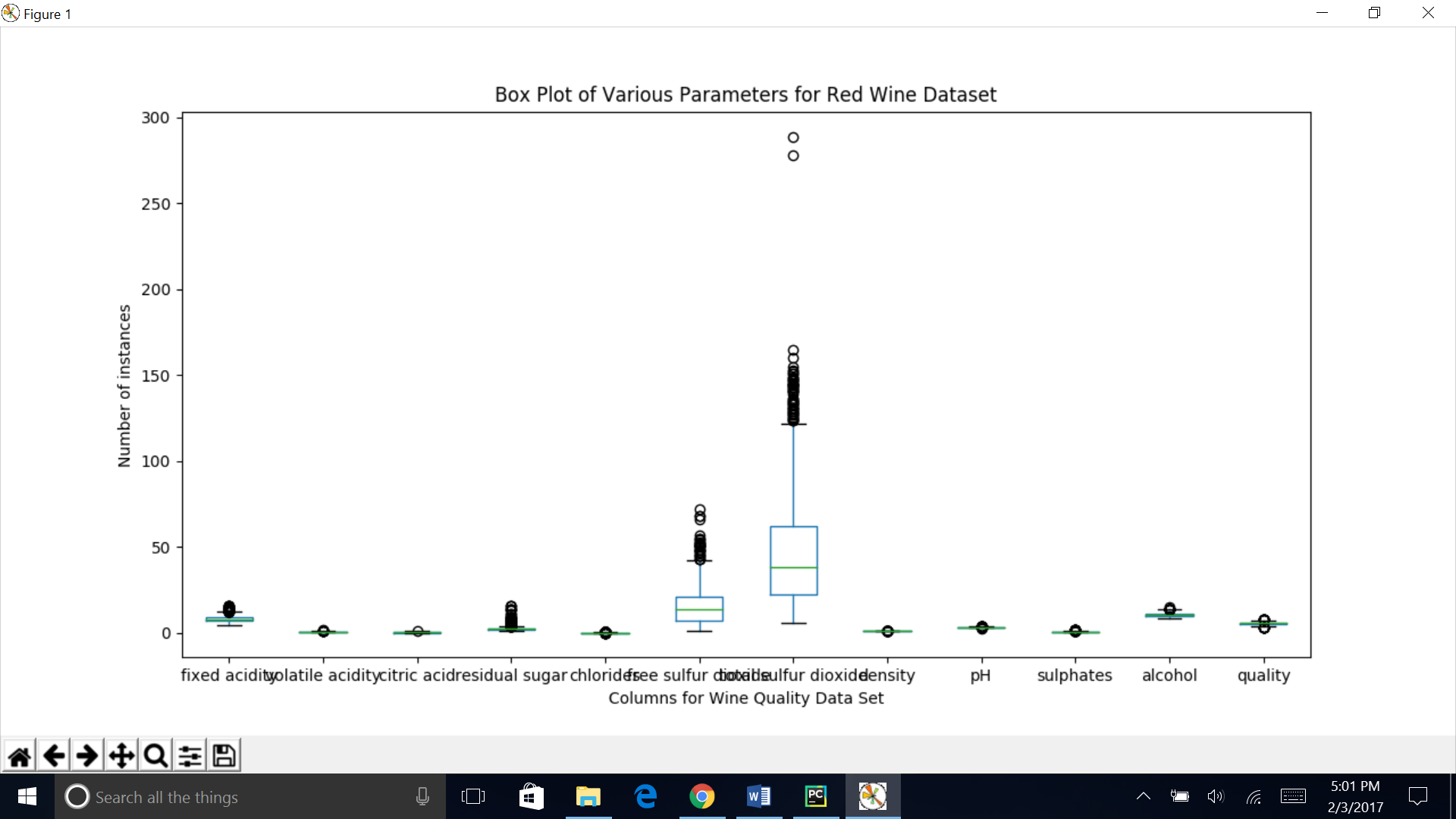
**Name : Venkatesh Suvarna**

**Student ID : 1001390986**

**Python Code :**

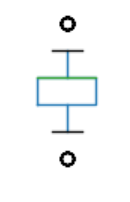
*#Necessary imports***import** pandas **as** pd  
**import** matplotlib.pyplot **as** plt  
  
*# Reading data locally*df = pd.read\_csv(**'/Users/Venkatesh Suvarna/PycharmProjects/DataMining\_Assignment2/winequality-red.csv'**,sep=**';'**)  
  
*#plotting boxplot*ax = df.plot(title = **'Box Plot of Various Parameters for Red Wine Dataset'**, kind = **'box'**)  
ax.set\_xlabel(**'Columns for Wine Quality Data Set'**)  
ax.set\_ylabel(**'Number of instances'**)  
plt.show(ax)

**Boxplot Output :**



**Analysis:**

Each of the boxplot represents analysis of each of the columns present in the Red – Wine dataset. For each of the boxplot, the lower blue edge represent the 25% quartile (Q1), the upper blue edge represent the 75% quartile(Q3), the green line represents the mean of the column values, the lower black line represents the lowest value of the column, the higher black line represents the highest value of the column. The main output column is the column quality with the screenshot below :



The above boxplot denotes that the lowest value of the column is 2, the highest value of the column is 8, the lower quartile is 4 and the upper quartile is 6, the mean is the same as quartile as the green and the blue line are overlapping. The dots are the outliers – with values of around 1 and 9.