#700743770

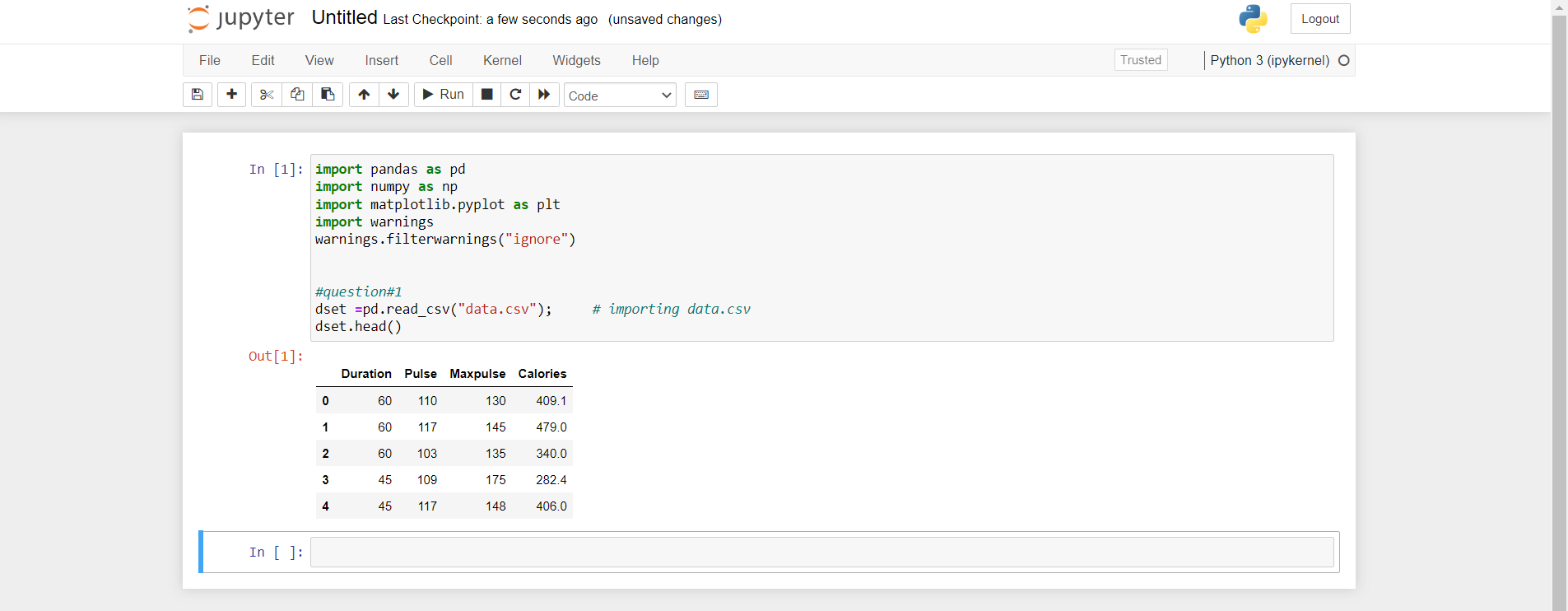
Sravanti Cherukuri Assignment #4

**Git hub link :** <https://github.com/sxc37701/ML_Assignments>

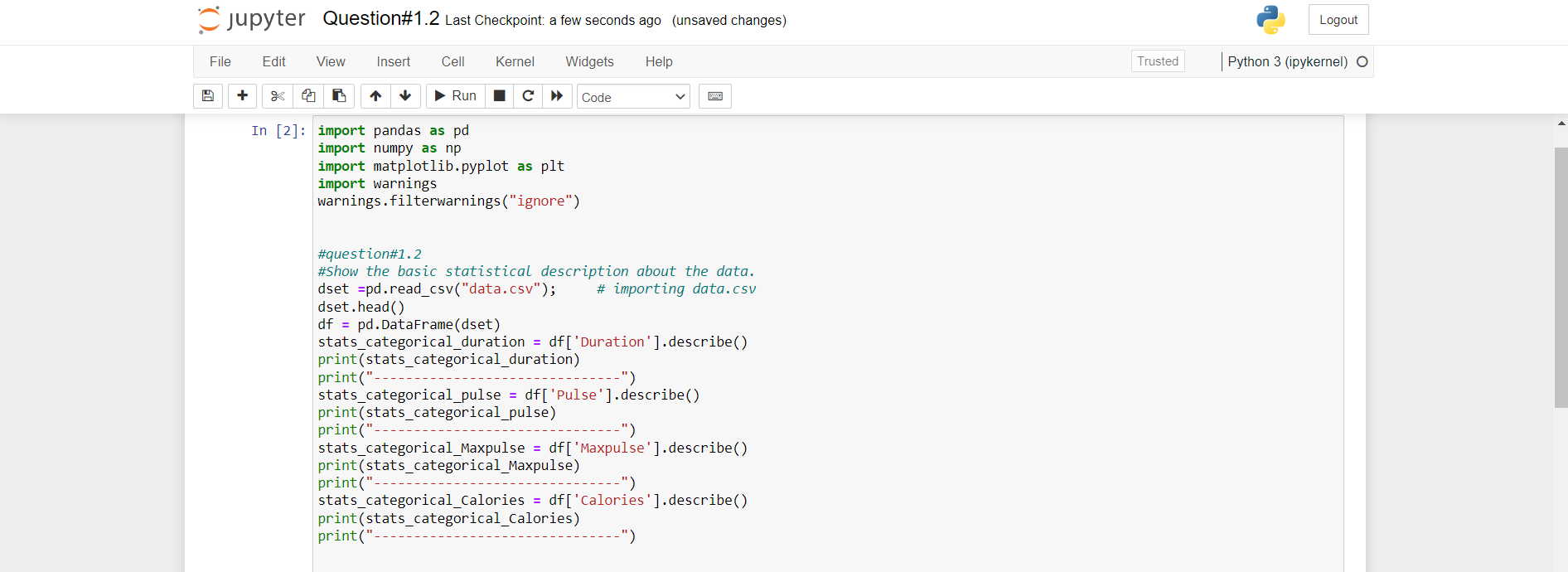
**Video Link:**https://drive.google.com/drive/folders/1CtEjxD967psXs6P3PzpRjSIn1j51gX8p

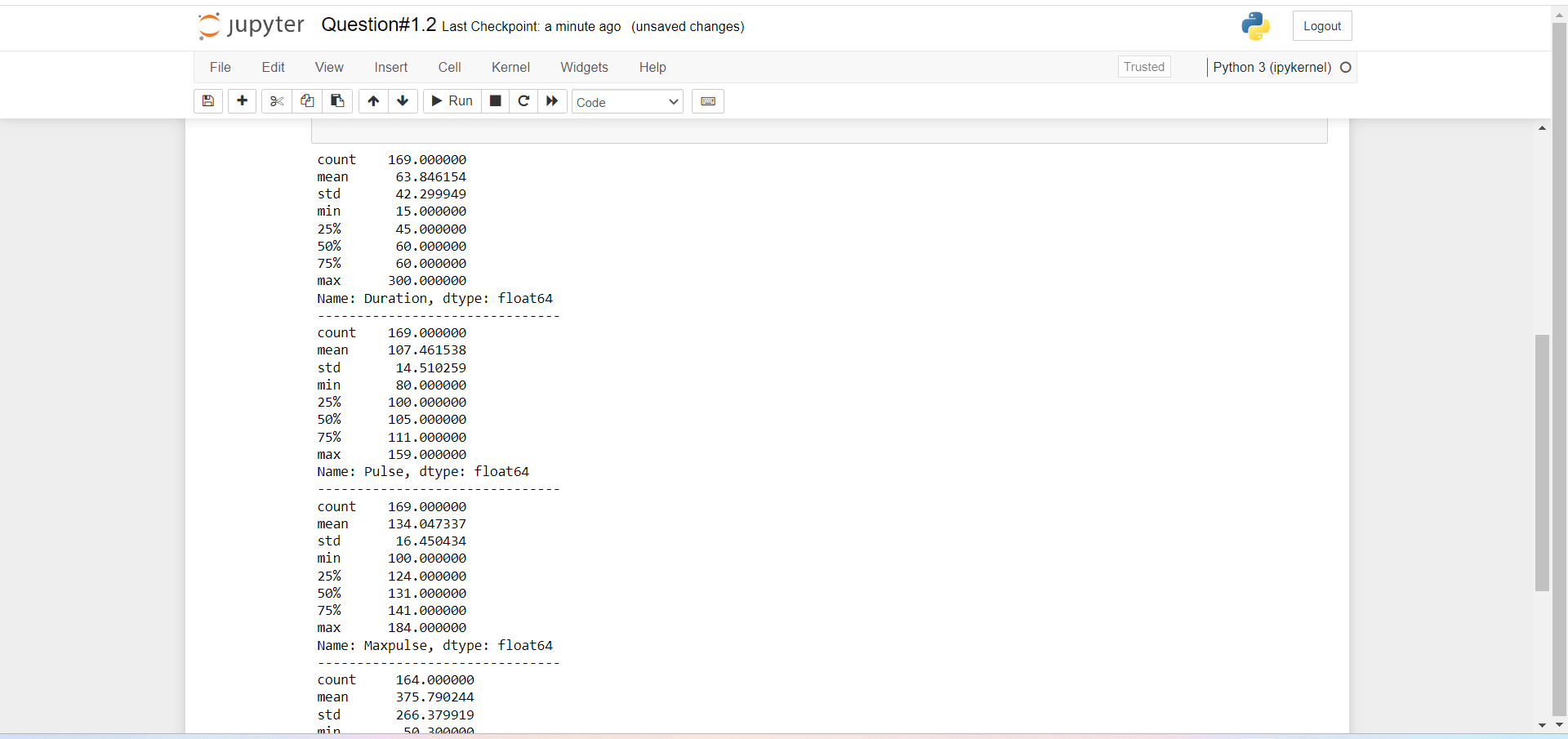
**1.Pandas**

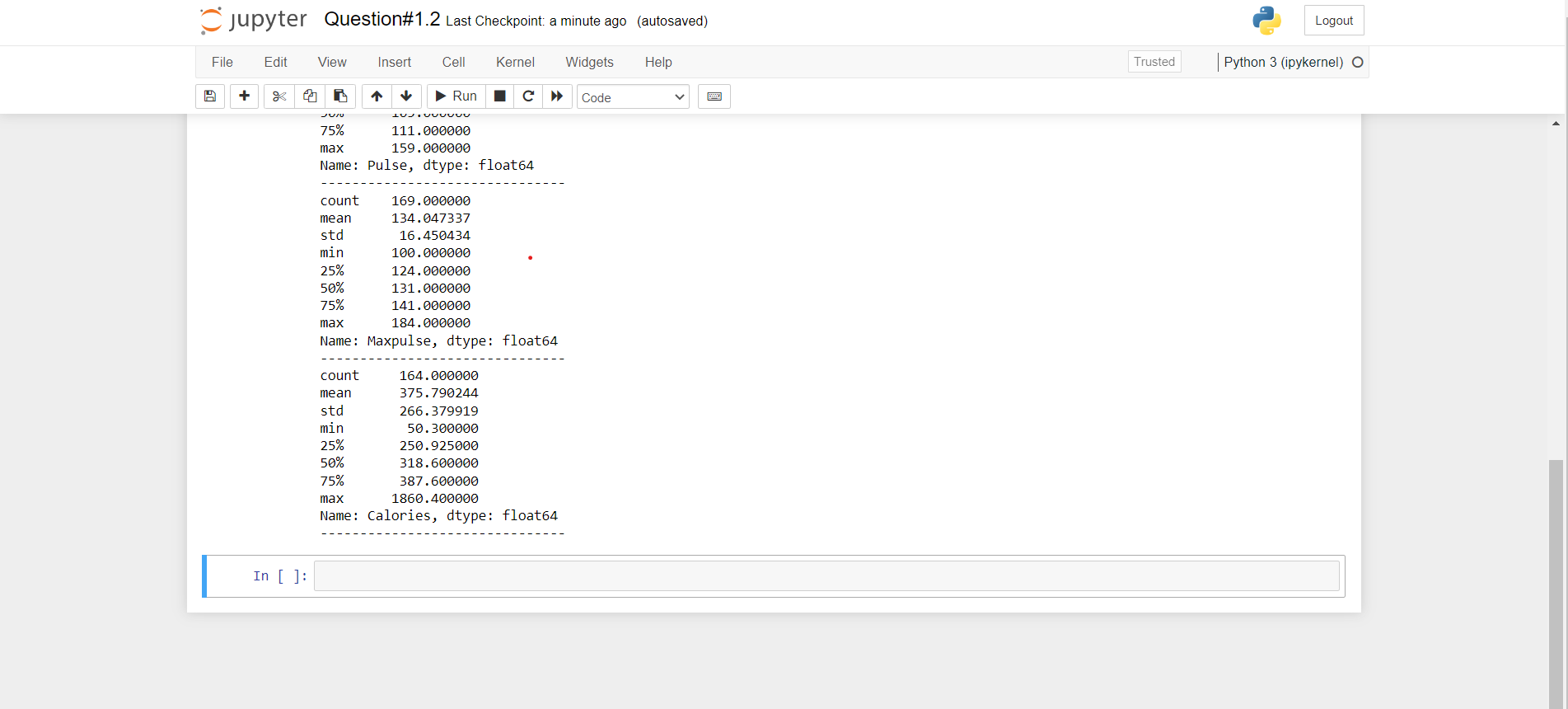
Question#1.1. Read the provided CSV file ‘data.csv’.



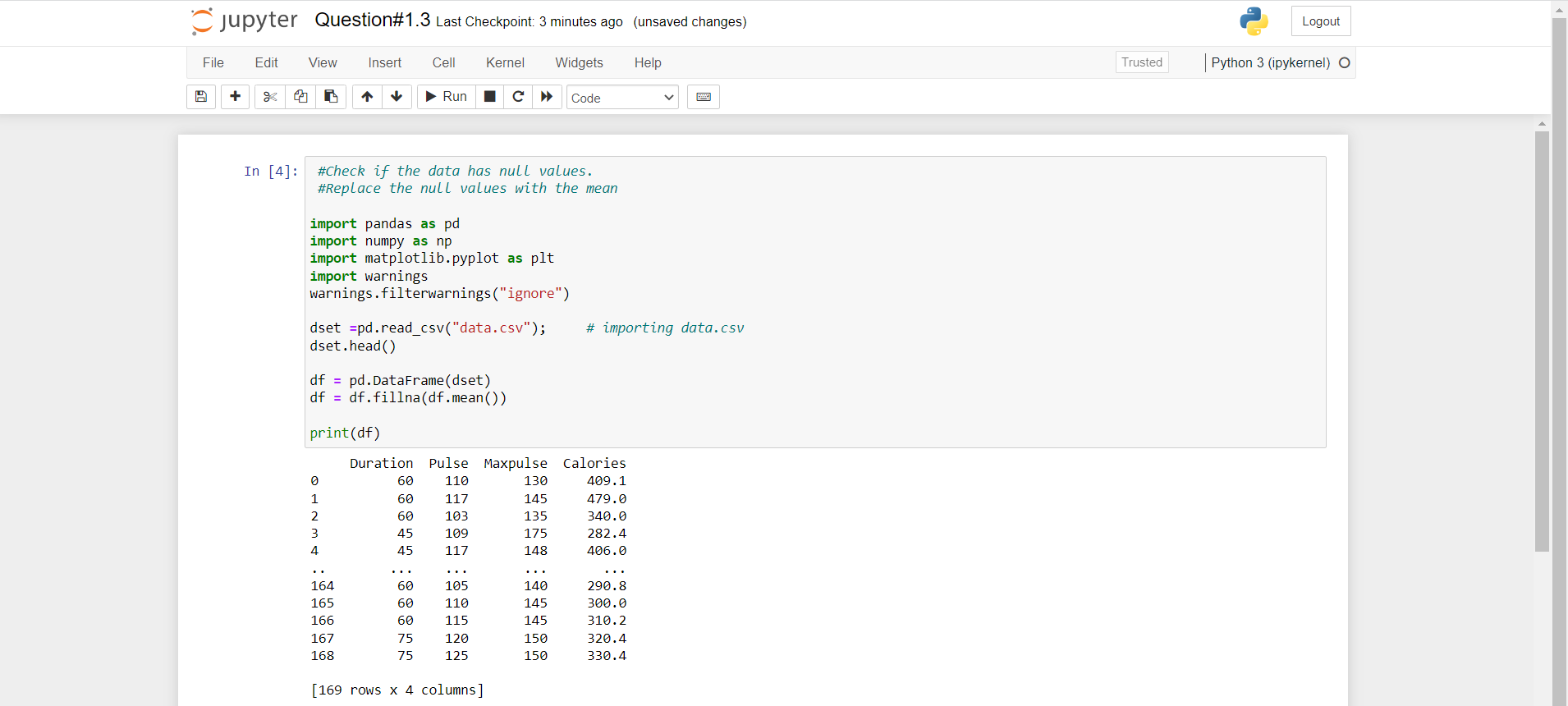
Question#1.2. Show the basic statistical description about the data.



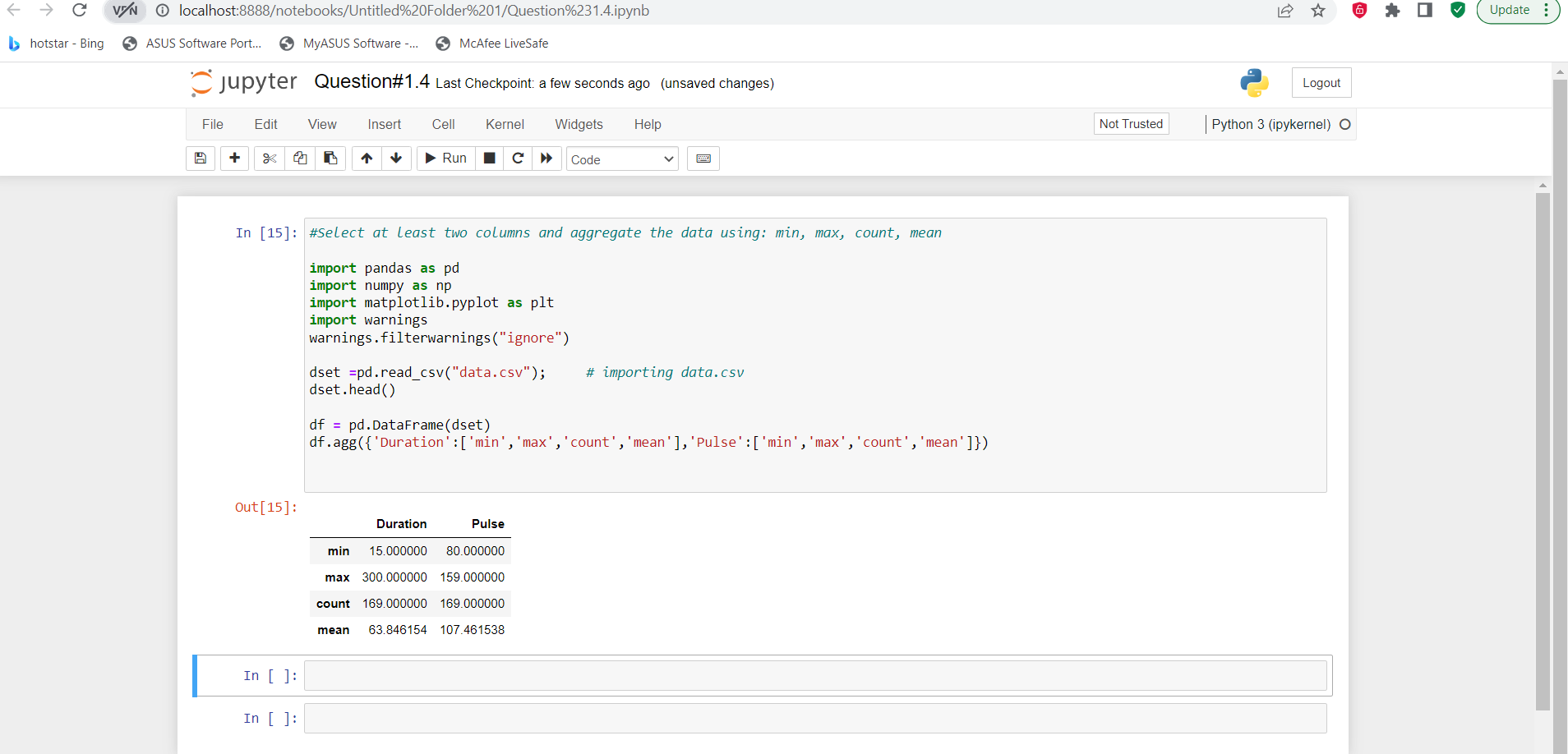




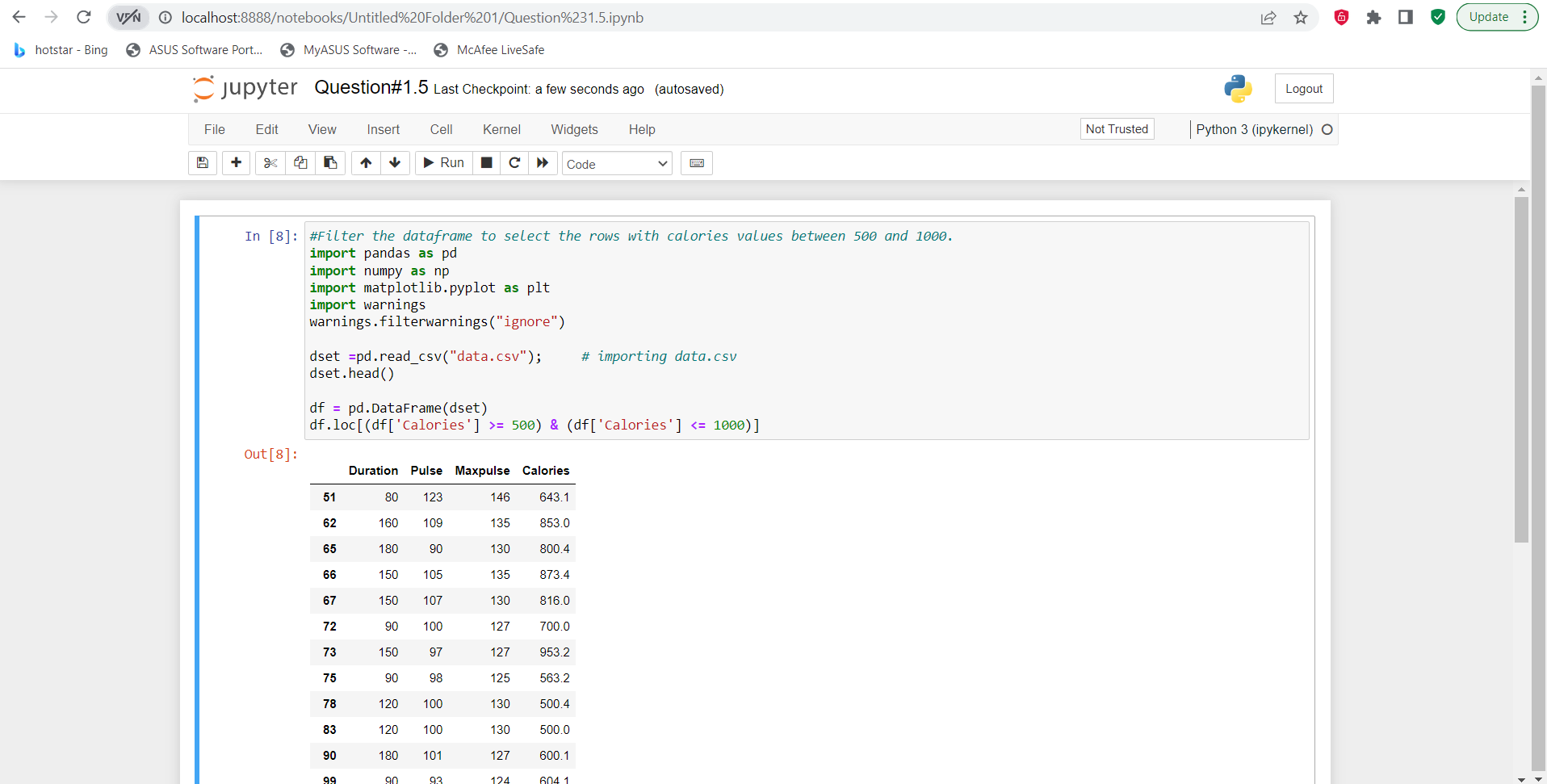
Question#1.3. Check if the data has null values.

1. Replace the null values with the mean

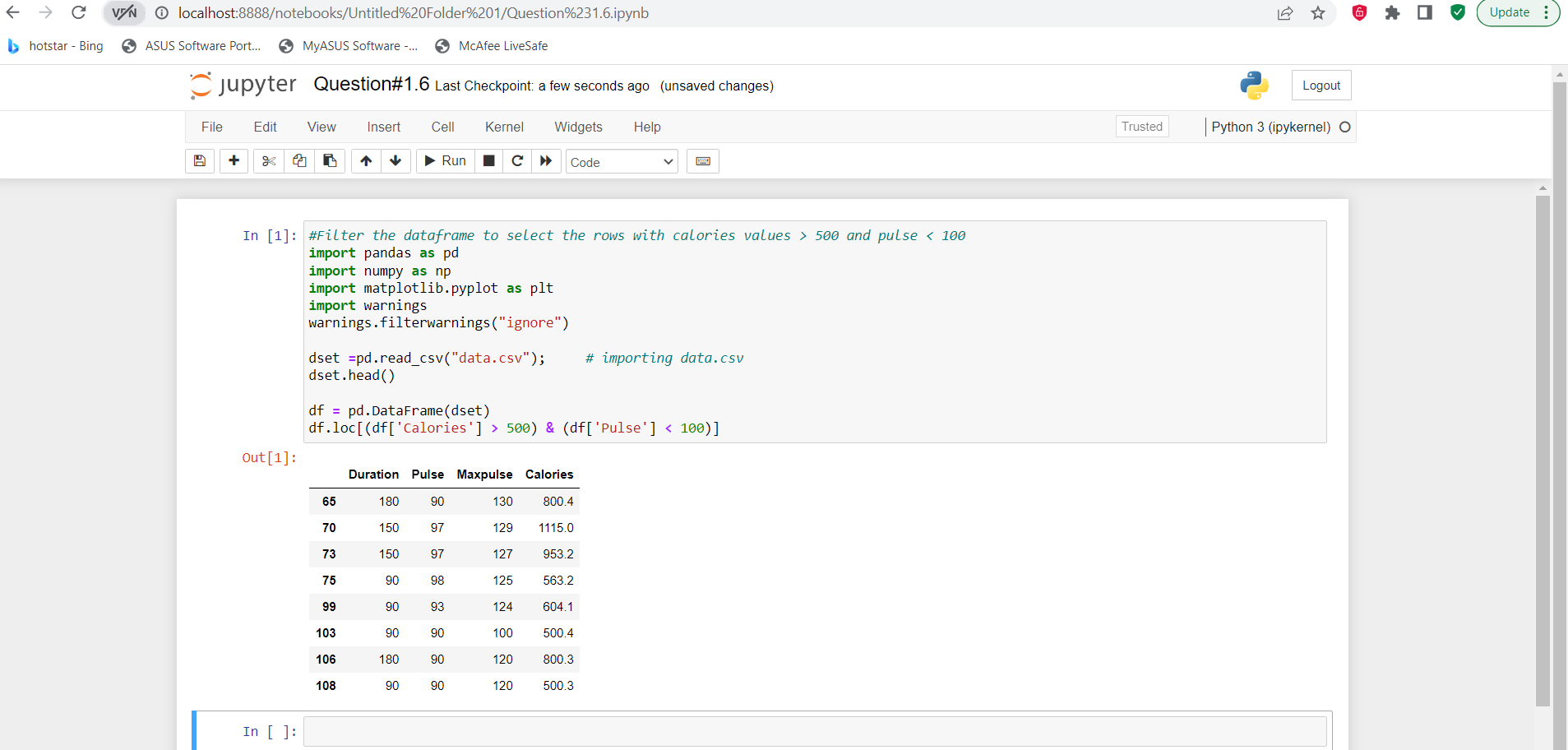
Question#1.4. Select at least two columns and aggregate the data using: min, max, count, mean.



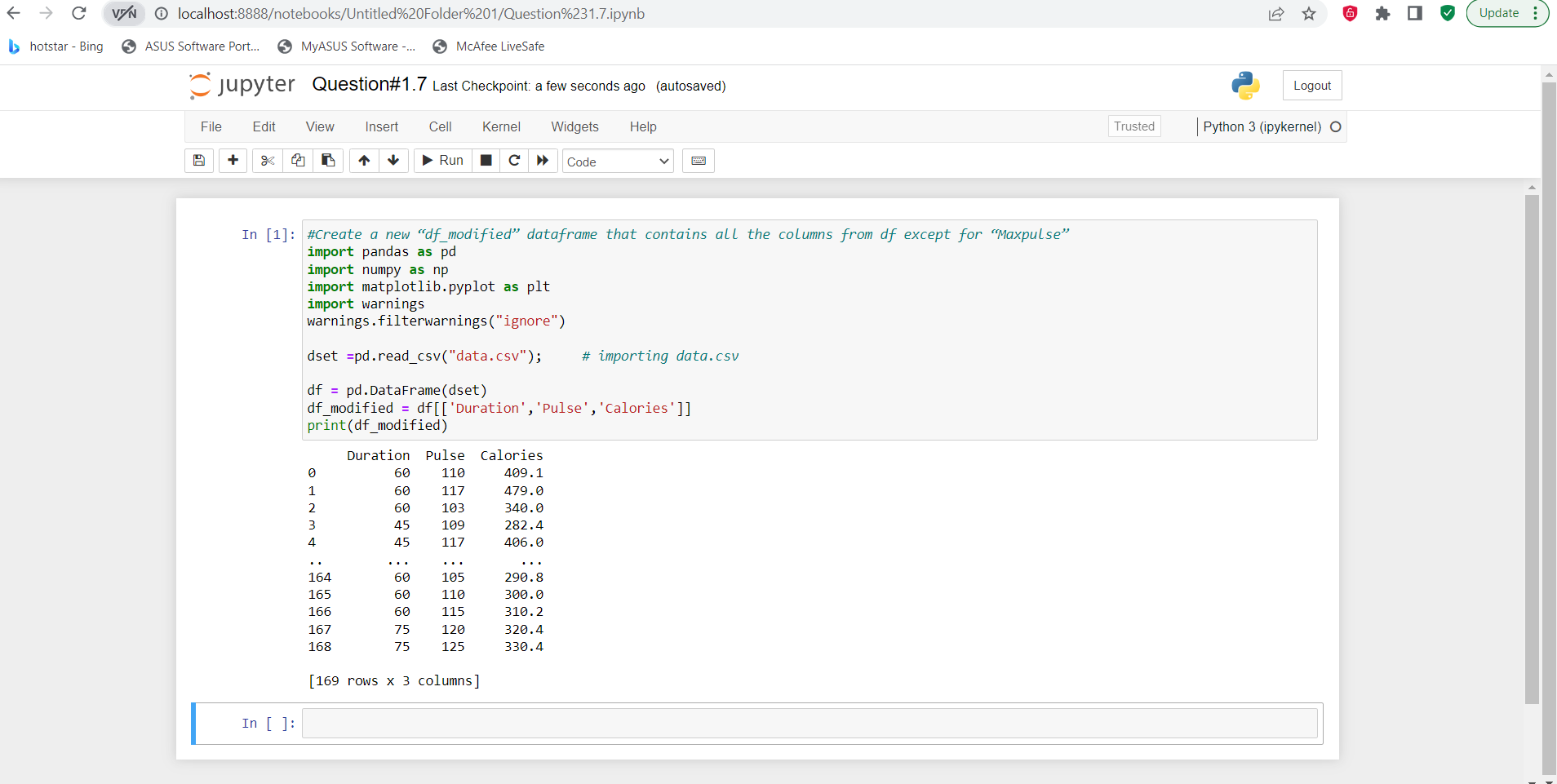
Question#5: Filter the data frame to select the rows with calories values between 500 and 1000.



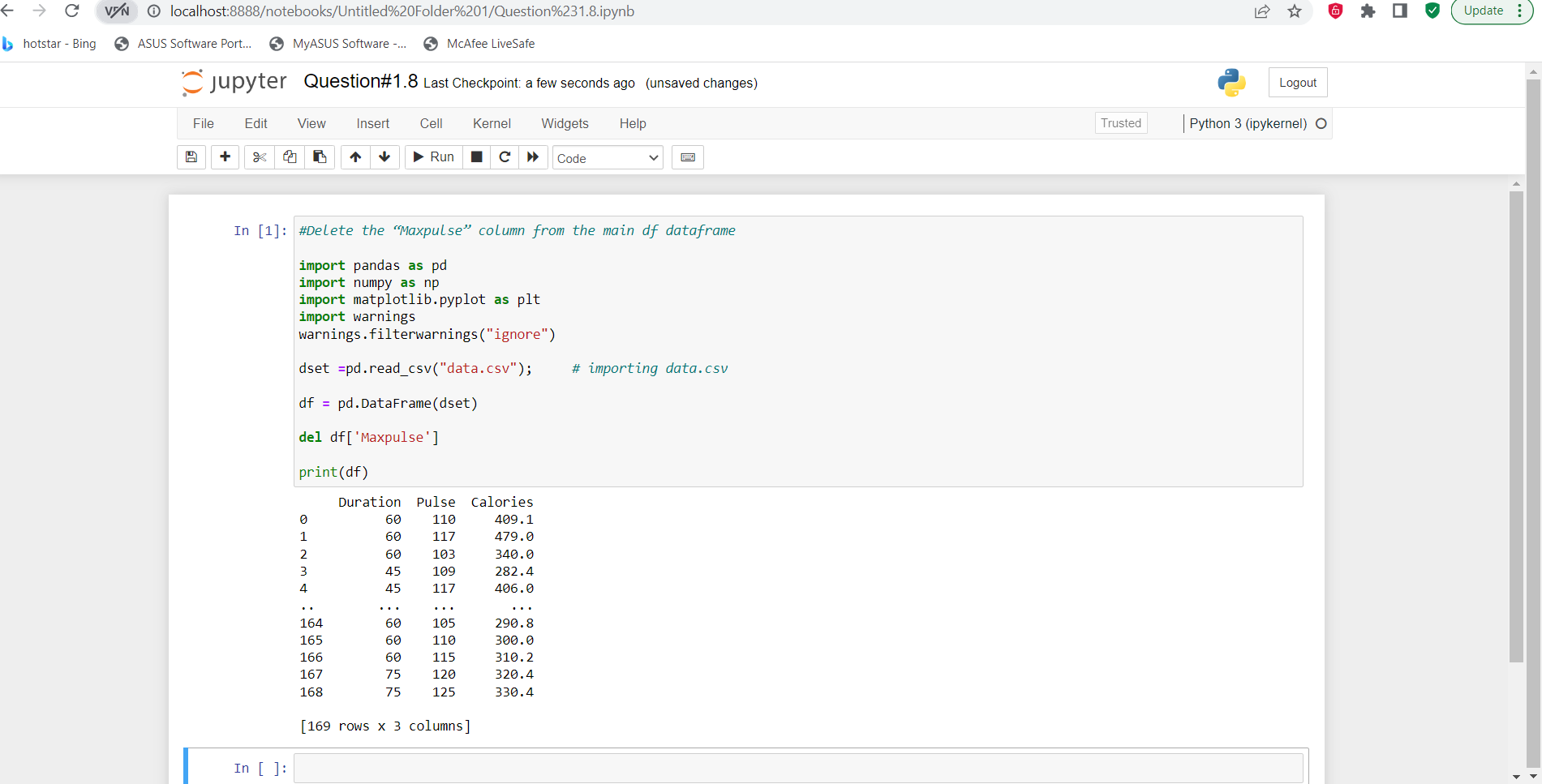
Question#6: Filter the dataframe to select the rows with calories values > 500 and pulse < 100.



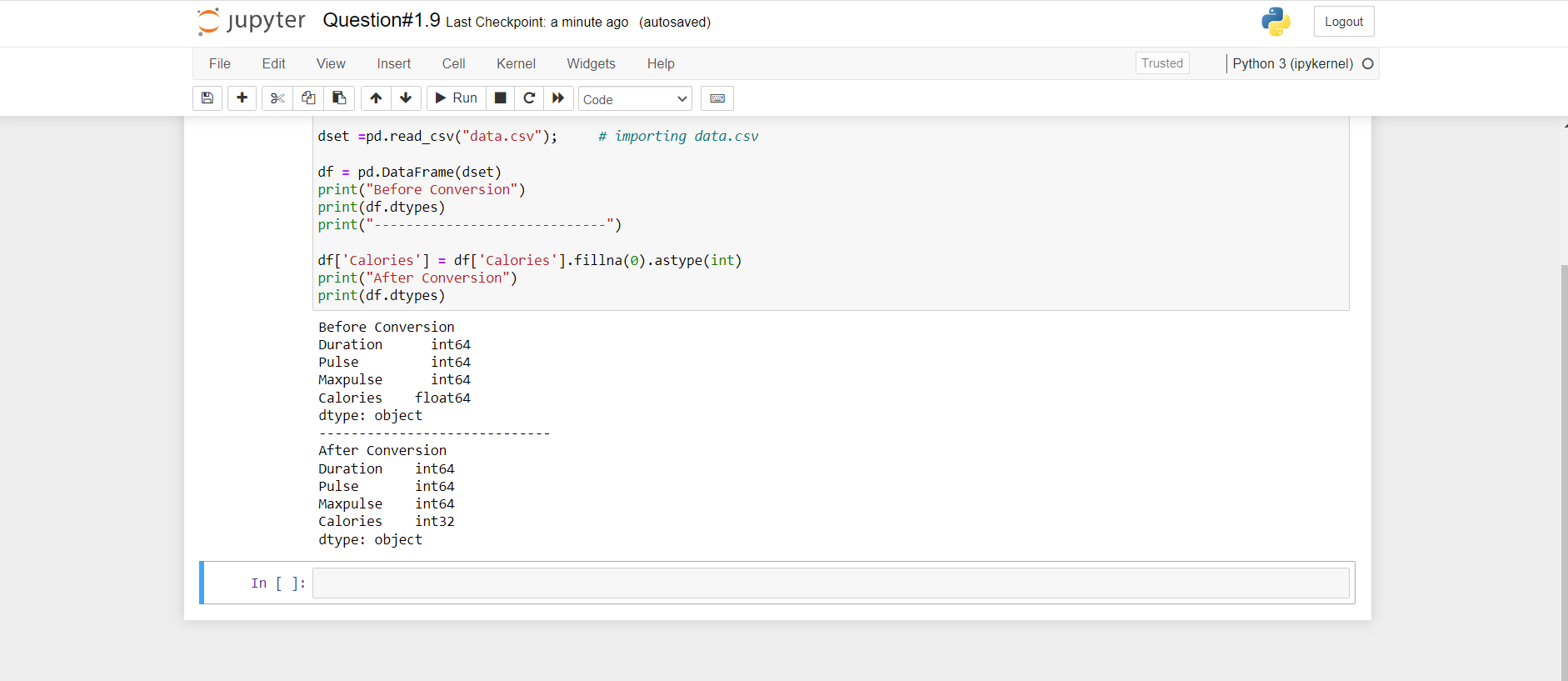
Question#7: Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.



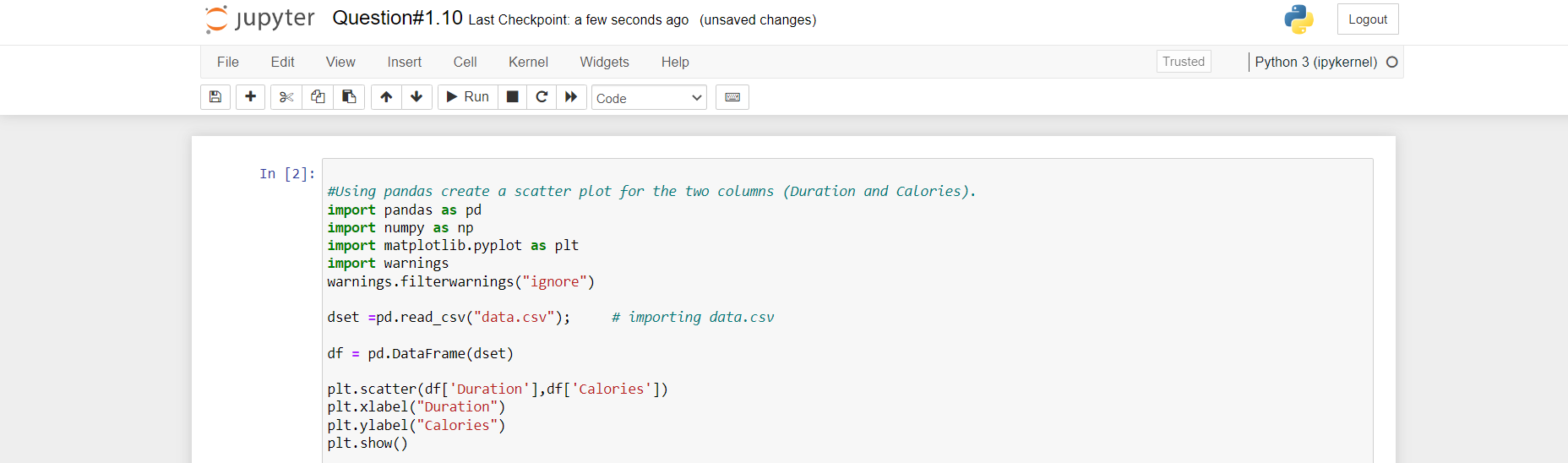
Question#8: Delete the “Maxpulse” column from the main df dataframe

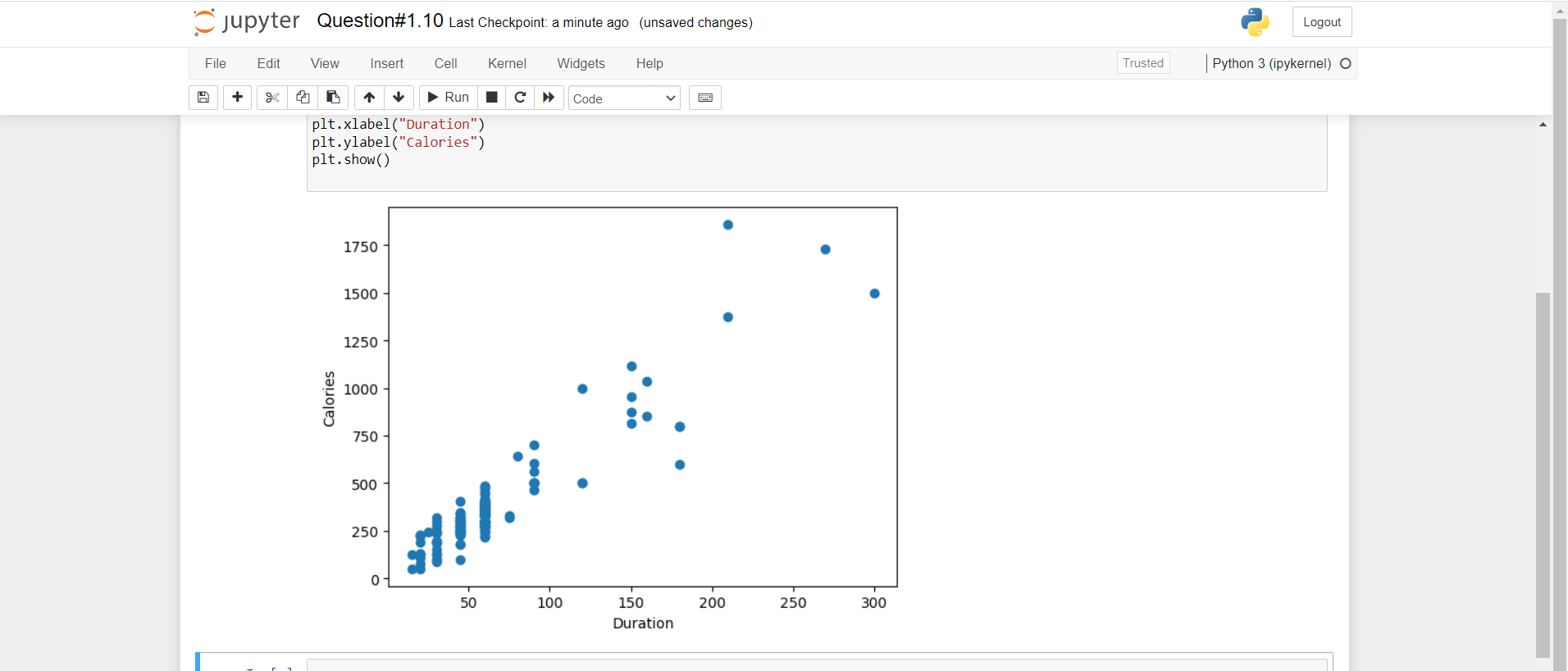


Question#9: Convert the datatype of Calories column to int datatype.



Question#10: Using pandas create a scatter plot for the two columns (Duration and Calories).





Question #2

1. (Titanic Dataset)

1. Find the correlation between ‘survived’ (target column) and ‘sex’ column for the Titanic use case in class.

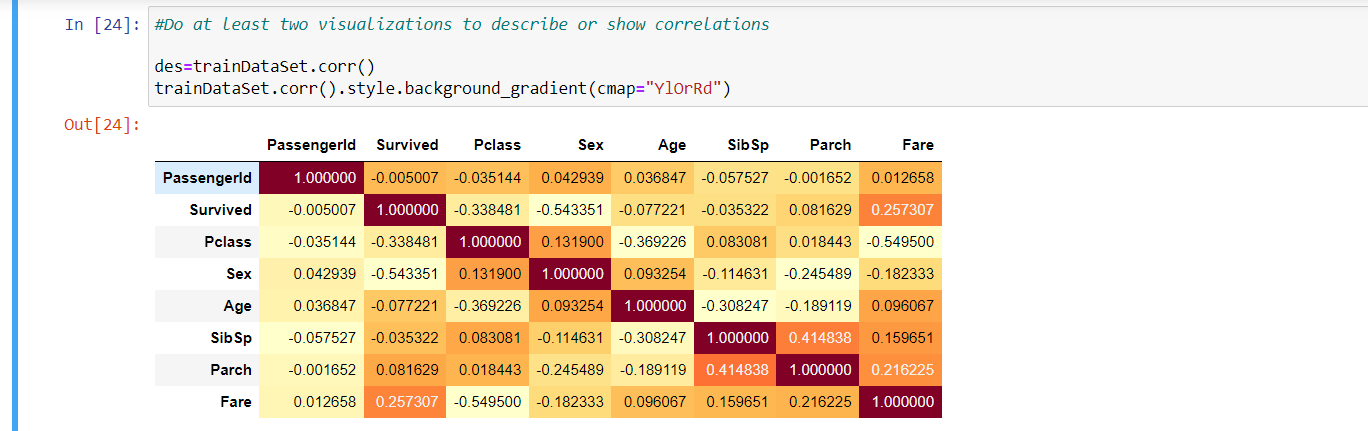
a. Do you think we should keep this feature?

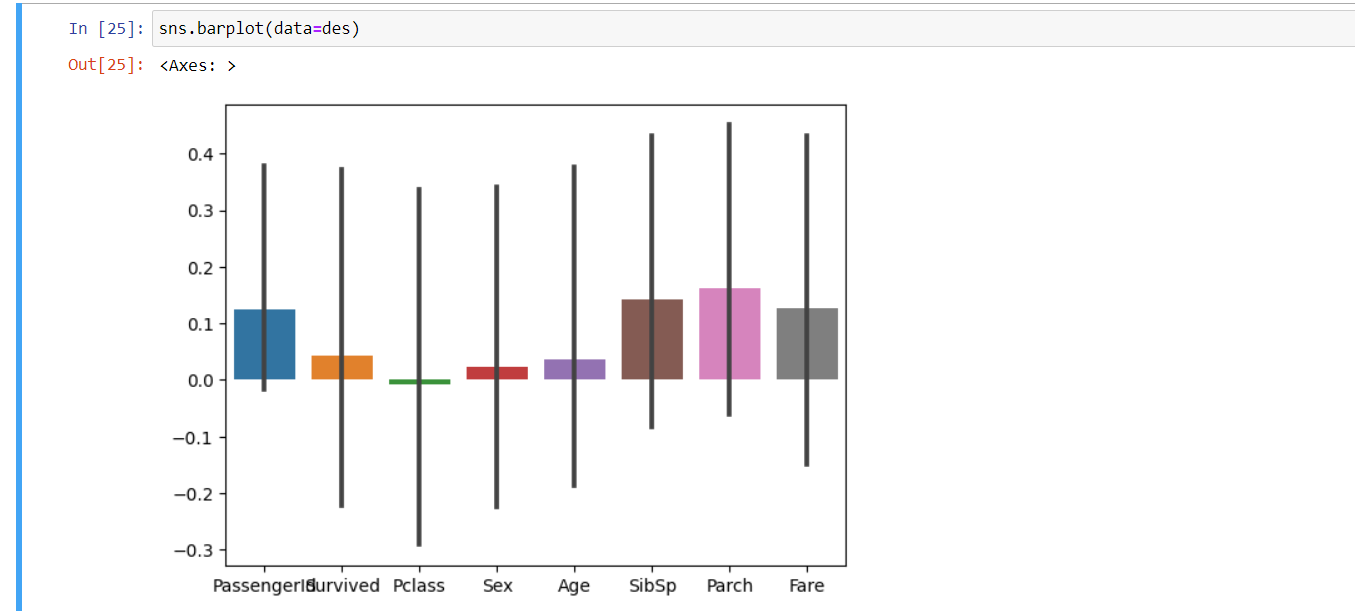


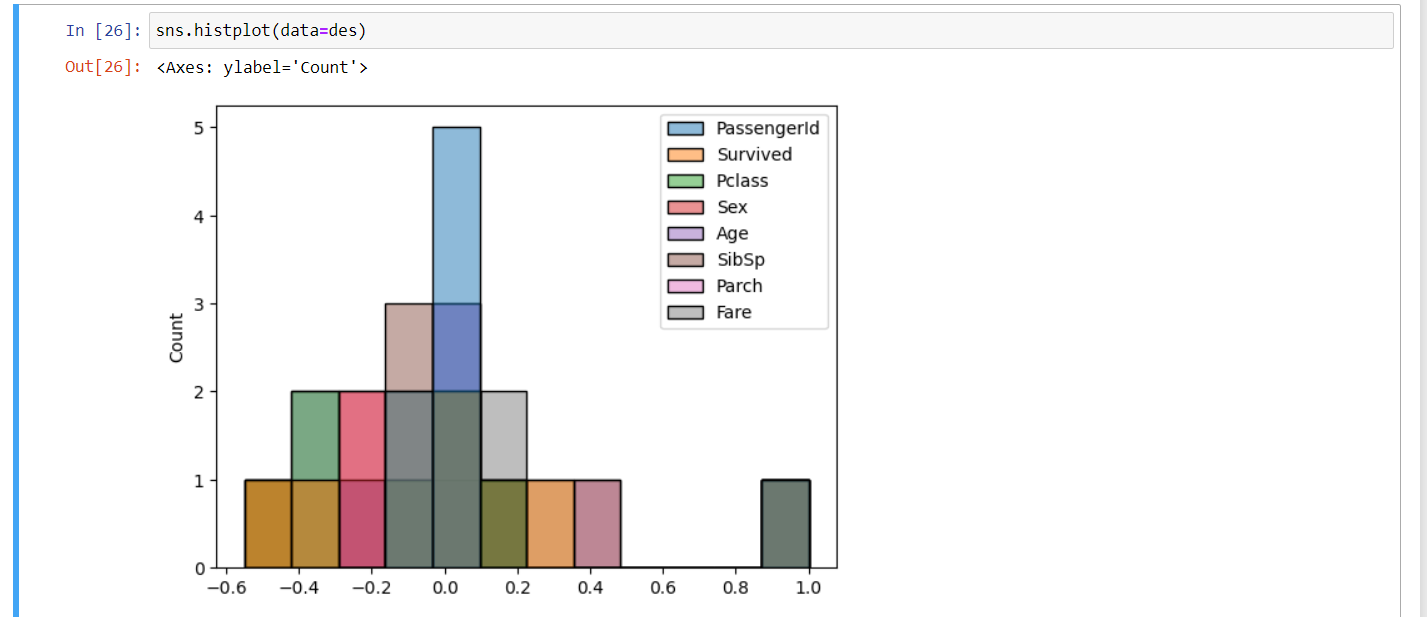
Correlation is a measure of the degree of linear association among a pair of variables.

Inverse Correlation occurs when the correlation coefficient is less than 0.This indicates that both variables move in opposite direction. So its **better to retain this feature** as it has -0.5 (Value should be between (-1 to 1)

2. Do at least two visualizations to describe or show correlations



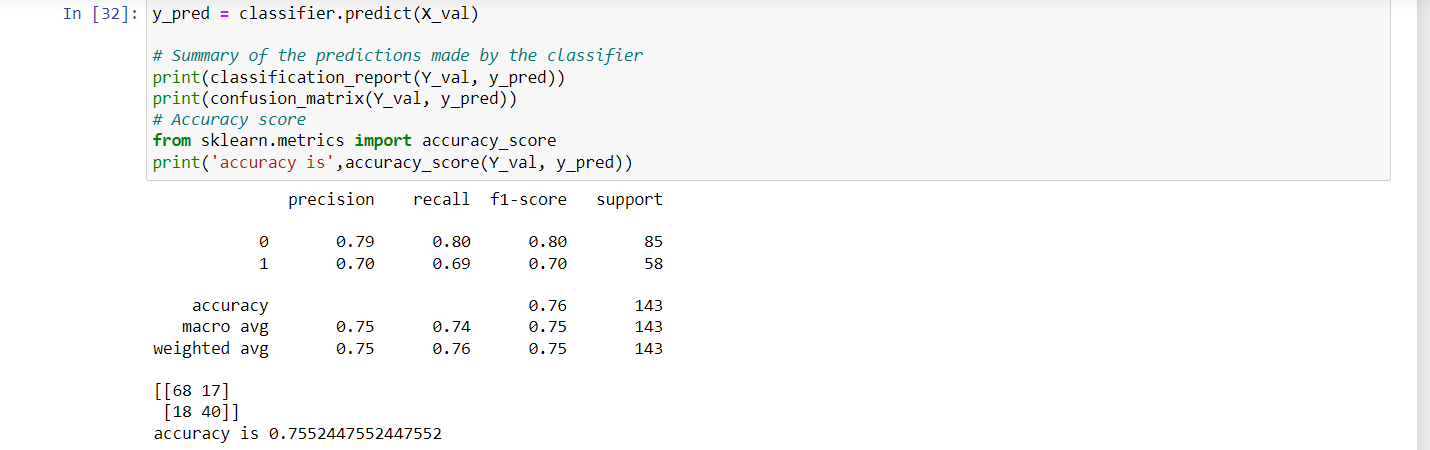




Question#3: Implement Naïve Bayes method using scikit-learn library and report the accuracy.

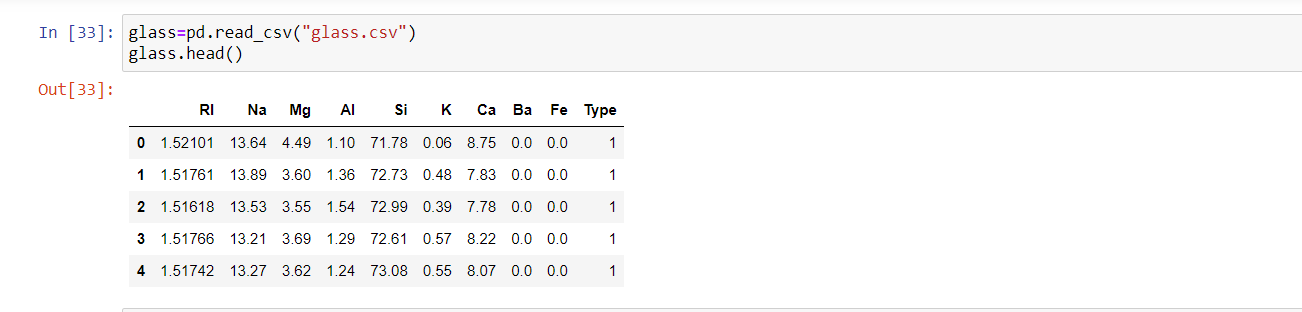


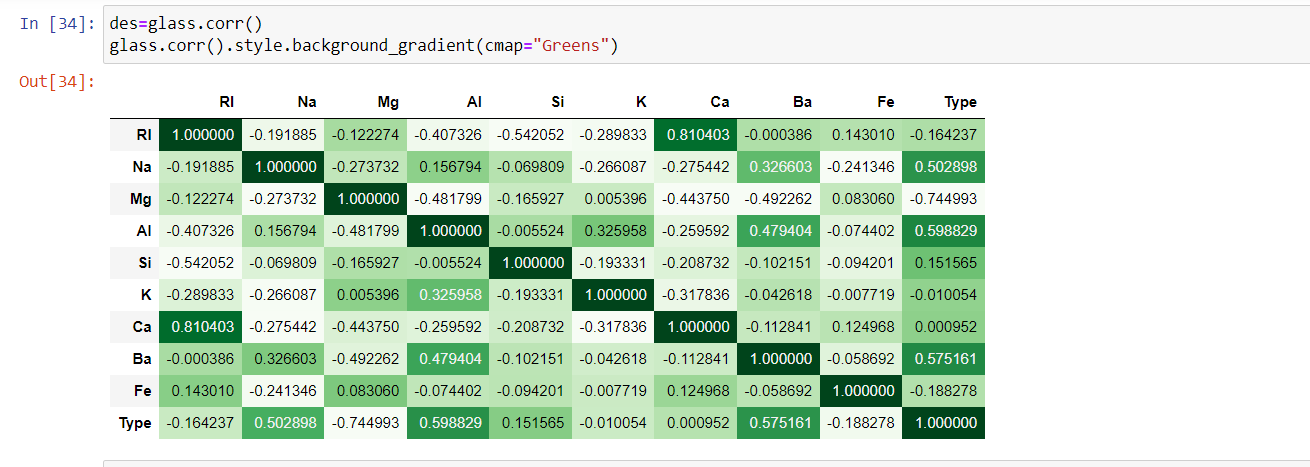




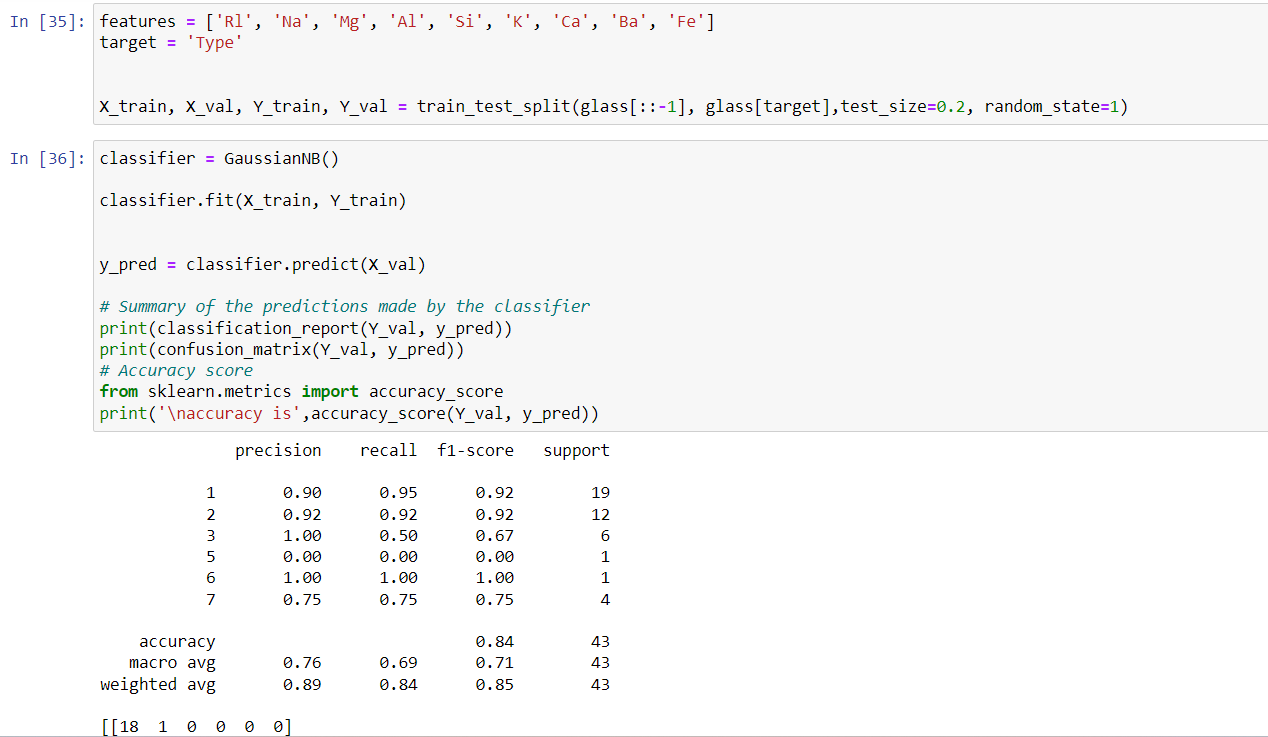
**2. (Glass Dataset)**

1. Implement Naïve Bayes method using scikit-learn library. a. Use the glass dataset available in Link also provided in your assignment.



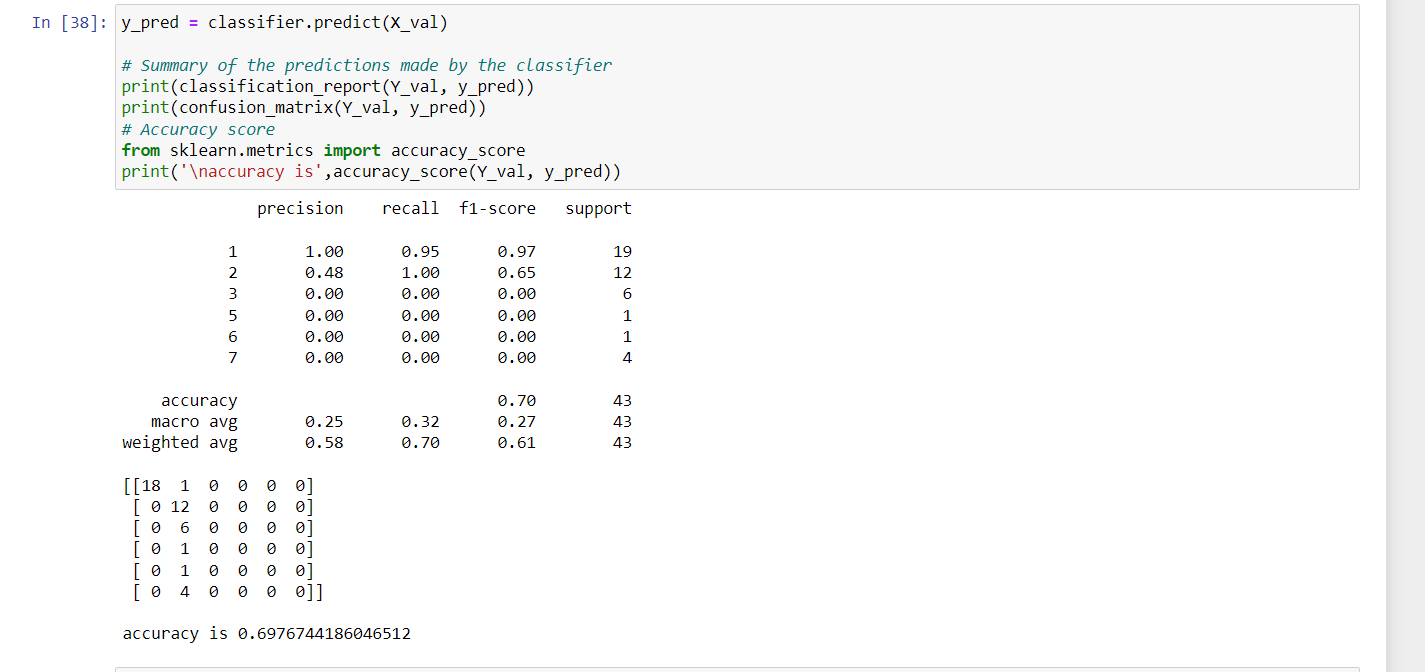


. b. Use train\_test\_split to create training and testing part.

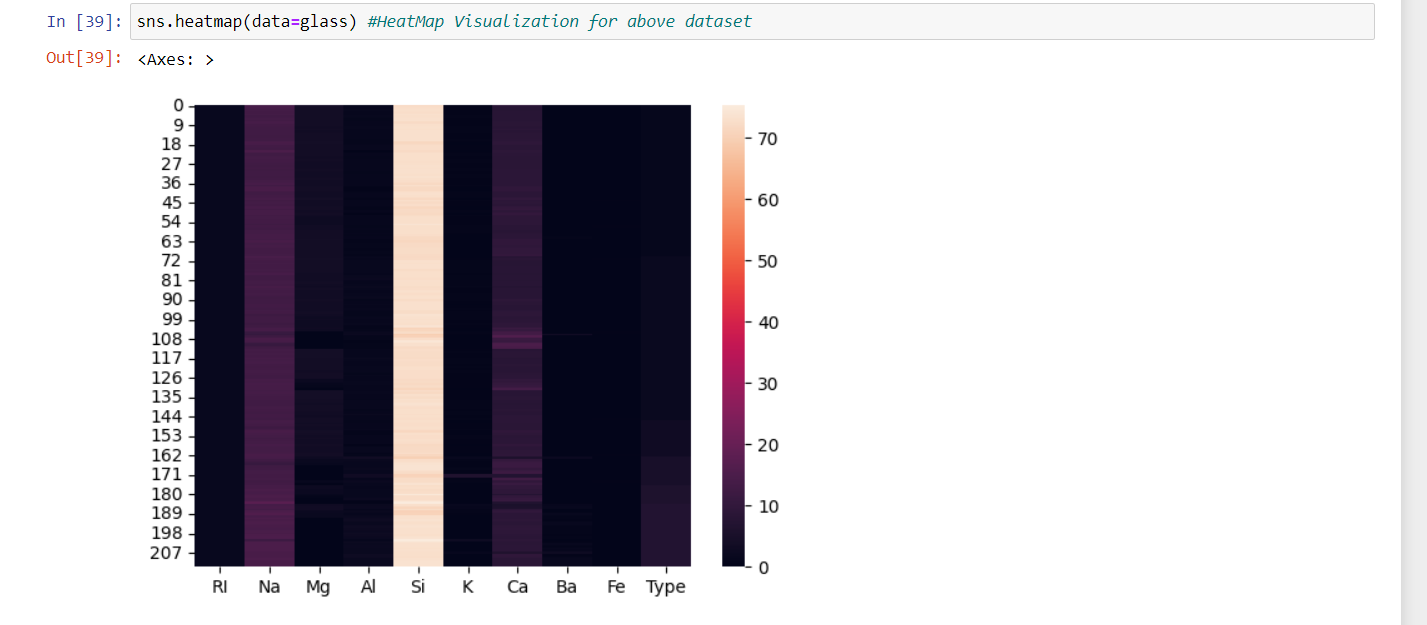


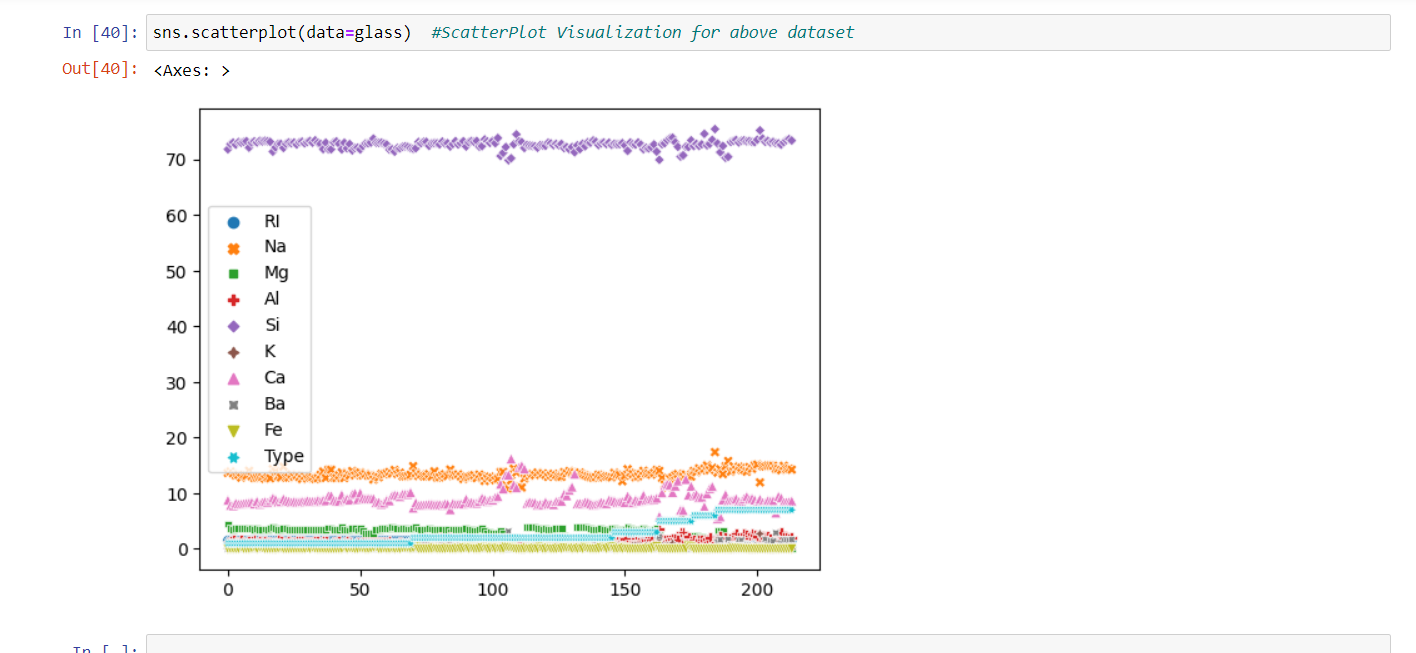
1. Implement linear SVM method using scikit library a. Use the glass dataset available in Link also provided in your assignment. b. Use train\_test\_split to create training and testing part.





Do at least two visualizations to describe or show correlations in the Glass Dataset.





Which algorithm you got better accuracy? Can you justify why?

The Naives Bayes classifier (0.837 = 83%)is more accurate than SVM(Support Vector Machine) (0.69 = 69%) according to above accuracy scores.

Each ML algorithm performs differently depending on number of variables.

We can evaluate different algorithms by evaluating the model and declare which one is the best one.