Machine Learning Assignme nt2

Name: Aditya Vardhan Nadupalli

Ucm Id: 700747773

Github source code link:

<https://github.com/AdityaVardhanNadupalli/MachineLearning_Assg_2.git>

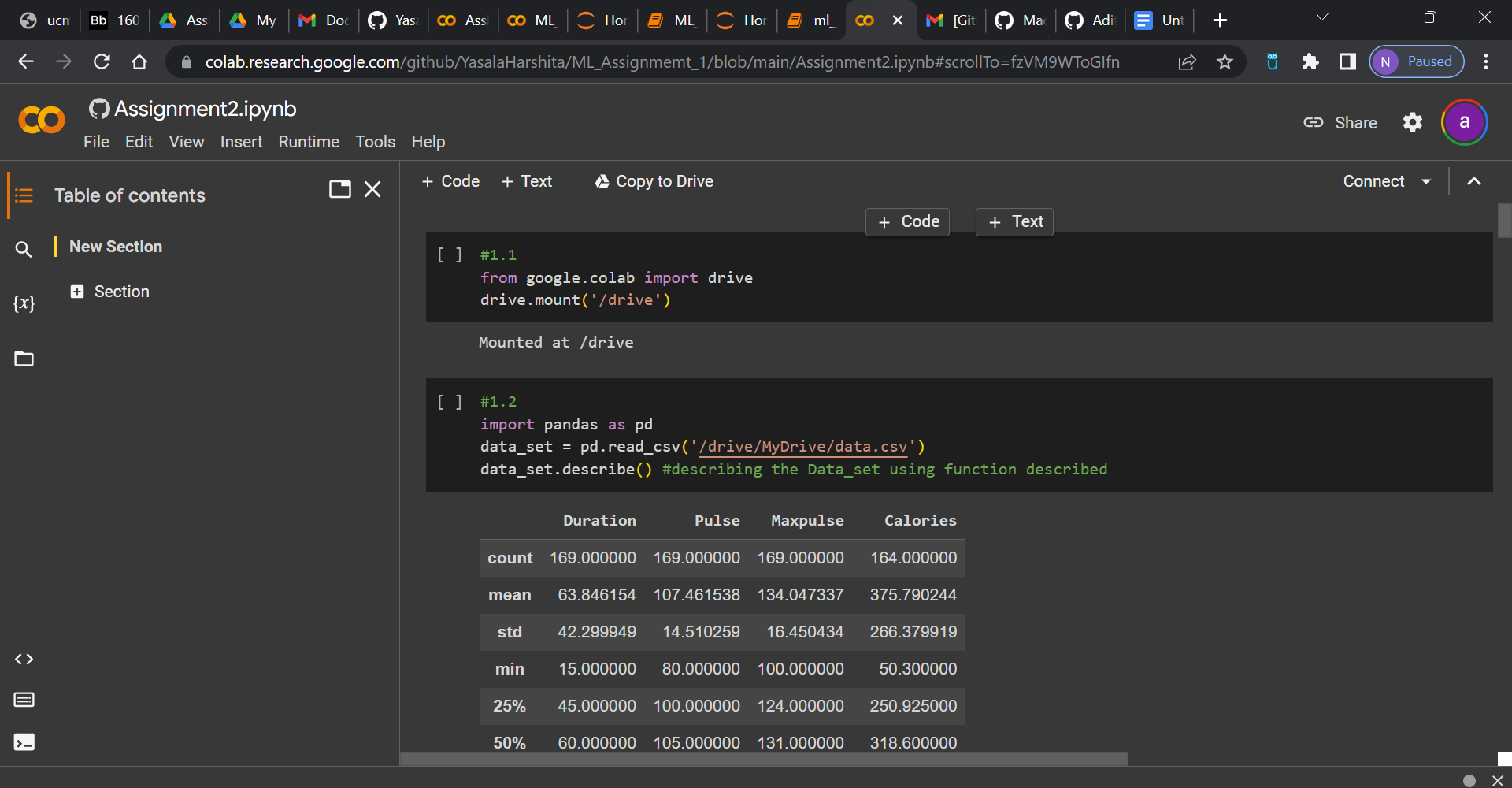
Drive link:

<https://drive.google.com/file/d/1OV34mf79E-1iRFCdtCT1n16iYaRKOXYl/view?usp=sharing>

1. Pandas

1. Read the provided CSV file ‘data.csv’.

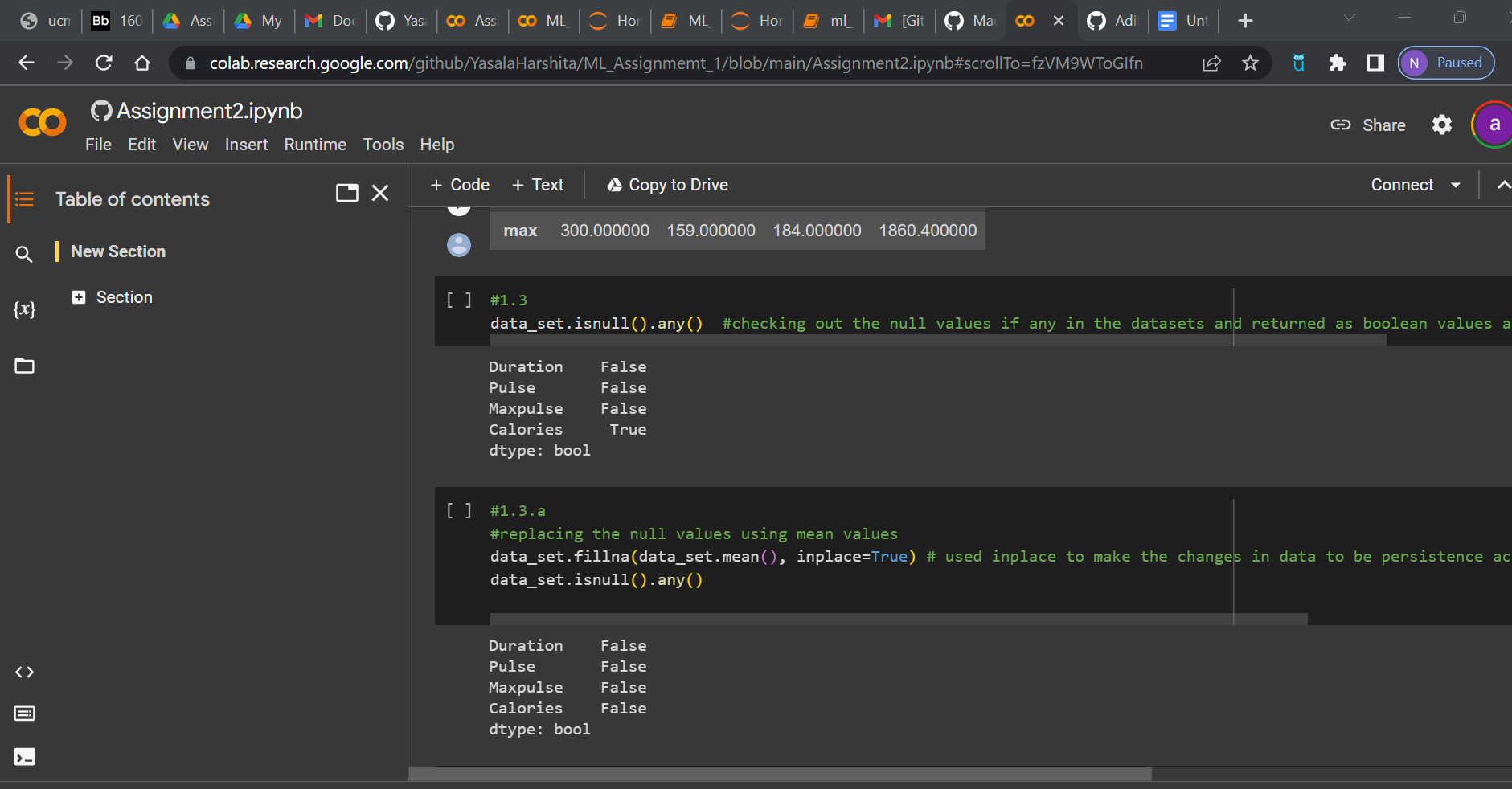
2. Show the basic statistical description about the data.

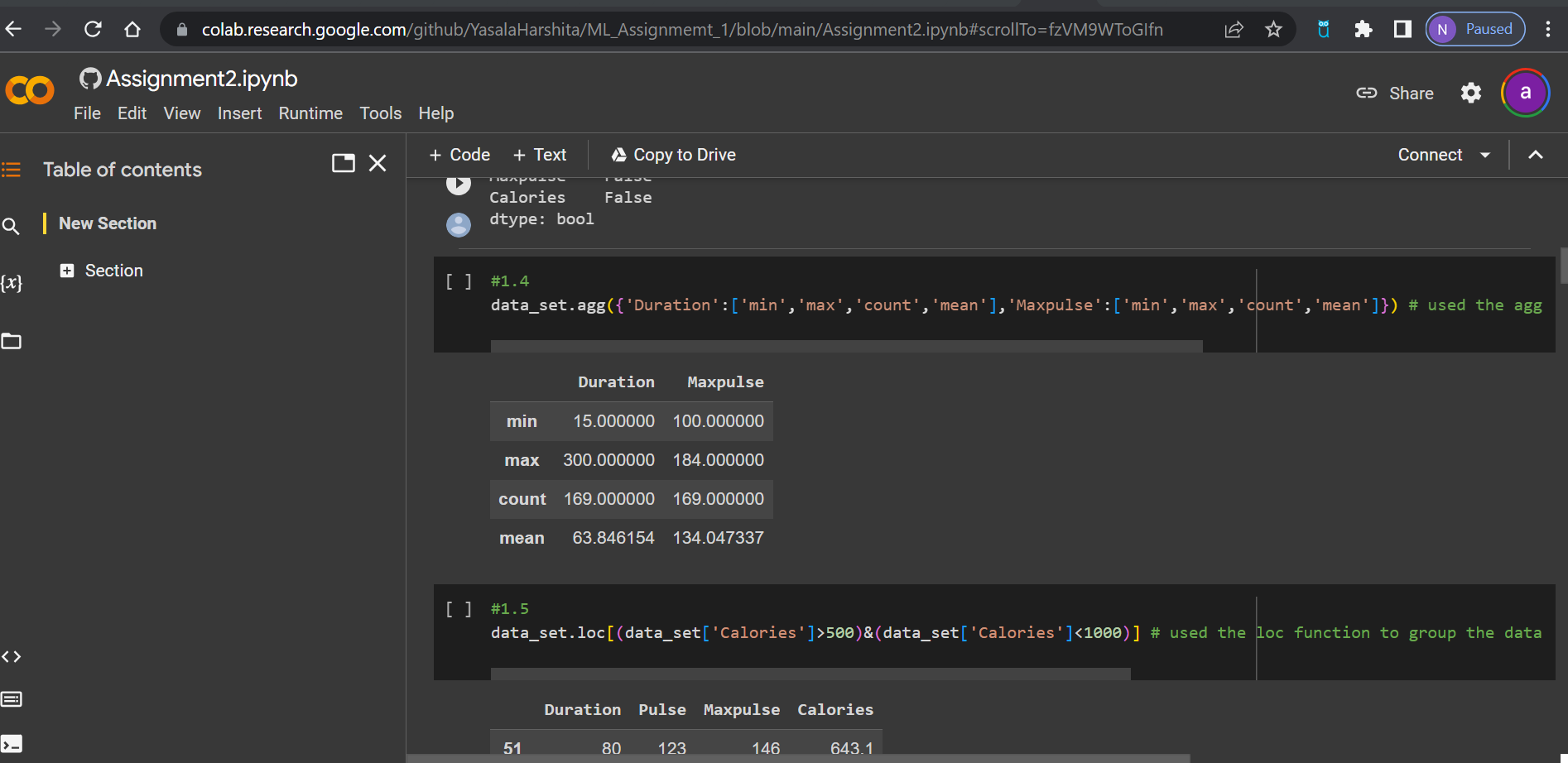


3. Check if the data has null values.

a. Replace the null values with the mean

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

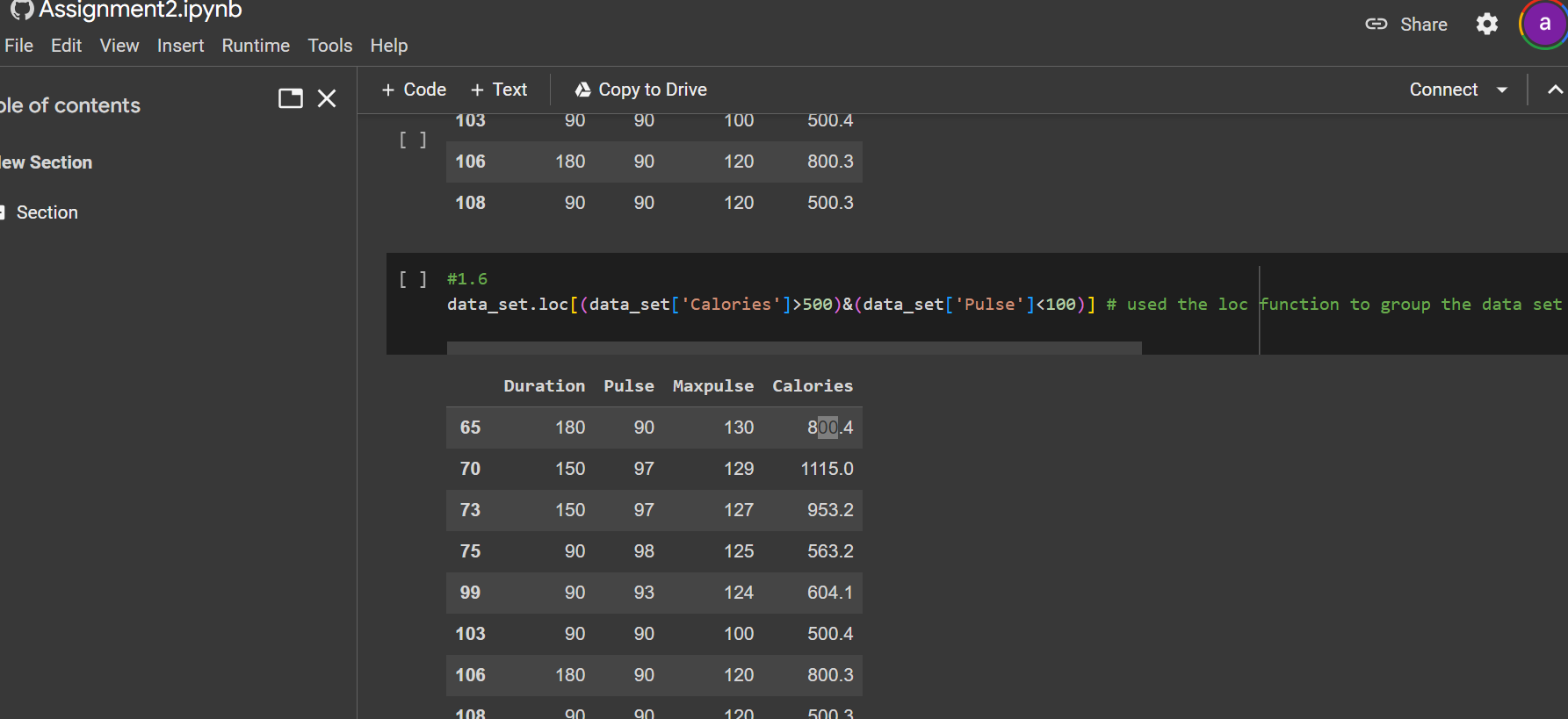


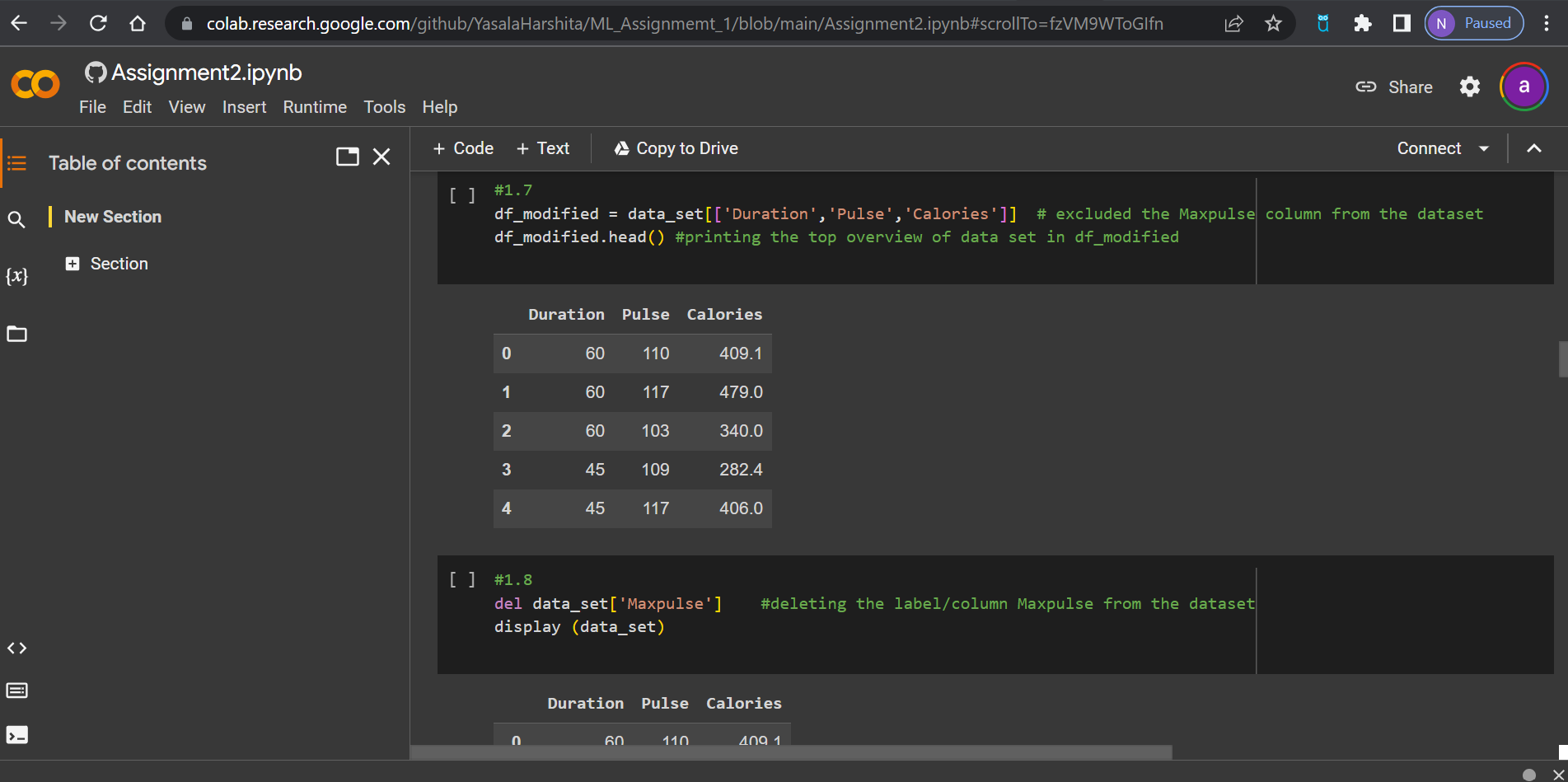


5. Filter the dataframe to select the rows with calories values between 500 and 1000.

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

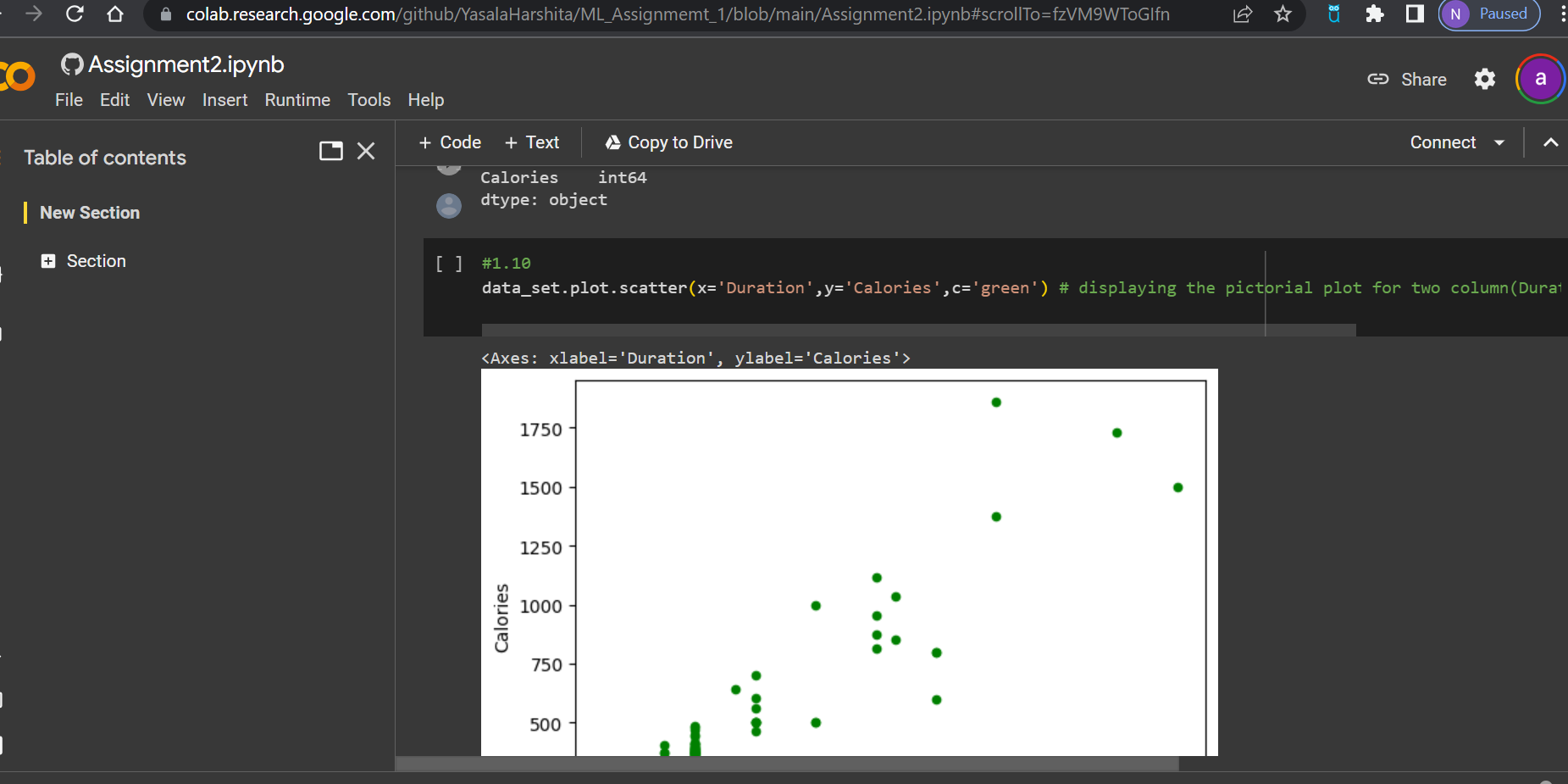
7. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”



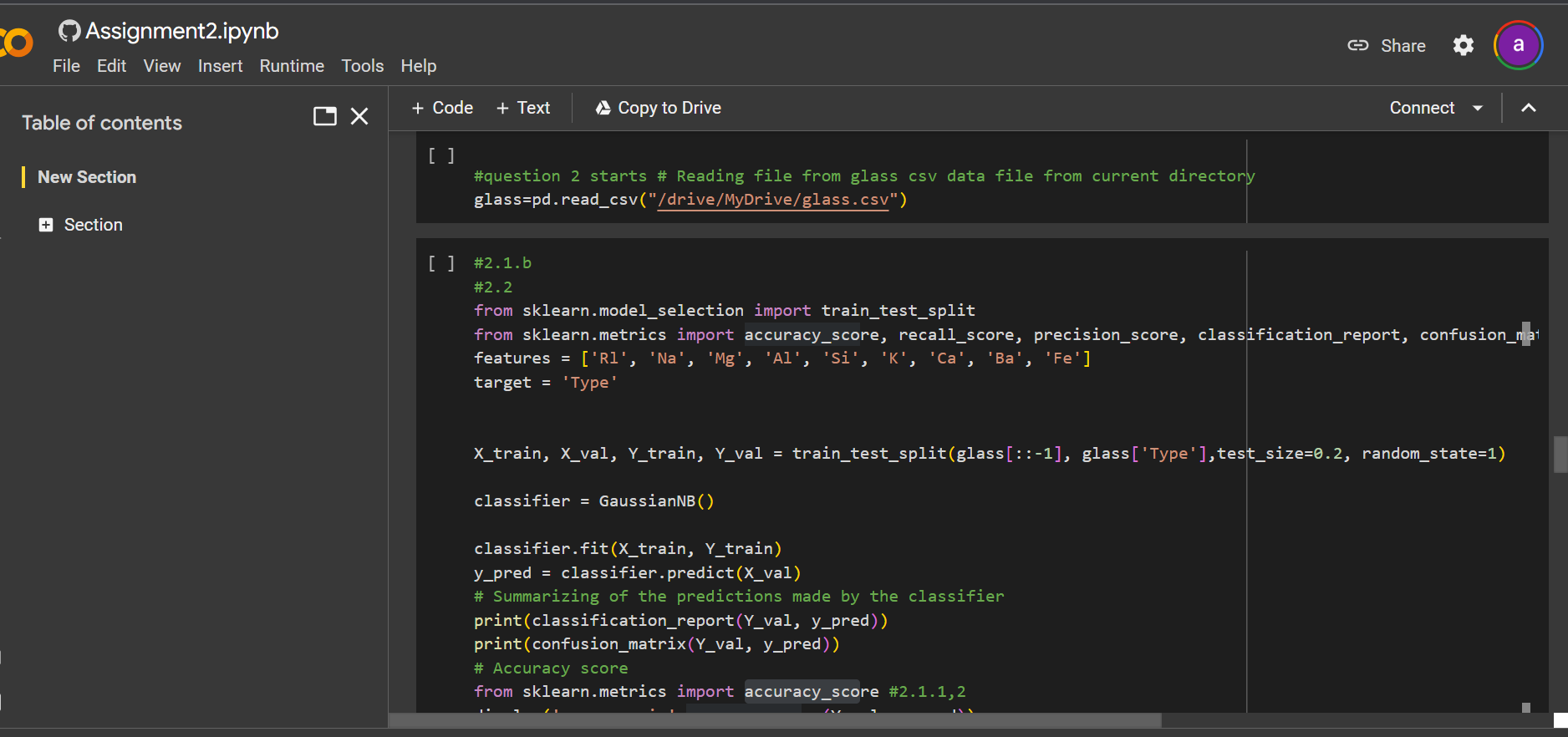


8. Delete the “Maxpulse” column from the main df dataframe

9. Convert the datatype of Calories column to int datatype

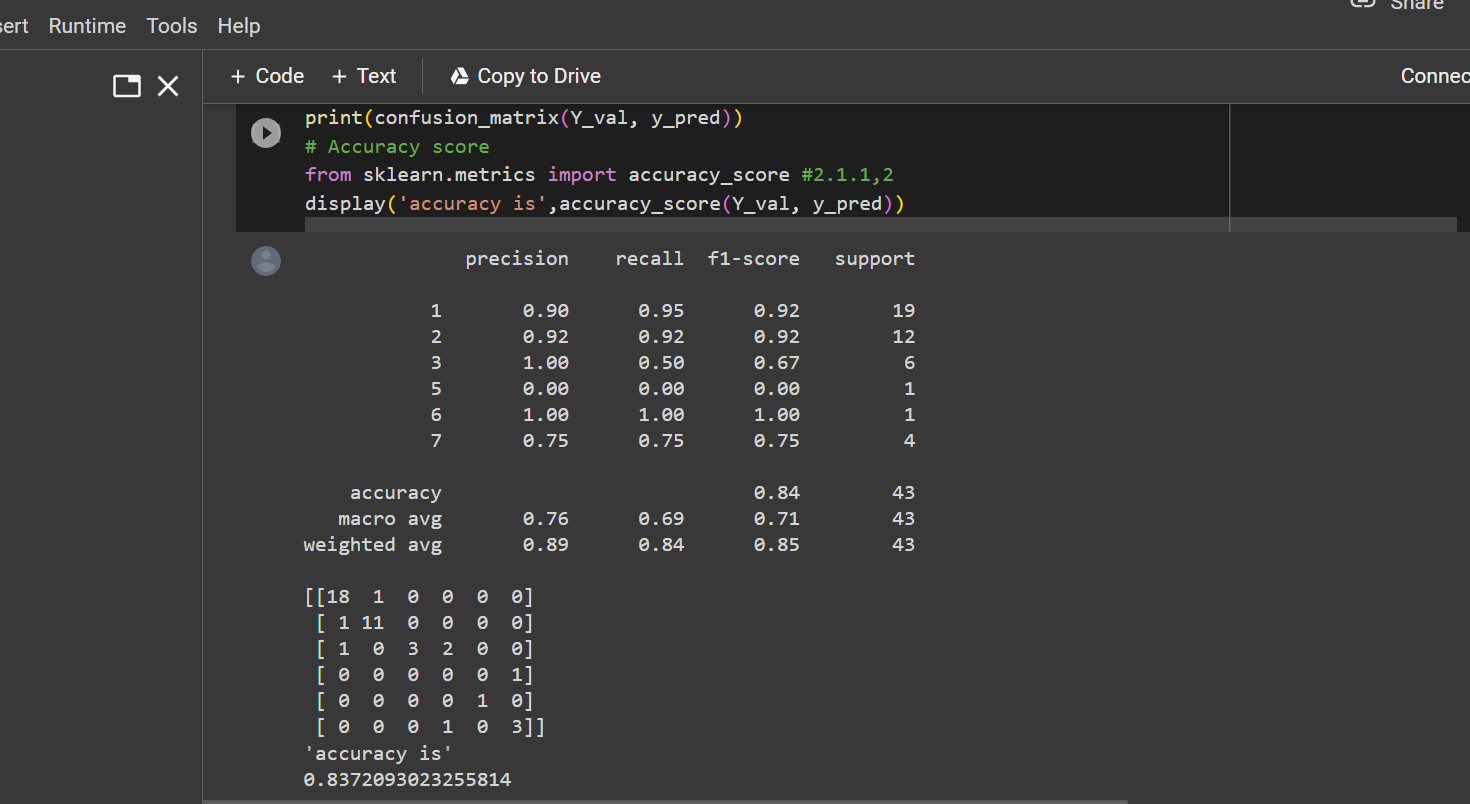


10. Using pandas create a scatter plot for the two columns (Duration and Calories).



a. Do you think we should keep this feature?

Yes, as By using correlation to demonstrate which variable has a high or low correlation with another variable, we can plot correlation matrices to show the strength of the relationship between the dependent and independent variables/columns



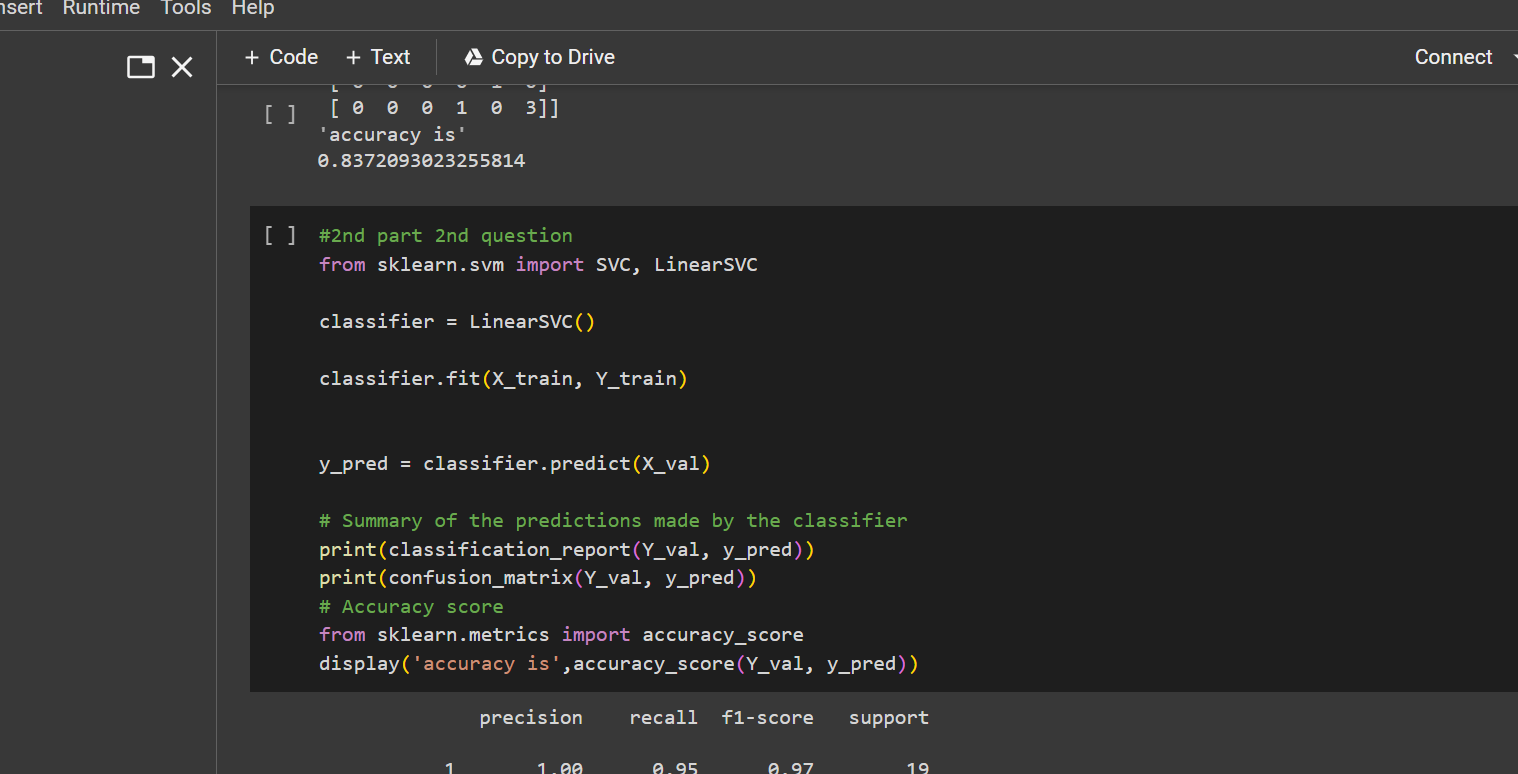
2. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)

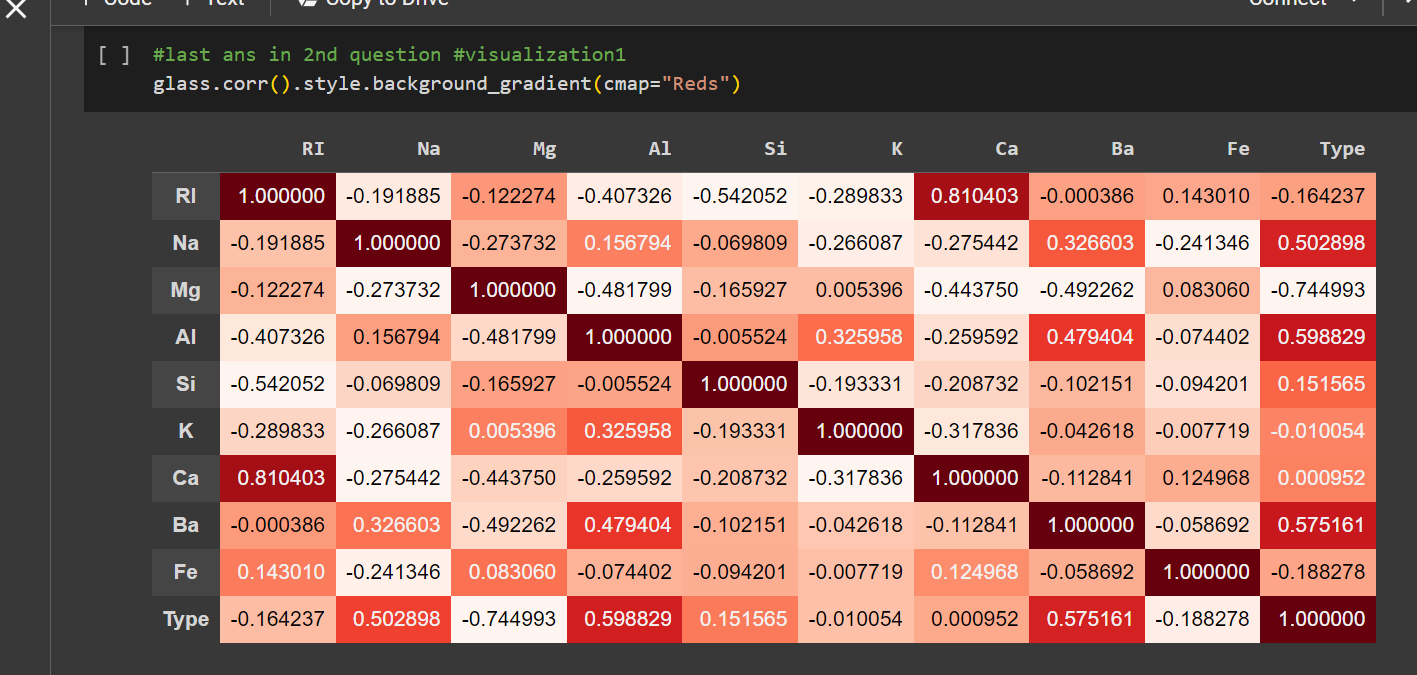
1. Implement linear SVM method using scikit library

a. Use the glass dataset available in Link also provided in your assignment.

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

2 . Evaluate the model on testing part using score







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

ANS: Naïve Bayes Justification: Logistic regression and SVM are discriminative models, but the Multinomial "Naive Bayes" in this case has higher accuracy in comparison to the SVM algorithm from the executions above. While Naive Bayes is a generative model, logistic regression and SVM are discriminative models. It is well known that a generative model can outperform a discriminative model when you have very little data, taking into account the parameter (Dataset) sensitivity and implementation pattern