Machine Learning (Assignment # 2)

Sravani Lankala

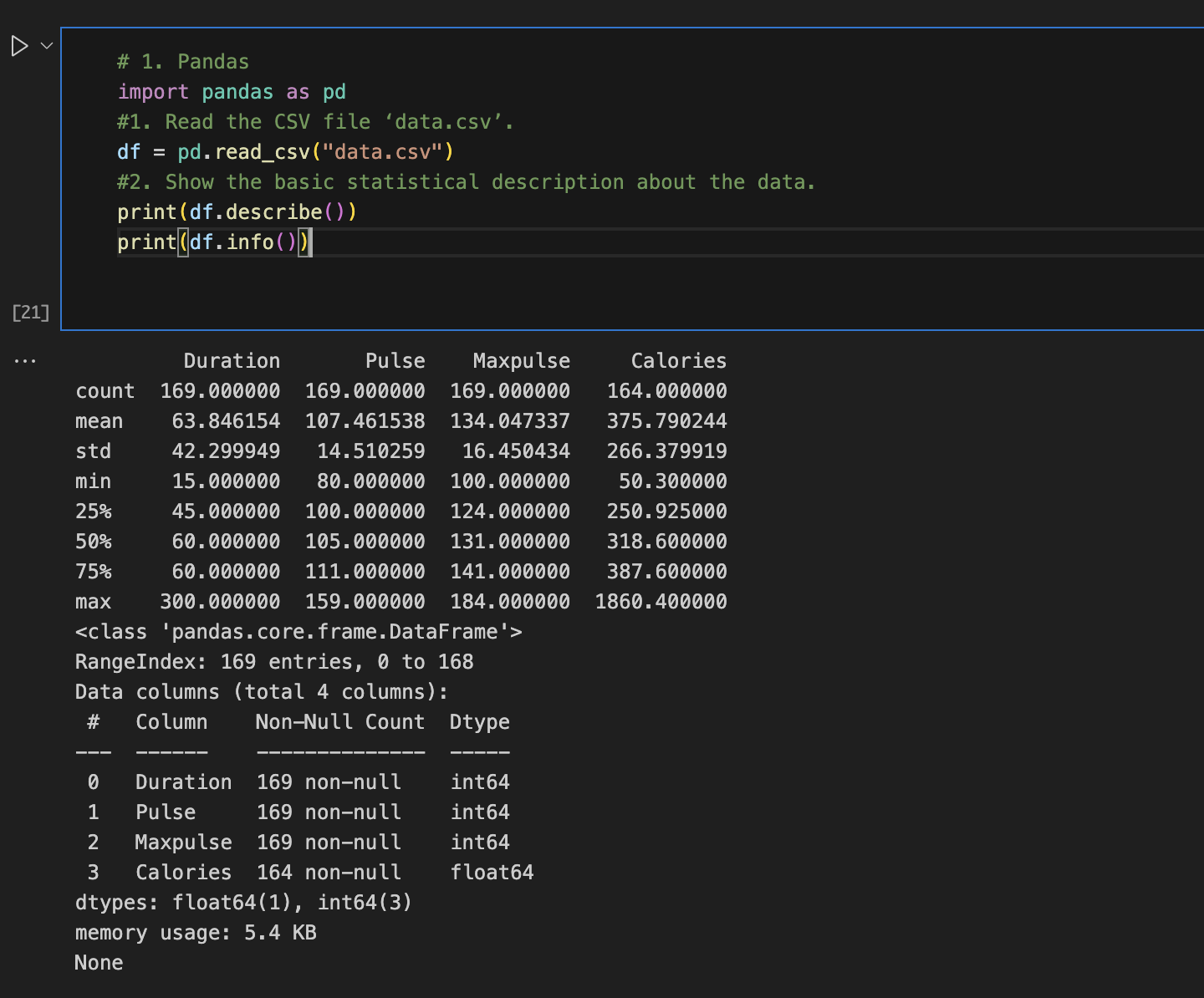
700746285

GitHub link : <https://github.com/sravanilankala/Assignment-2-ML-Summer>

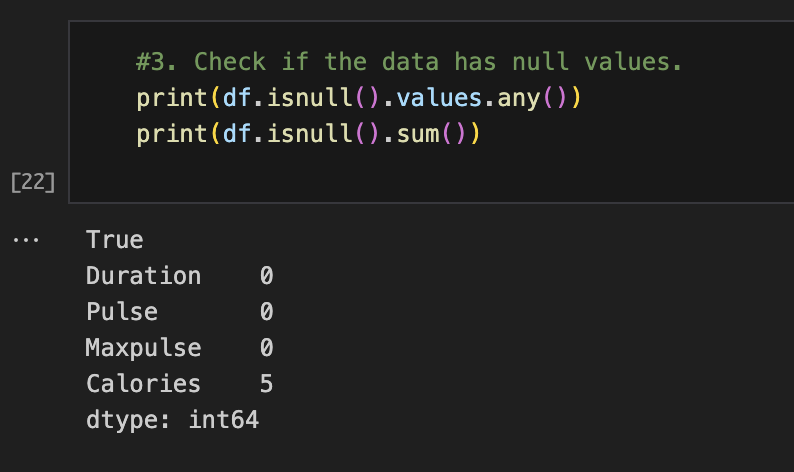
Video Link: <https://drive.google.com/file/d/1SAV3kUl3LOLCD6AOTOX0EXkP_EuSRuGz/view?usp=sharing>

**1. Pandas**

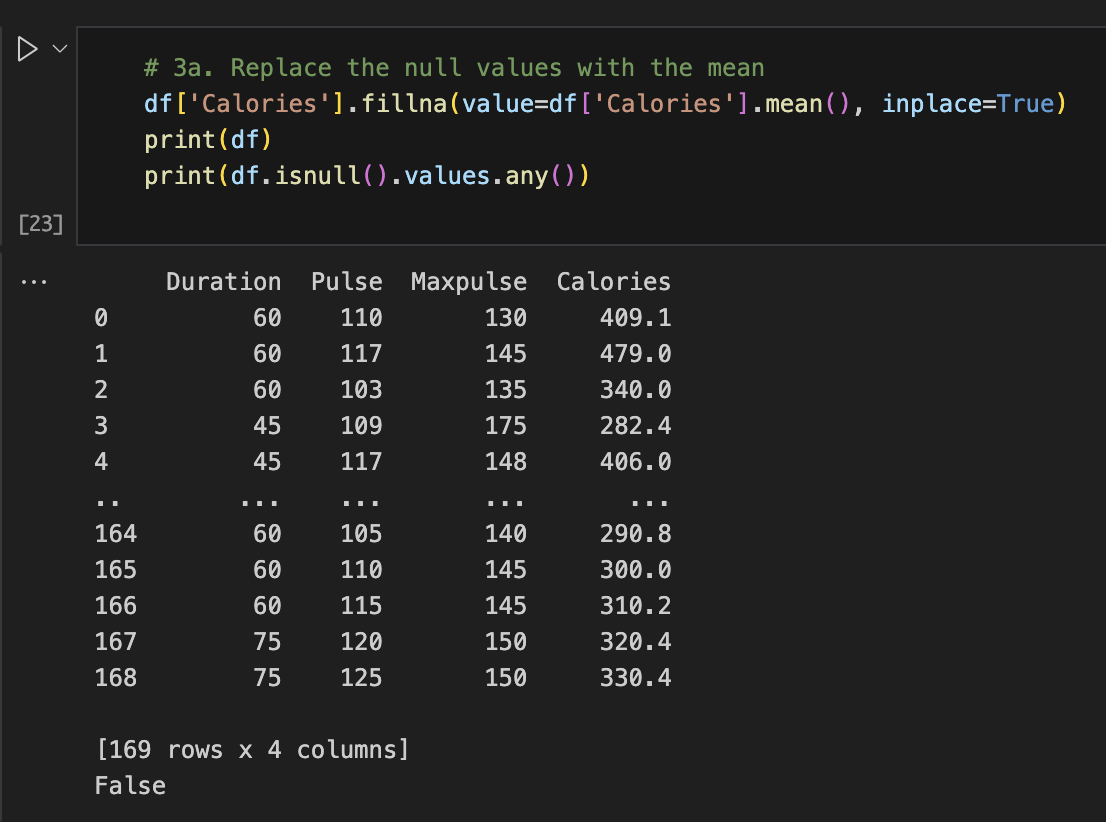
1. Read the CSV file ‘data.csv’.
2. Show the basic statistical description about the data.



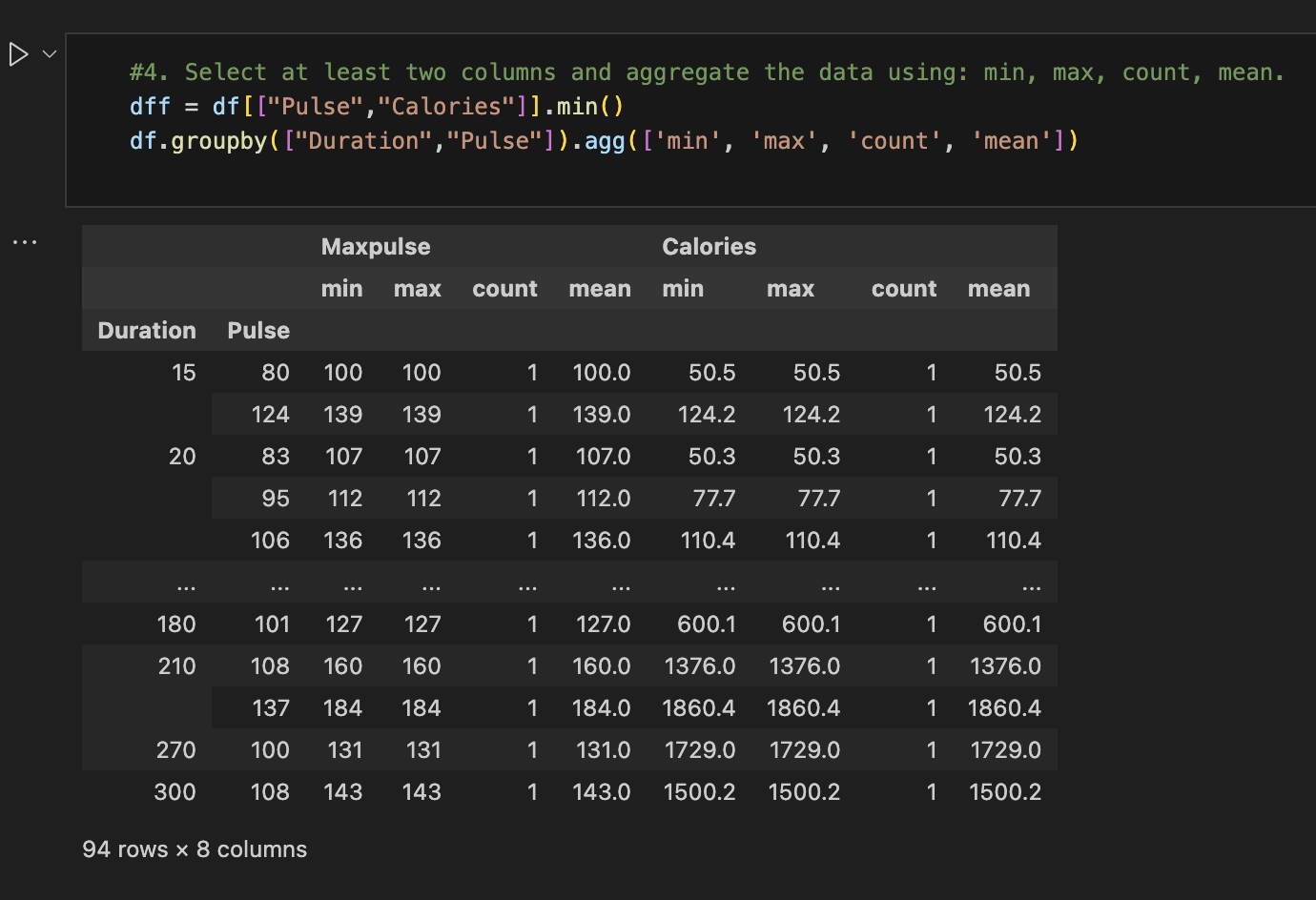
1. Check if the data has null values.



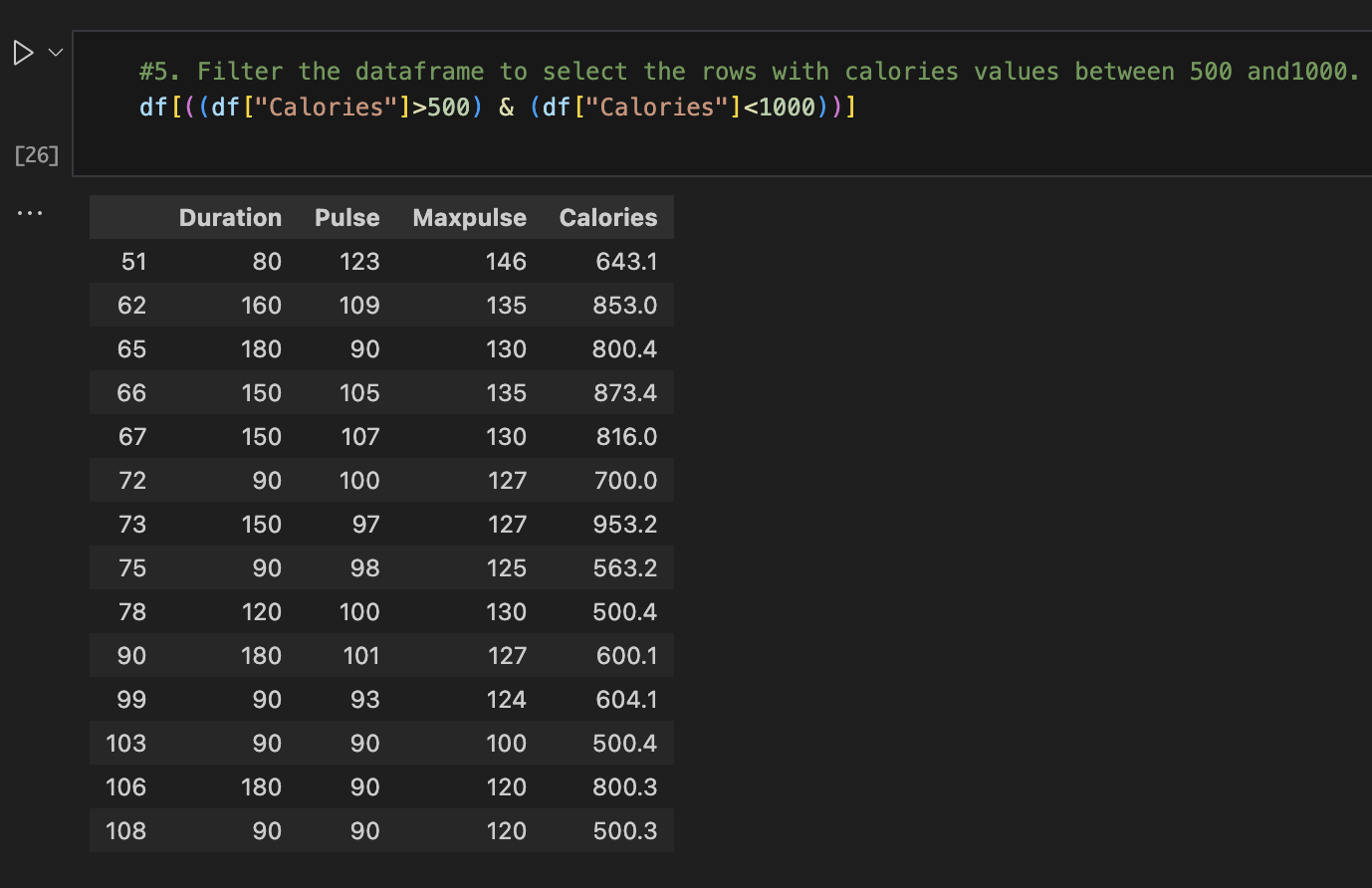
* 1. Replace the null values with the mean



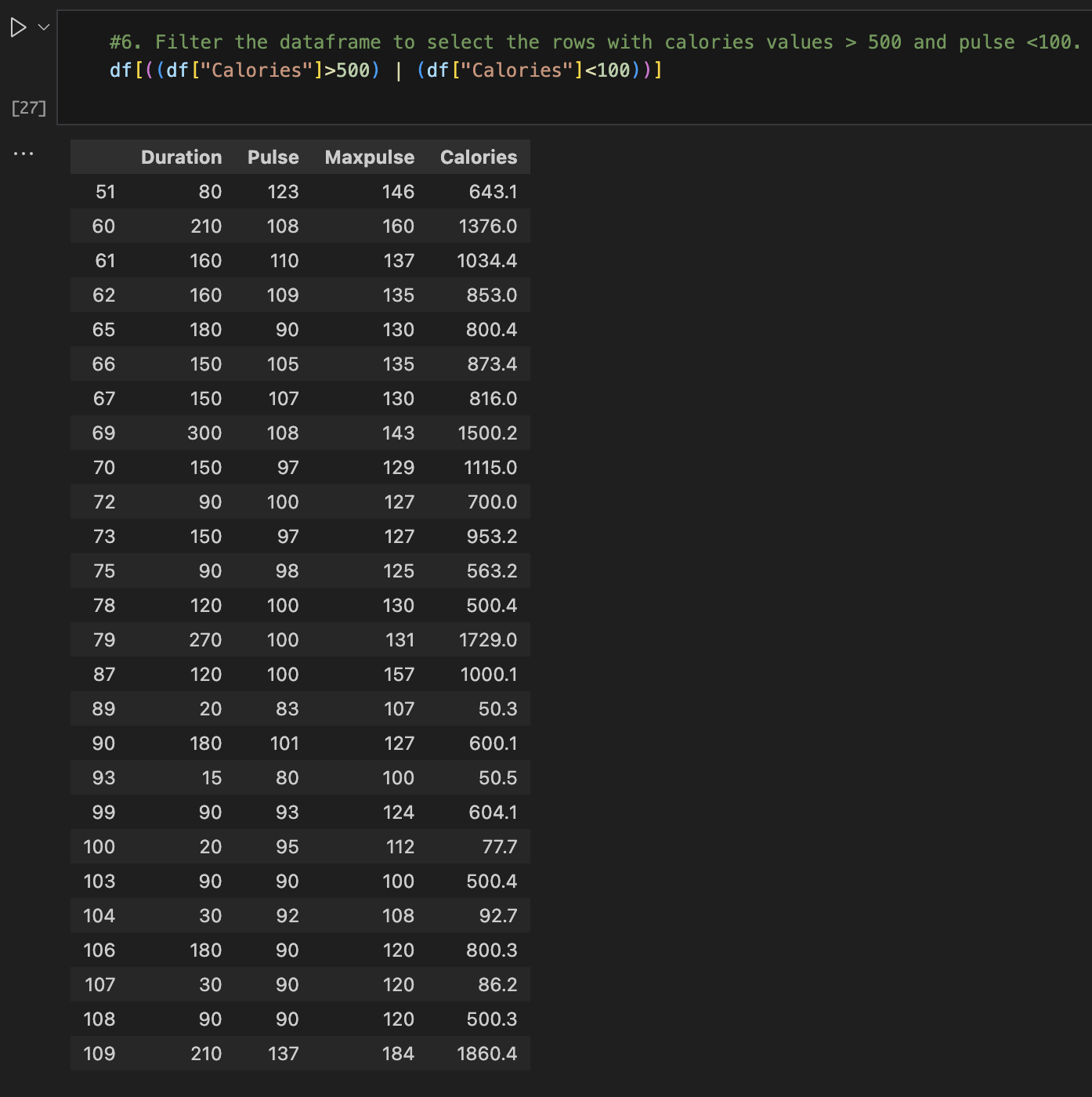
1. Select at least two columns and aggregate the data using: min, max, count, mean.



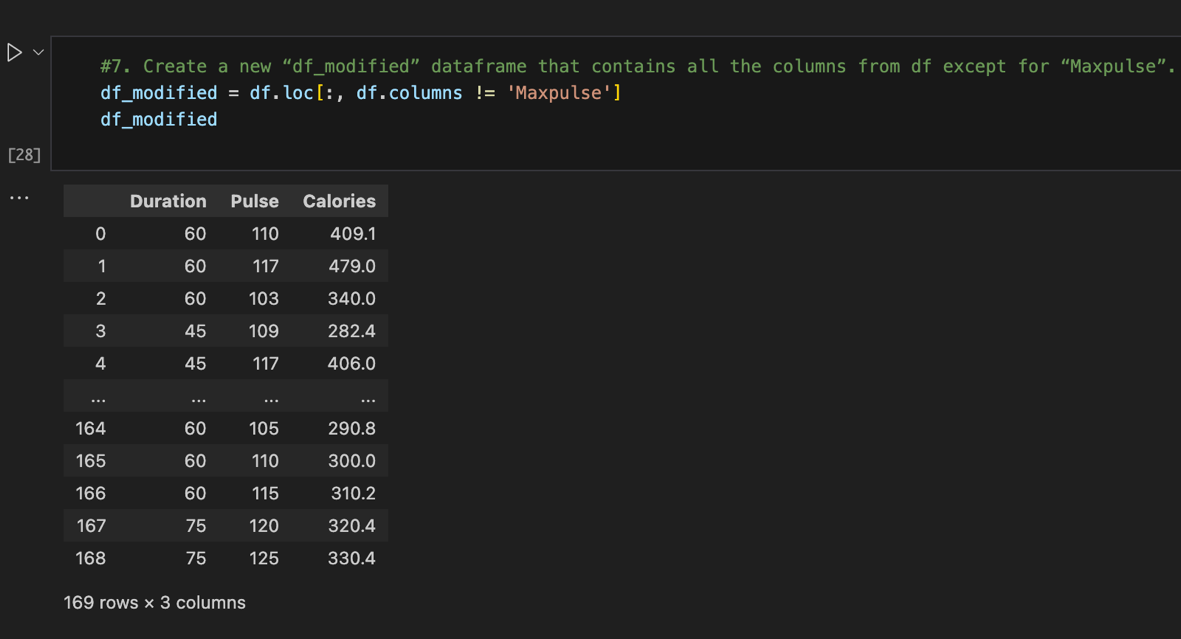
1. Filter the dataframe to select the rows with calories values between 500 and1000.



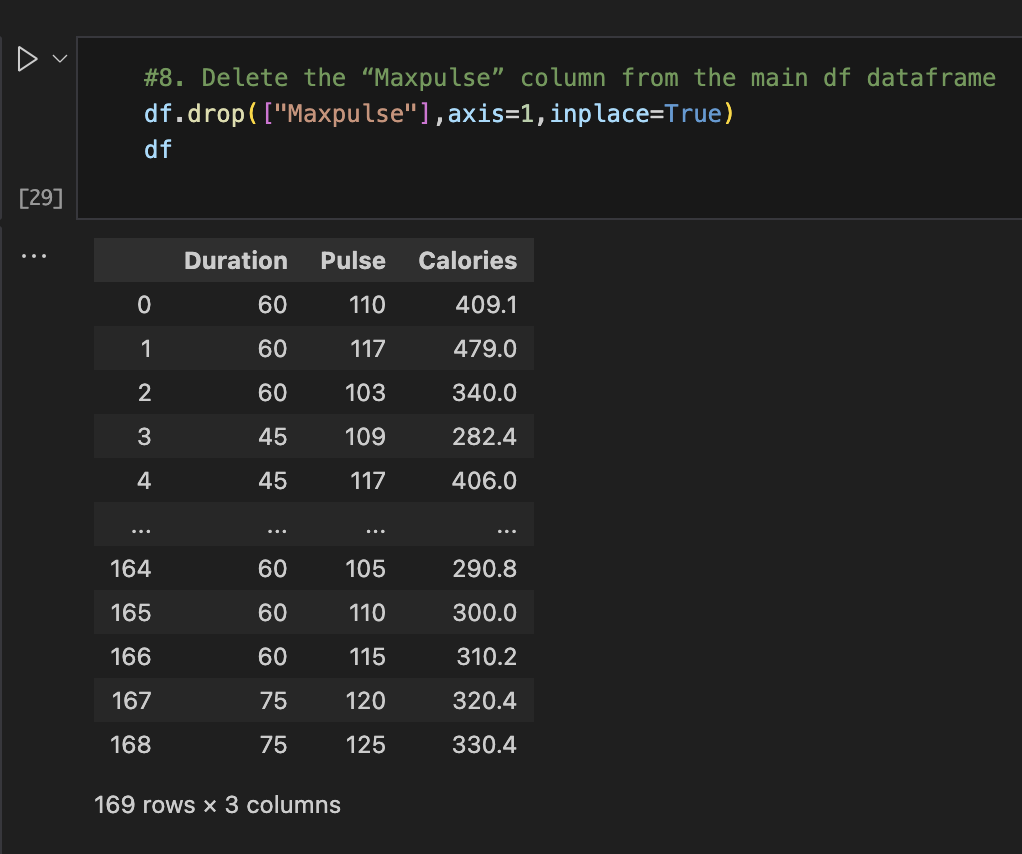
1. Filter the data frame to select the rows with calories values > 500 and pulse <100.



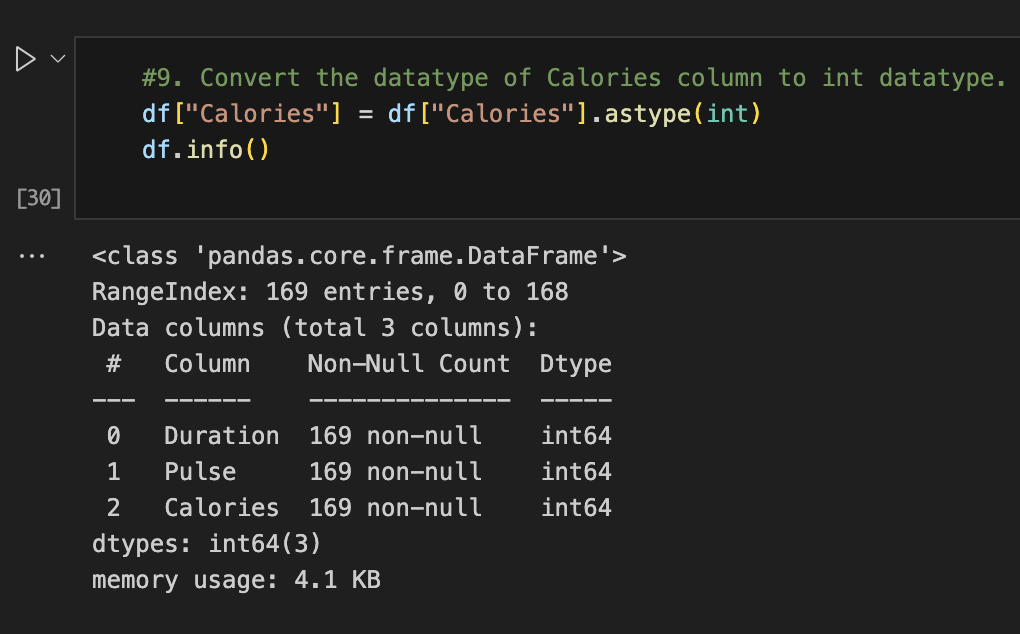
1. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.



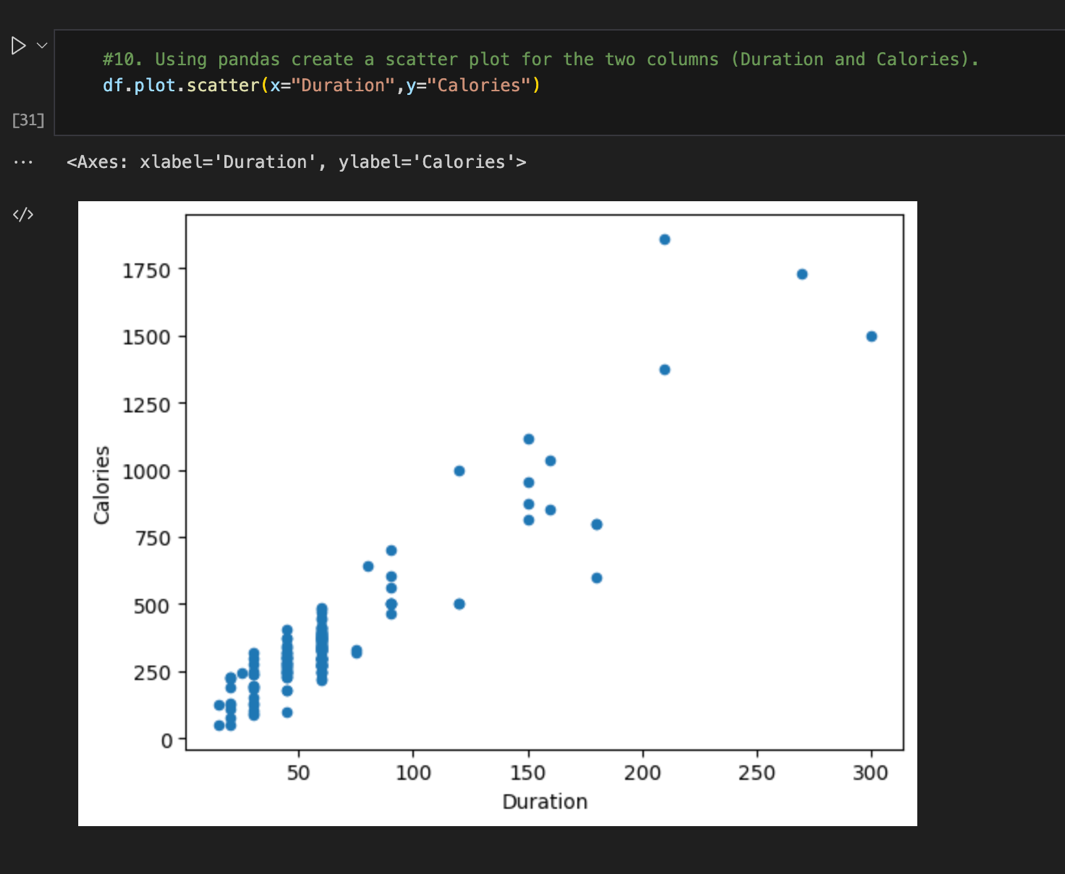
1. Delete the “Maxpulse” column from the main df data frame



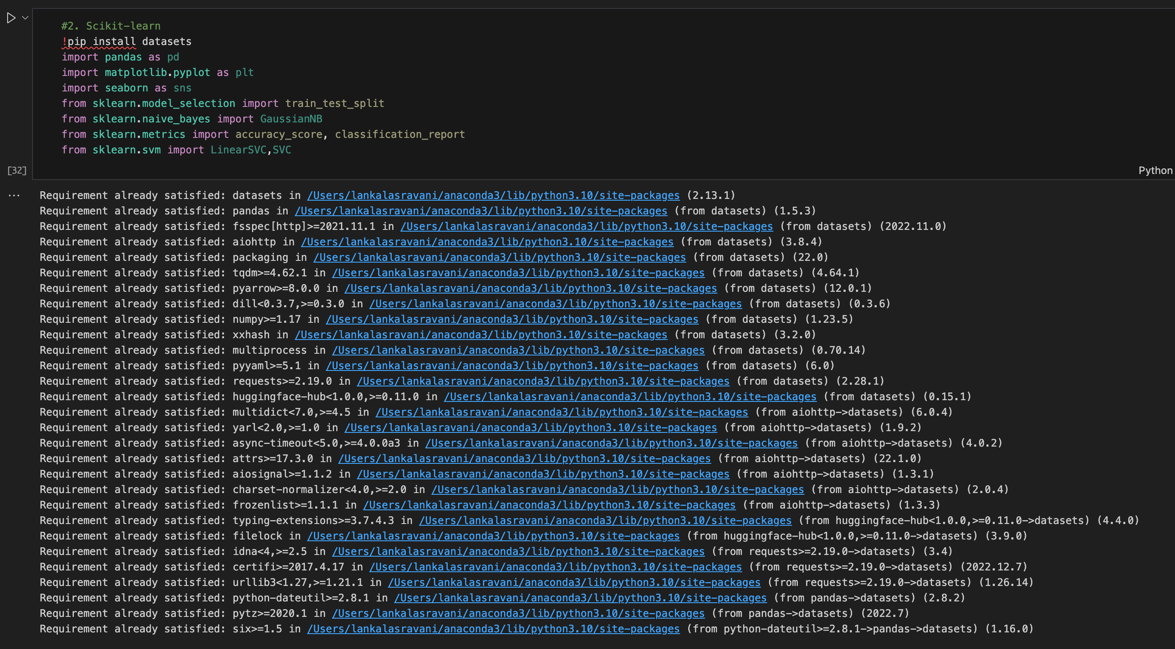
1. Convert the datatype of Calories column to int datatype.



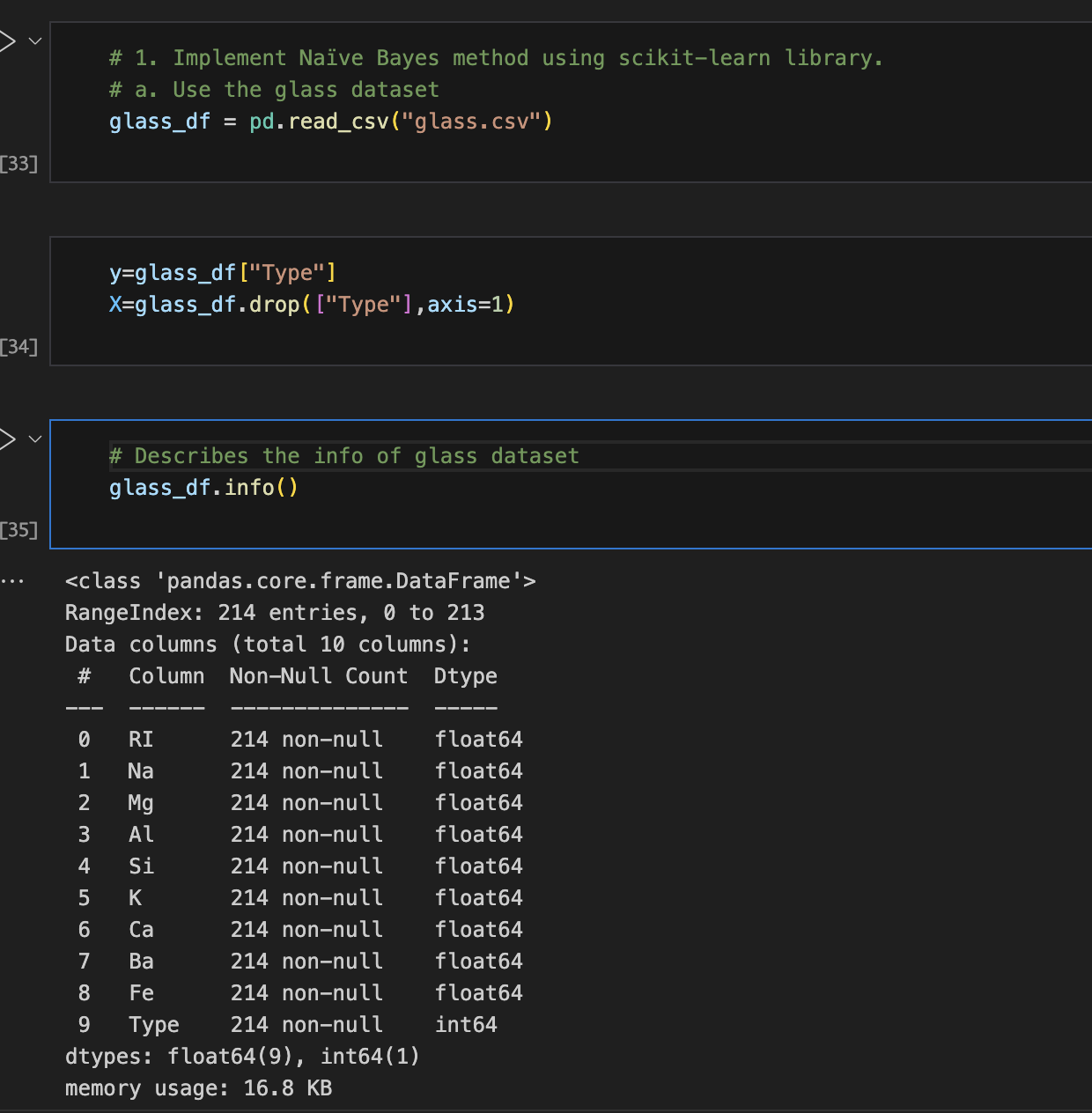
1. Using pandas create a scatter plot for the two columns (Duration and Calories).



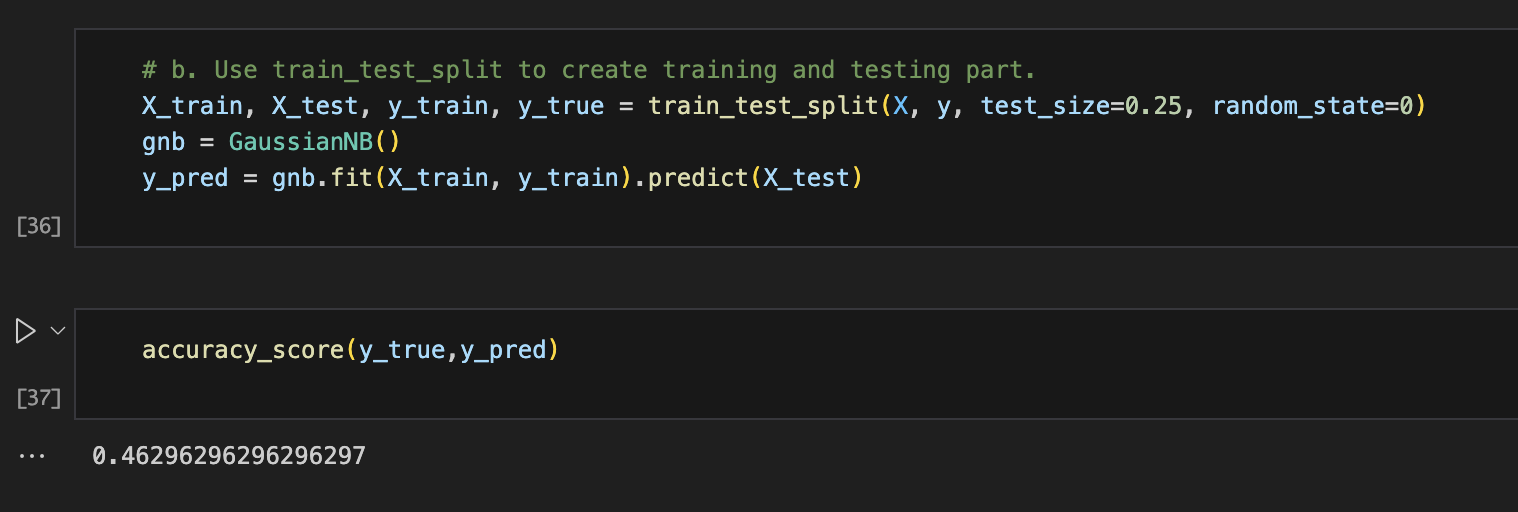
**2. Scikit-learn**



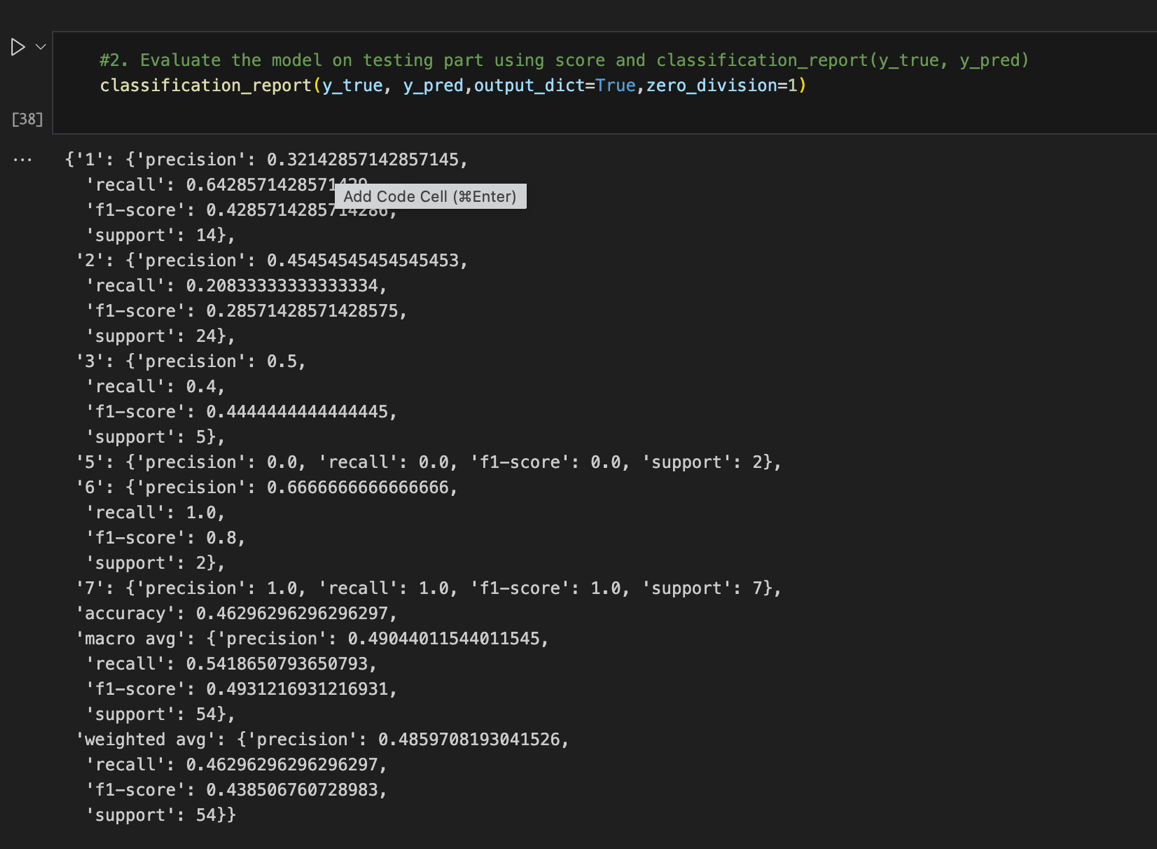
1. Implement Naïve Bayes method using scikit-learn library.
2. Use the glass dataset provided in assignment.



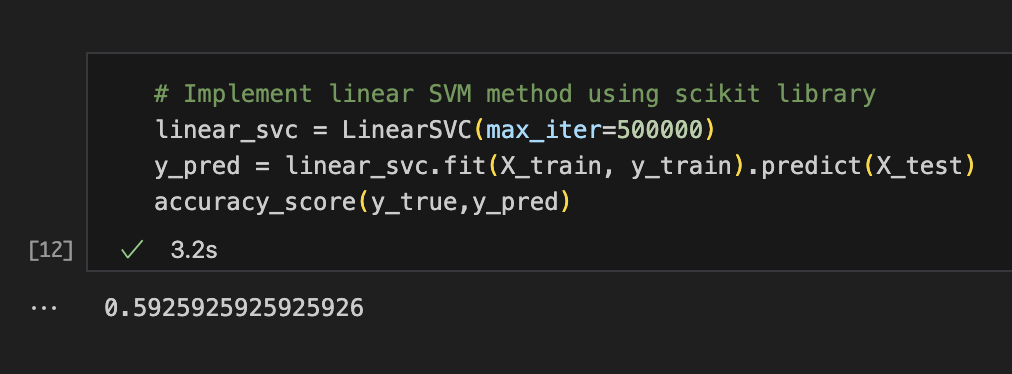
1. Use **train\_test\_split** to create training and testing part.



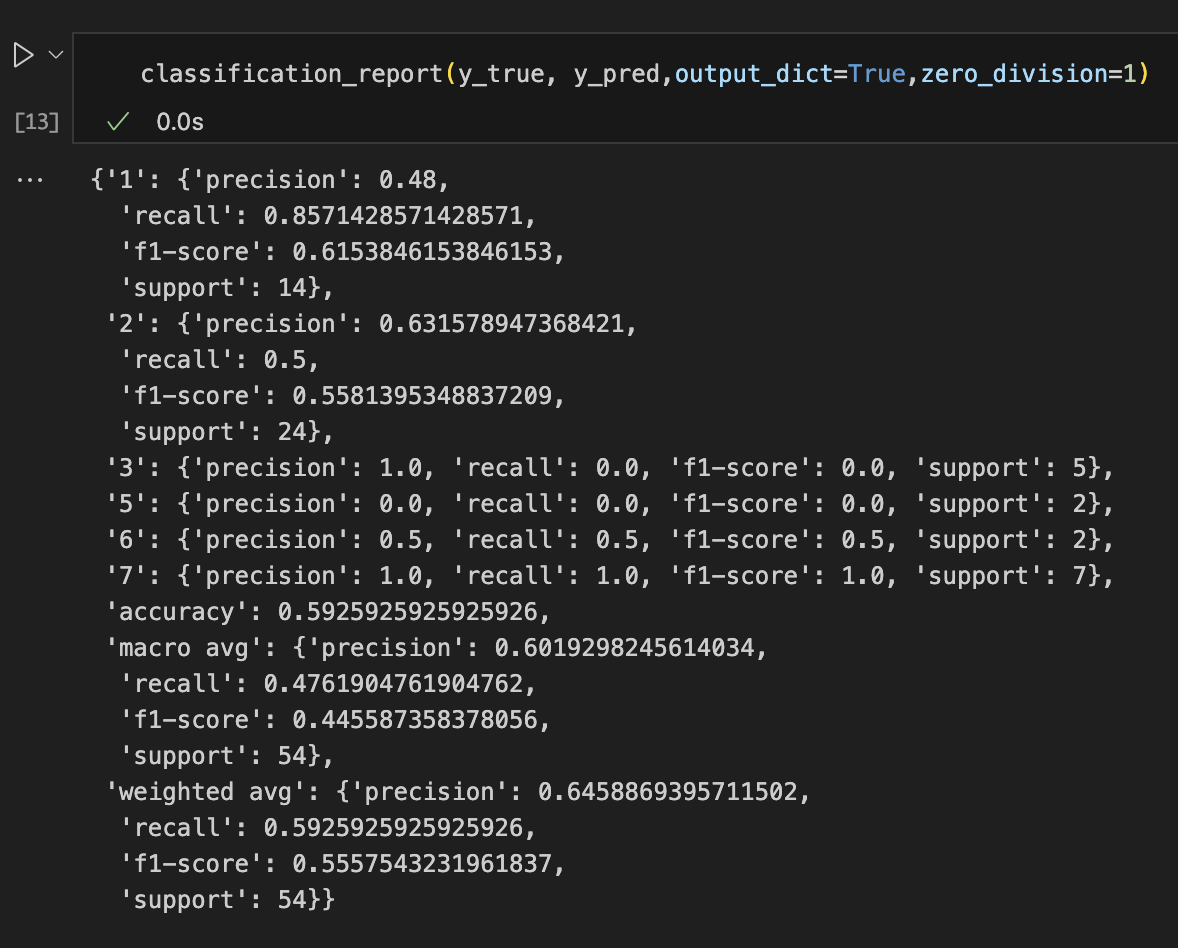
1. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)



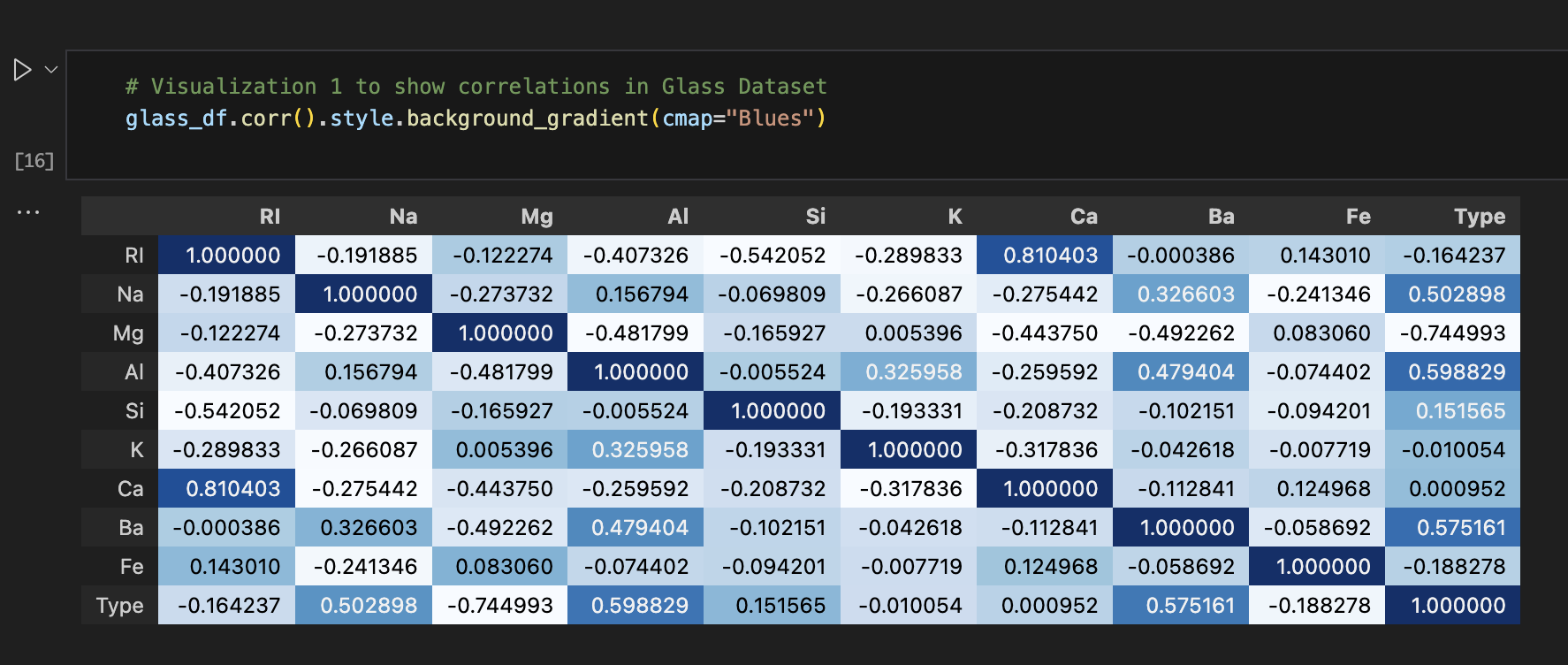
1. Implement linear SVM method using scikit library
   1. Use the glass dataset provided in assignment.
   2. Use **train\_test\_split** to create training and testing part.

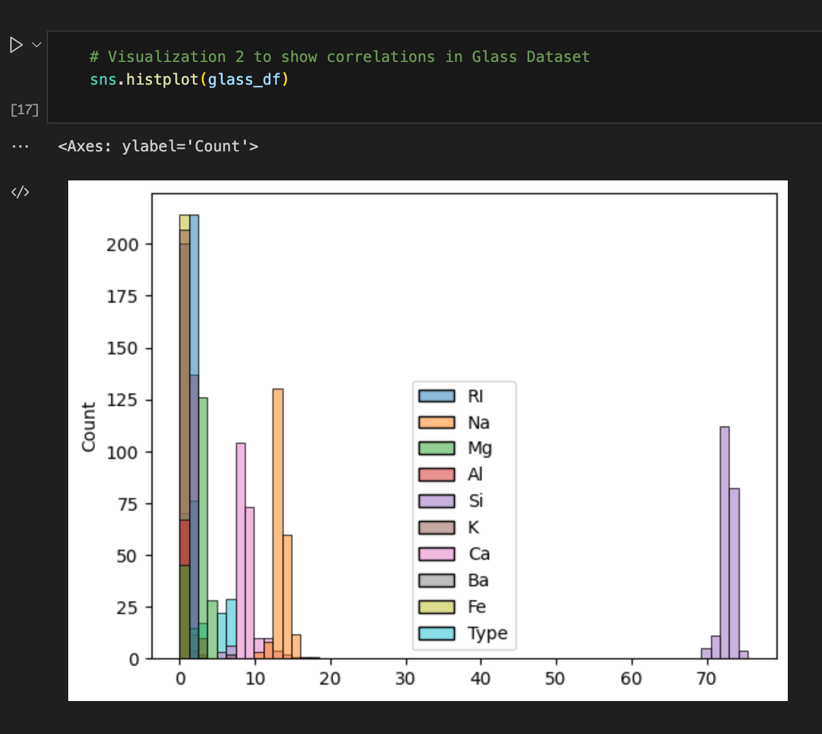


1. Evaluate the model on testing part using score and classification\_report(y\_true, y\_pred)



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





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

On Glass Dataset:

Linear SVM has the better accuracy than SVM and Naïve Bayes method

SVMs are different from other Naive Bayes algorithm because of the way they choose the decision boundary that maximizes the distance

from the nearest data points of all the classes. The maximum margin classifier or maximum margin hyper plane is the name of the decision

boundary generated by SVMs.