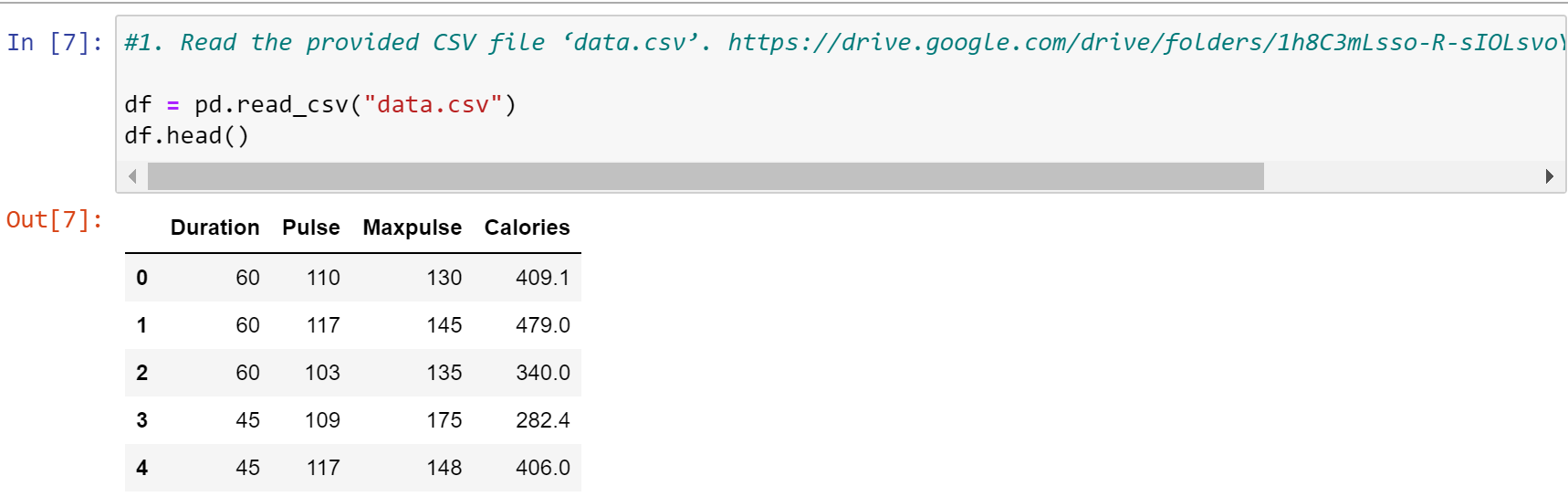
University of central Missouri

Machine learning

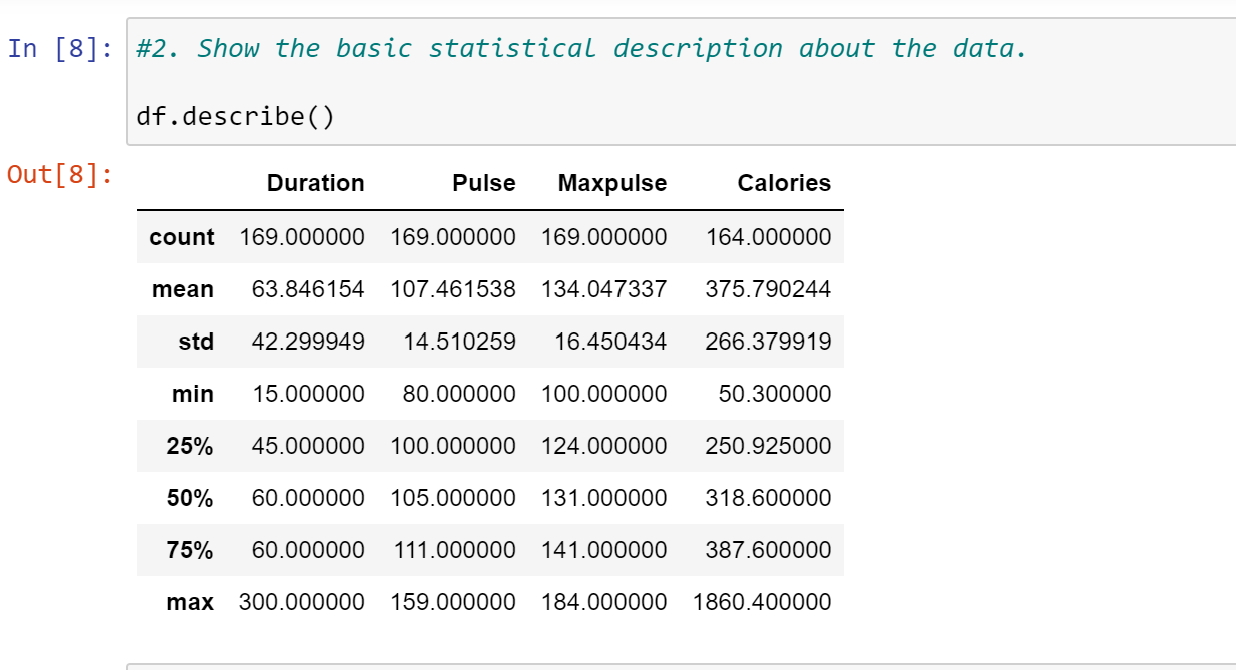
Sowmya Ala

700740199

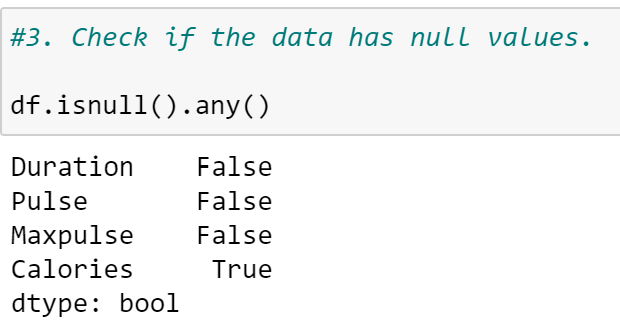
1. Pandas
2. Read the provided CSV file ‘data.csv’.



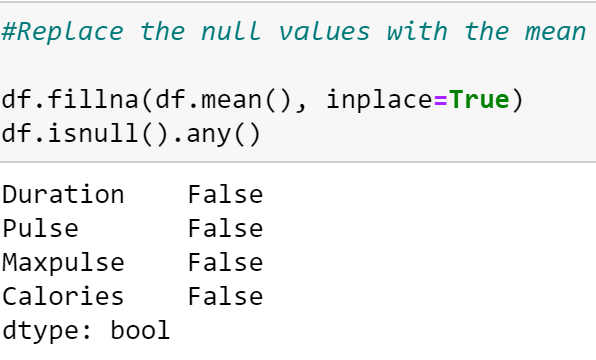
1. Show the basic statistical description about the data.



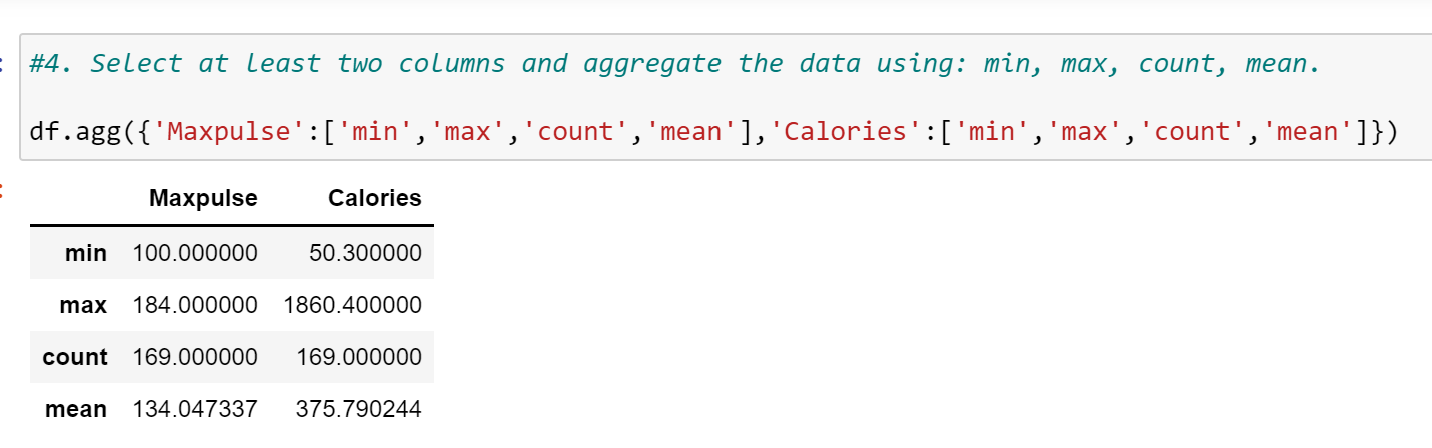
1. Check if the data has null values. a. Replace the null values with the mean



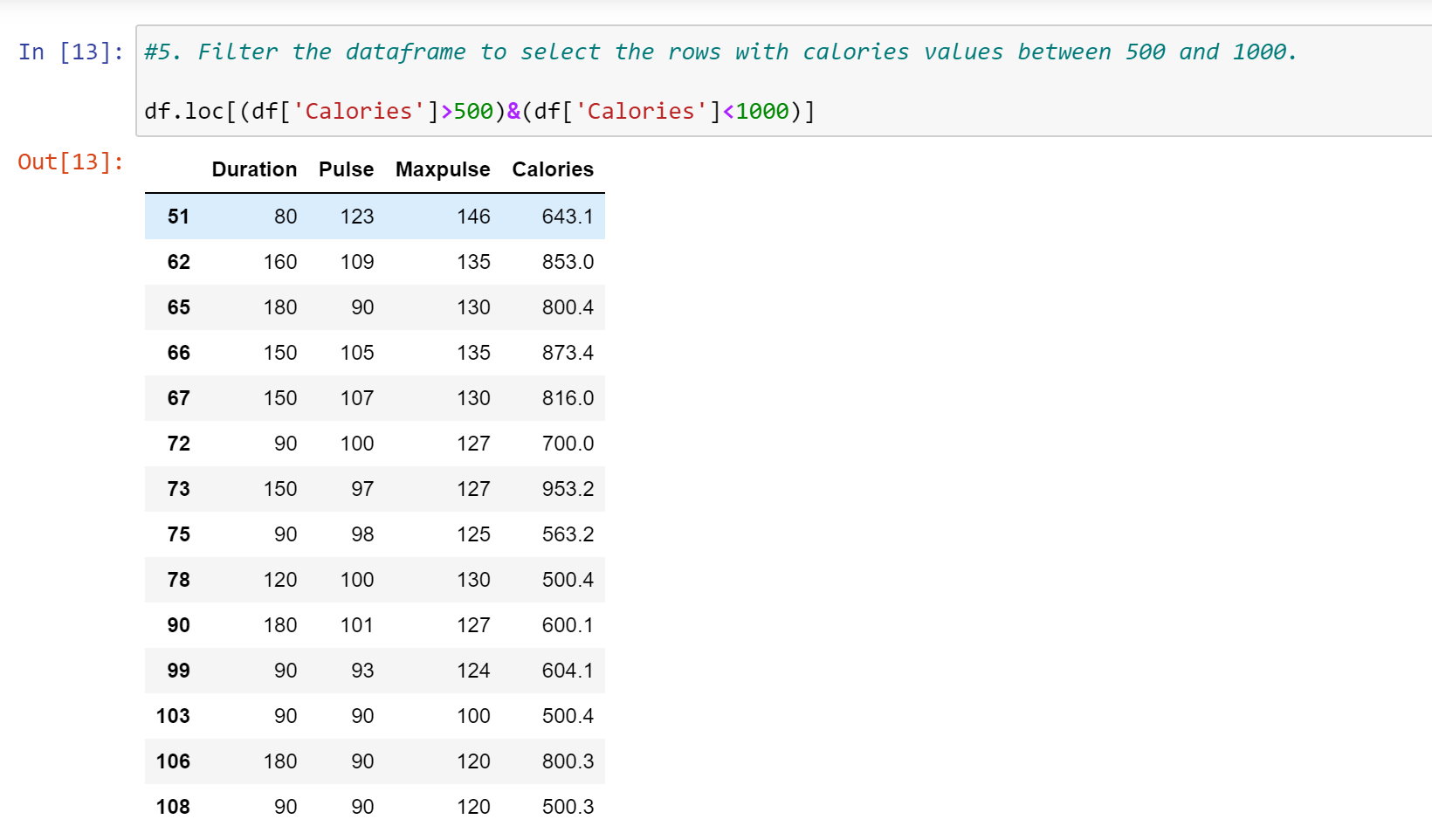
1. Select at least two columns and aggregate the data using: min, max, count, mean.



1. Filter the data frame to select the rows with calories values between 500 and 1000.



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

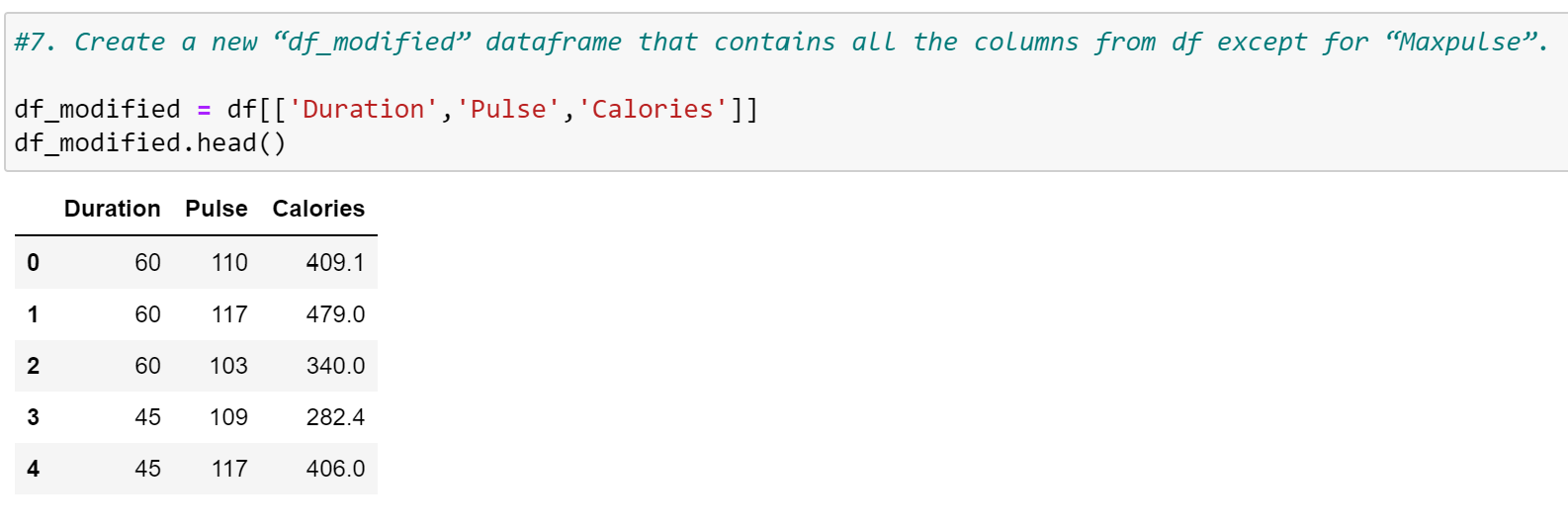


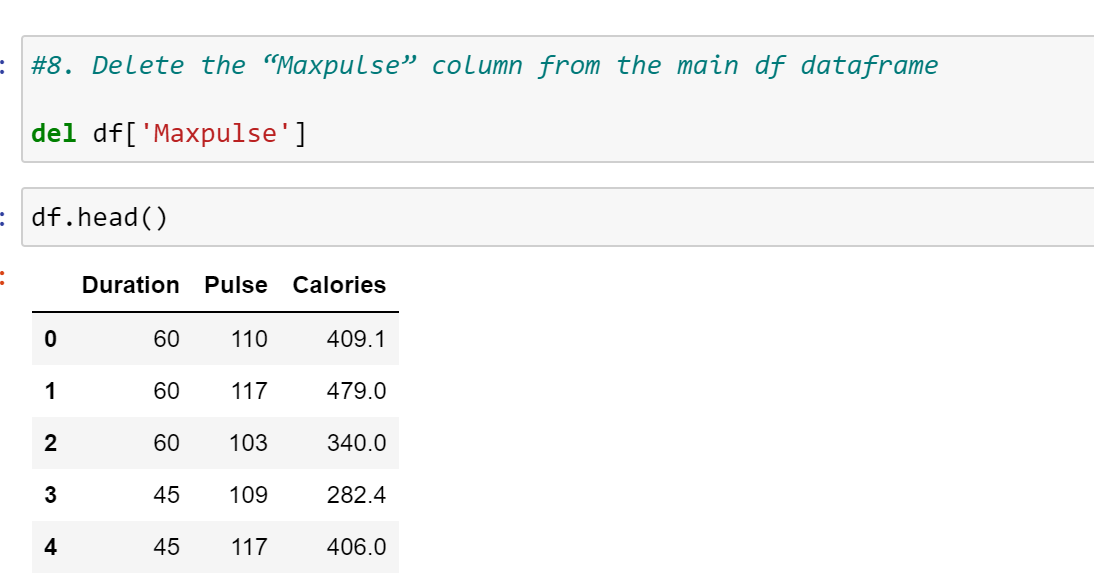
1. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.



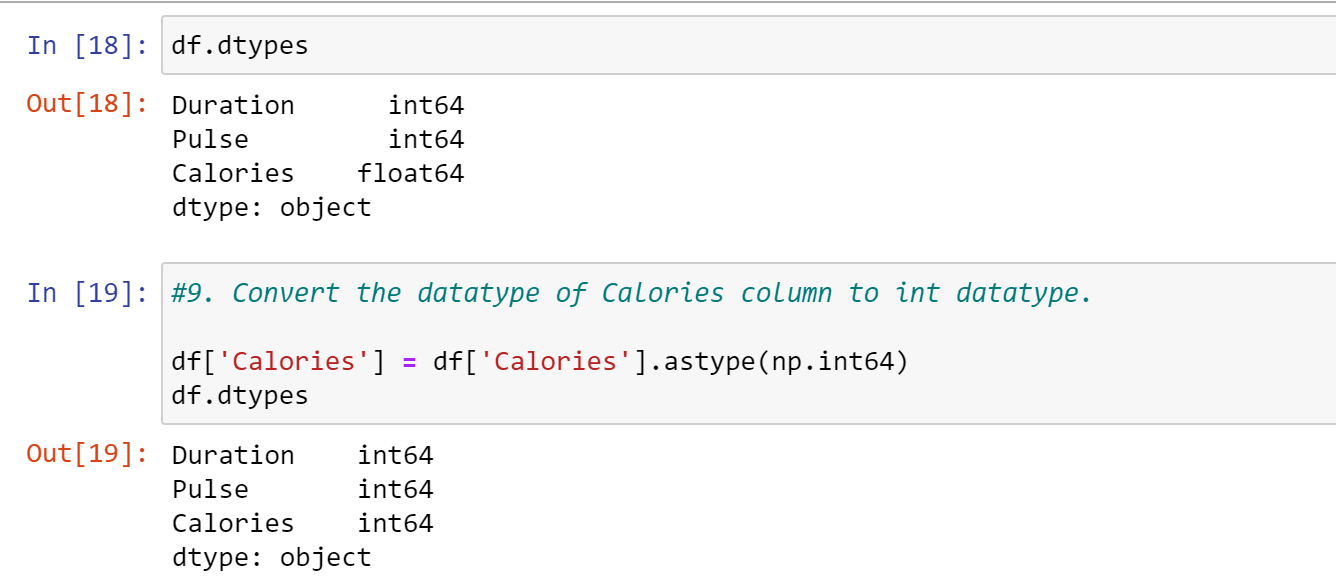
8. Delete the “Maxpulse” column from the main df dataframe

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

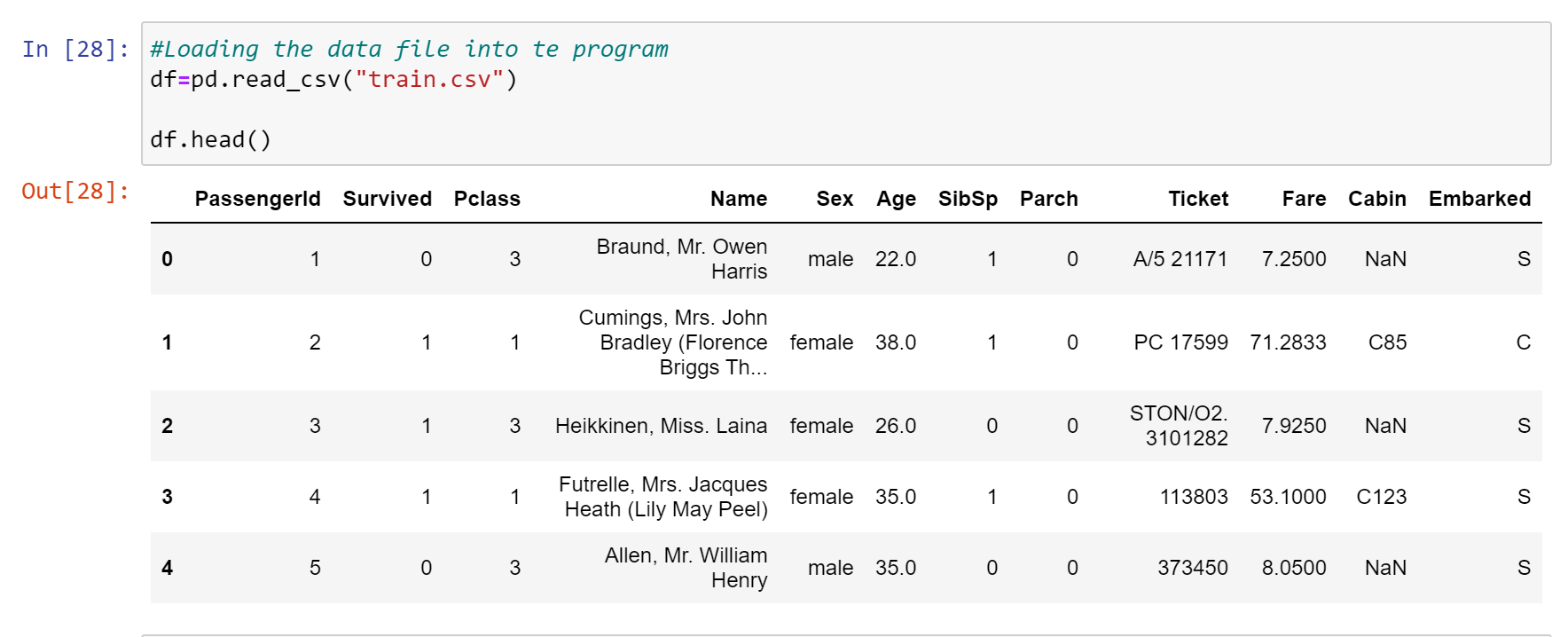




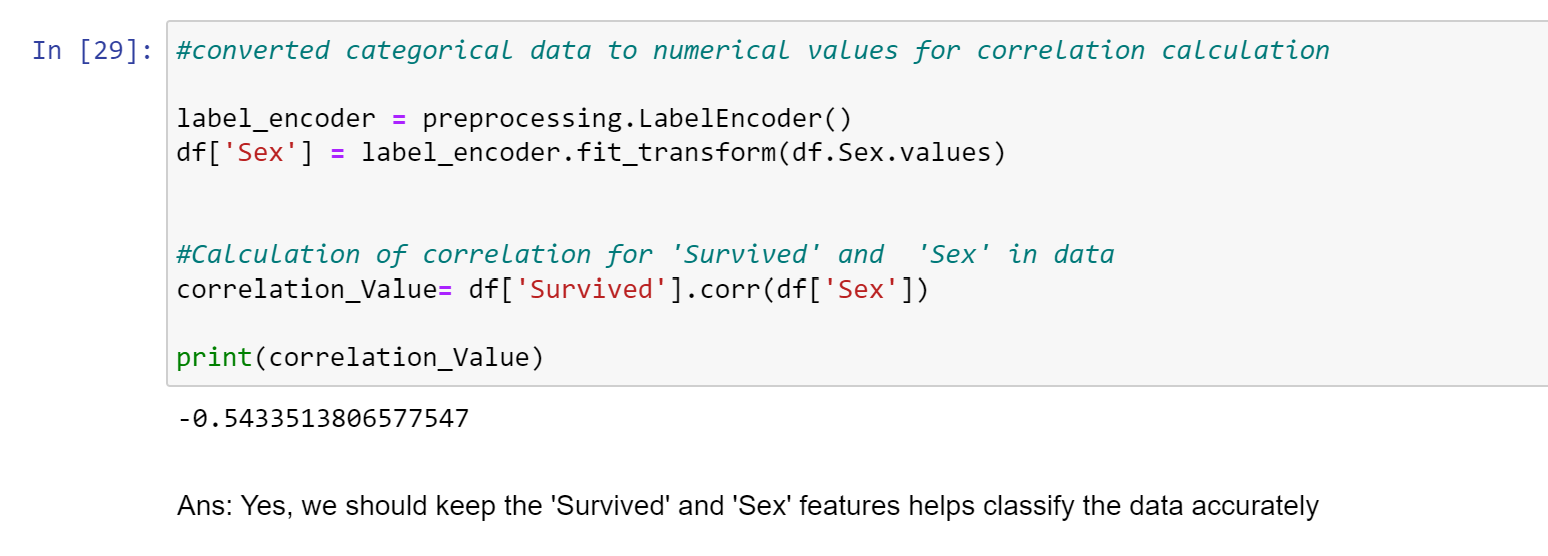
10. Using pandas create a scatter plot for the two columns (Duration and Calories).

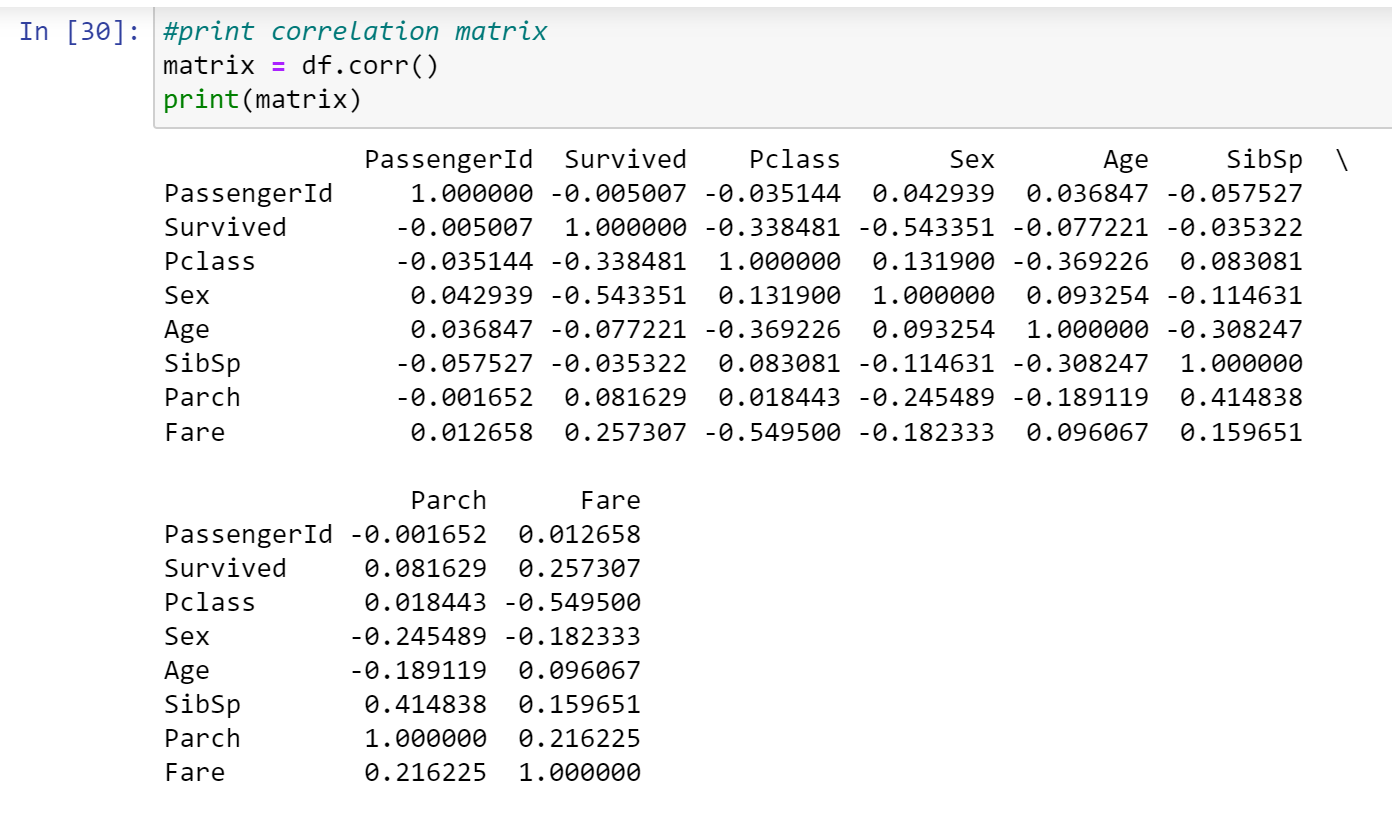


2.Titanic Dataset:

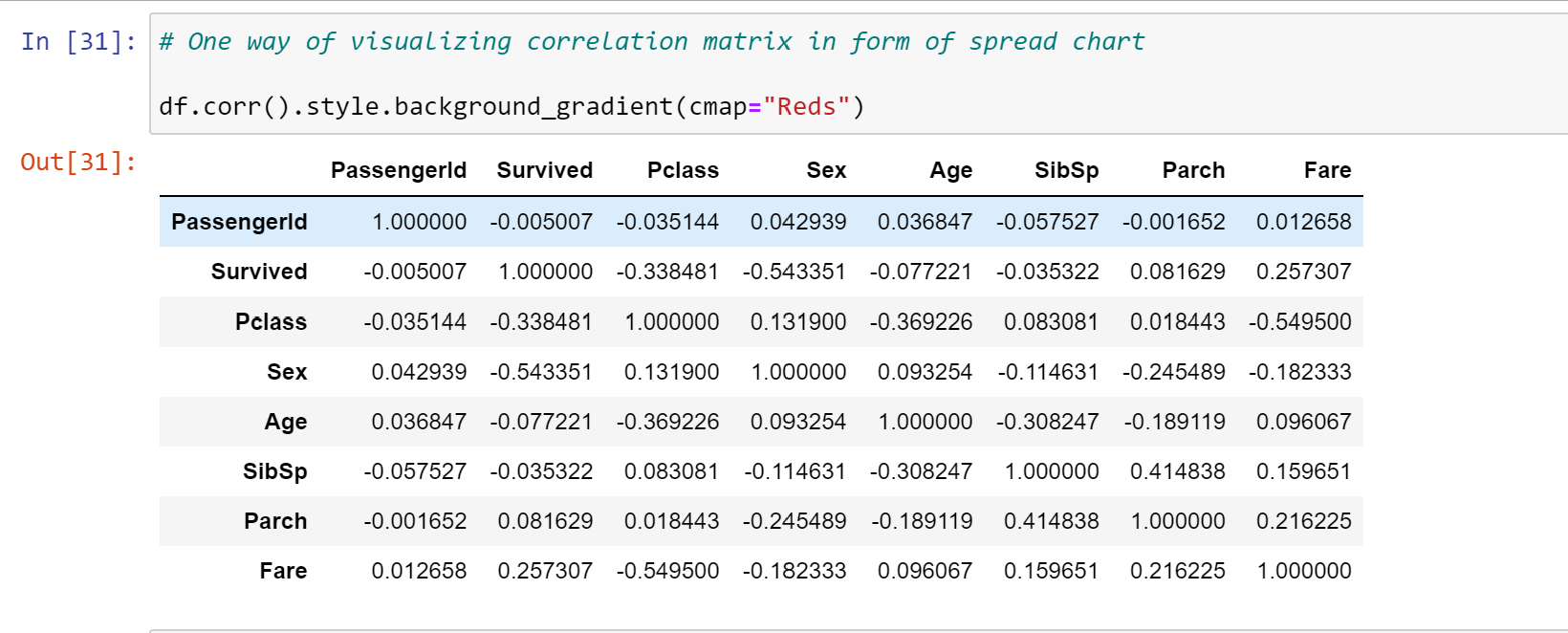


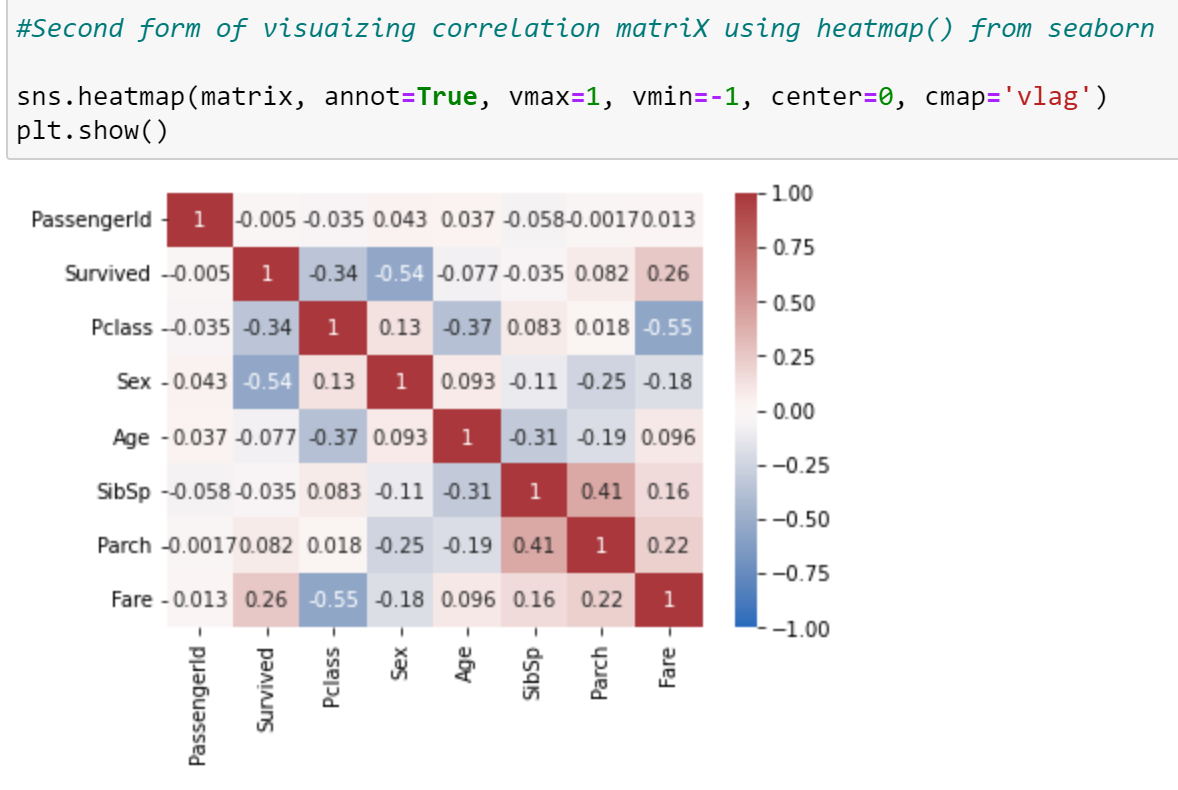
2.1. Find the correlation between ‘survived’ (target column) and ‘sex’ column for the Titanic use case in class. a. Do you think we should keep this feature?



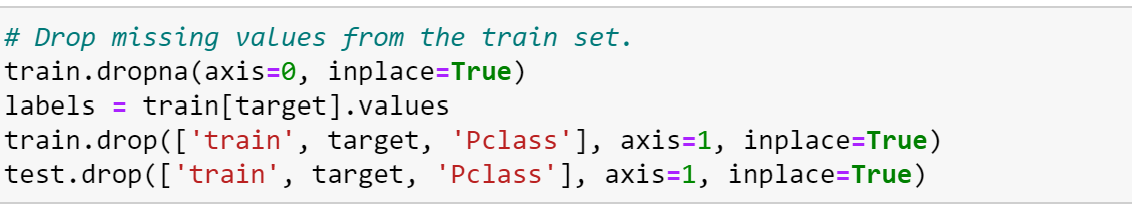


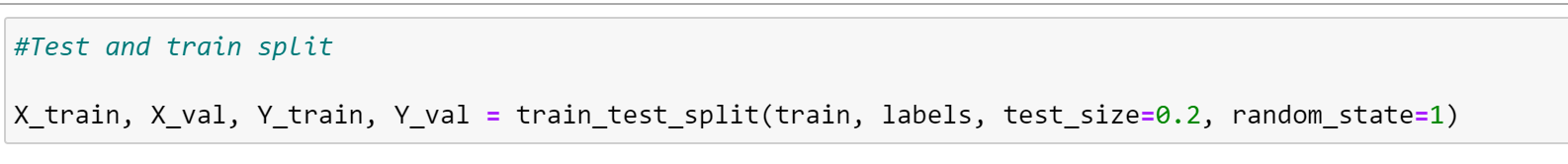
1. Do at least two visualizations to describe or show correlations.

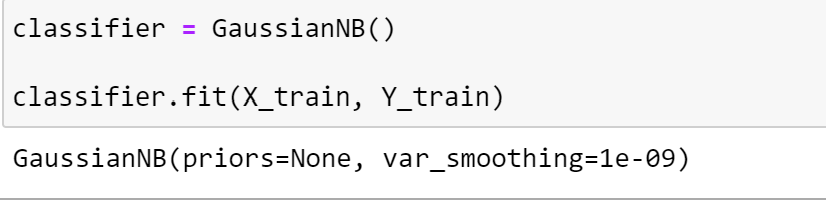


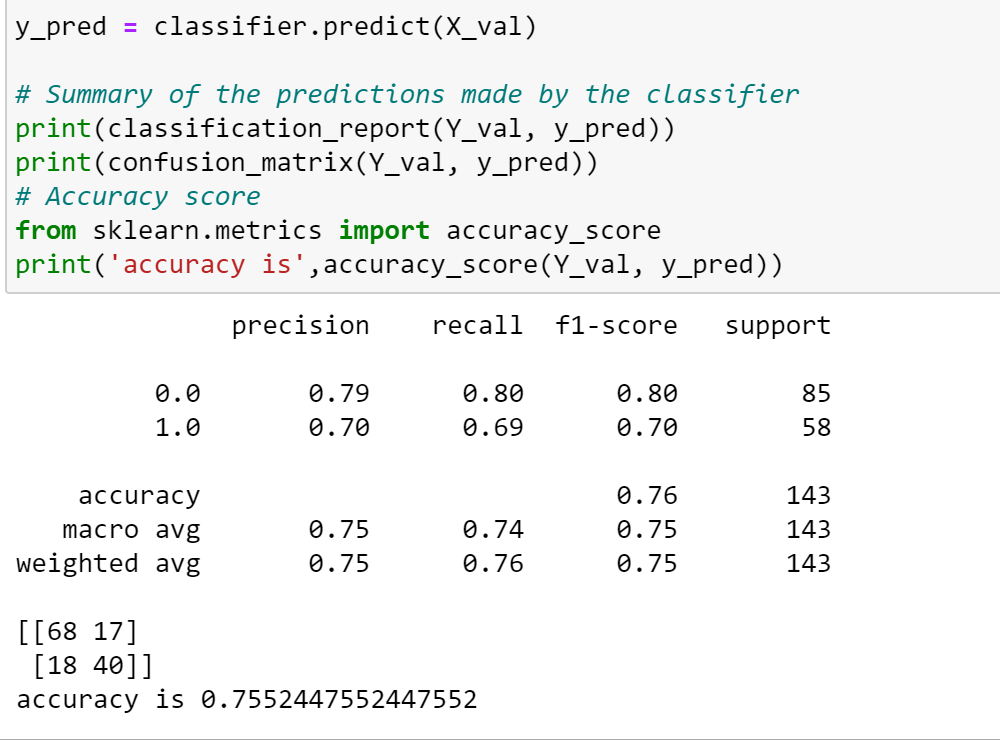


1. Implement Naïve Bayes method using scikit-learn library and report the accuracy.

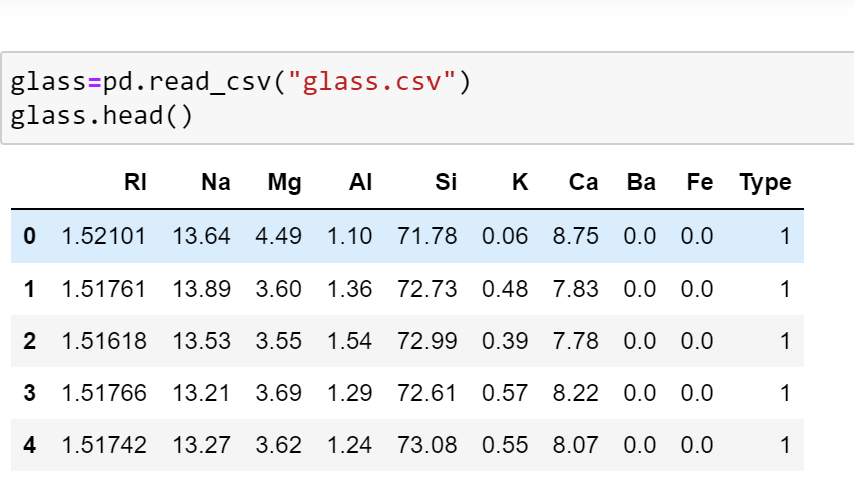
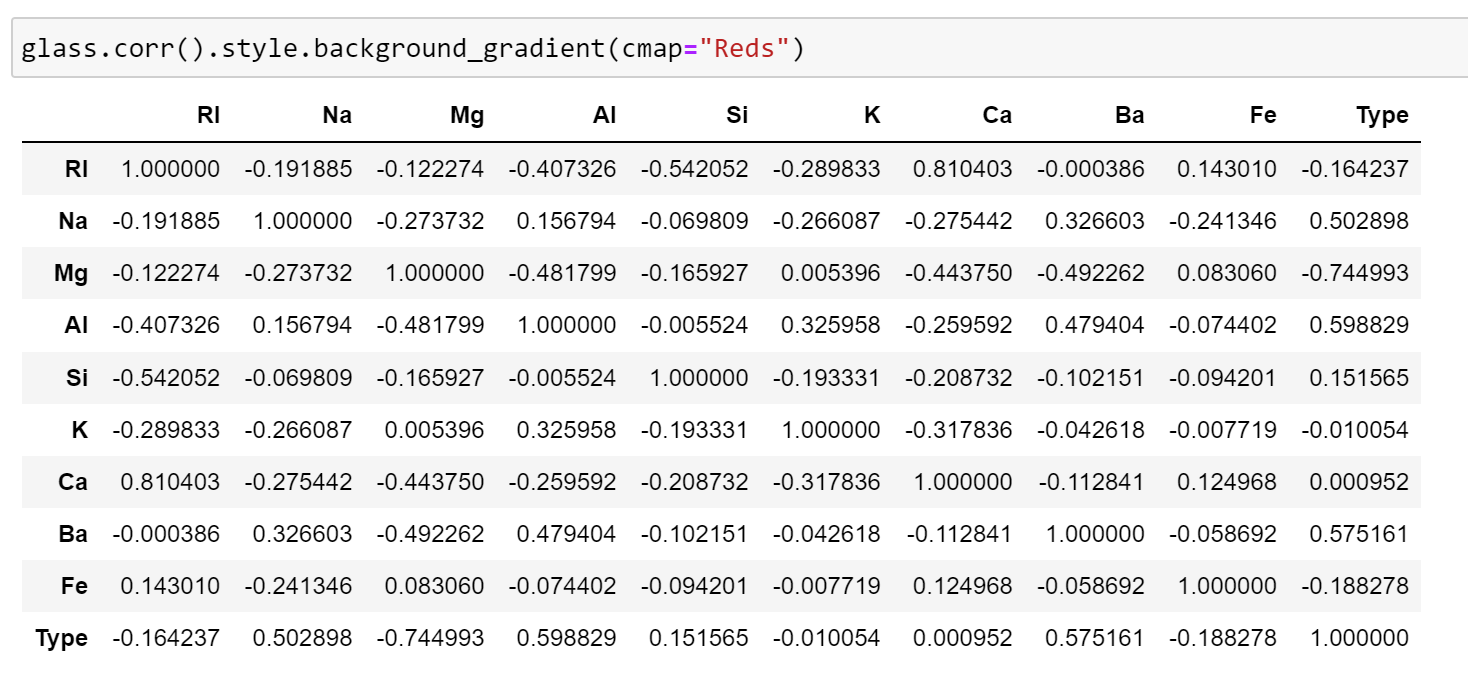
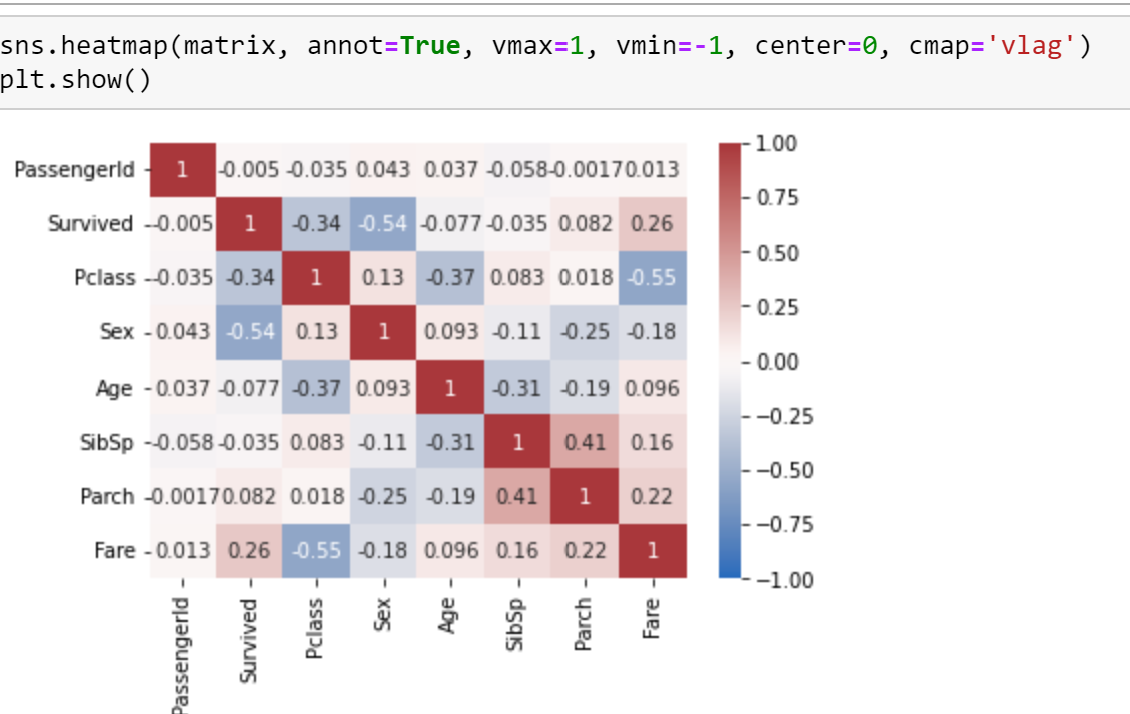
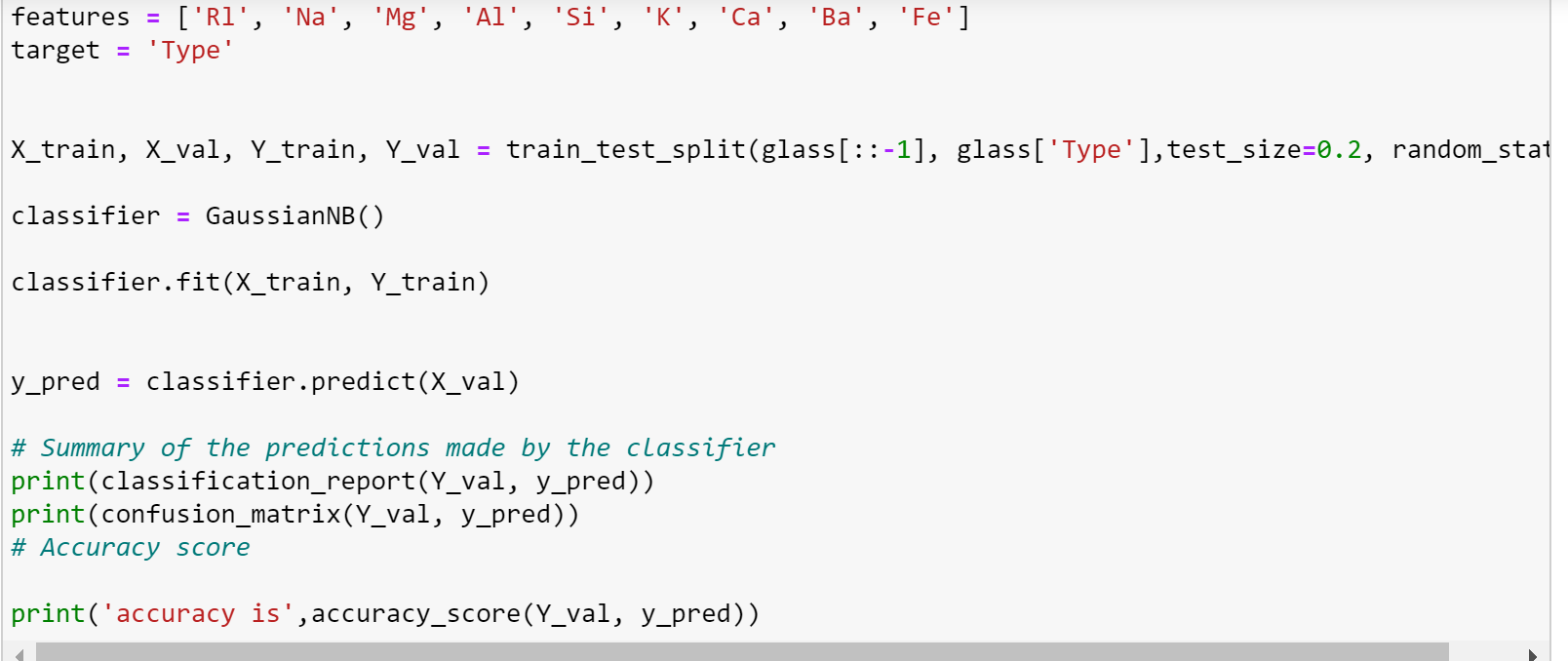
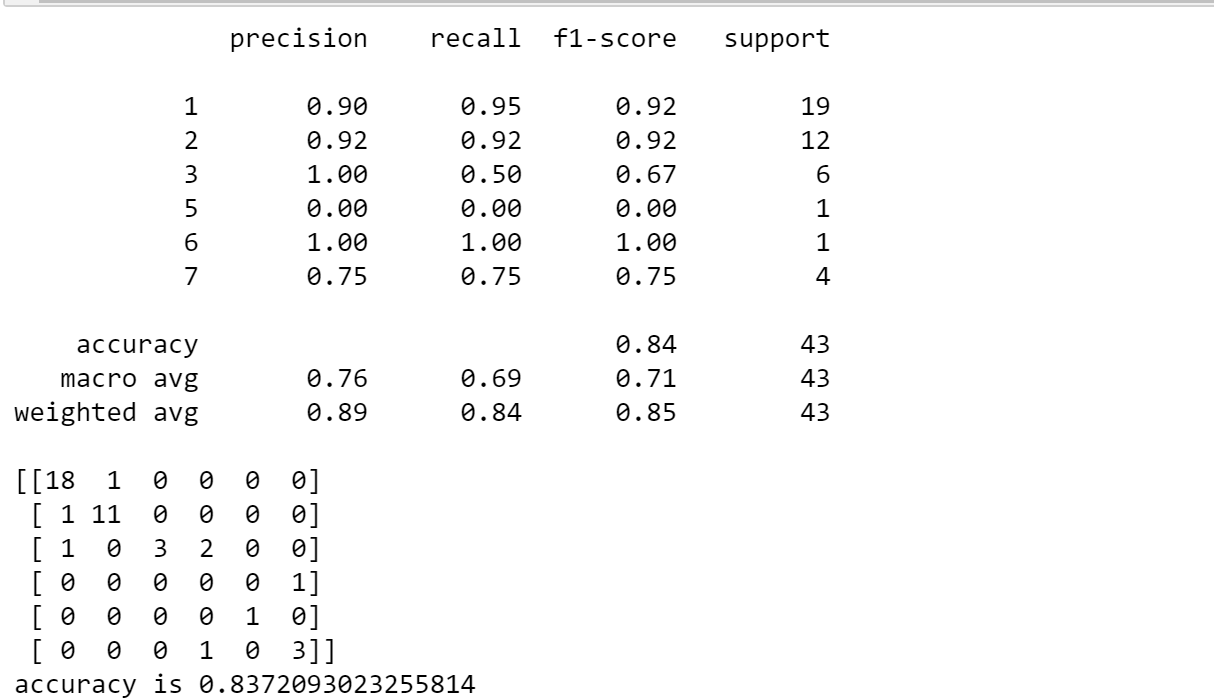
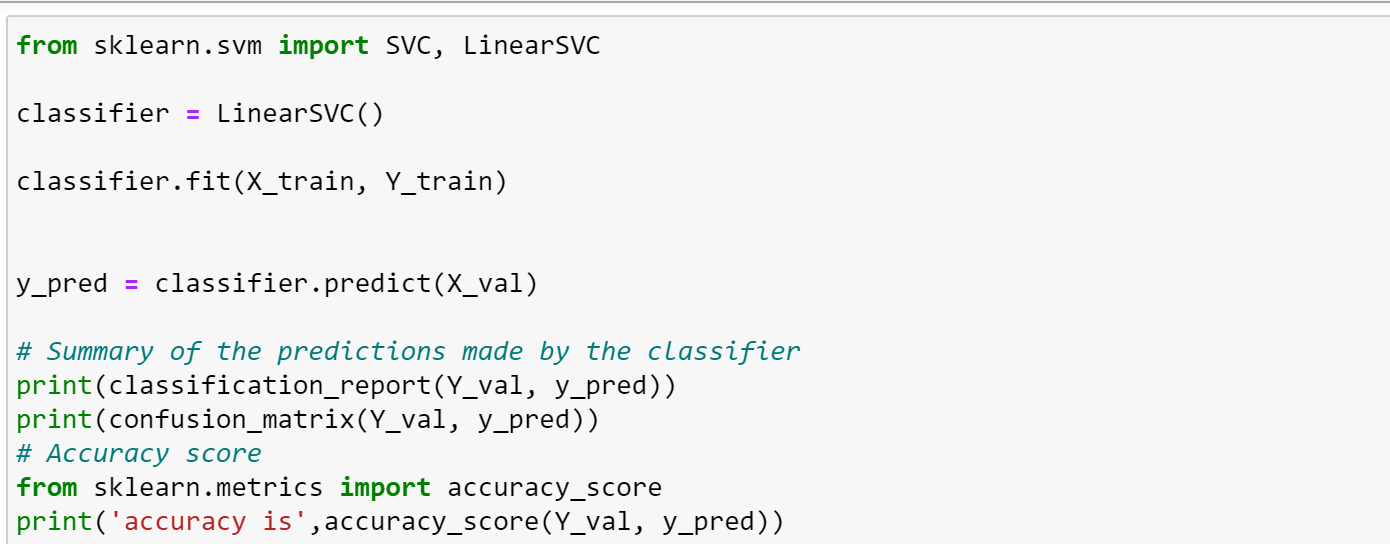
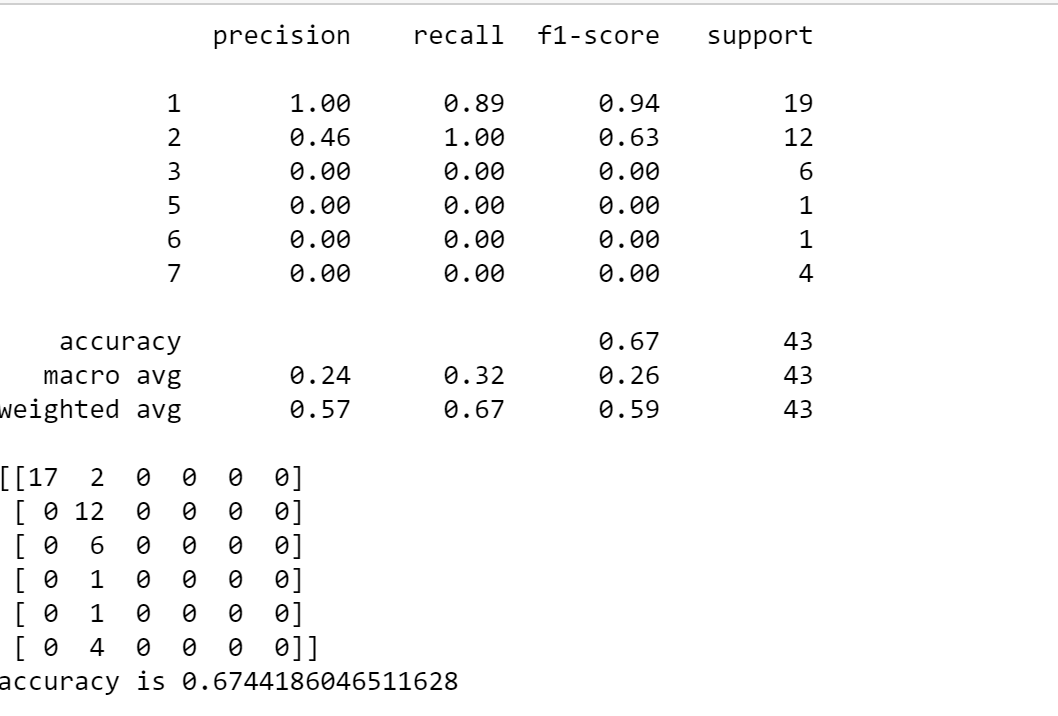








1. (Glass Dataset):
   * 1. Implement Naïve Bayes method using scikit-learn library.
2. Implement Naïve Bayes method using scikit-learn library. a. Use the glass dataset available in Link also provided in your assignment. b. Use train\_test\_split to create training and testing part. 2. Evaluate the 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?

**Justification:**

We got better accuracy for Naïve Bayes method which is 0.8372093023255814. Naive Bayes analysis works well with probabilistic concepts where as Linear SVM works better with linear regression logics. But to perform more accurately SVM requires large amounts of data to train and test the data. So, due to the amount of data Naive Bayes algorith gives better accuracy compared to Linear SVM.