# MACHINE LEARNING ASSIGNMENT4

GitHub Link : <https://github.com/Divakar24/Divakar-Assignment-4>

Video Link : https://github.com/Divakar24/Divakar-Assignment-4/blob/main/video1394567725.mp4

Question 1

1. Pandas

1. Read the provided CSV file ‘data.csv’. <https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>

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.

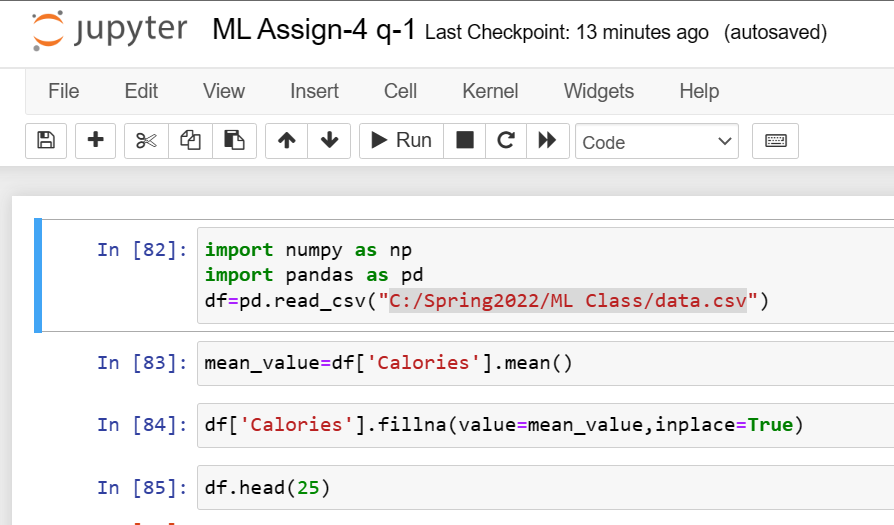
6. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.

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).



Table

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated

Graphical user interface, table

Description automatically generated

Chart, scatter chart

Description automatically generated

Question 2

1. (Titanic Dataset)

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

a. Do you think we should keep this feature?

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

3. Implement Naïve Bayes method using scikit-learn library and report the accurac

Text

Description automatically generated

Text

Description automatically generated with low confidence

Graphical user interface, text, application

Description automatically generated

Chart, bar chart

Description automatically generated

Graphical user interface, text

Description automatically generated

Table

Description automatically generated

Question 3

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.

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 and Do at least two visualizations to describe or show correlations in the Glass Dataset. Which algorithm you got better accuracy? Can you justify why?Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application, table

Description automatically generatedGraphical user interface, text, application, email

Description automatically generated

Chart

Description automatically generatedGraphical user interface, application

Description automatically generated  
Graphical user interface, text, application

Description automatically generated

Graphical user interface

Description automatically generated with low confidence

Table

Description automatically generated