Mannava Sravya

700747750

Semester: Spring

Course Id : CS 5710

CRN: 23922

GitHub Link: <https://github.com/Sravya-mannava/Assignment_4>

1. 1. Pandas

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

Graphical user interface, text, application, email

Description automatically generated

2. Show the basic statistical description about the data

Graphical user interface, text, application

Description automatically generated

3. Check if the data has null values.

a. Replace the null values with the mean

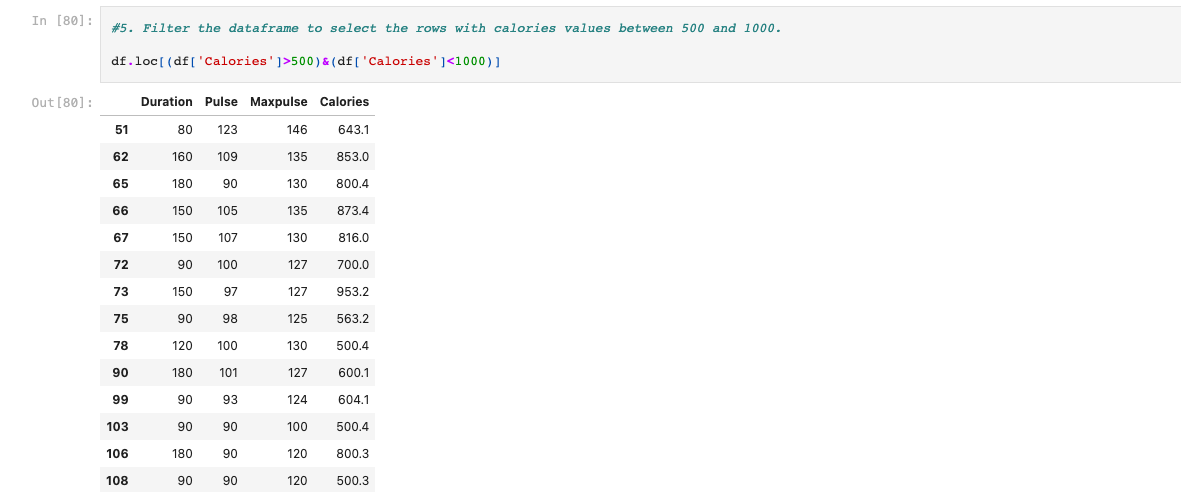
A picture containing application

Description automatically generated

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

Description automatically generated

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.

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface, text, application

Description automatically generated

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

Graphical user interface

Description automatically generated with medium confidence

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

Chart, scatter chart

Description automatically generated

2.(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?

Table

Description automatically generated

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

Chart

Description automatically generated

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.

Graphical user interface, application, Teams

Description automatically generated

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

Graphical user interface, text, application, Teams

Description automatically generated

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

Background pattern

Description automatically generated with medium confidence

1. Implement linear SVM method using scikit library

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

Graphical user interface, text, application, email

Description automatically generated

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

Graphical user interface, application

Description automatically generated with medium confidence

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

Chart, application

Description automatically generated