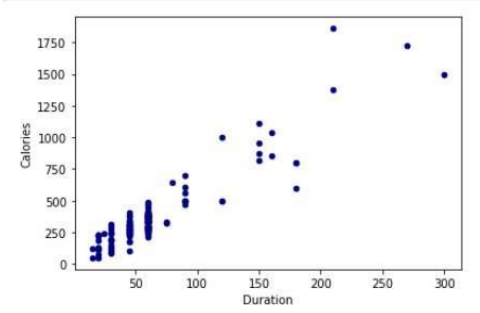
Fall 2023: CS5720 – NN &DL

In-Class Programming Assignment-4

1. Data Manipulation

a. Read the provided CSV file 'data.csv'.
b. https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing
c. Show the basic statistical description about the data.
d. Check if the data has null values.
    i. Replace the null values with the mean
e. Select at least two columns and aggregate the data using: min, max, count, mean.
f. Filter the dataframe to select the rows with calories values between 500 and 1000.
g. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.
h. Create a new "df_modified" dataframe that contains all the columns from df except for "Maxpulse".
i. Delete the "Maxpulse" column from the main df dataframe
j. Convert the datatype of Calories column to int datatype.
k. Using pandas create a scatter plot for the two columns (Duration and Calories).

Example



2. Linear Regression

a) Import the given "Salary_Data.csv"
b) Split the data in train_test partitions, such that 1/3 of the data is reserved as test subset.
c) Train and predict the model.
d) Calculate the mean_squared error
e) Visualize both train and test data using scatter plot.

** Follow the rubric guidelines.

Submission Guidelines:

1. Once finished document your code and make sure all parts if the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info.
3. Submit the assignment.
4. Present your work in class time to proof the execution and complete submission.

After class submission:

1. Once finished document your code and make sure all parts if the assignments are completed.
2. Push your code to your GitHub repo and update the ReadMe file, add your info.
3. Submit the assignment before the deadline.
4. Record a short video (1~3) minute, proof of execution and complete assignment.
5. Add video link to ReadMe file.