**Spring 2024: CS5720 Neural Networks & Deep Learning - ICP-4**

**Bhanu Chandrika Lakkimsetti (700747439)**

GitHub Link: <https://github.com/bhanuchandrika99/NNDL_ICP_4>

1. **Data Manipulation**
   * 1. Read the provided CSV file ‘data.csv’.
     2. <https://drive.google.com/drive/folders/1h8C3mLsso-R-sIOLsvoYwPLzy2fJ4IOF?usp=sharing>
     3. Show the basic statistical description about the data.

A screenshot of a computer

Description automatically generated

* + 1. Check if the data has null values.

1. Replace the null values with the mean

A screenshot of a computer

Description automatically generated

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

A screen shot of a computer

Description automatically generated

* + 1. Filter the dataframe to select the rows with calories values between 500 and 1000.

A screenshot of a computer

Description automatically generated

* + 1. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.

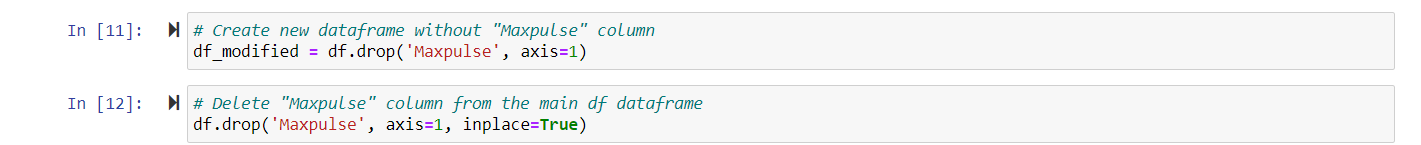
A screenshot of a computer

Description automatically generated

* + 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 dataframe



* + 1. Convert the datatype of Calories column to int datatype.
    2. Using pandas create a scatter plot for the two columns (Duration and Calories).

A screen shot of a graph

Description automatically generated

1. **Linear Regression**

* 1. Import the given “Salary\_Data.csv”
  2. 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.
  3. Calculate the mean\_squared error
  4. Visualize both train and test data using scatter plot.

A screenshot of a computer code

Description automatically generated  
A graph with red and black dots

Description automatically generated