**CS5720**

**Neural Networks & Deep Learning - ICP-4**

**Name: Sai Krishna Reddy Chevutukur**

**#700: 700743830**

**Username: SXC38300**

**Github Link**: <https://github.com/csk17/NNDL_Assignment4>

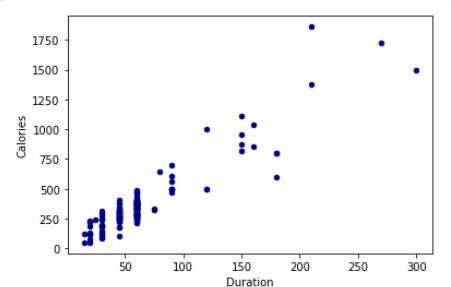
**1. Data Manipulation**

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.
4. Check if the data has null values.

i. Replace the null values with the mean

1. Select at least two columns and aggregate the data using: min, max, count, mean.
2. Filter the dataframe to select the rows with calories values between 500 and 1000.
3. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.
4. Create a new “df\_modified” dataframe that contains all the columns from df except for “Maxpulse”.
5. Delete the “Maxpulse” column from the main df dataframe
6. Convert the datatype of Calories column to int datatype.
7. Using pandas create a scatter plot for the two columns (Duration and Calories).

Example



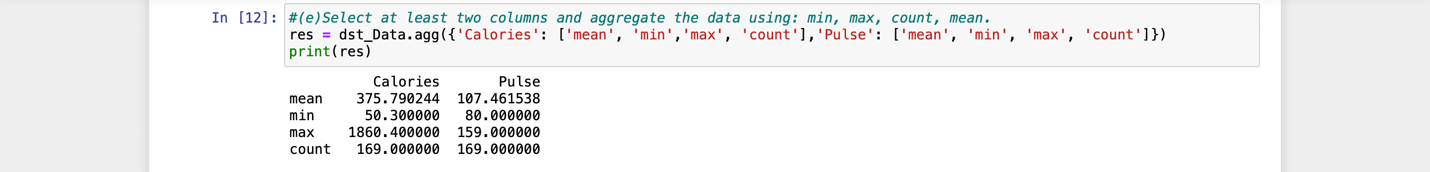
**Solution:**

* The provided CSV file ‘data.csv’ is read.
* The statistical description in the file is printed.
* The null values from the data frame are checked and replaced with the mean values.

**Graphical user interface

Description automatically generated with low confidenceGraphical user interface, text, application

Description automatically generatedTable

Description automatically generatedTable

Description automatically generated with medium confidenceGraphical user interface, text

Description automatically generatedGraphical user interface, application

Description automatically generated**

**2. Linear Regression**

1. a)  Import the given “Salary\_Data.csv”
2. b)  Split the data in train\_test partitions, such that 1/3 of the data is reserved as test subset.
3. c)  Train and predict the model.
4. d)  Calculate the mean\_squared error
5. e)  Visualize both train and test data using scatter plot.

**Solution:**

* The Salary\_Data.csv is imported.
* The data is splitted in the train & test partitions where 1/3 of the data is reserved for test subset.
* The data is trained & predicted.
* The mean squared error is calculated.
* The train & test data is visualized using scatter plot.

**Graphical user interface, text, application

Description automatically generated**

**Graphical user interface, application

Description automatically generated**