**In class Programming Assignment - 4**

**GitHub Link:** <https://github.com/Viditha2277/CS-5720/tree/main/Assignment%204>

**Problem 1:** To demonstrate Data Manipulation using Pandas in Python

**Input**: .csv file

**Solution:**

1. Reading the data.csv file using read\_csv() and showing basic statistical description of the data using describe().

2. Checking if there are any null values in the data using. If there are null values, replacing the values with mean of the column.

3. Aggregating the “Duration” and ‘Pulse” data using agg().

4. Displaying data that has “Calories” column values between 500 and 1000.

5. Displaying data that has “Calories” column values less than 500 and greater than 1000.

6. Displaying all the columns except “Maxpulse” using loc().

7. Deleting the “Maxpulse” column from the dataframe using drop().

8. Converting the column data type to int data type of “Calories” column using astype().

9. A scatter plot is drawn for the columns “Duration” and “Calories”.

**Code:**

A screen shot of a computer

Description automatically generated

A screen shot of a computer

Description automatically generated

**Problem 2:** To demonstrate Linear Regression on the Salary\_Data.csv file

**Input:** .csv file

**Solution:**

1. Reading the data.csv file using read\_csv() and showing basic statistical description of the data using describe().

2. Entire data is spilt into training and testing data using train\_test\_splot() method imported from sklearn module.

3. Linear regression model is imported from sklearn and loaded into “regressor” object. Then the training data and test data is used to tune the weights of the model.

4. Using predict() method, values are predicted for the test data.

5. Mean squared error is calculated between the predicted and actual values.

6. Scatter plots are drawn for training and testing data.

**Code:**

A screen shot of a computer

Description automatically generated

A black screen with white text

Description automatically generated