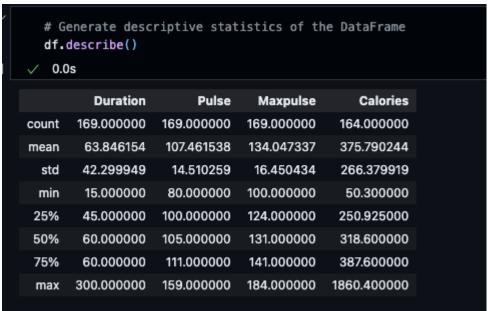
Machine Learning - CS5710 Assignment 1 Rishma Reddy Nalla 700752916

GitHub Link: https://github.com/RishmaReddy-Nalla/CS-5710/tree/main/Assignment

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

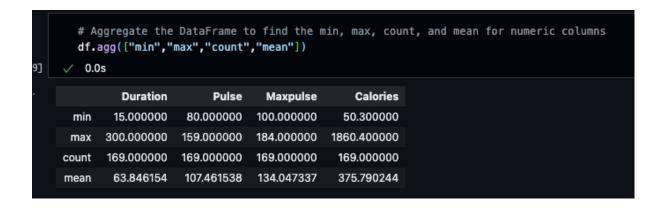
2. Show the basic statistical description about the data.



- 3. Check if the data has null values.
- a. Replace the null values with the mean

```
# Check for missing values in the DataFrame
   df.isnull()
 ✓ 0.0s
       Duration
                 Pulse
                        Maxpulse Calories
   0
          False
                 False
                             False
                                       False
   1
          False
                 False
                             False
                                       False
          False
   2
                 False
                             False
                                       False
   3
          False
                             False
                 False
                                       False
   4
          False
                 False
                             False
                                       False
 164
          False
                 False
                             False
                                       False
 165
          False
                 False
                             False
                                       False
 166
          False
                             False
                 False
                                       False
 167
          False
                 False
                             False
                                       False
 168
          False
                 False
                             False
                                       False
169 rows x 4 columns
   # Fill missing values with the mean of each column
   df = df.fillna(df.mean())
 ✓ 0.0s
   # Check if there are any missing values left in the DataFrame
   df.isna().any()
 ✓ 0.0s
Duration
             False
Pulse
             False
Maxpulse
             False
Calories
             False
dtype: bool
```

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



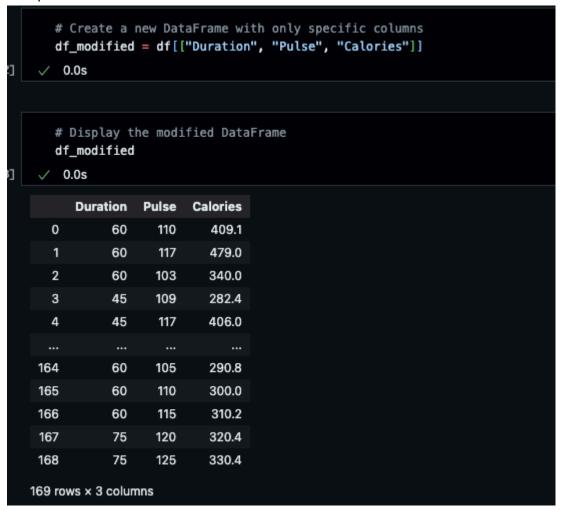
5. Filter the dataframe to select the rows with calories values between 500 and 1000

			rame for r > 500) &	
	0.0s			
	Duration	Pulse	Maxpulse	Calories
51	80	123	146	643.1
62	160	109	135	853.0
65	180	90	130	800.4
66	150	105	135	873.4
67	150	107	130	816.0
72	90	100	127	700.0
73	150	97	127	953.2
75	90	98	125	563.2
78	120	100	130	500.4
90	180	101	127	600.1
99	90	93	124	604.1
103	90	90	100	500.4
106	180	90	120	800.3
108	90	90	120	500.3

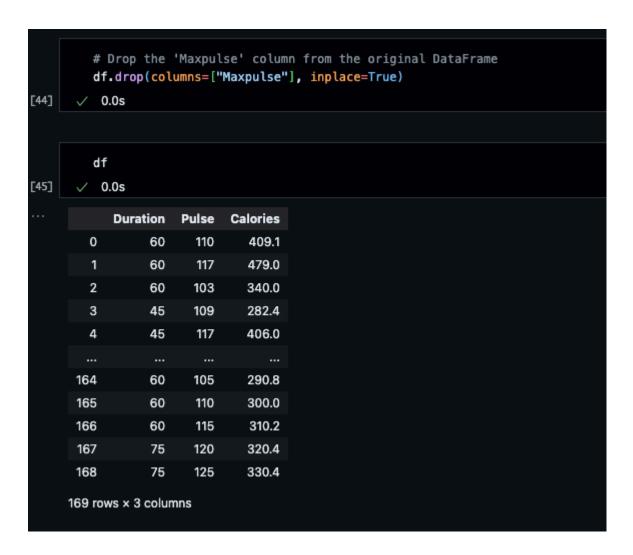
6. Filter the dataframe to select the rows with calories values > 500 and pulse < 100.



7. Create a new "df_modified" dataframe that contains all the columns from df except for "Maxpulse".



8. Delete the "Maxpulse" column from the main df dataframe



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

```
# Print data types before conversion
       print("Data Types before conversion")
       print(df.dtypes)
      # Convert the 'Calories' column to integer type
      df['Calories'] = df["Calories"].astype(int)
      # Print data types after conversion
      print("\n","Data Types after Conversion")
       print(df.dtypes)
-8] 🗸 0.0s
   Data Types before conversion
   Duration int64
   Pulse
               int64
   Calories
               int64
   dtype: object
    Data Types after Conversion
               int64
   Duration
   Pulse
               int64
   Calories
               int64
   dtype: object
```