MACHINE LEARNING

GITHUB LINK: [https://github.com/sparsha21/sparsha21](https://github.com/sparsha21/sparsha21%20)

RECORDING VIDEO LINK: <https://drive.google.com/file/d/1qwVv6_tNlxQcIpKF6mJa-n_nqv8sVLSl/view?usp=sharing>

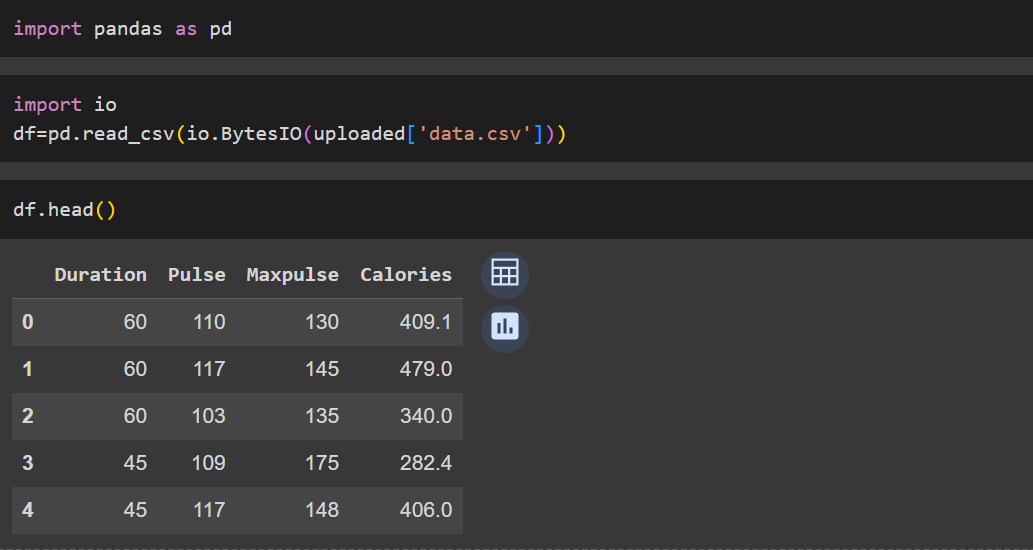
NAME:SPARSHA REDDY ADURI

700764939

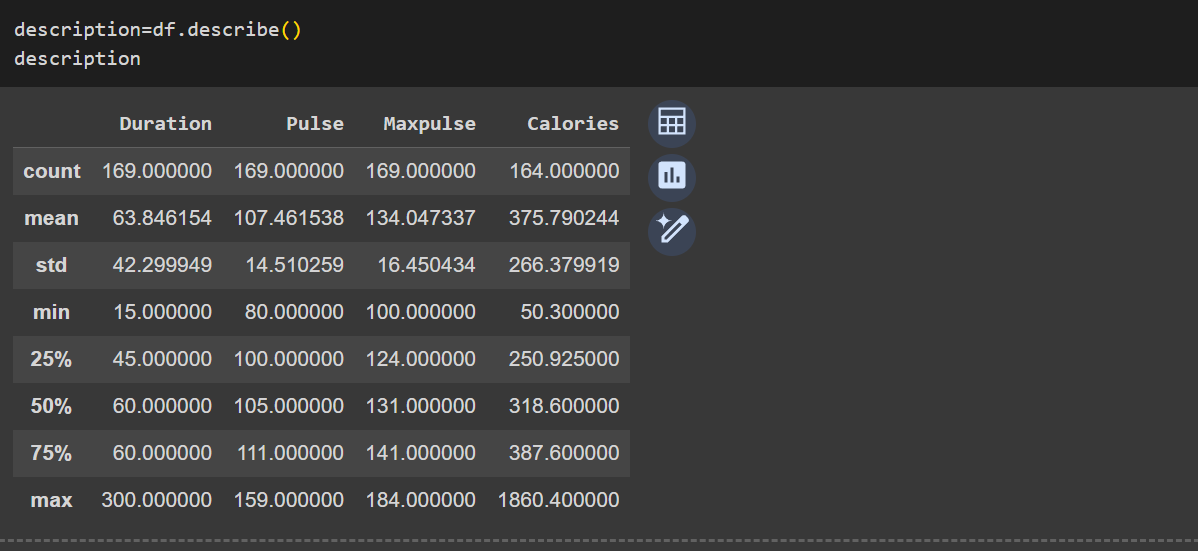
Q1. Read the provided CSV file ‘𝒅𝒂𝒕𝒂. 𝒄𝒔𝒗’.

[https://drive.google.com/file/d/1-Ir3AXK1A77A-qCDu5gGkAxv- nbmWlHO/view?usp=drive\_link](https://drive.google.com/file/d/1-Ir3AXK1A77A-qCDu5gGkAxv-%20%20%20nbmWlHO/view?usp=drive_link)

Here in the code, I have used 𝒓𝒆𝒂𝒅\_𝒄𝒔𝒗( ) and 𝒉𝒆𝒂𝒅( ) functions to read CSV file and to display the first 5 rows of the 𝐷𝑎𝑡𝑎𝐹𝑟𝑎𝑚𝑒 respectively.

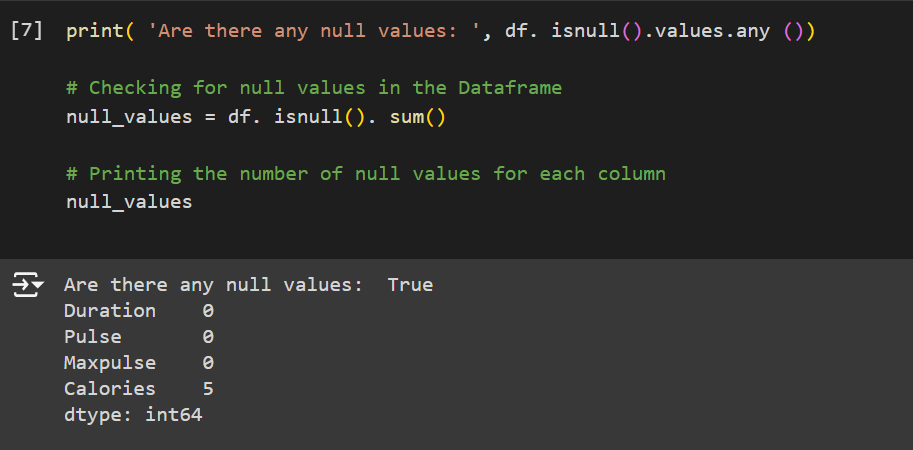


Q2. Here in the code, I have used the 𝒅𝒆𝒔𝒄𝒓𝒊𝒃𝒆( ) function to generate descriptive statistics of a 𝐷𝑎𝑡𝑎𝐹𝑟𝑎𝑚𝑒. Then the description is printed.

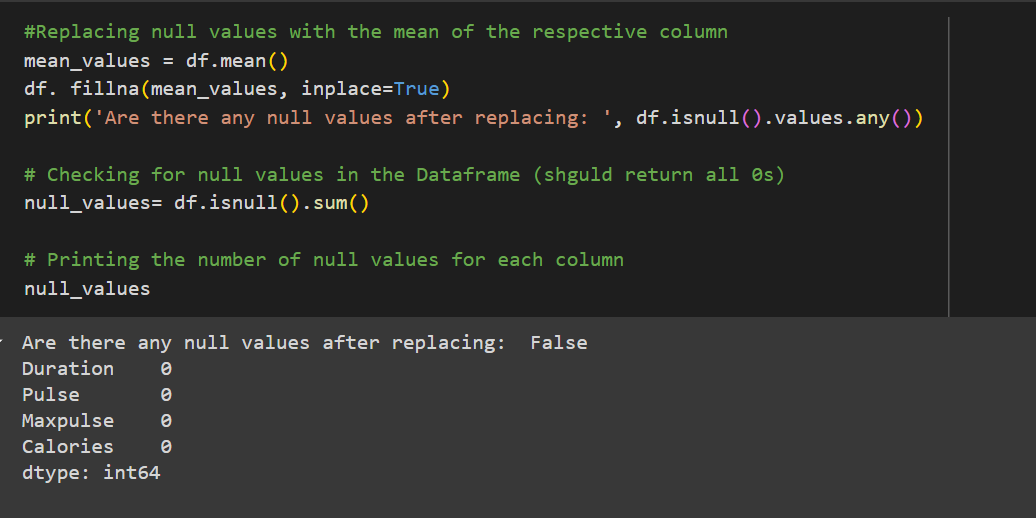


Q3. Here in the code, I have used the 𝒊𝒔𝒏𝒖𝒍𝒍( ) function to identify null values in the 𝐷𝑎𝑡𝑎𝐹𝑟𝑎𝑚𝑒, and the 𝒗𝒂𝒍𝒖𝒆𝒔. 𝒂𝒏𝒚( ) function to check if any of the values are null.

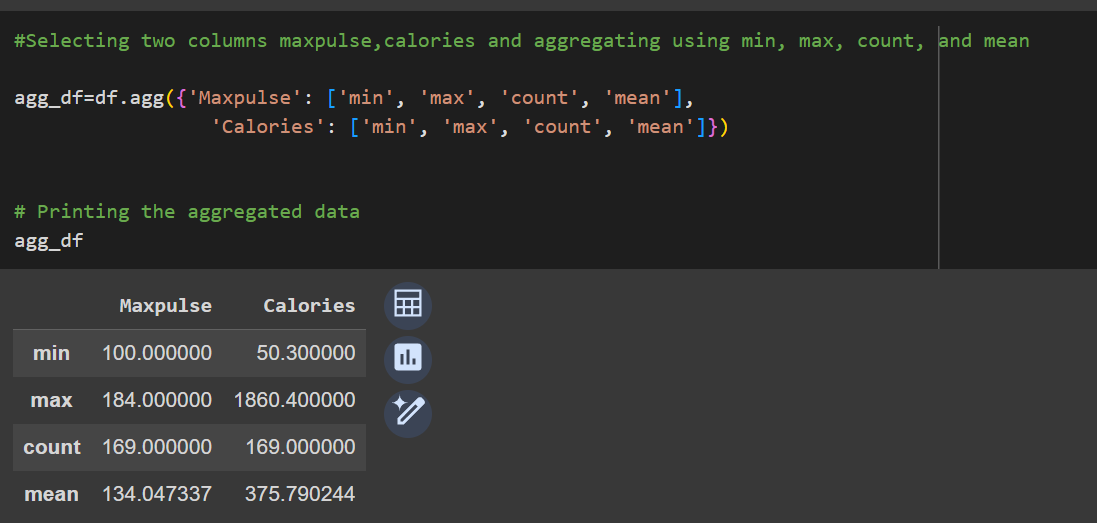
Then 𝒔𝒖𝒎( ) function is called on this Boolean DataFrame and displayed a new DataFrame containing the sum of 𝑻𝒓𝒖𝒆 values for each column in 𝒅𝒇.



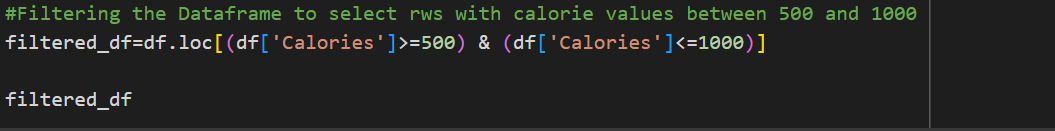
Q3(a). Replace the null values with the mean.



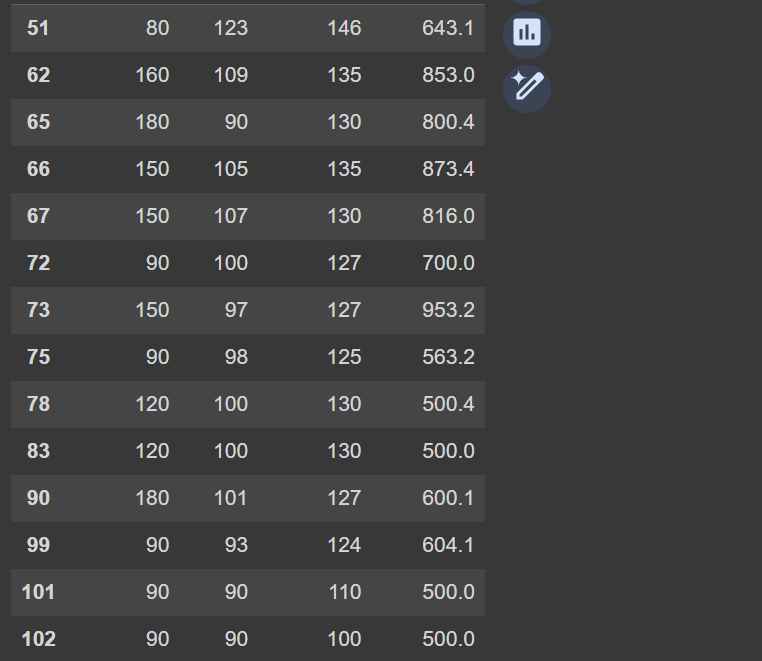
Q4. I have used 𝒂𝒈𝒈( ) function to perform the aggregation operations on the selected columns, with a dictionary as an argument, where the keys are the column names to be aggregated and the values are lists of aggregation functions to be applied to each column.



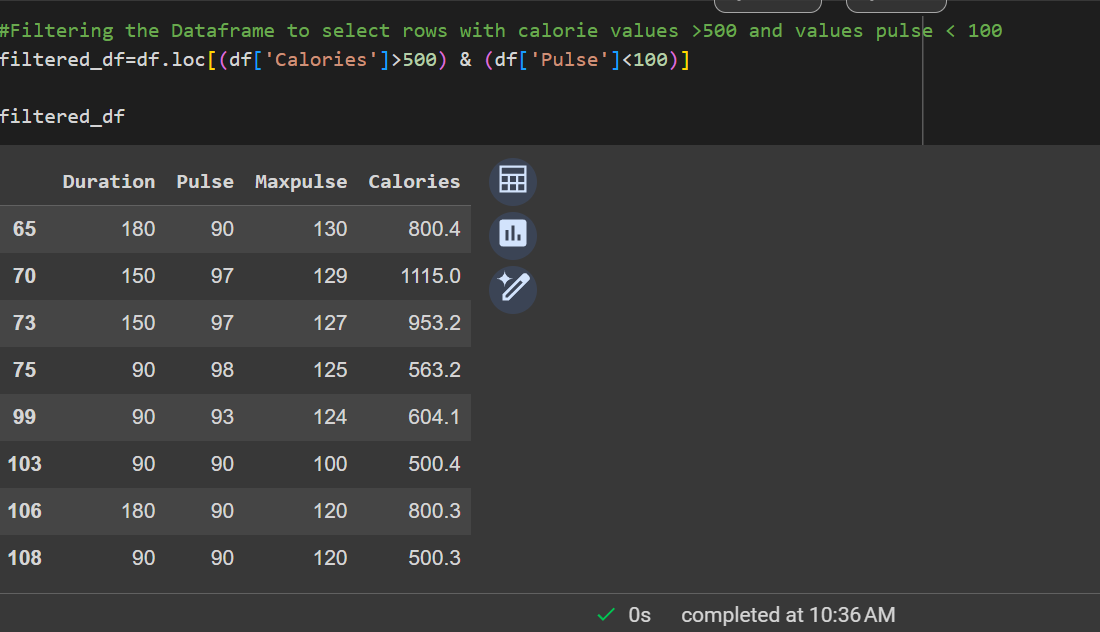
Q5. The 𝒍𝒐𝒄[ ] function is used to select rows based on a boolean condition. The condition is specified inside the square brackets of the  𝒍𝒐𝒄[ ] function using the '𝐶𝑎𝑙𝑜𝑟𝑖𝑒𝑠' column of 𝑑𝑓. Specifically, 𝒅𝒇[′𝑪𝒂𝒍𝒐𝒓𝒊𝒆𝒔′]  >=  𝟓𝟎𝟎 and 𝒅𝒇[′𝑪𝒂𝒍𝒐𝒓𝒊𝒆𝒔′]  <=  𝟏𝟎𝟎𝟎 are two separate boolean conditions, which are combined using the & operator to specify the filter condition.



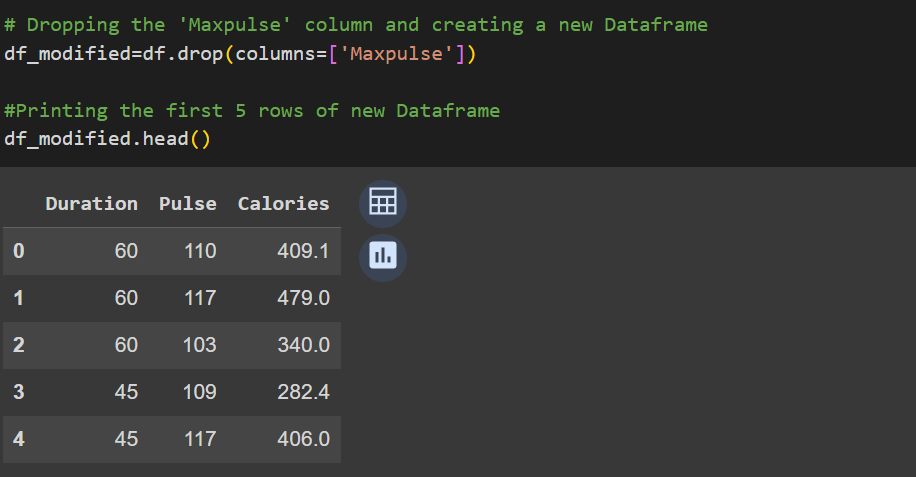
OUTPUT:



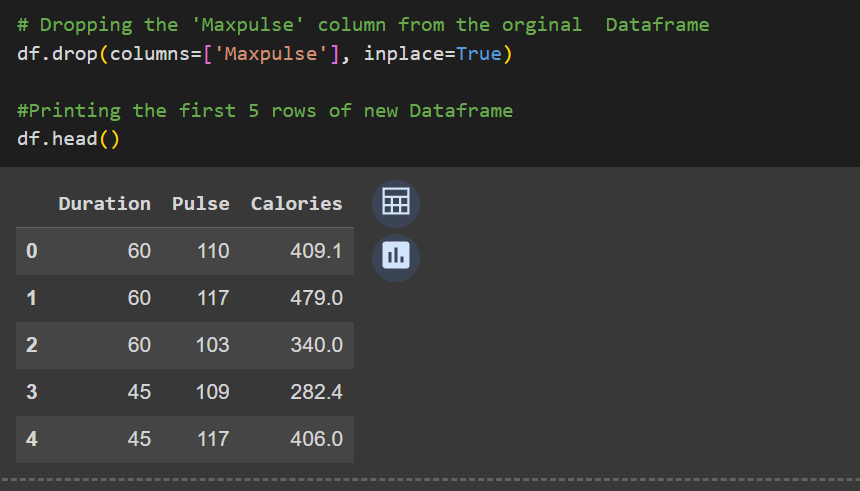
Q6. Filter the DataFrame to select the rows with calories values > 500 and pulse < 100.



Q7. Create a new “𝑑𝑓\_𝑚𝑜𝑑𝑖𝑓𝑖𝑒𝑑” dataframe that contains all the columns from 𝑑𝑓 except for “𝑀𝑎𝑥𝑝𝑢𝑙𝑠𝑒”.



Q8. Delete the “𝑀𝑎𝑥𝑝𝑢𝑙𝑠𝑒” column from the main 𝑑𝑓 dataframe.



Q9. Convert the datatype of 𝐶𝑎𝑙𝑜𝑟𝑖𝑒𝑠 column to 𝑖𝑛𝑡 datatype.

