

Name: Gautam Gupta

Title: Health report related to calories Burned

Python project (Jupyter notebook)

1. Import necessary libraries:

``pandas`` for data manipulation and analysis.

``numpy`` for numerical operations.

``seaborn`` for creating informative and aesthetically pleasing statistical graphics.

``matplotlib.pyplot`` for plotting.

2. Set up environment:

``%matplotlib inline`` tells Jupyter Notebook (a common environment for Python code) to display plots directly within the notebook.

``import warnings`` allows you to control how warning messages are handled.

3. Load the data:

``import statsmodels.api as sm`` imports a library for statistical modeling.

``df=pd.read_csv(...)`` reads a CSV file containing data about cardio activities. The path to the file is specified.

4. Store the data:

``df`` is a variable that stores the data read from the CSV file. It's now ready to be used for analysis or visualization.

1. Data Preparation:

It starts by cleaning up a DataFrame (a table-like data structure) by removing unnecessary columns like "Friend's Tagged," "Route Name," etc.

It examines the data types of the remaining columns (e.g., dates, distances, heart rate, etc.).

2. Filtering Data:

The next three lines create new DataFrames:

``df_run = df [df ['Type'] == 'Running'].copy()``: Creates a DataFrame named 'df_run' containing only the rows where the 'Type' column has the value 'Running'.

``df_walk = df [df ['Type'] == 'Walking'].copy()``: Creates a DataFrame named 'df_walk' containing only the rows where the 'Type' column has the value 'Walking'.

``df_run = df [df ['Type'] == 'Cycling'].copy()``: Creates a DataFrame named 'df_run' containing only the rows where the 'Type' column has the value 'Cycling'.

3. Viewing the Null Data:

``pd.isnull(df).sum()``: This checks for missing values in the dataset and counts how many are present for each column.

4. Remove Null Values [df=df.dropna()]

: Identify and eliminate any rows in your DataFrame that contain null (missing) values. This helps ensure that your analysis is based on complete data.

5. Remove Duplicate Values (unique): Check for and remove any duplicate rows in your DataFrame. This step is crucial for maintaining the integrity of your data and preventing biased analysis.

6. Change Data Type

[df['Type']=df['Type'].replace(['Other'],'Unicycling')]: Adjust the data types of specific columns as needed. For instance, you might want to convert a column of numbers stored as strings into integers, or change a text column into a categorical type for more efficient memory usage.

By following these steps, we can clean and prepare your data effectively for further analysis.

1. Data Description function :

The first line `df['Average Heart Rate (bpm)'].describe()` takes a column named 'Average Heart Rate (bpm)' from a DataFrame called 'df' and calculates some basic statistics:

Count: The number of valid heart rate readings.

Mean: The average heart rate.

Standard Deviation (std): A measure of how spread out the heart rate values are.

Minimum (min): The lowest heart rate recorded.

25th Percentile: The value below which 25% of the heart rates fall.

50th Percentile (Median): The middle heart rate value.

75th Percentile: The value below which 75% of the heart rates fall.

Maximum (max): The highest heart rate recorded.

2. Groupby function:

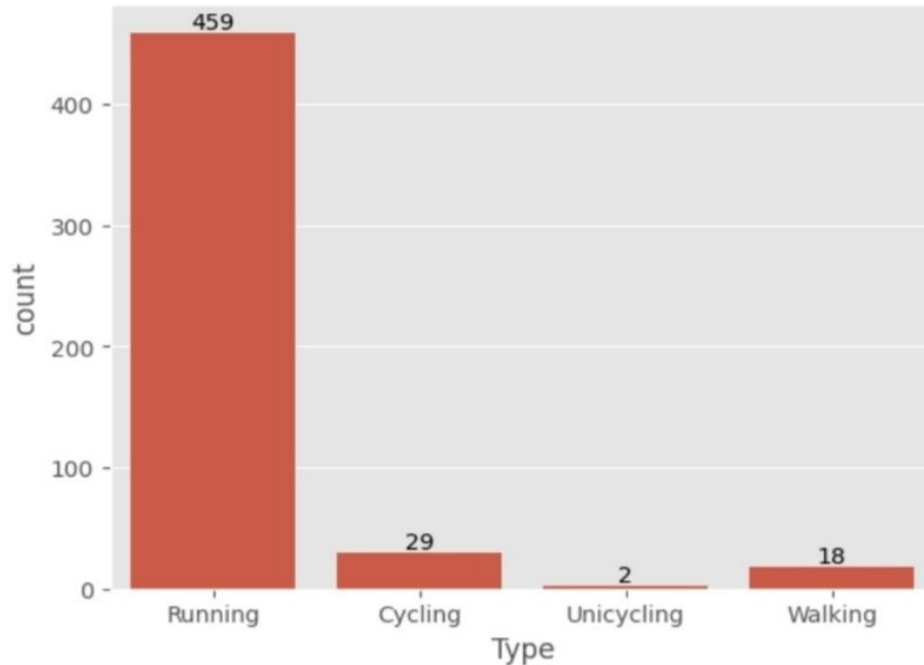
This groups the data by the exercise type ('Type') and calculates the sum of 'Calories Burned' and 'Distance (km)' for each type.

```
groupby_type=df.groupby(['Type']) [['Calories Burned', 'Distance (km)']].sum()`:
```

- Cycling: The lowest calorie burn, but the largest distance traveled.
- Running: The highest calorie burn, and the second largest distance traveled.
- Unicycling: The second lowest calorie burn, and the shortest distance traveled.
- Walking: The second highest calorie burn, and the second shortest distance traveled.

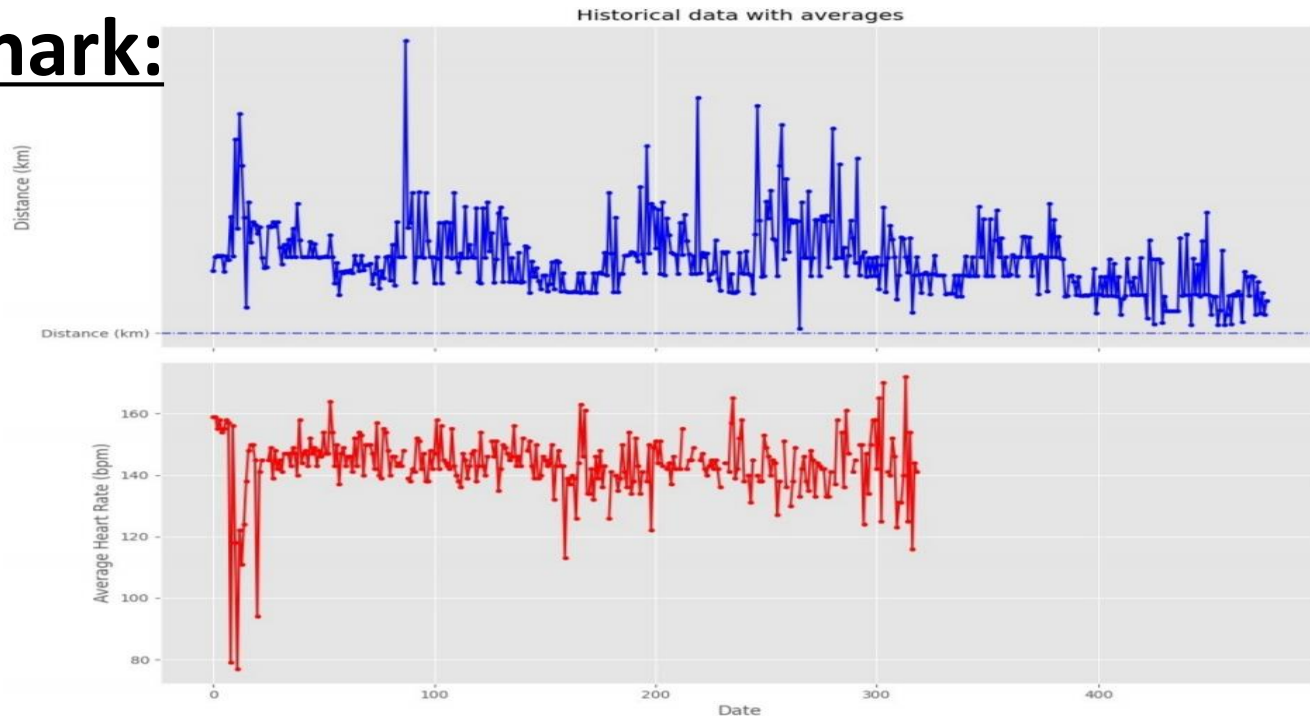
It seems like running burns the most calories, but that could be due to a longer time spent running compared to the other activities. To get a better sense of the calorie burn per unit of distance, you could calculate calories burned per kilometer.

Bars chark:



This bar chart shows the number of times each type of activity was recorded. Running is the most frequent activity at 459 times, followed by cycling at 29. Unicycling is the least frequent at 2, and walking is in the middle at 18.

Lines Chart:



The graphic shows the trend of running distance and the average heart rate over time.

The blue line represents the trend of running distance, which is increasing over time. This suggests that the runner is getting better at running longer distances.

The red line represents the average heart rate, which is relatively stable over time. This suggests that the runner is not overexerting themselves and is maintaining a consistent effort level.

The graph suggests that the runner is improving their running performance over time. They are running longer distances while maintaining a steady heart rate, which indicates that they are becoming more efficient.

Data:

`Total_distance_run`: This variable stores the total distance you've run (5224 km).

`number_of_shoes_used`: This variable stores the total number of running shoes you've used (7 pairs).

`forrest_total_distance`: This variable stores the total distance Forrest Gump is assumed to have run (24700 km).

Calculations:

1. `average_shoes_lifetime`: Calculates the average distance a pair of your running shoes lasts. It divides the `total_distance_run` by `number_of_shoes_used` (resulting in 746.29 km per pair).

2. 'shoes_for_forrest_run': Calculates the number of shoe pairs Forrest Gump would need. It divides ``forrest_total_distance`` by `average_shoes_lifetime``.

3. Output: Finally, the code prints a message indicating the calculated number of shoe pairs Forrest Gump would need (rounded to the nearest whole number, which is 33).

The data you provided summarizes different activities along with the calories burned and distances covered for each:

1. **Cycling**:

- **Calories Burned**: 16,196
- **Distance**: 680.58 km

2. **Running**:

- **Calories Burned**: 6,727,389
- **Distance**: 5,224.50 km

3. **Unicycling**:

- **Calories Burned**: 2,239
- **Distance**: 34.45 km

4. **Walking**:

- **Calories Burned**: 2,795,414
- **Distance**: 33.45 km

Explanation:

- **Calories Burned**: This value indicates the total energy expenditure associated with each activity. Running has the highest calories burned, which is typical due to its intensity and duration.
- **Distance**: This represents the total distance traveled for each activity. Running also has a significant distance, reflecting a sustained effort over time.

Insights:

- Running is the most calorie-intensive activity here, likely due to the longer distance covered.
- Cycling, while it has a lower calorie burn than running, covers a substantial distance.
- Unicycling and walking have lower calorie burns and distances compared to the other activities, indicating less overall energy expenditure or shorter durations.

This data can help in understanding the effectiveness of different exercises for calorie burning and fitness goals.

Thank you..