Prep Insta Week 8 Captsone Project

Fitbit Consumer Behavior Analysis

Done by Manoj Raju Pericherla (mperiche@gitam.in)

Summary :-

The business task is to analyze the FitBit Fitness Tracker App data, which was collected through a distributed survey on Amazon Mechanical Turk between March 12, 2016, and May 12, 2016. The dataset includes personal tracker data from thirty eligible Fitbit users, covering minute-level output for physical activity, heart rate, and sleep monitoring. The data is distributed across 18 different files, such as dailyActivity, dailyCalories, hourlySteps, etc.

Data Sources :-

Given datasets include :-

- **Daily Activity merged.csv**
- **Daily Calories merged.csv**
- **Daily Intensities merged.csv**
- **Daily Steps merged.csv**
- **Heart Rate Seconds merged.csv**

- **Hourly Calories merged.csv**
- **Hourly Intensities merged.csv**
- **Hourly Steps merged.csv**
- **Minute Calories Narrow merged.csv**
- **Minute Calories Wide merged.csv**
- **Minute METs Narrow merged.csv**
- **Minute Sleep merged.csv**
- **Minute Steps Narrow merged.csv**
- **Minute Steps Wide merged.csv**
- **Sleep_day merged.csv**
- **Weight LogInfo merged.csv**

Cleaning and Manipulation of Data :-

Merged Datasets :-

In Colab

- **df1 (dailyactivity_merged.csv)**
- **merged_df = [df2(hourlycalories_merged.csv) + df3(hourlyintensities_merged.csv) + df4(hourlysteps_merged.csv)]**
- **df5 (sleepday_merged.csv)**
- **df6 (weightloginfo_merged.csv)**

- **df7(heartrate_merged.csv)**

In Tableau

- **Per Min Avg Calories = [minutecaloriesNarrow_merged.csv + minutecalorieswide_merged.csv]**
- **Per Min Avg Intensity = [minuteintensitiesNarrow_merged.csv + minuteintensitieswide_merged.csv]**
- **Avg Metabolic Equivalent = [minuteMETsNarrow_merged.csv]**
- **Per Hour Avg Steps = [minutestepsNarrow_merged.csv + minutestepsWide_merged]**

Cleaning steps involved :-

- Checking for missing values
- Removing duplicates
- Correcting Data types

A Summary of the analysis :-

The analysis of the FitBit Fitness Tracker App data involved a comprehensive exploration of various aspects, including daily activity, hourly activity, minutes activity, sleep data, weight logs, and heart rate. The process employed Python for data cleaning, transformation, and analysis, along with the use of Pandas Profiling for in-depth dataset exploration.

1. Daily Activity Merged (EDA, Plots):

Explored trends in daily activity, identifying patterns and variations.

Generated exploratory data analysis (EDA) plots to visualize key metrics.

Extracted insights related to user engagement and overall activity patterns.

2. Hourly Activity Merged (EDA, Plots):

Analyzed hourly activity data to understand user behaviors throughout the day.

Utilized EDA plots to visualize hourly trends and identify peak activity periods.

Extracted insights regarding user activity patterns at different times of the day.

3. Minutes Activity Merged (EDA, Plots) - Dashboard (Filters - Daily, Hourly, Minutes):

Conducted detailed analysis at the minute level, providing granular insights into user activities.

Developed a comprehensive dashboard with filters for daily, hourly, and minute-level data.

Enabled users to interactively explore and understand activity patterns with flexibility.

4. Sleep Data (EDA, Plots):

Explored sleep data to identify trends in sleep patterns among users.

Utilized EDA plots to visualize sleep duration, quality, and variations over time.

Extracted insights into sleep behavior and potential correlations with other factors.

5. Weight_Log (EDA, Plots):

Conducted exploratory analysis on weight logs to understand user weight trends.

Visualized weight changes over time and identified potential factors influencing fluctuations.

Extracted insights related to weight management and user progress.

6. Heart Rate (EDA, Plots):

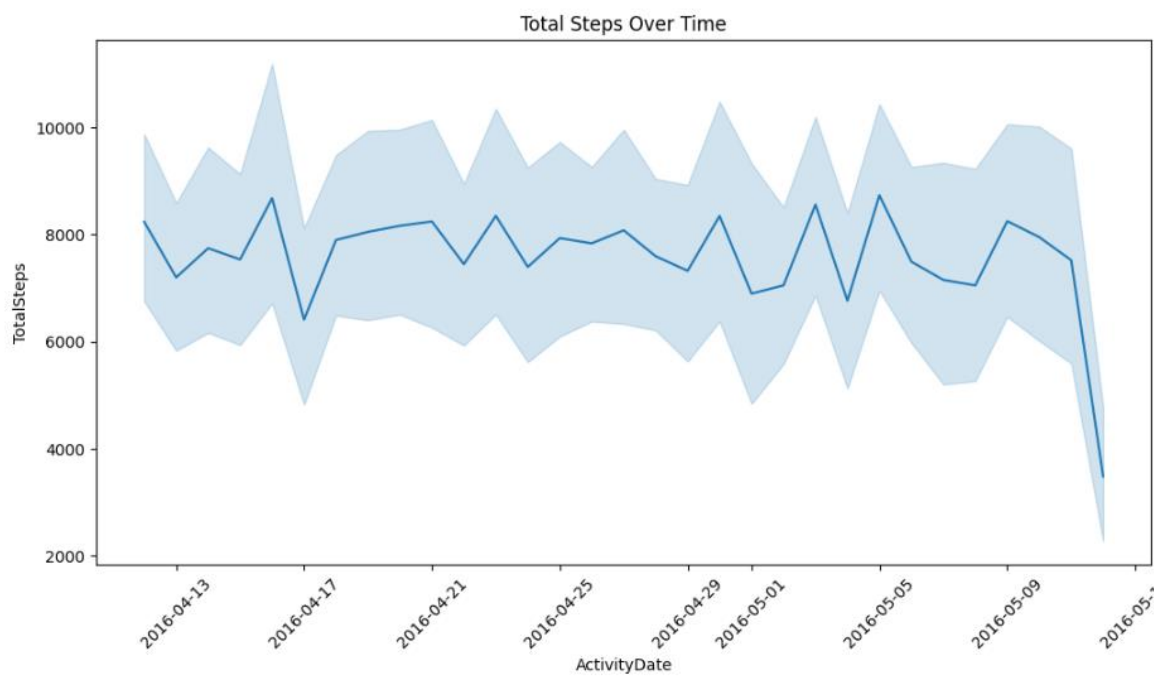
Analyzed heart rate data to understand variations in users' heart rates.

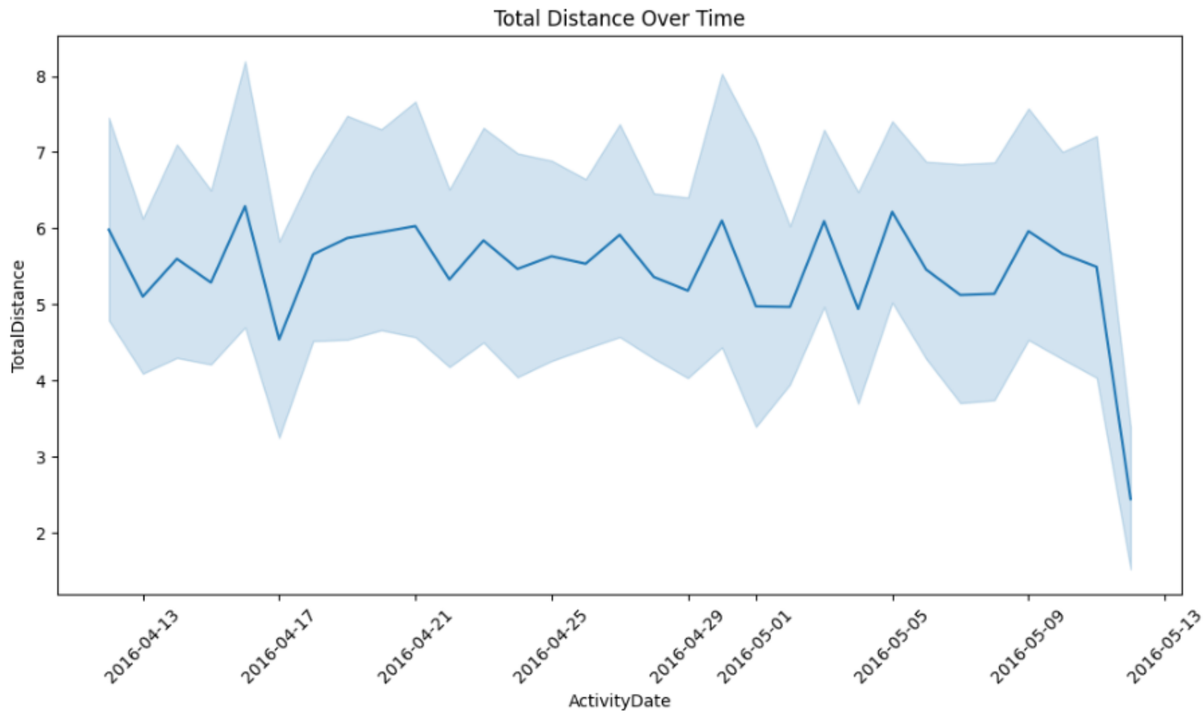
Utilized EDA plots to visualize trends, peak periods, and potential correlations with activities.

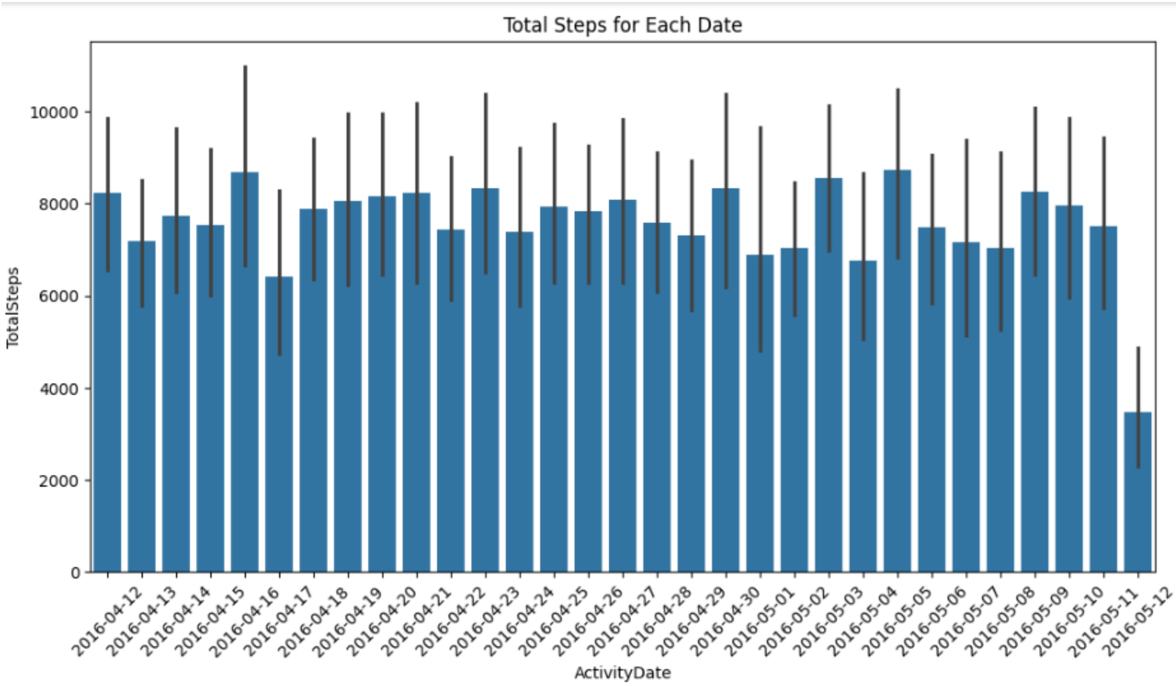
Extracted insights related to cardiovascular health and user response to physical activity.

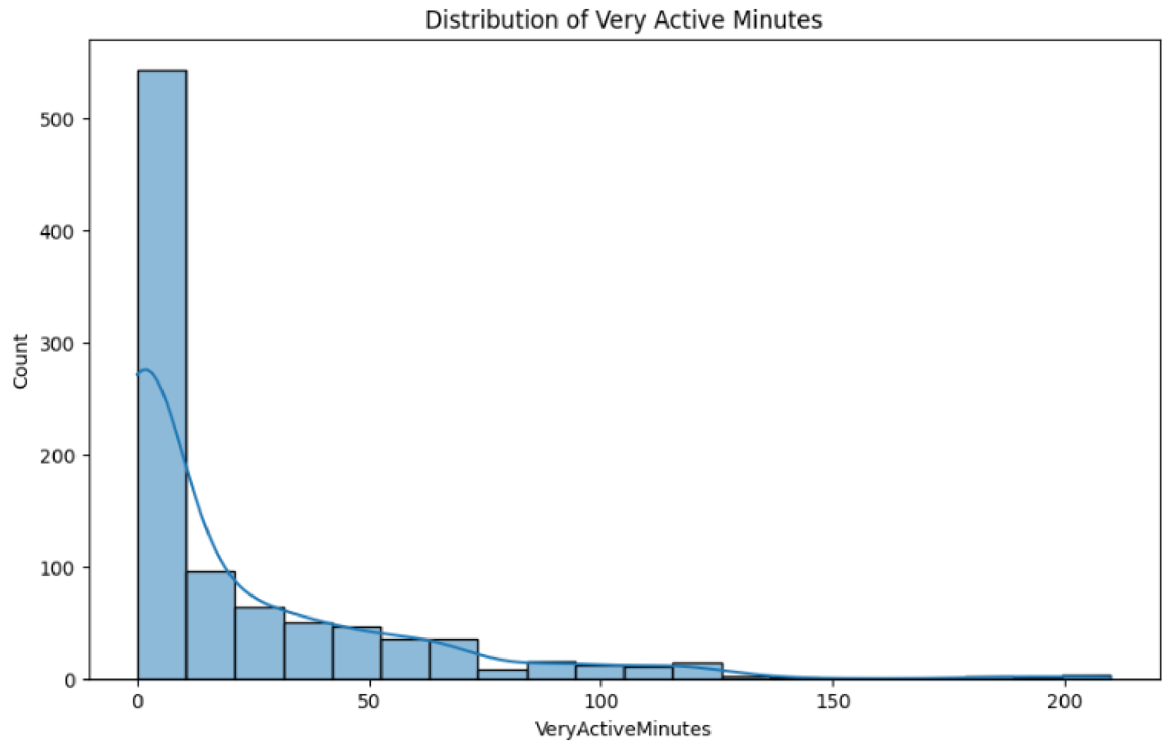
Supporting Visualizations and Key Findings :-

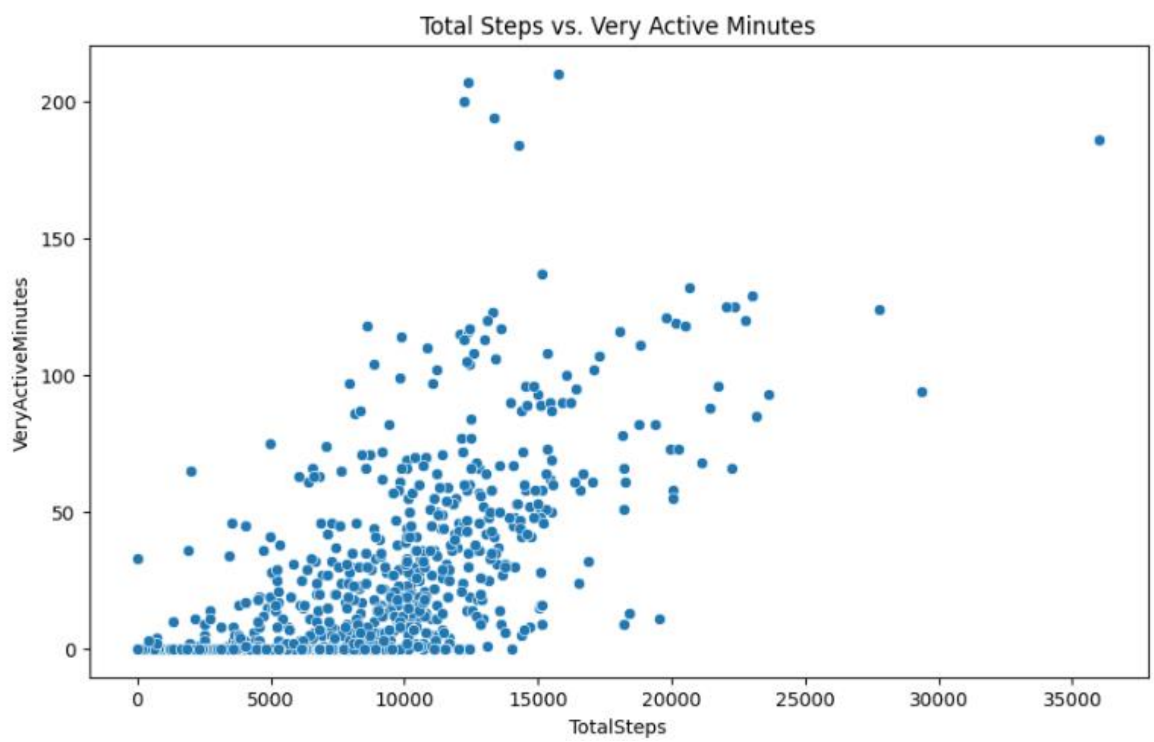
- **df1 (dailyactivity_merged.csv)**

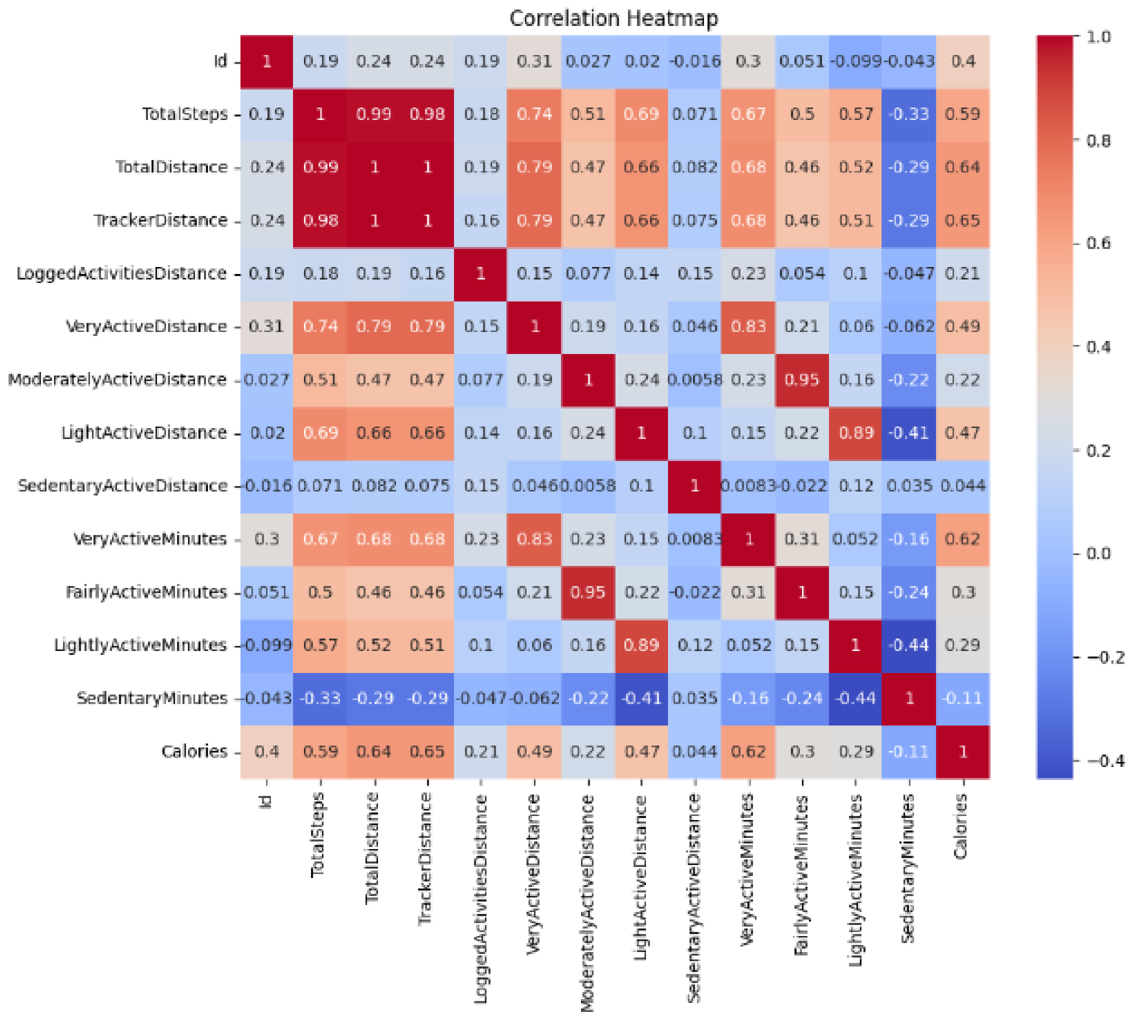


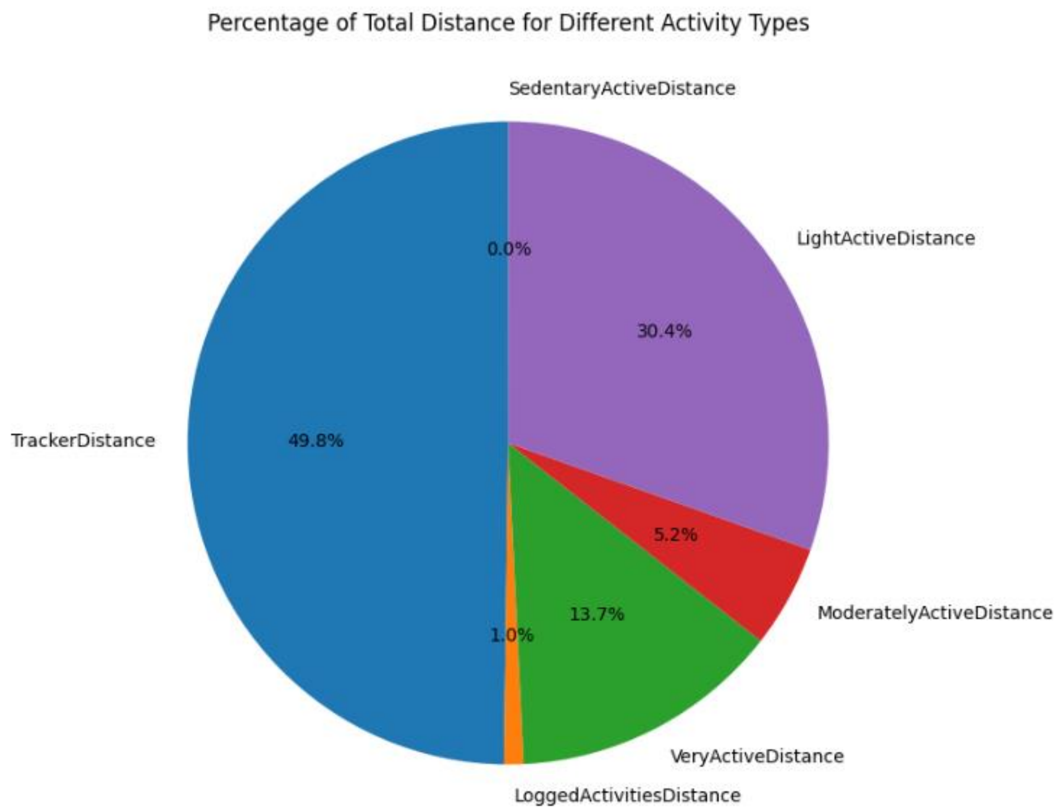

Total Distance Over Time
TotalDistance
8
7
6
5
4
3
2
2016-04-13
2016-04-17
2016-04-21
2016-04-25
2016-04-29
2016-05-01
2016-05-05
2016-05-09
2016-05-13
ActivityDate


Total Steps for Each Date
TotalSteps
10000
8000
6000
4000
2000
0
2016-04-12
2016-04-13
2016-04-14
2016-04-15
2016-04-16
2016-04-17
2016-04-18
2016-04-19
2016-04-20
2016-04-21
2016-04-22
2016-04-23
2016-04-24
2016-04-25
2016-04-26
2016-04-27
2016-04-28
2016-04-29
2016-04-30
2016-05-01
2016-05-02
2016-05-03
2016-05-04
2016-05-05
2016-05-06
2016-05-07
2016-05-08
2016-05-09
2016-05-10
2016-05-11
2016-05-12
ActivityDate

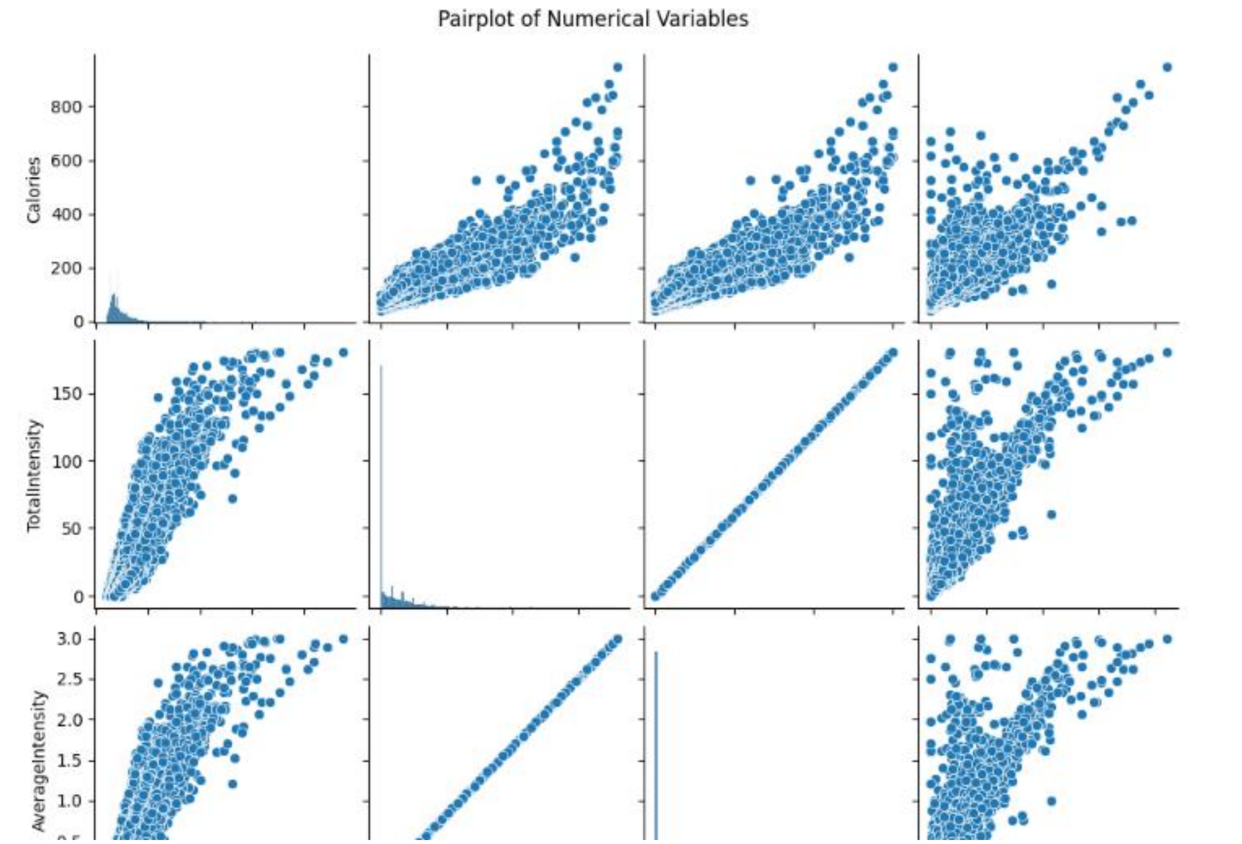

Distribution of Very Active Minutes
Count
VeryActiveMinutes
0
100
200
300
400
500
0
50
100
150
200

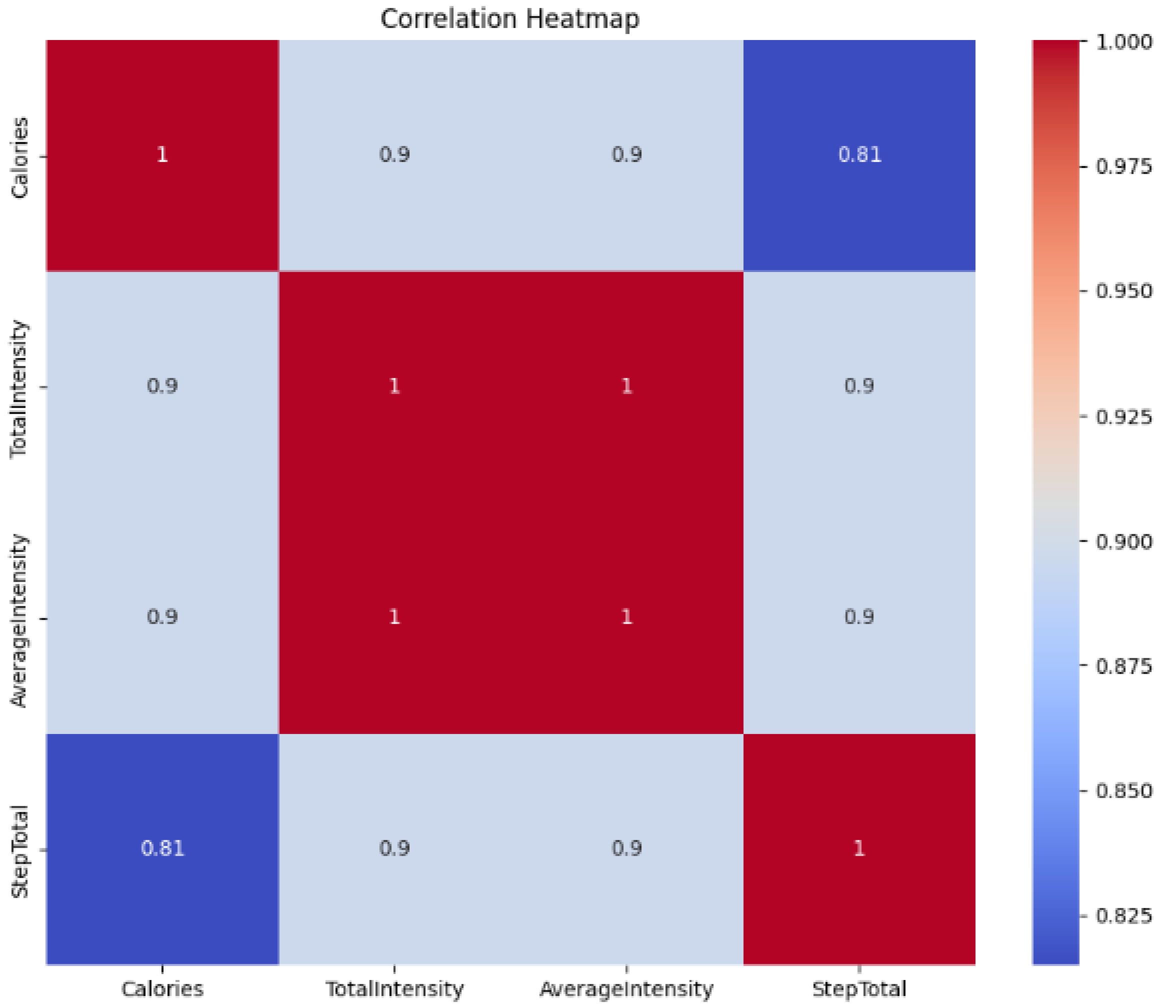

Total Steps vs. Very Active Minutes
VeryActiveMinutes
TotalSteps
0
50
100
150
200
0
5000
10000
15000
20000
25000
30000
35000


Correlation Heatmap
Id 1 0.19 0.24 0.24 0.19 0.31 0.027 0.02 -0.016 0.3 0.051 -0.099 -0.043 0.4
TotalSteps 0.19 1 0.99 0.98 0.18 0.74 0.51 0.69 0.071 0.67 0.5 0.57 -0.33 0.59
TotalDistance 0.24 0.99 1 1 0.19 0.79 0.47 0.66 0.082 0.68 0.46 0.52 -0.29 0.64
TrackerDistance 0.24 0.98 1 1 0.16 0.79 0.47 0.66 0.075 0.68 0.46 0.51 -0.29 0.65
LoggedActivitiesDistance 0.19 0.18 0.19 0.16 1 0.15 0.077 0.14 0.15 0.23 0.054 0.1 -0.047 0.21
VeryActiveDistance 0.31 0.74 0.79 0.79 0.15 1 0.19 0.16 0.046 0.83 0.21 0.06 -0.062 0.49
ModeratelyActiveDistance 0.027 0.51 0.47 0.47 0.077 0.19 1 0.24 0.0058 0.23 0.95 0.16 -0.22 0.22
LightActiveDistance 0.02 0.69 0.66 0.66 0.14 0.16 0.24 1 0.1 0.15 0.22 0.89 -0.41 0.47
SedentaryActiveDistance -0.016 0.071 0.082 0.075 0.15 0.046 0.0058 0.1 1 0.0083 -0.022 0.12 0.035 0.044
VeryActiveMinutes 0.3 0.67 0.68 0.68 0.23 0.83 0.23 0.15 0.0083 1 0.31 0.052 -0.16 0.62
FairlyActiveMinutes 0.051 0.5 0.46 0.46 0.054 0.21 0.95 0.22 -0.022 0.31 1 0.15 -0.24 0.3
LightlyActiveMinutes -0.099 0.57 0.52 0.51 0.1 0.06 0.16 0.89 0.12 0.052 0.15 1 -0.44 0.29
SedentaryMinutes -0.043 -0.33 -0.29 -0.29 -0.047 -0.062 -0.22 -0.41 0.035 -0.16 -0.24 -0.44 1 -0.11
Calories 0.4 0.59 0.64 0.65 0.21 0.49 0.22 0.47 0.044 0.62 0.3 0.29 -0.11 1
Id TotalSteps TotalDistance TrackerDistance LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

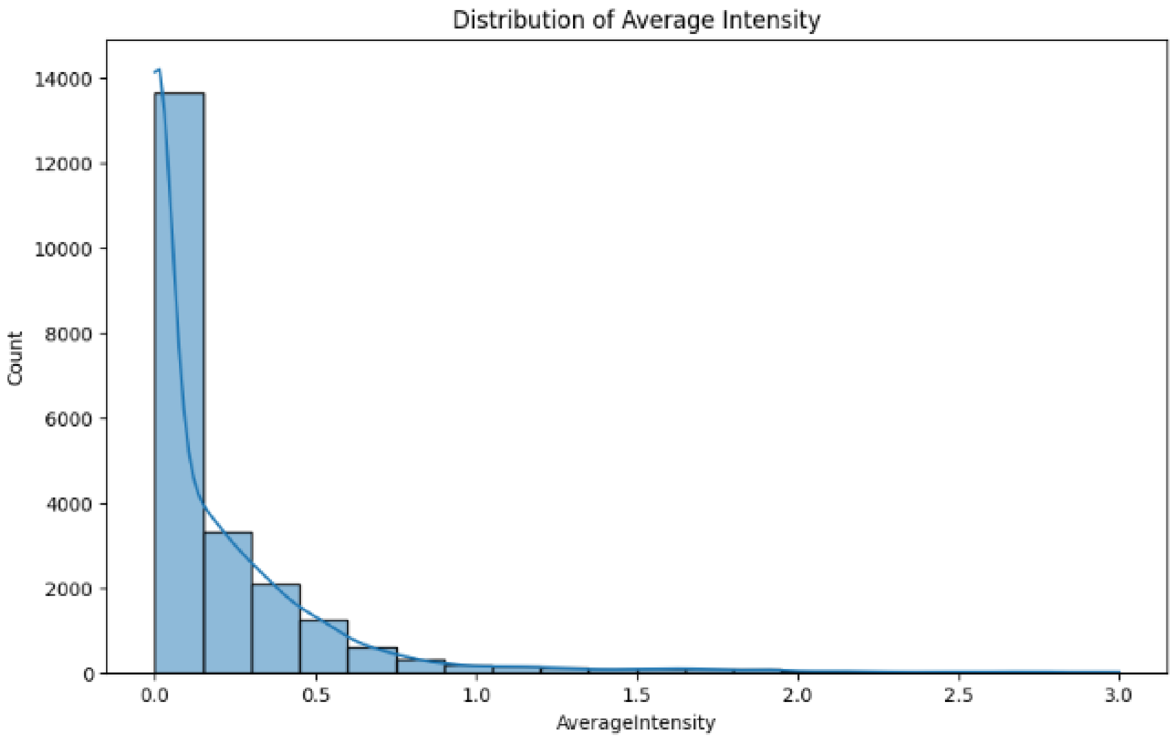


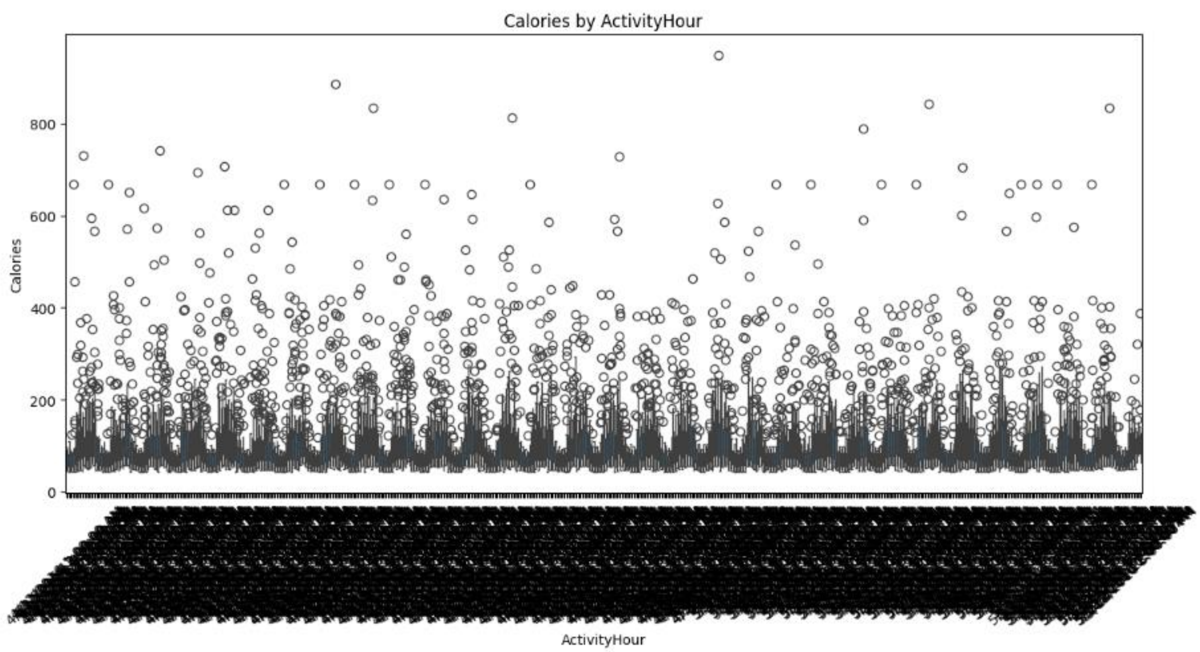
- **merged_df = [df2(hourlycalories_merged.csv) + df3(hourlyintensities_merged.csv) + df4(hourlysteps_merged.csv)]**


Pairplot of Numerical Variables
Calories
TotalIntensity
AverageIntensity
800
600
400
200
0
150
100
50
0
3.0
2.5
2.0
1.5
1.0

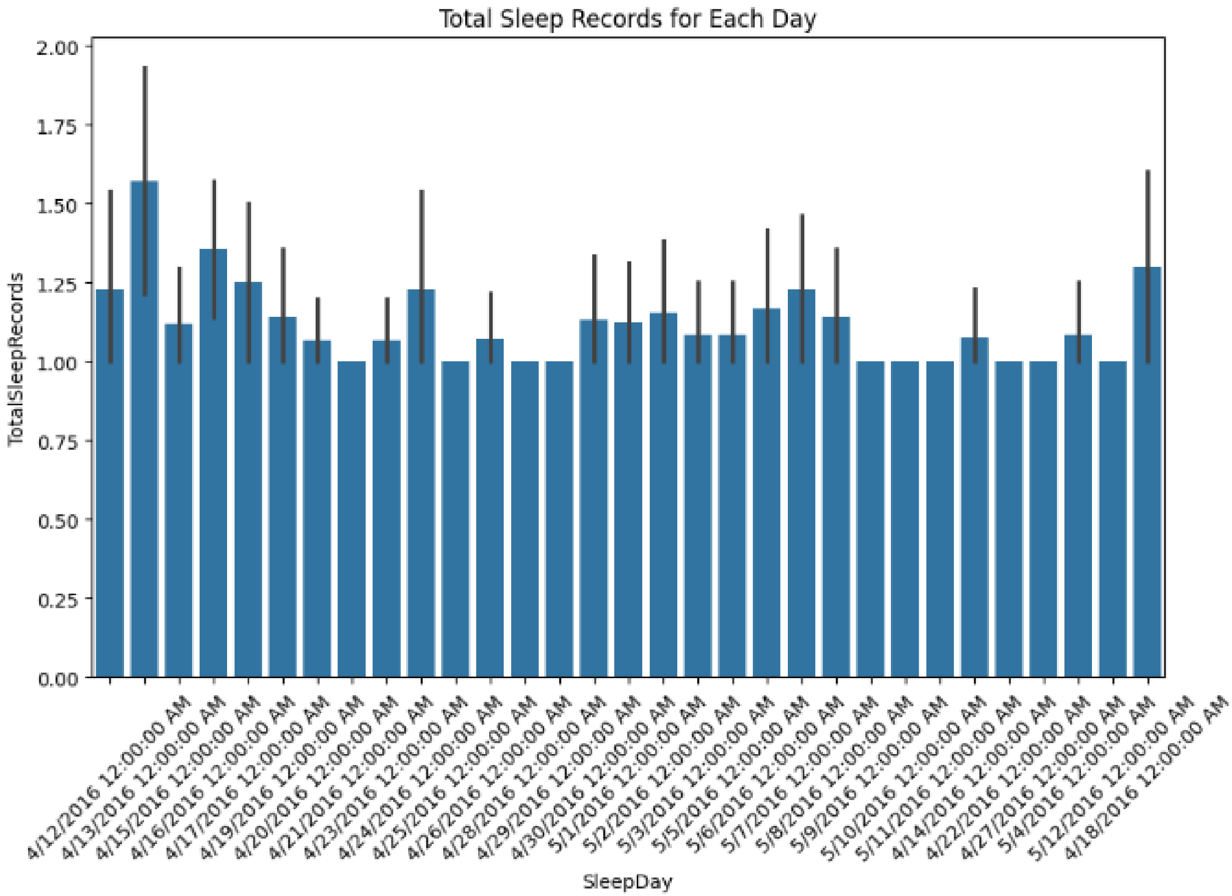


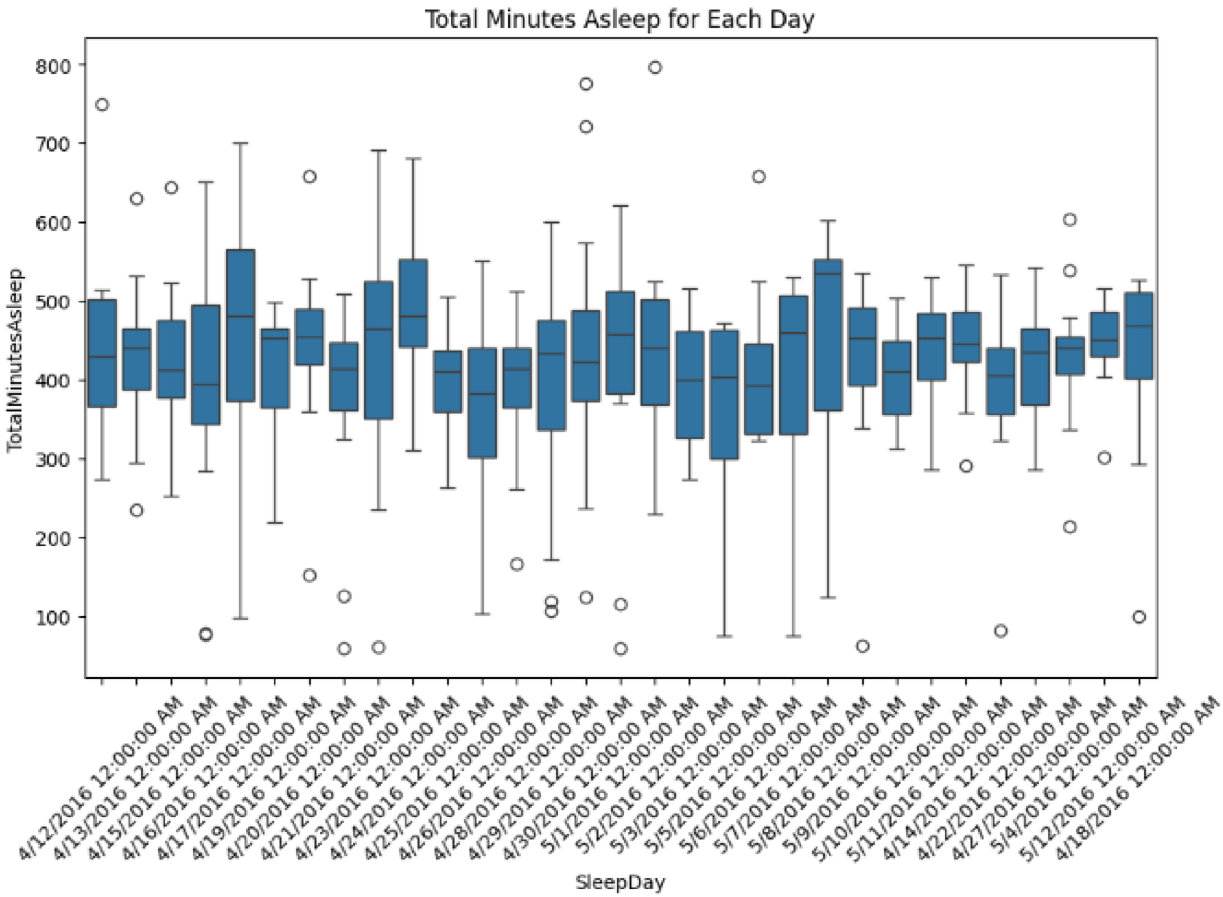
Correlation Heatmap
Calories
TotalIntensity
AverageIntensity
StepTotal
1
0.9
0.9
0.81
0.9
1
1
0.9
0.9
1
1
0.9
0.81
0.9
0.9
1
Calories
TotalIntensity
AverageIntensity
StepTotal
1.000
0.975
0.950
0.925
0.900
0.875
0.850
0.825

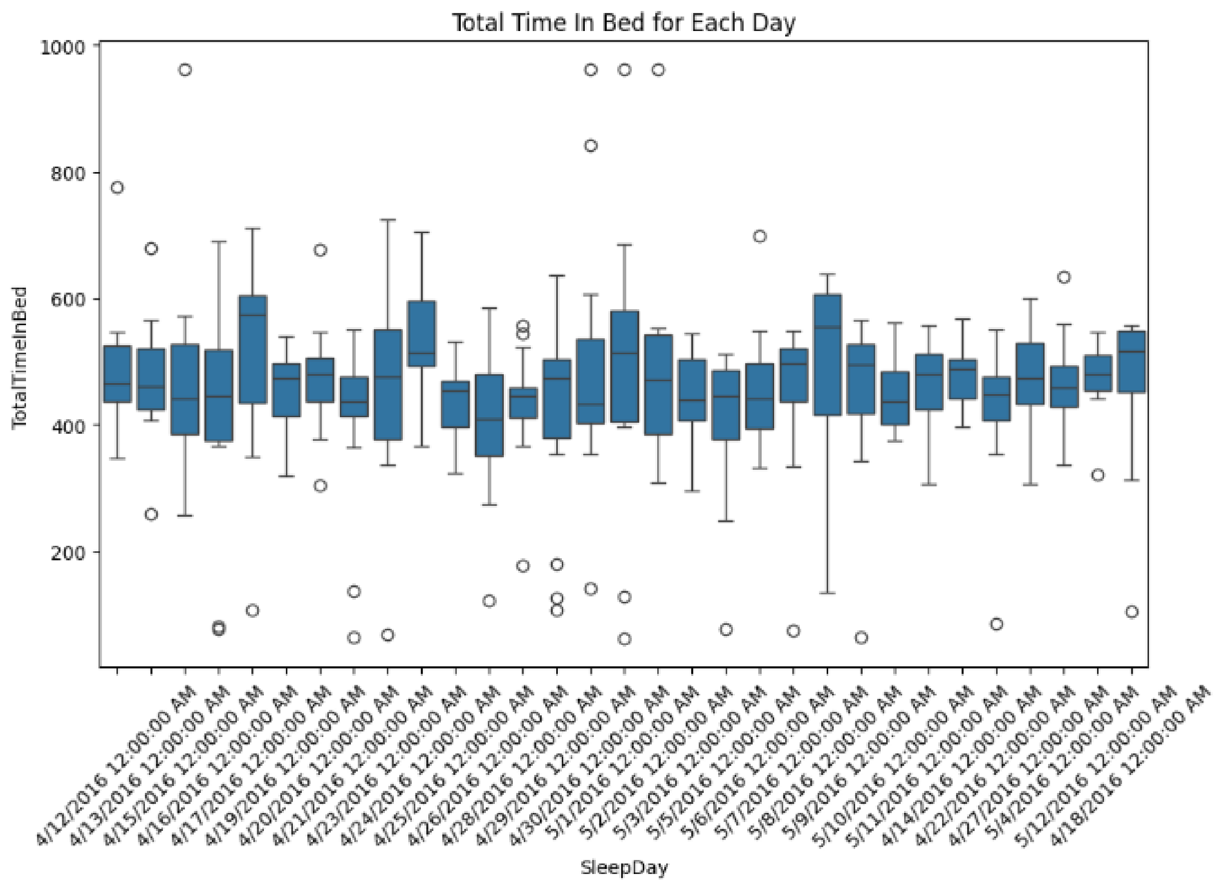




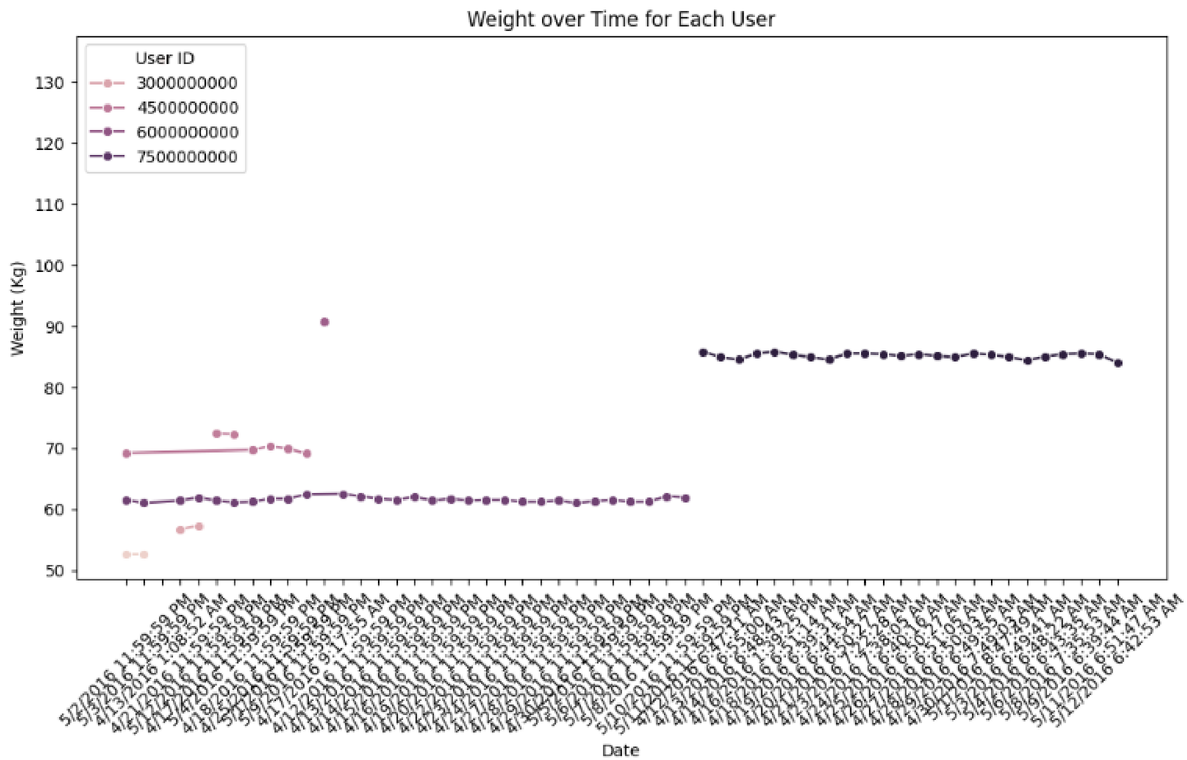
- **df5 (sleepday_merged.csv)**

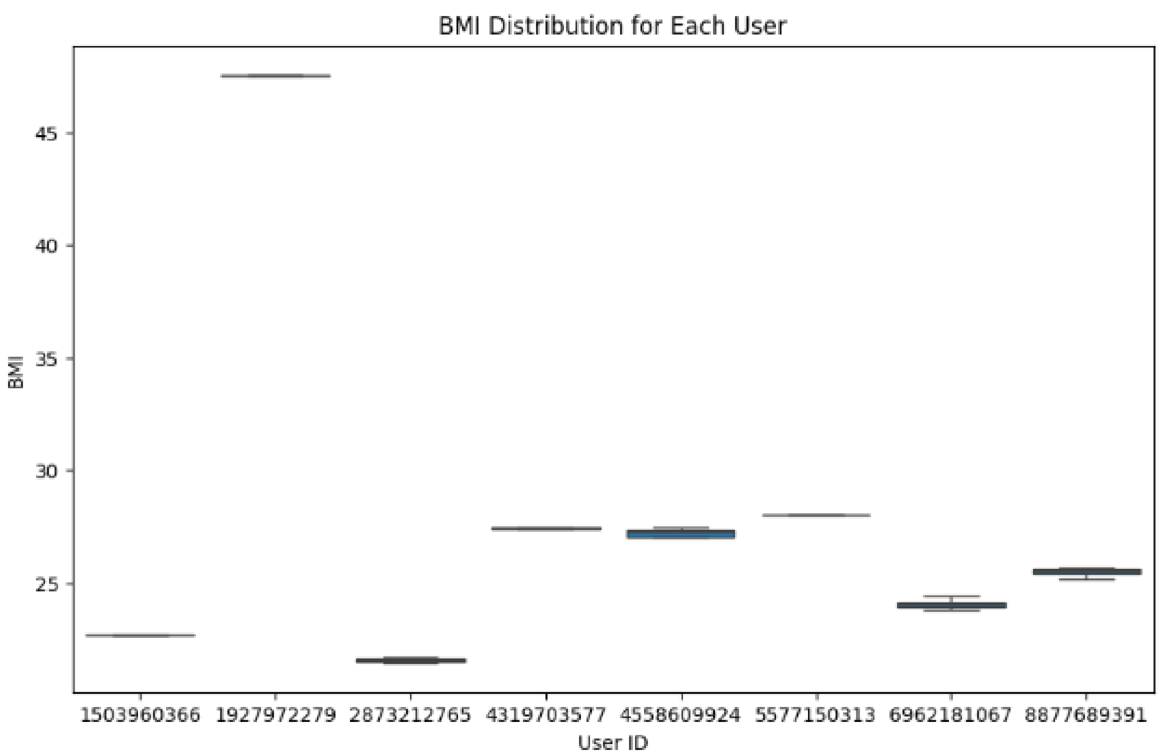

Total Sleep Records for Each Day
TotalSleepRecords
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
SleepDay
4/12/2016 12:00:00 AM
4/13/2016 12:00:00 AM
4/15/2016 12:00:00 AM
4/16/2016 12:00:00 AM
4/17/2016 12:00:00 AM
4/19/2016 12:00:00 AM
4/20/2016 12:00:00 AM
4/21/2016 12:00:00 AM
4/23/2016 12:00:00 AM
4/24/2016 12:00:00 AM
4/25/2016 12:00:00 AM
4/26/2016 12:00:00 AM
4/28/2016 12:00:00 AM
4/29/2016 12:00:00 AM
4/30/2016 12:00:00 AM
5/1/2016 12:00:00 AM
5/2/2016 12:00:00 AM
5/3/2016 12:00:00 AM
5/5/2016 12:00:00 AM
5/6/2016 12:00:00 AM
5/7/2016 12:00:00 AM
5/8/2016 12:00:00 AM
5/9/2016 12:00:00 AM
5/10/2016 12:00:00 AM
5/11/2016 12:00:00 AM
4/14/2016 12:00:00 AM
4/22/2016 12:00:00 AM
4/27/2016 12:00:00 AM
5/4/2016 12:00:00 AM
5/12/2016 12:00:00 AM
4/18/2016 12:00:00 AM

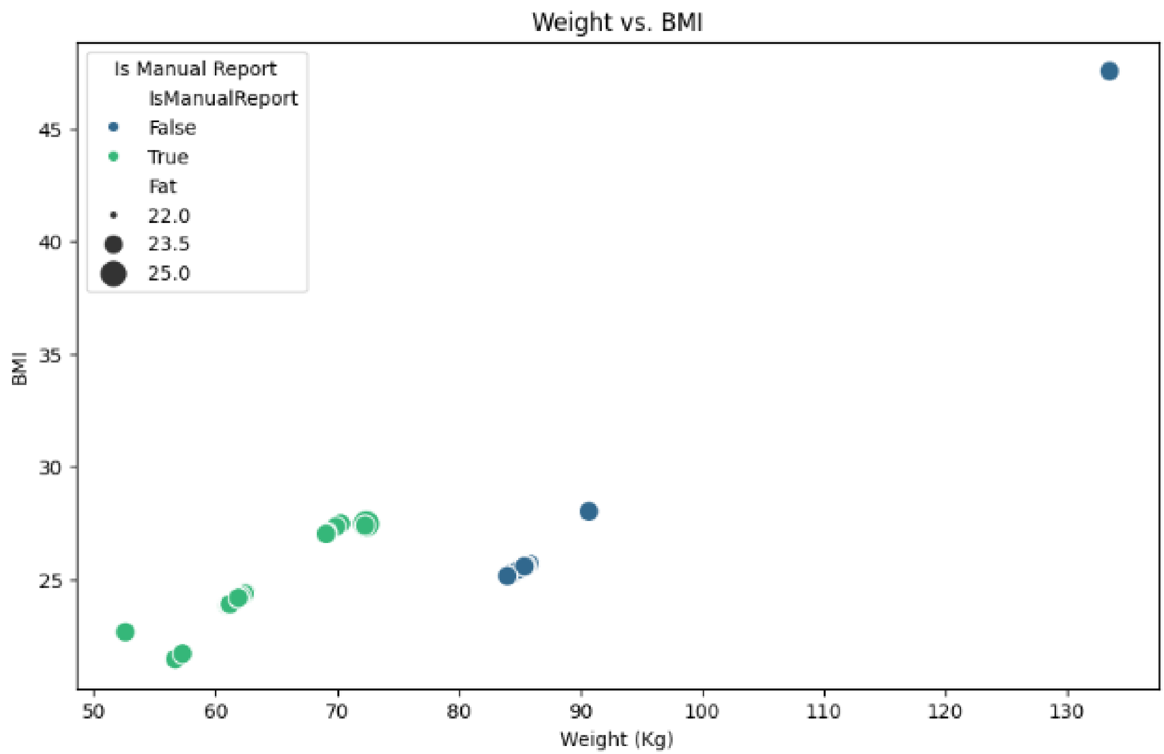

Total Minutes Asleep for Each Day
TotalMinutesAsleep
100
200
300
400
500
600
700
800
SleepDay
4/12/2016 12:00:00 AM
4/13/2016 12:00:00 AM
4/15/2016 12:00:00 AM
4/16/2016 12:00:00 AM
4/17/2016 12:00:00 AM
4/19/2016 12:00:00 AM
4/20/2016 12:00:00 AM
4/21/2016 12:00:00 AM
4/23/2016 12:00:00 AM
4/24/2016 12:00:00 AM
4/25/2016 12:00:00 AM
4/26/2016 12:00:00 AM
4/28/2016 12:00:00 AM
4/29/2016 12:00:00 AM
4/30/2016 12:00:00 AM
5/1/2016 12:00:00 AM
5/2/2016 12:00:00 AM
5/3/2016 12:00:00 AM
5/5/2016 12:00:00 AM
5/6/2016 12:00:00 AM
5/7/2016 12:00:00 AM
5/8/2016 12:00:00 AM
5/9/2016 12:00:00 AM
5/10/2016 12:00:00 AM
5/11/2016 12:00:00 AM
4/14/2016 12:00:00 AM
4/22/2016 12:00:00 AM
4/27/2016 12:00:00 AM
5/4/2016 12:00:00 AM
5/12/2016 12:00:00 AM
4/18/2016 12:00:00 AM



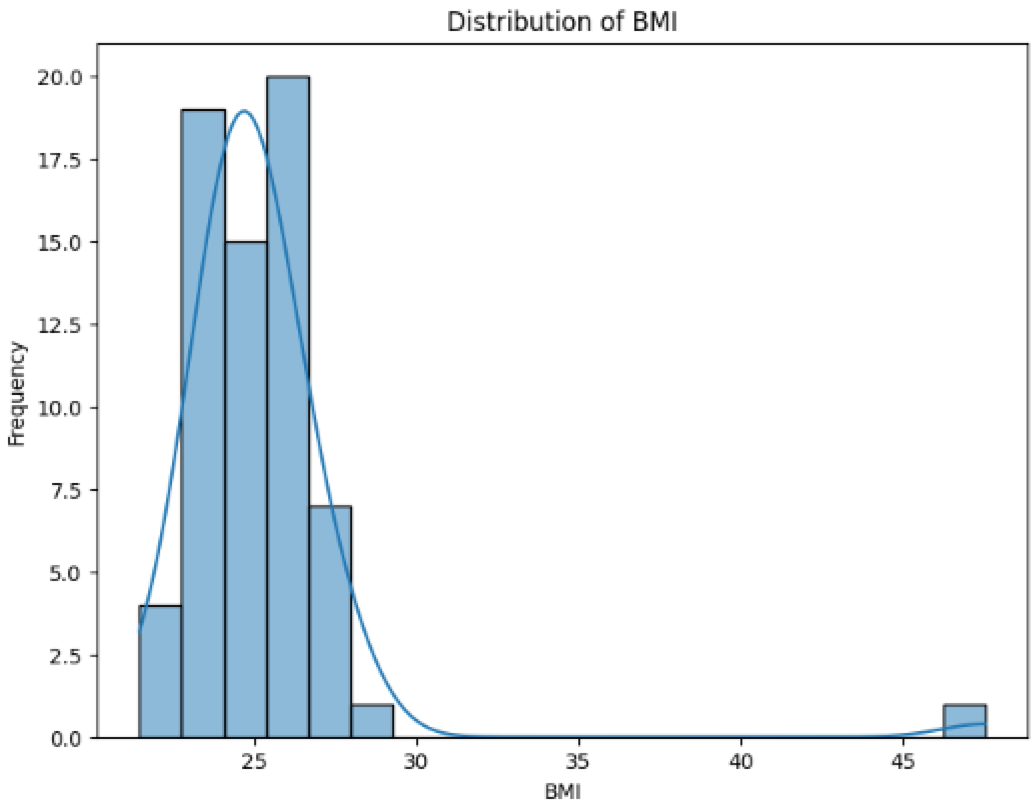
- **df6 (weightloginfo_merged.csv)**


Weight over Time for Each User
User ID
3000000000
4500000000
6000000000
7500000000
Weight (Kg)
130
120
110
100
90
80
70
60
50
Date

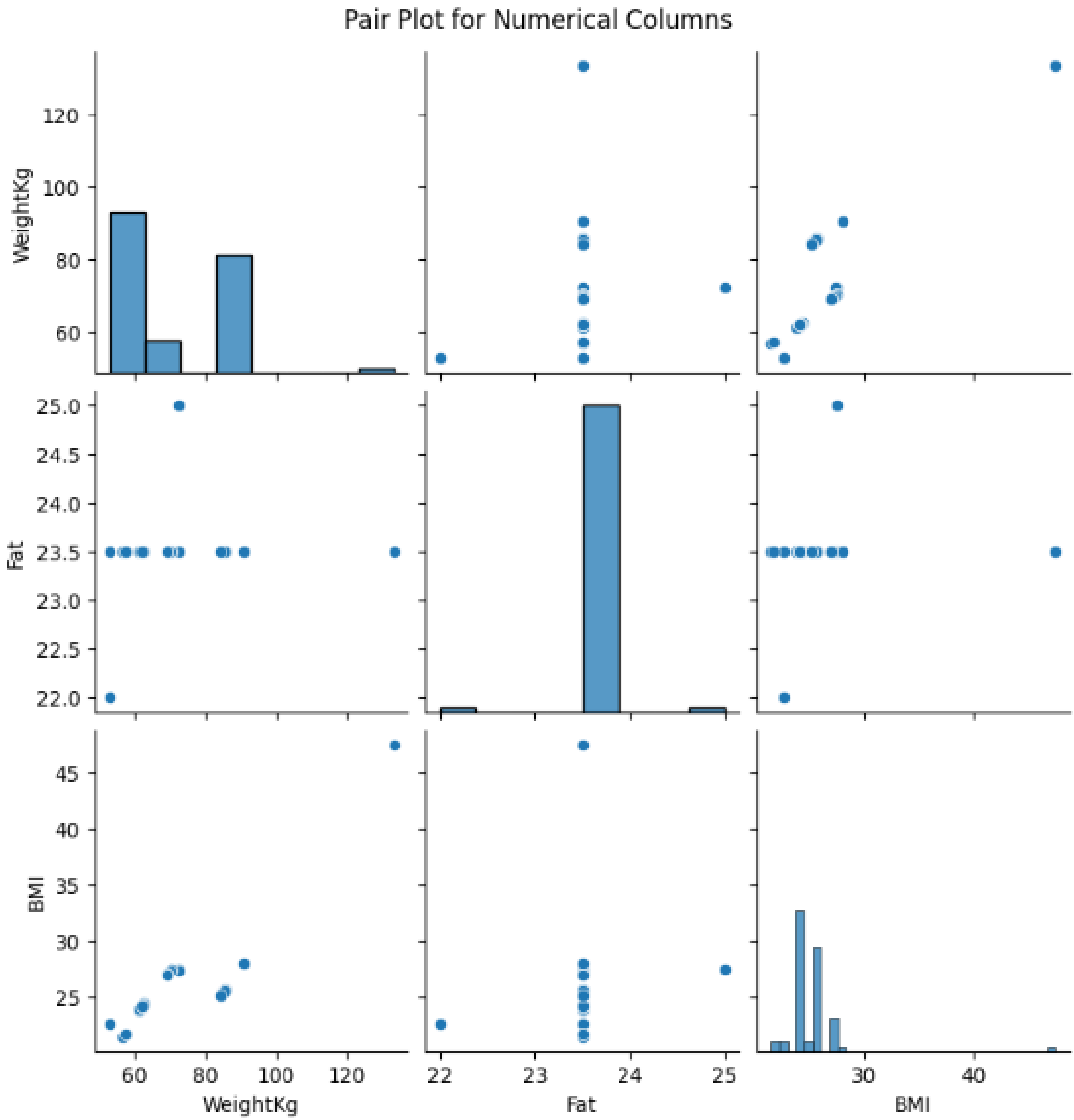

BMI Distribution for Each User
BMI
45
40
35
30
25
1503960366
1927972279
2873212765
4319703577
4558609924
5577150313
6962181067
8877689391
User ID

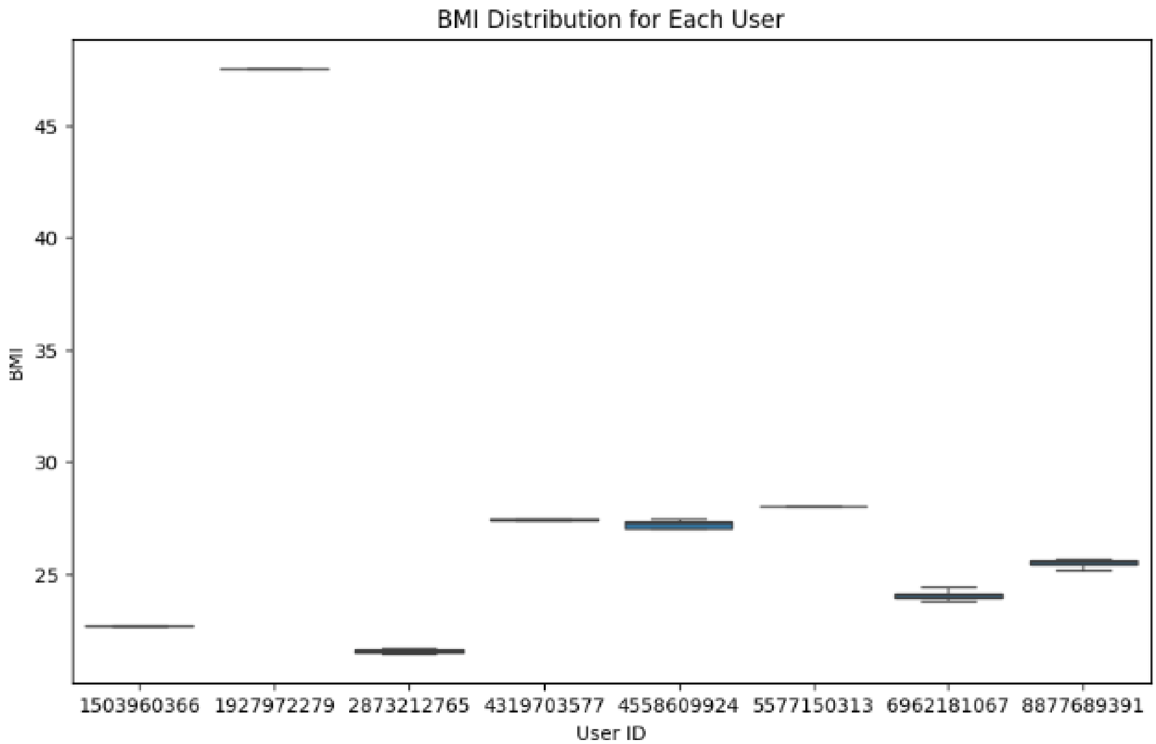


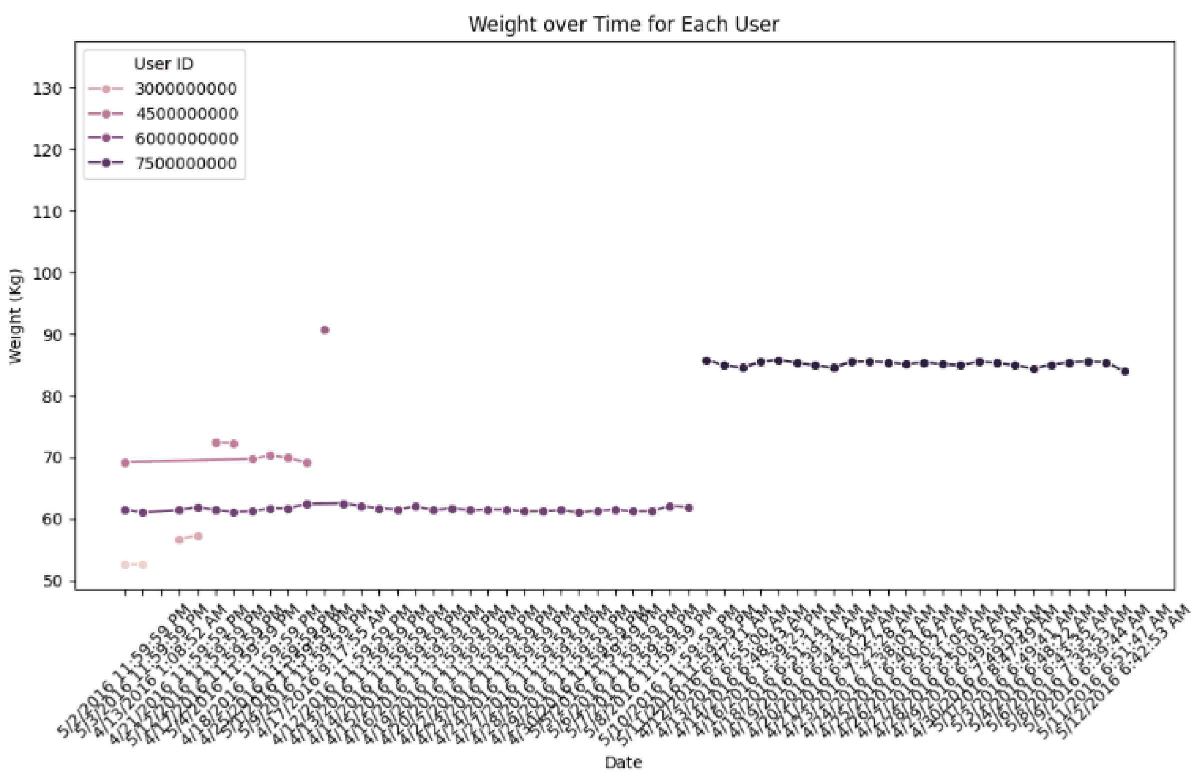
- **<u>df7(heartrate_merged.csv)</u>**



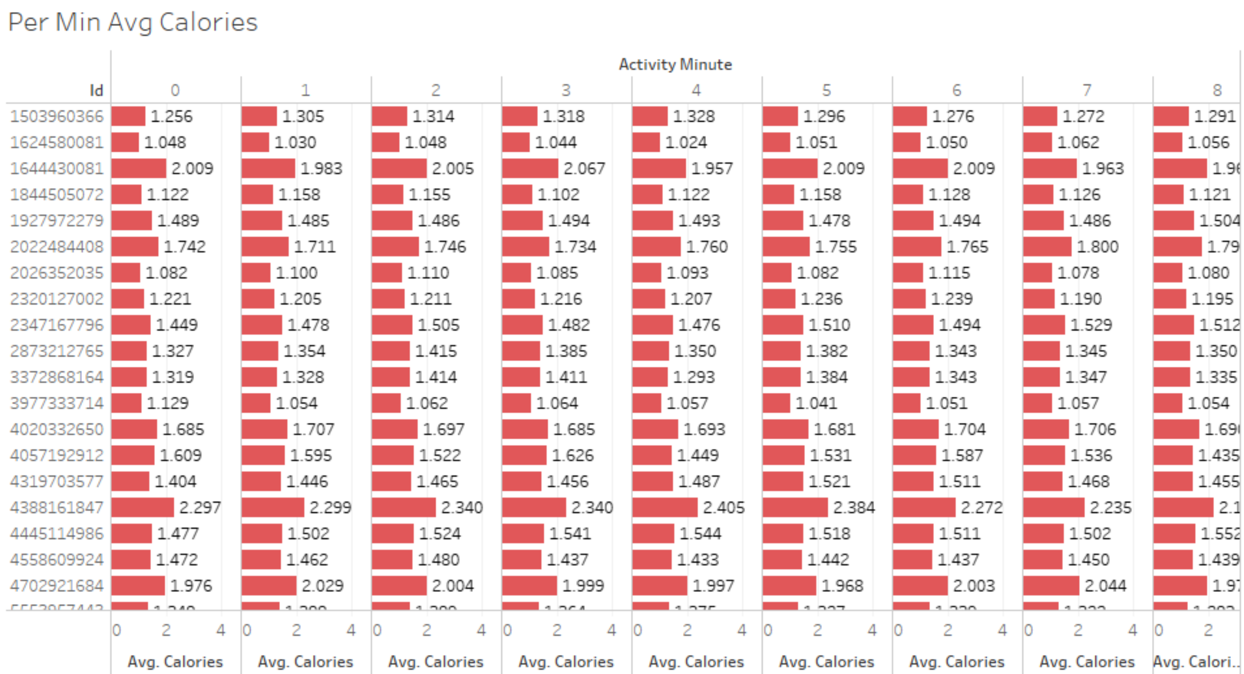
Distribution of BMI
Frequency
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
25
30
35
40
45
BMI


Pair Plot for Numerical Columns
WeightKg
Fat
BMI
120
100
80
60
25.0
24.5
24.0
23.5
23.0
22.5
22.0
45
40
35
30
25
60
80
100
120
WeightKg
22
23
24
25
Fat
30
40
BMI

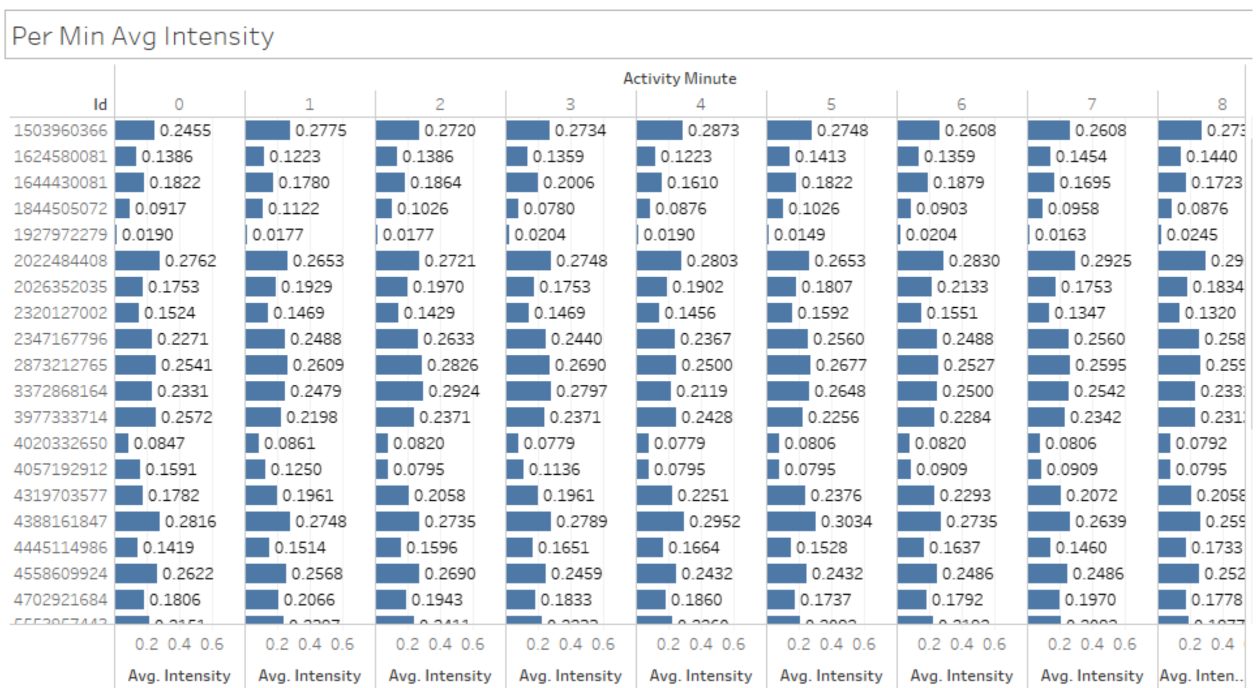

BMI Distribution for Each User
BMI
45
40
35
30
25
1503960366
1927972279
2873212765
4319703577
4558609924
5577150313
6962181067
8877689391
User ID


Weight over Time for Each User
User ID
3000000000
4500000000
6000000000
7500000000
Weight (Kg)
130
120
110
100
90
80
70
60
50
Date

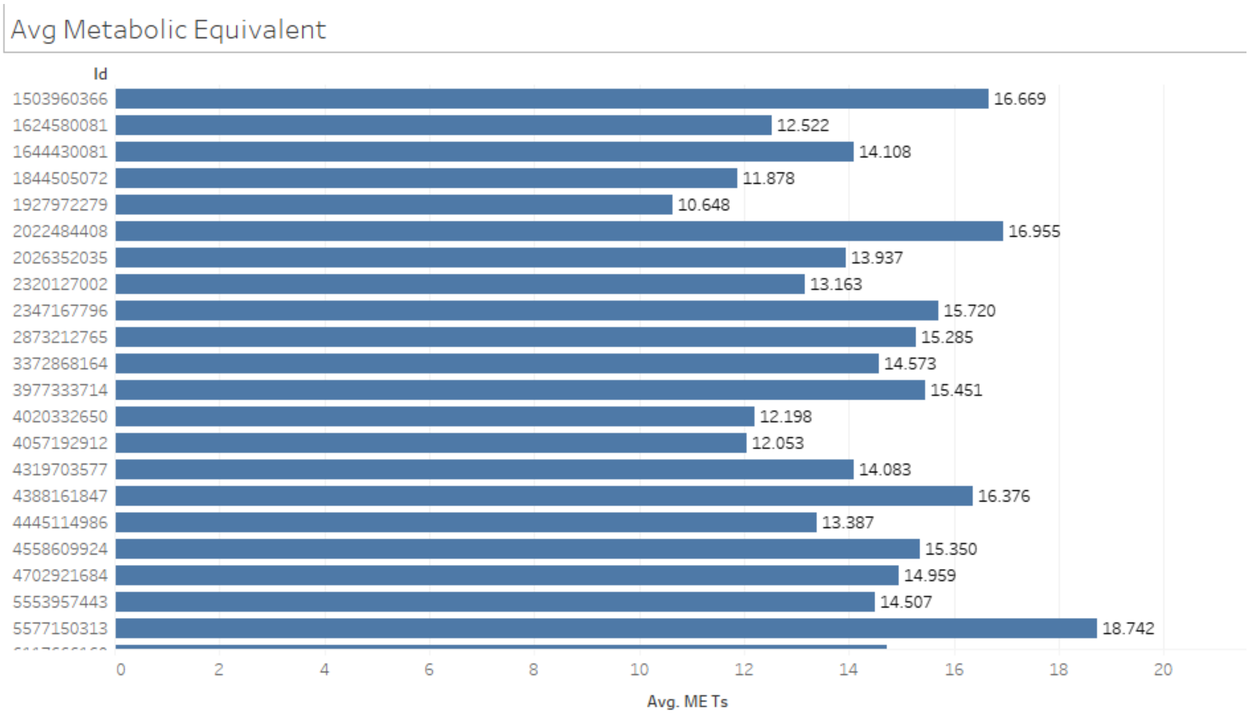
- **<u>Per Min Avg Calories = [minutecaloriesNarrow_merged.csv + minutecalorieswide_merged.csv]</u>**



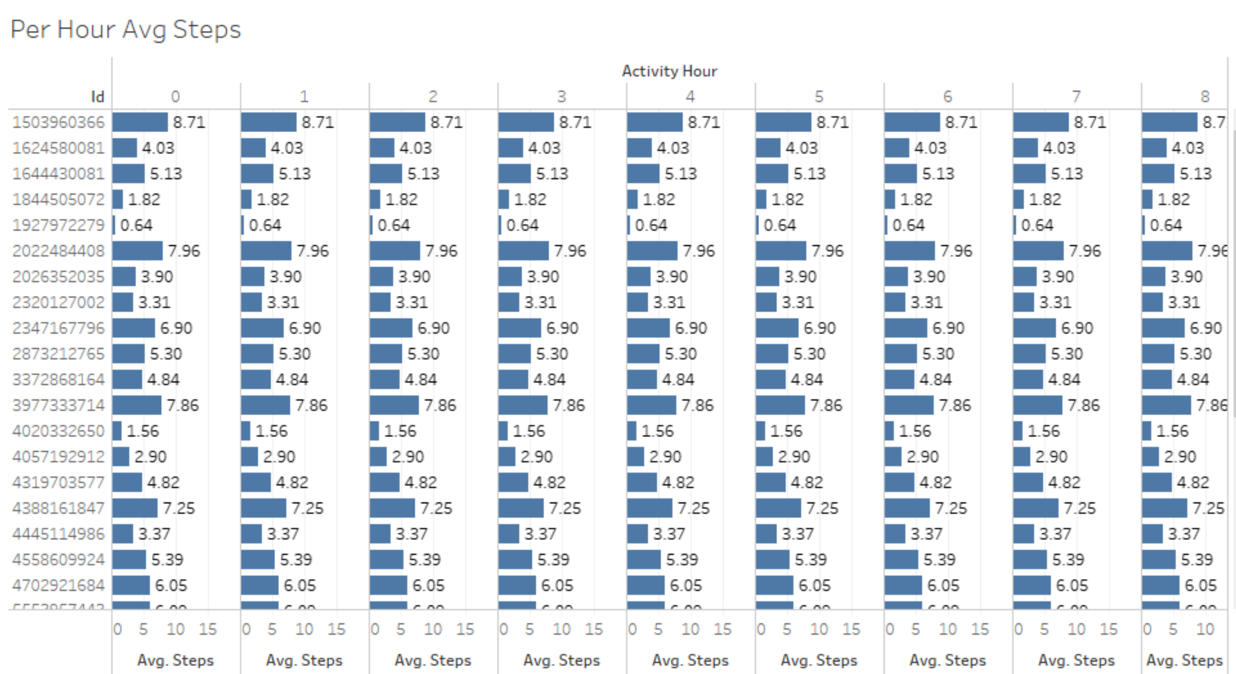
- **<u>Per Min Avg Intensity = [minuteintensitiesNarrow_merged.csv + minuteintensitieswide_merged.csv]</u>**



- **Avg Metabolic Equivalent = [minuteMETsNarrow_merged.csv]**



- **Per Hour Avg Steps = [minutestepsNarrow_merged.csv + minutestepsWide_merged]**



Key Findings:

Identified peak activity periods during the day and variations in daily, hourly, and minute-level activities.

Discovered trends in sleep patterns, potentially correlating with other lifestyle factors.

Explored weight management trends and variations over time.

Analyzed heart rate data for insights into cardiovascular health and responses to physical activity.

Content Recommendations :-

1)User Engagement Campaigns:

Develop targeted campaigns to encourage user engagement during identified peak activity periods. Consider offering challenges, rewards, or promotions aligned with these high-engagement times.

2)Sleep Optimization Features:

Introduce features within the app aimed at optimizing sleep patterns. This could include personalized sleep recommendations, sleep quality tracking, and tips for improving sleep based on the identified trends.

3)Weight Management Support:

Implement features or content within the app to support users in their weight management journey. This could involve providing nutritional guidance, fitness routines, and personalized insights based on weight log trends.

4)Heart Health Education:

Launch educational content or notifications within the app to raise awareness about cardiovascular health. Provide tips on maintaining a healthy heart rate, recognizing peak periods, and incorporating suitable activities.

5)Interactive Dashboards for Users:

Enhance user experience by integrating interactive dashboards into the app. Users should be able to explore their own data with filters for daily, hourly, and minute-level activities, enabling them to gain deeper insights into their fitness patterns.

6)Personalized Recommendations:

Implement a recommendation engine that suggests personalized fitness goals, challenges, or content based on users' specific activity patterns, sleep behavior, weight management goals, and heart rate trends.

7)Community Engagement Initiatives:

Foster a sense of community among FitBit users by introducing features that allow users to share achievements, participate in challenges together, and exchange tips and advice. This community-building strategy can enhance user motivation.

8)Data-Driven Marketing Campaigns:

Leverage the insights gained from the analysis to create data-driven marketing campaigns. Tailor promotions, advertisements, and content based on user preferences, behaviors, and identified trends.

9)Wellness Challenges and Rewards:

Introduce wellness challenges that span across various aspects like daily activity, sleep, and weight management. Reward users for achieving milestones, fostering a sense of accomplishment and motivation.

10)Integration with Health Professionals:

Explore partnerships or integrations with health professionals or wellness experts. Provide users with the option to receive personalized advice or consultation based on their fitness data, enhancing the overall health and wellness experience.