PrepInsta Winter Internship week 8 (Capstone Project) : Analysis of Fitbit Fitness Tracker App Data

Introduction :-

The aim of this report is to analyse the Fitbit Fitness Tracker App data to gain insights into consumer behaviour and usage patterns, with a focus on informing marketing strategies. The analysis was conducted using Python programming language for data cleaning, transformation, visualisation, and analysis.

Why Python for Visualization :-

Python was chosen over Tableau for visualisation due to its flexibility, customization options, and ability to handle complex data manipulation and analysis tasks. Python's libraries such as Matplotlib and Seaborn offer extensive capabilities for creating a wide range of visualisations, allowing for deeper exploration and interpretation of the data.

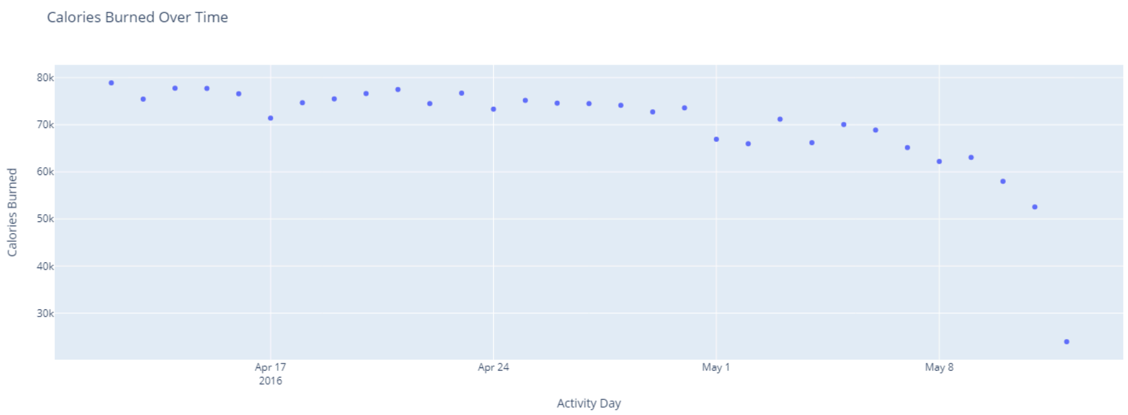
Insights and Visualizations:

Below are key insights that i have derived after analysis of each type of dataset

1. Daily Activity Analysis:

I have initially merged daily activity datasets into one single to form master dataset for the daily activity to derive more meaningful insights and analysis.

1.Calories Burned over TIme :



Observation : One common trend observed in the daily activity data is that the calories burned at the start of the month tend to be higher compared to the end of the month. As the month progresses, there is a gradual decrease in the number of calories burned by users.

Insight:

This trend suggests that users might begin the month with higher motivation and engagement towards their fitness goals, resulting in more vigorous physical activity and consequently, higher calorie expenditure. However, as the month progresses, factors such as fatigue, busy schedules, or a decrease in motivation could lead to a decline in physical activity levels and hence, a decrease in calories burned towards the end of the month.

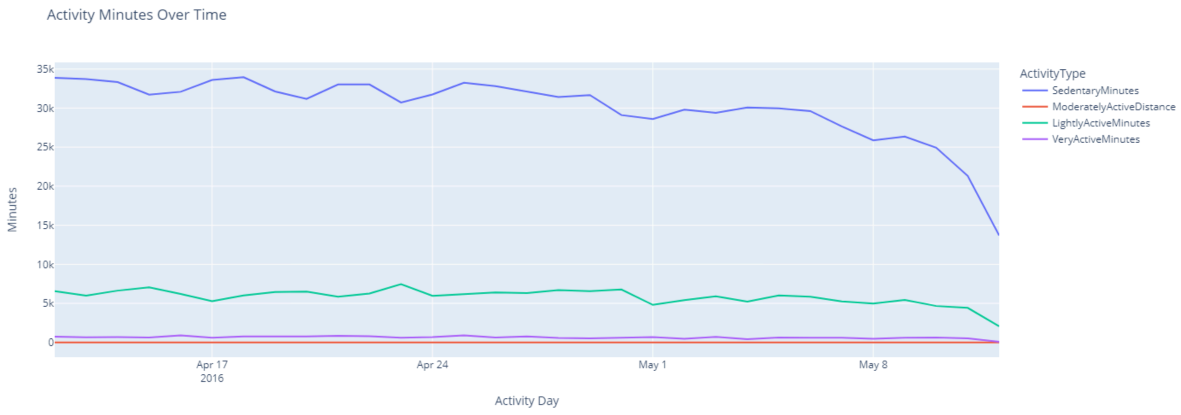
Reasoning:

Several factors may contribute to the higher count of sedentary minutes:

> Occupational and Lifestyle Factors: Users may have sedentary jobs or lifestyles that involve prolonged periods of sitting, such as desk-based work or commuting by car.

Sociocultural Norms: Cultural norms and societal expectations may influence sedentary behaviour, with activities like watching TV, using electronic devices, or socialising often involving minimal physical activity.
Time Constraints: Users may perceive limited opportunities or time constraints to engage in more active forms of physical activity due to work commitments, family responsibilities, or other obligations.

2.Activity Minutes (by type) over time :



Observation:

In the daily activity data, a notable trend is the higher count of sedentary minutes compared to other activity levels such as very active, lightly active, and fairly active.

Insight:

The analysis reveals that users tend to accumulate a higher number of sedentary minutes compared to other activity levels. Sedentary behaviour, characterised by prolonged periods of sitting or low levels of physical activity, appears to be more prevalent among users than engaging in more active forms of physical activity.

Reasoning:

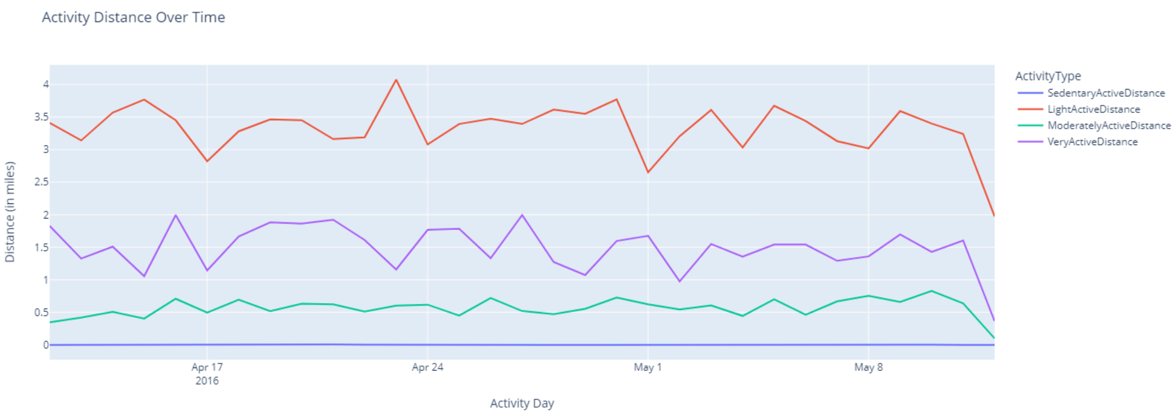
Several factors may contribute to the higher distance covered during lightly active periods:

Daily Commuting: Users may engage in lightly active activities such as walking or cycling during their daily commute to work, school, or other destinations, contributing to the overall distance covered.

Leisurely Activities: Lightly active periods may also include leisurely walks, strolls, or outdoor activities that involve moderate physical exertion but are not categorised as very active or vigorous.

Incidental Movement: Lightly active periods may encompass incidental movement throughout the day, such as household chores, errands, or casual walking, which collectively contribute to a greater distance covered.

3.Activity Distance travelled(by type) over time



Observation:

A distinct trend observed in the daily activity data is that the distance covered during lightly active periods is higher compared to other activity levels such as very active, fairly active, and sedentary.

Insight:

The analysis indicates that users cover a greater distance during lightly active periods compared to other activity levels. Lightly active periods, characterized by activities such as walking at a moderate pace, may contribute significantly to the overall distance covered by users throughout the day.

Reasoning:

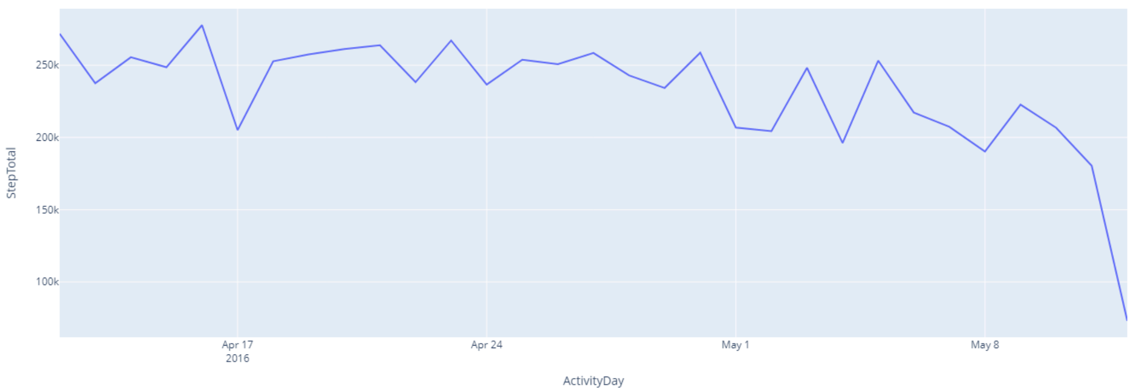
Several factors may contribute to the higher distance covered during lightly active periods:

Daily Commuting: Users may engage in lightly active activities such as walking or cycling during their daily commute to work, school, or other destinations, contributing to the overall distance covered.

Leisurely Activities: Lightly active periods may also include leisurely walks, strolls, or outdoor activities that involve moderate physical exertion but are not categorized as very active or vigorous.

Incidental Movement: Lightly active periods may encompass incidental movement throughout the day, such as household chores, errands, or casual walking, which collectively contribute to a greater distance covered.

4. Total Steps over the time :



Observation:

There is a gradual decrease in the total number of steps taken by users as the month progresses, with fewer steps recorded towards the end of the month.

Insight:

The analysis indicates a declining trend in the total number of steps taken by users over the course of the month. This decline may be attributed to various factors such as decreasing motivation, fatigue, or competing priorities that impact users' ability to maintain consistent levels of physical activity.

Reasoning:

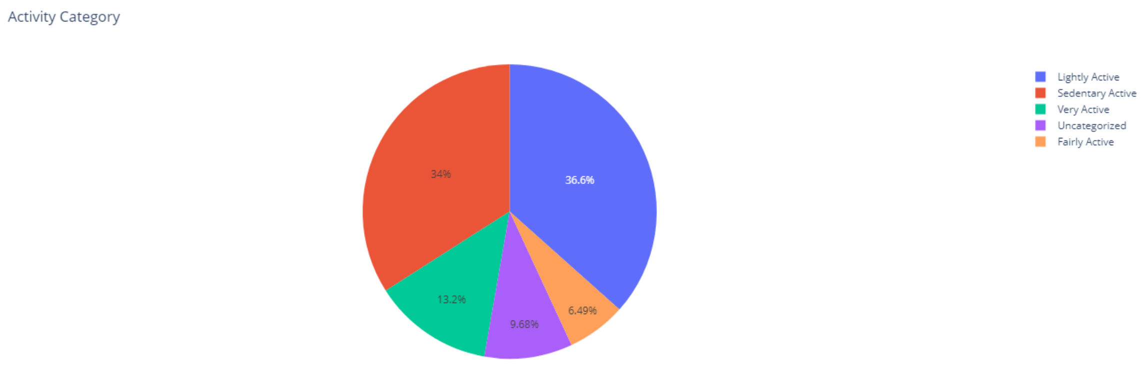
Several factors may contribute to the gradual decrease in total steps:

Fatigue and Habituation: Users may experience fatigue or habituation to their activity levels over time, leading to a decrease in motivation to maintain higher step counts.

Seasonal Variation: External factors such as weather conditions or seasonal changes may influence users' outdoor activities and contribute to fluctuations in step counts.

Workload and Time Constraints: Increasing workloads or time constraints towards the end of the month may limit users' opportunities for physical activity, resulting in lower step counts.

5. Distribution of users based activity category



Observation:

The majority of users fall into the lightly active category, followed by sedentary, very active, uncategorized, and fairly active categories.

Insight:

The analysis reveals a distribution of users across different activity categories, with the lightly active category having the highest number of users. This suggests that a significant portion of users engage in moderate levels of physical activity, while smaller proportions fall into other activity categories.

Reasoning:

The distribution of users across activity categories may be influenced by various factors such as:

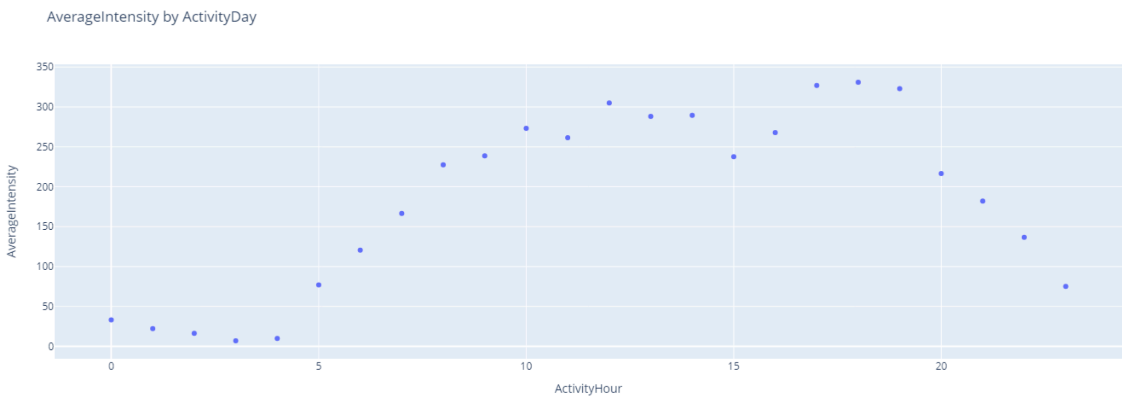
Lifestyle Preferences: Users may have varying preferences and habits related to physical activity, leading to differences in their activity levels and categorization.

Activity Tracking Accuracy: Differences in activity tracking accuracy or device usage patterns may affect the categorization of users into activity levels.

Sociodemographic Factors: Sociodemographic factors such as age, gender, occupation, and lifestyle choices may influence users' activity levels and their distribution across activity categories.

2. Hourly Activity Analysis:

1. Average Intensity by hour



Observation:

The average intensity of physical activity follows a pattern where it starts low at the beginning of the day, gradually increases, peaks during midday or early afternoon, and then gradually decreases towards the end of the day.

Insight:

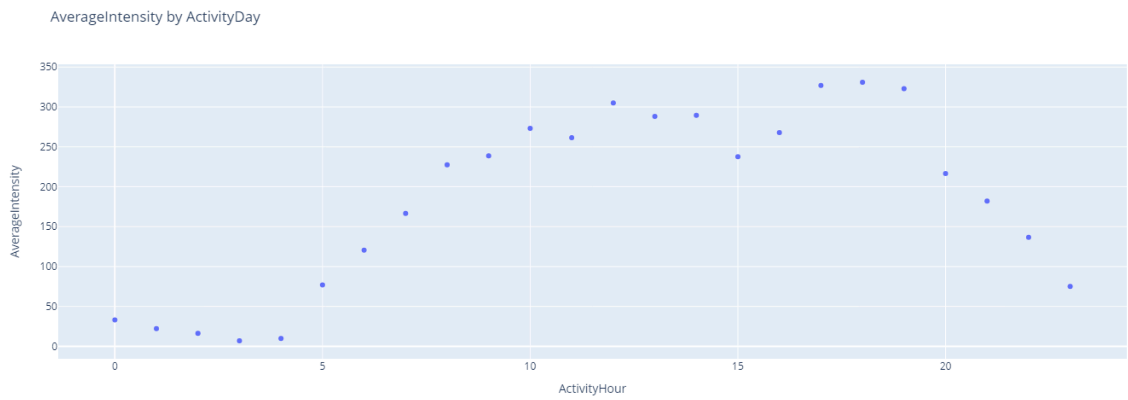
The analysis of average intensity by hour reveals a cyclical pattern throughout the day, with fluctuations corresponding to users' activity levels and energy expenditure. This pattern reflects typical human activity patterns, with higher levels of intensity during daytime hours and lower levels during nighttime hours.

Reasoning:

The observed pattern of average intensity by hour can be attributed to various factors such as:

> **Circadian Rhythms**: Human activity levels are influenced by circadian rhythms, with peak activity occurring during the daytime when individuals are most alert and energetic.
>
> **Work and Leisure Patterns**: Users' daily schedules, including work, school, and leisure activities, influence their activity levels and intensity throughout the day.
> **Physical and Mental Fatigue**: Changes in intensity levels may also reflect fluctuations in users' physical and mental fatigue, with energy levels typically higher in the morning and declining towards the evening.

2.Average Calories burned by hour



Observation:

The average number of calories burned follows a similar pattern to average intensity by hour, with higher calorie expenditure during peak activity hours and lower expenditure during low-intensity periods.

Insight:

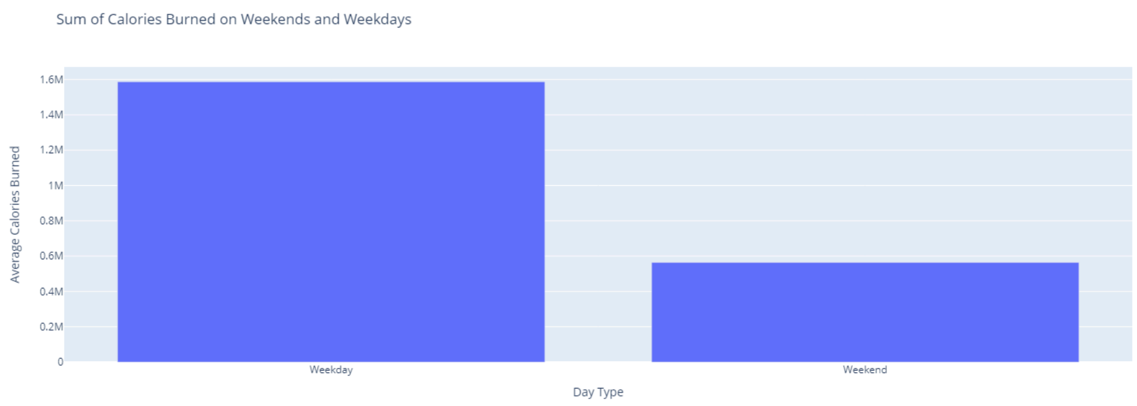
The analysis of average calories burned by hour reflects fluctuations in users' energy expenditure throughout the day, with higher calorie burn during periods of increased physical activity and lower burn during periods of rest or sedentary behavior.

Reasoning:

The observed pattern of average calories burned by hour can be attributed to similar factors as average intensity by hour, including circadian rhythms, work and leisure patterns, and physical and mental fatigue. Additionally, differences in activity types and intensity levels may contribute to variations in calorie expenditure across different hours of the day.

3. Minutes Activity Analysis:

1. Sum of calories burned per hour on weekdays comparing to weekend



Observation:

The sum of calories burned per hour is higher on weekdays compared to weekends, indicating higher overall energy expenditure during the weekdays.

Insight:

The analysis of calories burned per hour reveals a difference in energy expenditure between weekdays and weekends, with users typically burning more calories during weekdays when they are engaged in work, school, and other weekday activities.

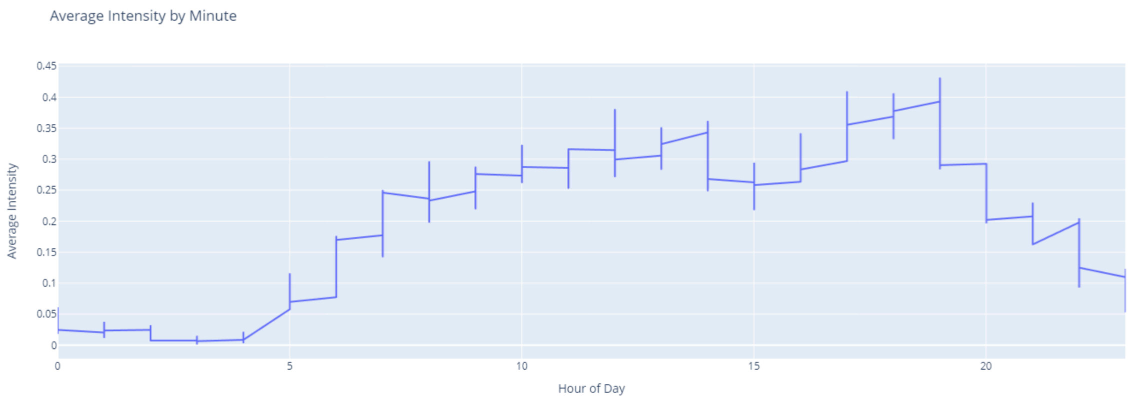
Reasoning:

Several factors may contribute to the higher sum of calories burned per hour on weekdays compared to weekends:

Work and School Activities:
Weekdays are typically characterized by higher levels of structured activities such as work, school, and commuting, which involve increased physical movement and energy expenditure.
Routine and Schedule: Users may follow a more structured routine during weekdays, including regular meal times, exercise routines, and daily activities, which contribute to higher overall energy expenditure.

Leisure and Relaxation: Weekends may be associated with more leisurely activities, relaxation, and downtime, which may involve lower levels of physical activity and energy expenditure compared to weekdays.

2.Average Intensity by Minute :



Observation:

The average intensity of physical activity varies throughout the day, with fluctuations observed in intensity levels across minutes.

Insight:

The analysis of average intensity by minute reveals a pattern where intensity levels are lower during the initial hours of the day, gradually increase as the day progresses, peak during midday or early afternoon, and then gradually decrease towards the end of the day.

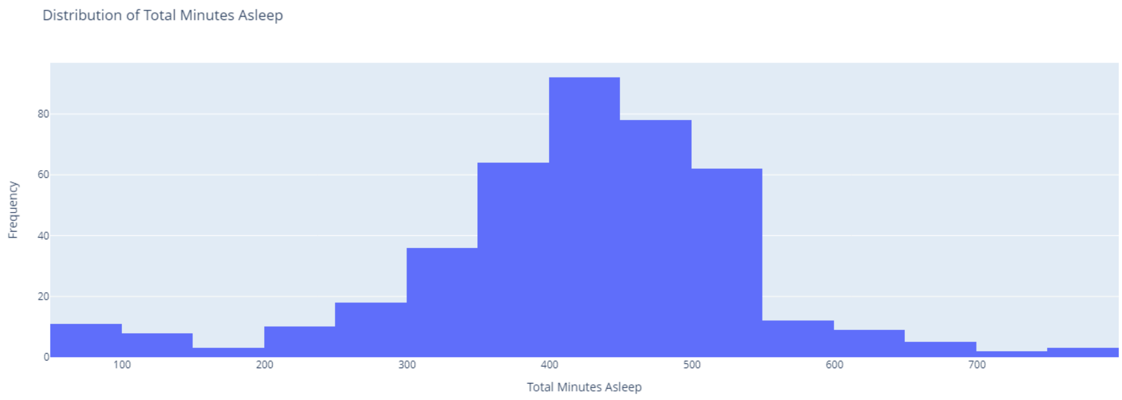
Reasoning:

The observed pattern of average intensity by minute can be attributed to various factors such as:

> **Circadian Rhythms**: Human activity levels are influenced by circadian rhythms, with peak activity occurring during the daytime when individuals are most alert and energetic, leading to higher intensity levels during midday hours.
>
> **Work and Leisure Patterns**: Users' daily schedules, including work, school, and leisure activities, influence their activity levels and intensity throughout the day, with differences in activity patterns contributing to fluctuations in intensity levels.
>
> **Physical and Mental Fatigue**: Changes in intensity levels may also reflect fluctuations in users' physical and mental fatigue, with energy levels typically higher in the morning and declining towards the evening, resulting in variations in intensity levels across minutes.

4. Sleep Data Analysis:

1.Distribution of Total Minutes Slept :



Observation:

The majority of users have total sleep durations falling within the range of 400-449 minutes, with the highest number of users in this category.

Insight:

The analysis of total minutes slept reveals a concentration of users within a specific range of sleep durations, with the majority falling within the 400-449 minutes range. This suggests a common sleep duration pattern among users, with variations around this central tendency.

Reasoning:

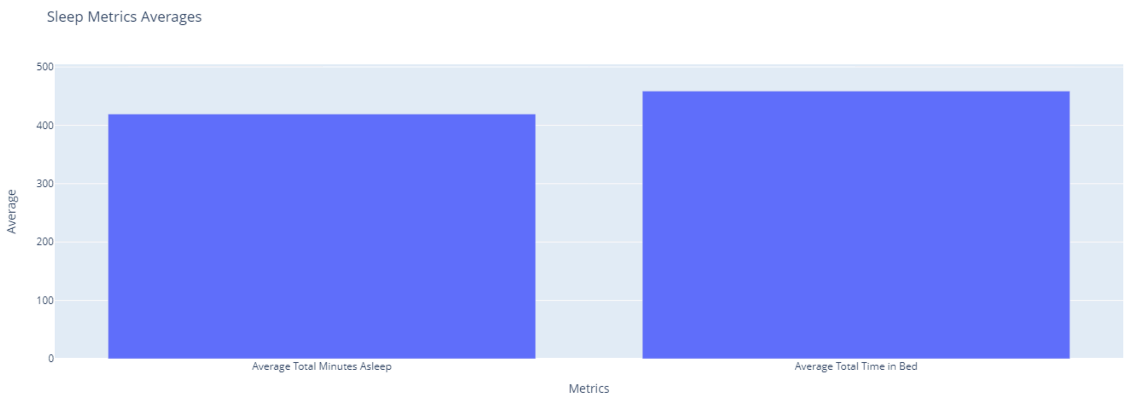
Several factors may contribute to the distribution of total minutes slept among users:

Sleep Needs and Preferences: Individuals have varying sleep needs and preferences, with some requiring more or less sleep than others to feel rested and refreshed.

Sleep Quality and Efficiency: Differences in sleep quality and efficiency may influence the total duration of sleep recorded by users, with factors such as sleep disturbances, environmental conditions, and lifestyle habits impacting overall sleep patterns.

Health and Wellbeing: Users' overall health and wellbeing, including factors such as stress levels, physical activity levels, and medical conditions, may affect their sleep duration and quality.

2. Total time in bed vs Total time asleep :



Observation:
The total time spent in bed is slightly higher than the total time taken to fall asleep, indicating a short period of time spent awake in bed before falling asleep.

Insight:

The analysis of total time spent in bed versus the total time taken to fall asleep reveals a pattern where users typically spend slightly more time in bed than the time it takes them to fall asleep. This suggests a relatively short latency period between getting into bed and falling asleep for most users.

Reasoning:

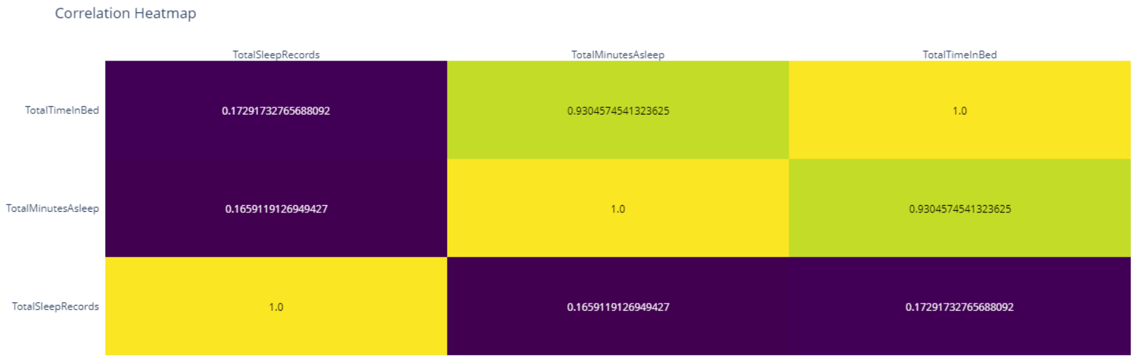
Several factors may contribute to the observed relationship between total time in bed and total time to fall asleep:

Sleep Latency: The time taken to fall asleep, also known as sleep latency, varies among individuals and can be influenced by factors such as stress levels, bedtime routines, and environmental conditions.

Sleep Efficiency: Users' sleep efficiency, which reflects the percentage of time spent asleep while in bed, may affect the relationship between total time in bed and total time to fall asleep.

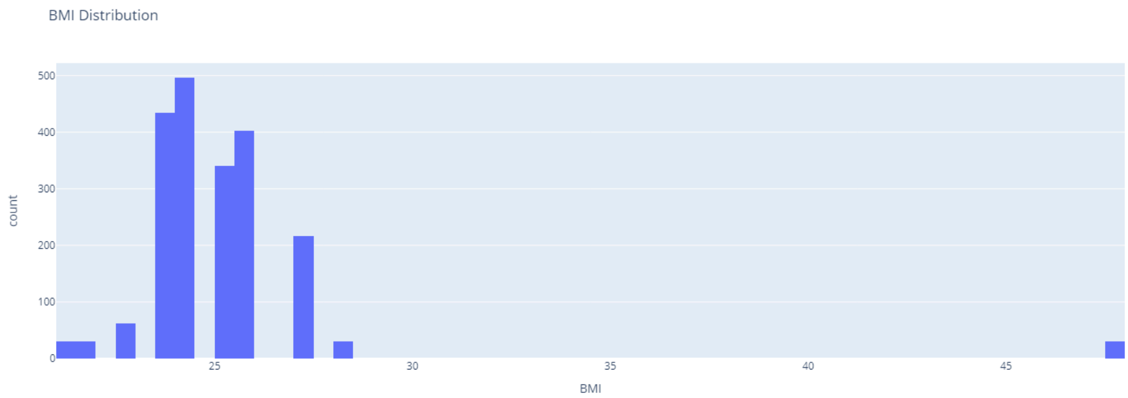
Sleep Architecture: Individual differences in sleep architecture, including sleep stages and cycles, may impact the time it takes users to fall asleep and the overall duration of time spent in bed.

3. Correlation Heatmap :



5. Weight Log Analysis :

1.Distribution of BMI Categories :



Observation:

The highest number of users falls into the BMI category range of 24-25, indicating a concentration of users within this range.

Insight:

The analysis of BMI distribution reveals a peak in the number of users within the BMI category range of 24-25, suggesting a common BMI profile among users.

Reasoning:

Several factors may contribute to the observed distribution of BMI categories among users:

> **Population Demographics:** Users' BMI distribution may reflect the demographics of the population being analyzed, with certain BMI ranges being more prevalent within specific demographic groups.
>
> **Health Status and Lifestyle**: Differences in users' health status, lifestyle choices, dietary habits, and physical activity levels may influence their BMI and contribute to variations in BMI distribution.
>
> **Genetic and Environmental Factors:** Genetic predispositions and environmental factors such as access to healthcare, socioeconomic status, and cultural influences may also play a role in shaping users' BMI profiles.

2. Categorization of weight status :

Observation:
Users are categorized into normal, overweight, and obese weight status categories, with approximately 50% classified as normal weight, 47.7% as overweight, and the remaining percentage as obese.

Insight:

The analysis of weight status categorization reveals a distribution of users across different weight status categories, with a relatively equal proportion of users classified as normal weight and overweight, and a smaller percentage classified as obese.

Reasoning:

The observed distribution of weight status categories among users may be influenced by various factors such as:

BMI Calculation Method: Users' weight status categorization is based on their

BMI calculated from height and weight measurements, using established criteria for defining normal weight, overweight, and obesity.

Population Characteristics: Users' weight status distribution may reflect the characteristics of the population being analyzed, including demographic factors, lifestyle behaviors, and health status.

Health and Lifestyle Factors: Differences in users' health behaviors, dietary habits, physical activity levels, and access to healthcare services may contribute to variations in weight status distribution.

3. Distribution of Calories Burned Across Weight Status Categories :

Observation:

Users in the overweight weight status category have the highest calorie expenditure, followed by users in the normal weight and obese categories.

Insight:

The analysis of calorie expenditure across weight status categories reveals variations in energy expenditure among users, with higher calorie burn observed among users classified as overweight compared to those classified as normal weight or obese.

Reasoning:

Several factors may contribute to the observed distribution of calorie expenditure across weight status categories:

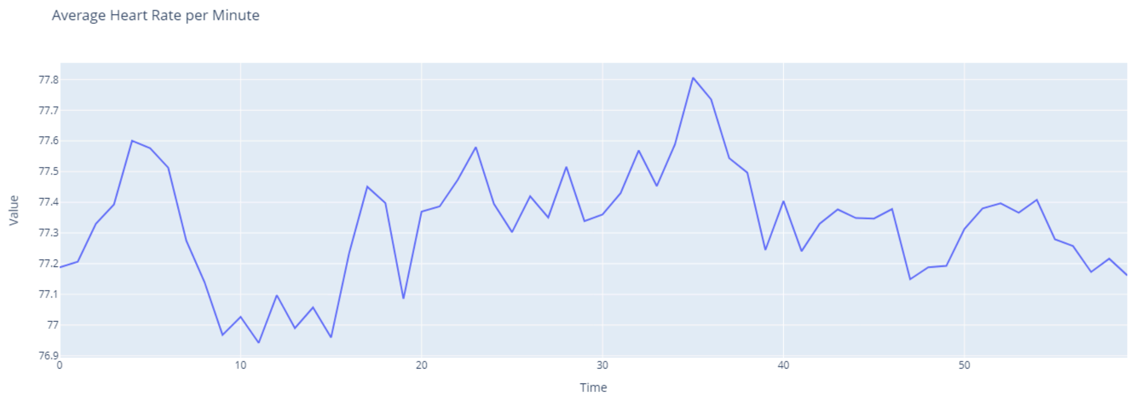
Metabolic Differences: Differences in metabolic rate, energy expenditure, and thermogenesis among individuals with different weight statuses may influence their calorie expenditure during physical activity and daily tasks.

Physical Activity Levels: Variations in physical activity levels, exercise intensity, and lifestyle behaviours among users with different weight statuses can impact their overall calorie burn and energy expenditure.

Body Composition: Variations in body composition, including muscle mass, fat distribution, and lean body mass, may contribute to differences in energy expenditure and calorie burn among users with different weight statuses.

6. Heart Rate Analysis:

11. Average Heart Rate per Minute :



Observation:

The average heart rate per minute varies in the range of 76.9 to 77.8 across users.

Insight:

The analysis of average heart rate per minute reveals a consistent range of values across users, with variations within a narrow range. This suggests a stable average heart rate pattern among users during the analyzed period.

Reasoning:

Several factors may contribute to the observed range of average heart rate per minute:

Individual Variability: Heart rate is influenced by various factors including age, fitness level, medication use, and overall health status, resulting in individual variability in heart rate patterns.

Physical Activity Levels: Differences in users' physical activity levels and exercise intensity may impact their average heart rate, with higher activity levels generally associated with higher average heart rates.

Resting Heart Rate: Resting heart rate, which reflects the heart's activity during periods of rest and relaxation, also contributes to the overall average heart rate per minute.

Conclusion:

Behavioral Patterns:
Users exhibit varying behavioral patterns in their usage of the Fitbit app, including fluctuations in physical activity levels, sleep duration, and heart rate throughout the day and month.

Activity Trends:
There is a consistent trend of higher activity levels, such as calories burned and steps taken, at the beginning of the month, which gradually decrease as the month progresses. Additionally, users tend to accumulate more sedentary minutes compared to other activity levels.

User Segmentation: The majority of users fall into the lightly active category, followed by sedentary, very active, uncategorized, and fairly active categories, indicating a diverse user base with varying activity levels.

Health Metrics: Users' health metrics, including BMI distribution and weight status categorization, reveal insights into their overall health and wellbeing, with implications for weight management and cardiovascular health.

Marketing Strategies:

- Targeted Engagement Campaigns: Implement targeted engagement campaigns at the beginning of each month to capitalise on users' higher motivation and activity levels, encouraging them to set and achieve fitness goals.

❖ Behavioural Nudges: Utilise behavioural science principles to design nudges and prompts within the app to encourage users to reduce sedentary behaviour and increase physical activity levels throughout the day.

❖ Personalised Recommendations: Offer personalised recommendations and content based on users' activity levels, health metrics, and individual goals to promote sustained engagement and adherence to healthy habits.

❖ Community Challenges: Organise community challenges and competitions within the app to foster social support, motivation, and accountability among users, driving increased participation and activity levels.

❖ Health Education Campaigns: Develop educational campaigns and resources within the app to raise awareness about the importance of maintaining a healthy weight, managing stress, improving sleep quality, and optimising overall health and wellbeing.

❖ Premium Features: Offer premium features or subscription-based services within the app that provide advanced analytics, personalised coaching, and tailored recommendations to users seeking more in-depth insights and guidance.

❖ Partnerships and Integrations: Explore partnerships with health and wellness organisations, fitness professionals, and healthcare providers to offer integrated solutions and services that complement the Fitbit app, enhancing its value proposition for users.

❖ Feedback and Support: Solicit feedback from users through surveys, polls, and user feedback mechanisms within the app to continuously improve and enhance user experience, features, and functionalities.