Analysis of FitBit Fitness Tracker App Data

PREPARED BY:

SHIVANG BHATNAGAR

SHIVANGBHATNAGARR@GMAIL.COM

APRIL 3, 2024

TABLE OF CONTENTS

INTRODUCTION

The FitBit Fitness Tracker App has revolutionized the way individuals monitor and manage their physical activity, sleep patterns, and overall well-being. As consumers increasingly prioritize health and wellness, the FitBit app serves as a valuable tool for tracking daily activity levels, setting fitness goals, and monitoring progress over time. In this report, we delve into the wealth of data collected by the FitBit app to uncover insights that can inform marketing strategies, enhance user engagement, and optimize the app's features to better meet the needs of its diverse user base.

With the proliferation of wearable technology and the growing emphasis on personal health and fitness, understanding how consumers interact with the FitBit app is paramount. By analyzing user behavior, trends, and preferences, we aim to provide actionable insights that will enable marketing teams to tailor their messaging, target key demographics more effectively, and develop compelling campaigns that resonate with FitBit users.

This analysis spans a comprehensive dataset encompassing user activity, sleep metrics, demographic information, and more, collected over a specified time period. By leveraging advanced analytical techniques, we seek to uncover patterns, correlations, and trends within the data, offering valuable insights into user engagement, feature usage, and broader market trends.

Through this exploration of FitBit app data, we endeavor to not only understand current user behaviors and preferences but also to anticipate future trends and opportunities. By aligning marketing strategies with the needs and preferences of FitBit users, we can enhance the app's relevance, drive user acquisition and retention, and ultimately, empower individuals to lead healthier, more active lifestyles.

In the following sections of this report, we will delve into the data sources and cleaning process, conduct exploratory data analysis to uncover key insights, and present actionable recommendations to guide marketing strategy and decision-making.

DATA SOURCES

The analysis in this report draws upon a total of 18 datasets, each offering unique insights into different aspects of user activity, health, and engagement within the FitBit app. These datasets include:

1. **Sleep Metrics**: Data on users' sleep duration, sleep stages, and sleep quality scores.
2. **Heart Rate Data**: Measurements of users' heart rates captured by FitBit devices.
3. **Weight Measurements**: Information on users' weight, BMI, and weight fluctuations over time.
4. **Hourly Activity Data**: Records of users' activity levels, steps taken, and calories burned on an hourly basis.
5. **Minute-Level Activity Data**: Granular details of users' activity levels and intensities recorded at minute intervals.
6. **Daily Activity Logs**: Summaries of users' daily activity, including steps taken, calories burned, and active minutes.

These datasets collectively provide a comprehensive view of users' behavior, preferences, and health metrics within the FitBit ecosystem. By analyzing this wealth of data, we aim to identify key trends, patterns, and correlations that can inform marketing strategies, product enhancements, and user engagement initiatives.

DATA CLEA NING AND MANIPULATION

Data cleaning and manipulation are essential steps in preparing the FitBit dataset for analysis. This process involves identifying and resolving inconsistencies, missing values, and formatting issues to ensure the integrity and quality of the data. Here's an overview of the steps involved in data cleaning and manipulation:

1. Handling Missing Values:
   - Identify columns with missing values and decide on appropriate strategies for handling them.
   - Options include imputation (replacing missing values with a suitable estimate), deletion of rows or columns with missing values, or flagging missing values for further analysis.
2. Data Type Conversion:
   - Ensure that variables are stored in appropriate data types for analysis.
   - Convert date/time variables to datetime format for time series analysis.
   - Convert categorical variables to the categorical data type if applicable.
3. Removing Duplicate Entries:
   - Identify and remove any duplicate rows from the dataset to avoid redundancy and ensure accuracy in analysis.
4. Handling Outliers:
   - Identify outliers in the data and decide on appropriate strategies for handling them.
   - Options include removing outliers, transforming the data, or using robust statistical methods that are less sensitive to outliers.
5. Data Aggregation and Reshaping:
   - Aggregate data at different levels (e.g., daily, weekly) to analyze trends and patterns over time.

ANALYSIS

DAILY ACTIVITY DATASET

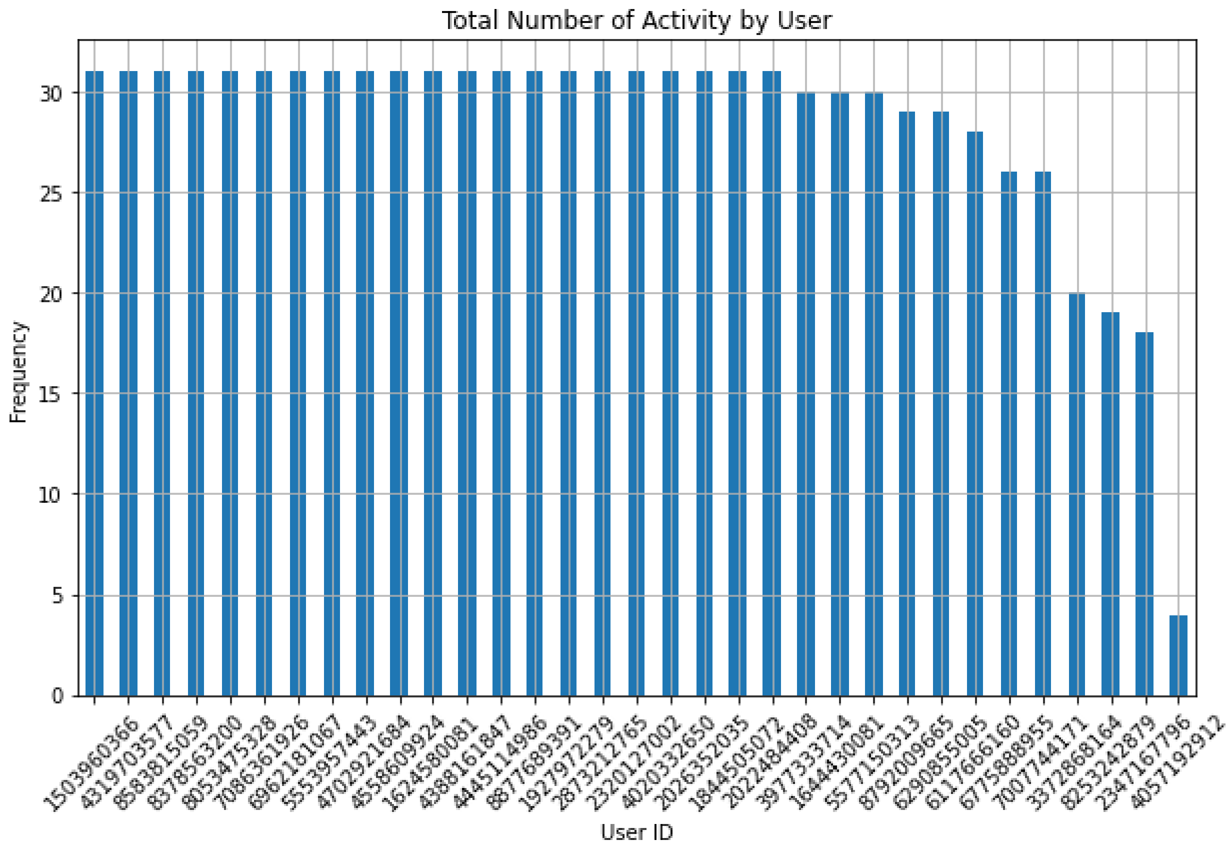
OVERVIEW OF THE DATASET

The daily activity dataset contains user-specific records of daily activity metrics such as steps taken, distance covered, calories burned, and active minutes.

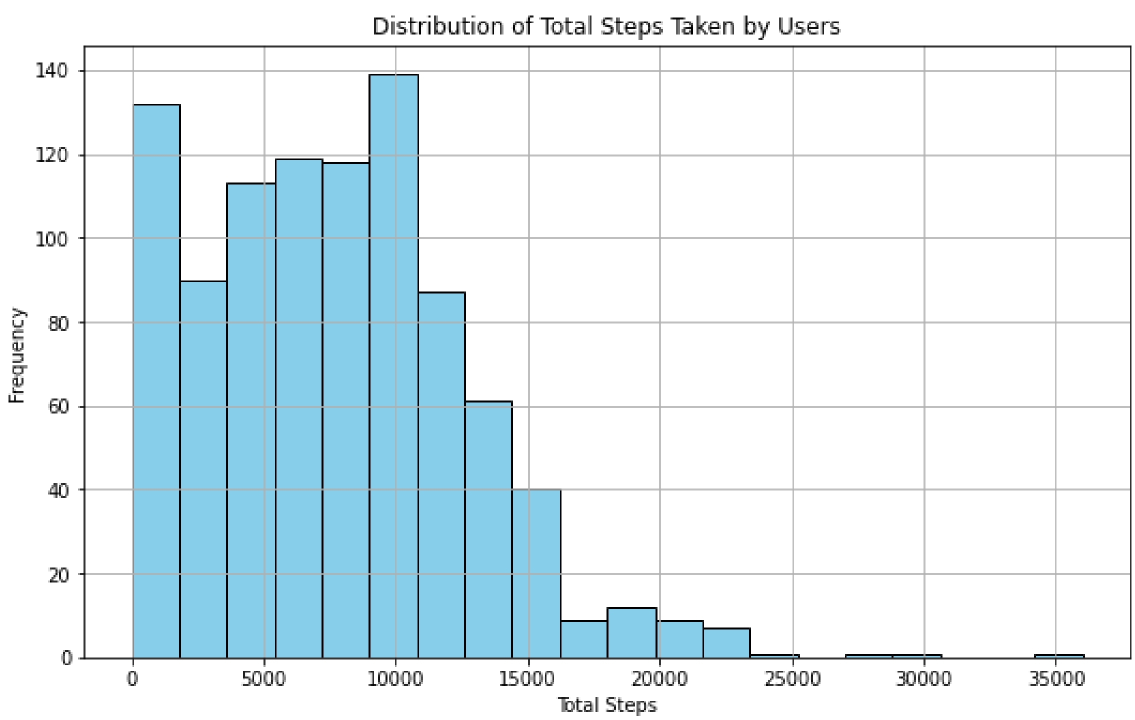
	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMinutes	Calories
0	1503960366	4/12/2016	13162	8.50	8.50	0.0	1.88	0.55	6.06	0.0	25	13	328	728	1985
1	1503960366	4/13/2016	10735	6.97	6.97	0.0	1.57	0.69	4.71	0.0	21	19	217	776	1797
2	1503960366	4/14/2016	10460	6.74	6.74	0.0	2.44	0.40	3.91	0.0	30	11	181	1218	1776
3	1503960366	4/15/2016	9762	6.28	6.28	0.0	2.14	1.26	2.83	0.0	29	34	209	726	1745
4	1503960366	4/16/2016	12669	8.16	8.16	0.0	2.71	0.41	5.04	0.0	36	10	221	773	1863

INSIGHTS

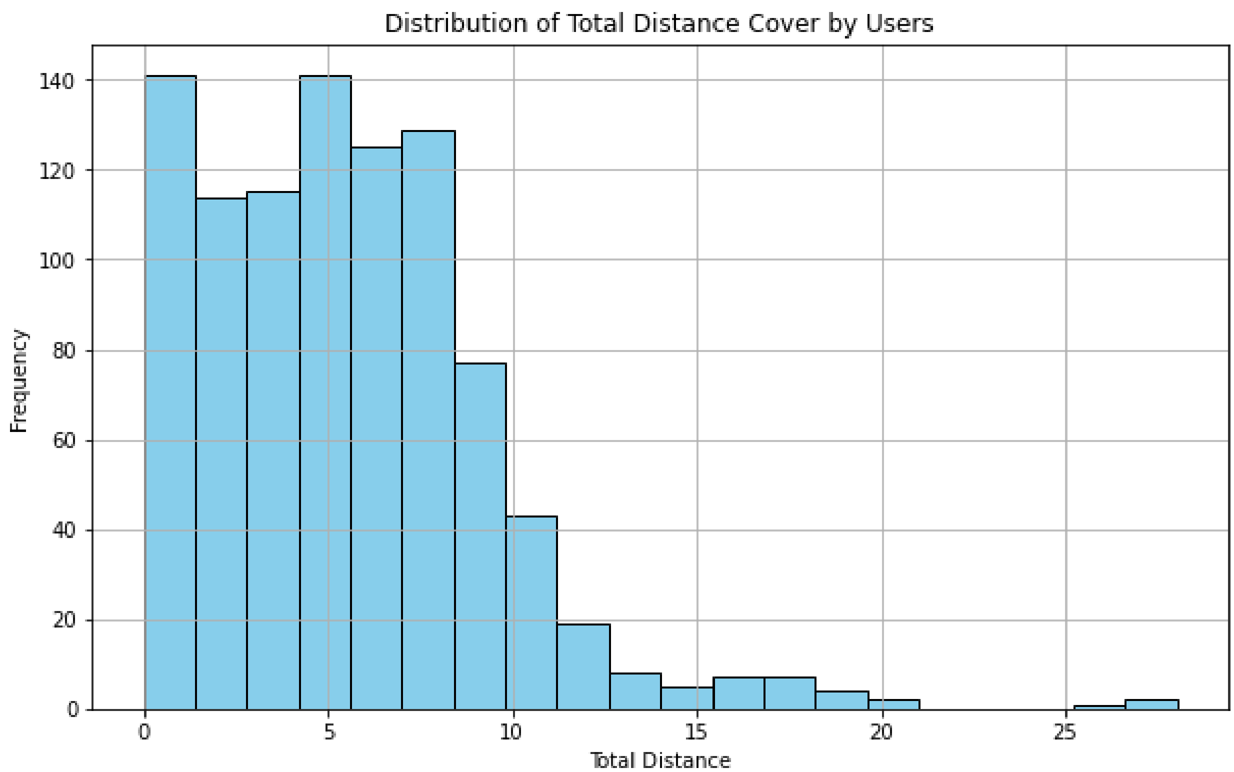
1. We have grouped the data of users to find the total numbers of days on which the user performs some activity. From the figure below, we can see that most of the users performs or log their daily activity data for the whole time period of experiment (13 April 2016-13 May 2016) i.e 30-31 days. While there are some users as well who does not log their daily activity data. We can see the down trend in the daily activity log data.



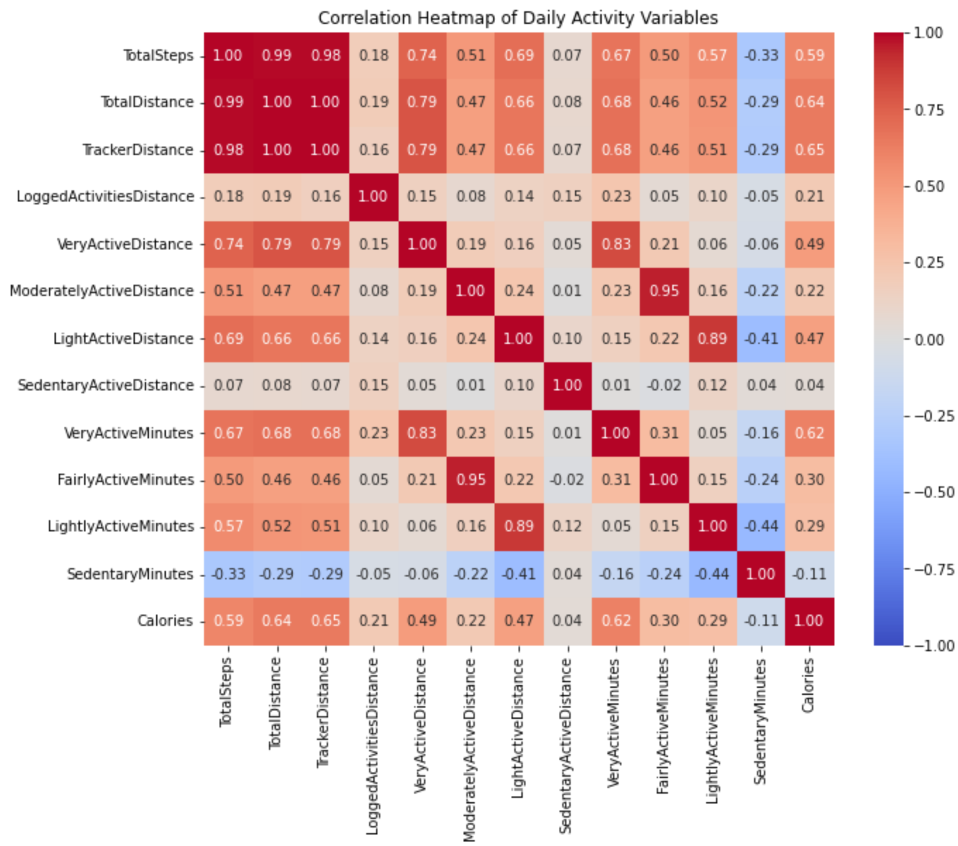
2. The graph shows that most users take between 10,000 and 15,000 steps per day. There is a smaller group of users who take either less than 10,000 steps or more than 15,000 steps per day.



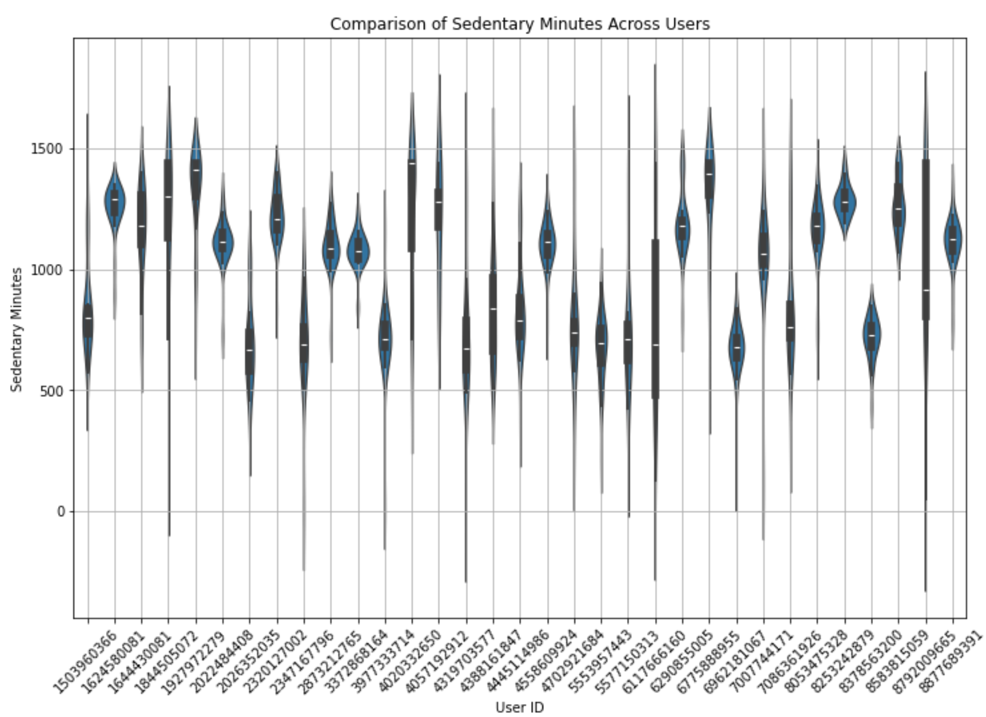
3. The majority of users appear to travel between 5 and 15 kilometers. This suggests that the users may be using this mode of transportation for short trips, such as commuting to work or running errands.



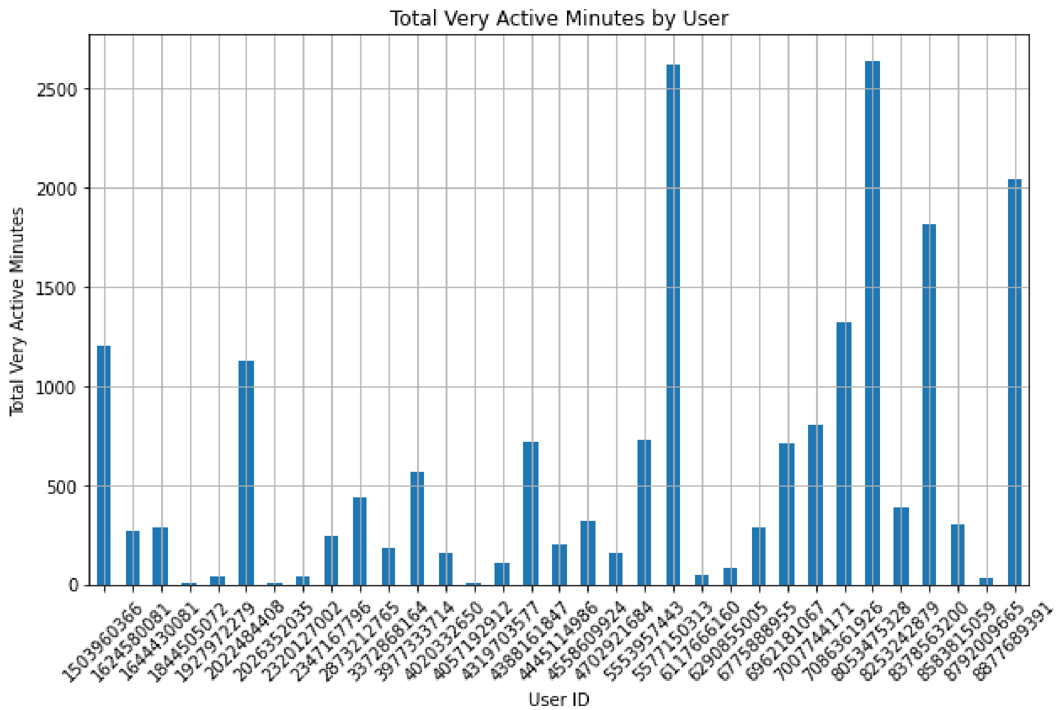
4. There is a very strong positive correlation between all the step-based metrics (TotalSteps, TotalDistance, TrackerDistance) and all the minute-based metrics (VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes). This means that people who take more steps tend to also spend more time being active. There is a very strong negative correlation between SedentaryMinutes and all the other activity metrics. This means that people who are more active tend to spend less time sedentary.



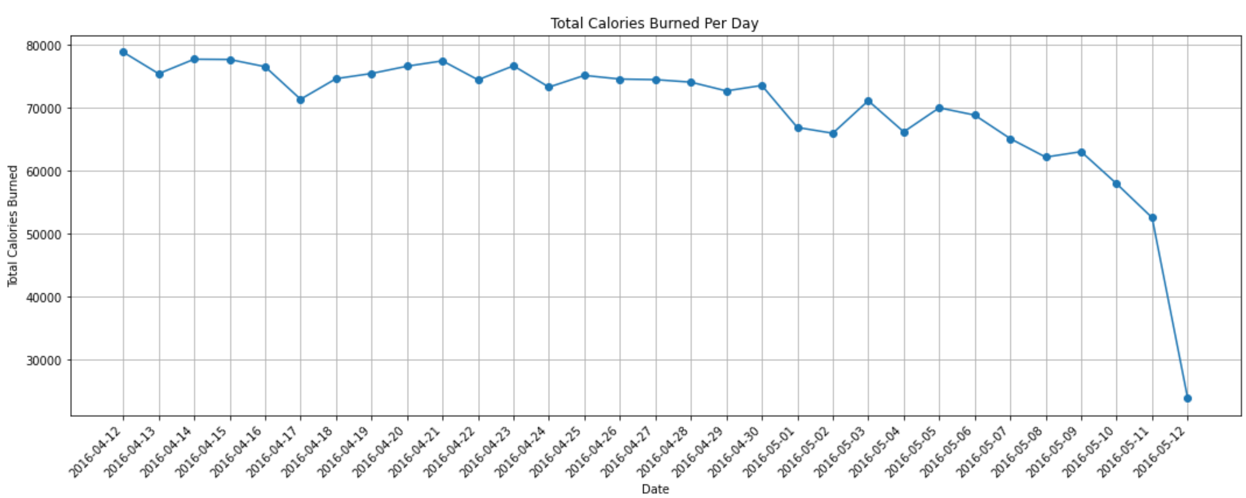
5. There is a wide range of sedentary minutes among users. Some users have sedentary minutes as low as 1503960366, while others have sedentary minutes as high as 8877689391.



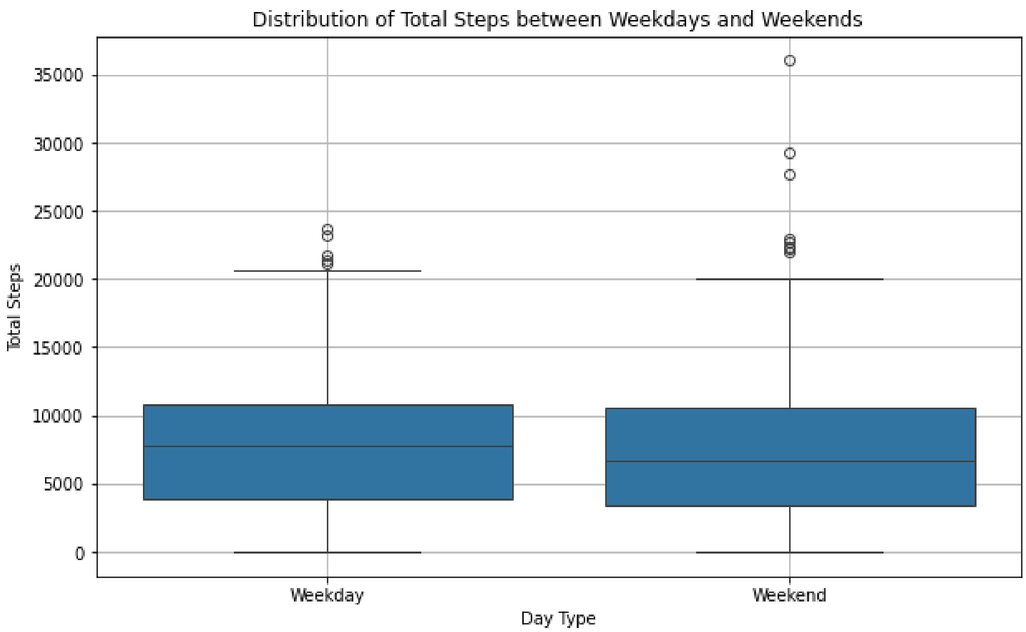
6. Now we have aggregated the data to see the total number of minutes during which the users are very active. From the graph below, we can see that very few users have cross the mark of 2000 minutes and most of the users are very active for total of below 500 minutes.



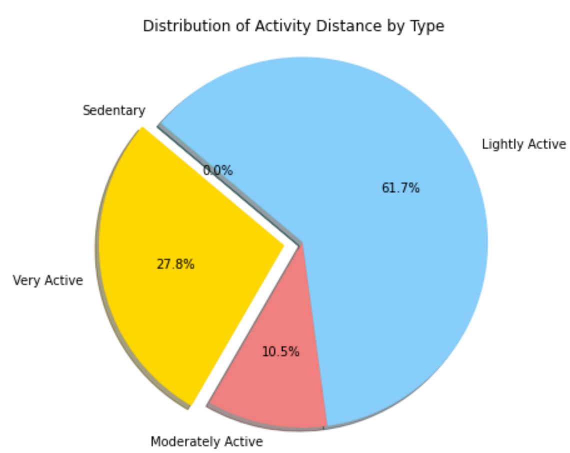
7. There is a significant variation in the total calories burned per day by users. Some days show calorie burn as low as 20,000, while others show days as high as 80,000.It appears there is no clear trend or pattern in the data over the 40 days. The calorie expenditure seems to fluctuate throughout the time period.



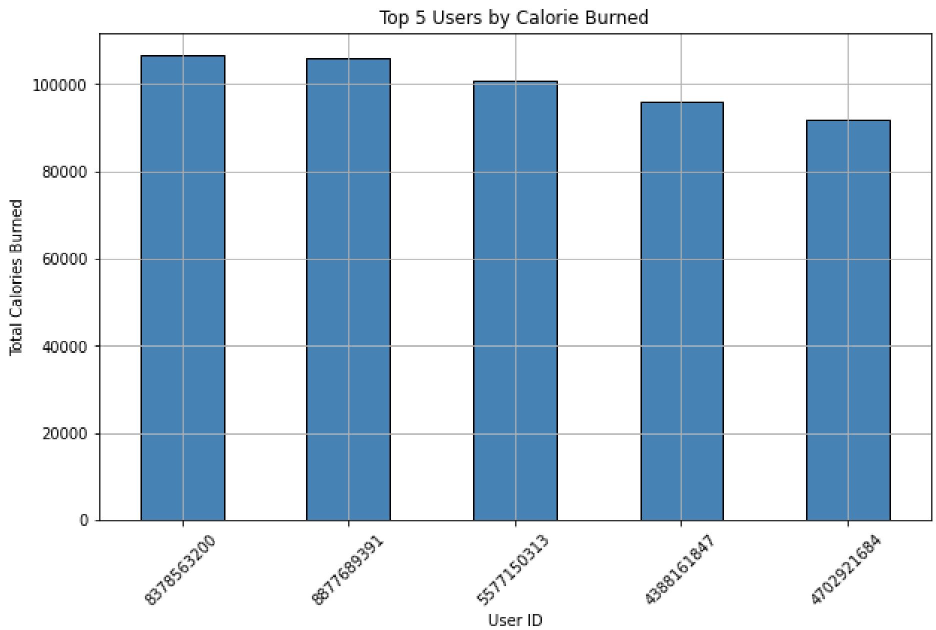
8. The graph shows that most users take more steps on weekdays than on weekends. This suggests that people tend to be more active during the week, possibly due to work or commuting. The distribution of steps on weekdays appears to be centered around 10,000 to 15,000 steps. There is a smaller group of users who take either less than 10,000 steps or more than 15,000 steps on weekdays. The distribution of steps on weekends appears to be centered around 5,000 to 10,000 steps. There is a larger group of users who take fewer than 10,000 steps on weekends compared to weekdays.

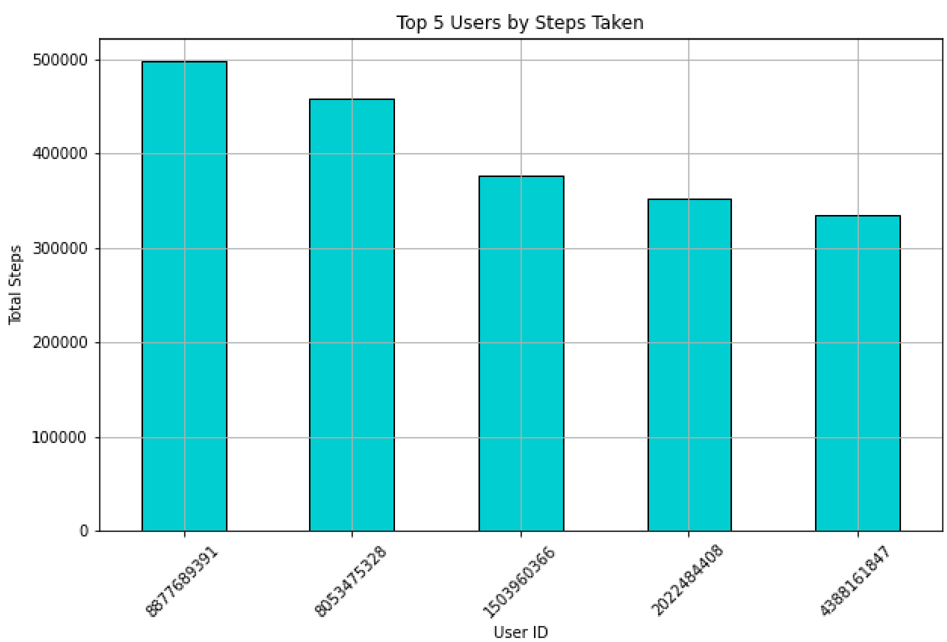


9. The largest portion of the pie chart (61.7%) is labeled "Sedentary". This suggests that users spend the majority of their time engaged in sedentary activities, which means they are likely not moving much. The second-largest slice (27.8%) is labeled "Lightly Active". This suggests that users spend a significant amount of time in activities that require slight movement. The remaining slices, labeled "Moderately Active" (10.5%) and "Very Active" (0.0%) are much smaller. This suggests that users spend a much smaller portion of their time in activities classified as moderately or very active.

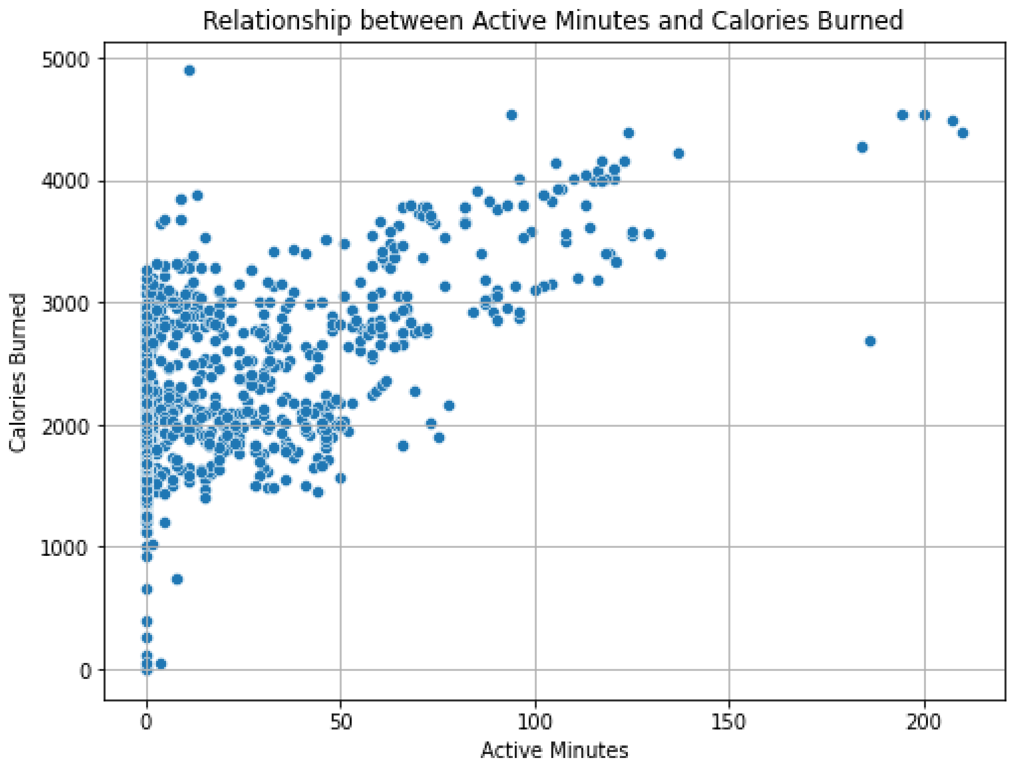


10. The table shows the top 5 users who burned the most calories. The user with the ID 8378563200 burned the most calories, followed by the user with the ID 8877689391, and so on.





11. These is a positive correlation between active minutes and calories burned by the users. As we can see, calories burn increase with increase in active minutes. So users with more active minutes have burn more calories as compares to others. But there are some users as well who have burn more calories in less minutes of activity.



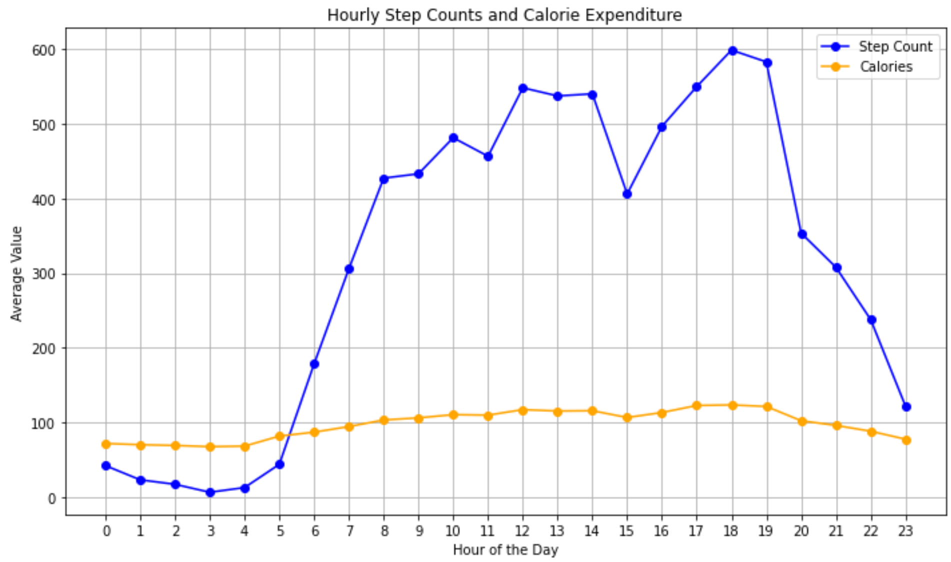
HOURLY DATASET

OVERVIEW OF THE MERGED DATASET

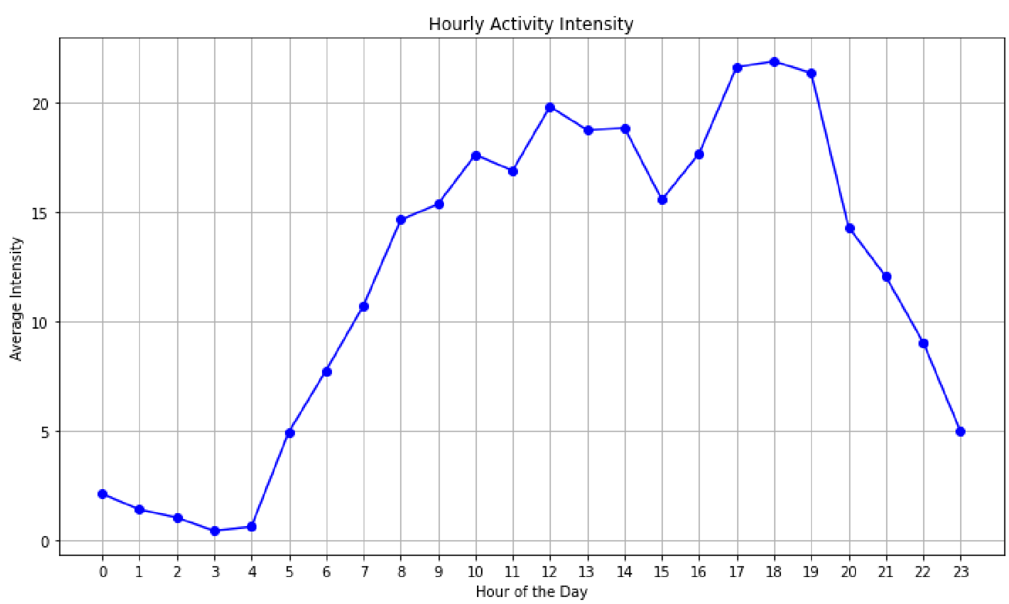
	Id	ActivityHour	Calories	StepTotal	TotalIntensity	AverageIntensity	DayofWeek
0	1503960366	2016-04-12 00:00:00	81	373	20	0.333333	Tuesday
1	1503960366	2016-04-12 01:00:00	61	160	8	0.133333	Tuesday
2	1503960366	2016-04-12 02:00:00	59	151	7	0.116667	Tuesday
3	1503960366	2016-04-12 03:00:00	47	0	0	0.000000	Tuesday
4	1503960366	2016-04-12 04:00:00	48	0	0	0.000000	Tuesday

INSIGHTS

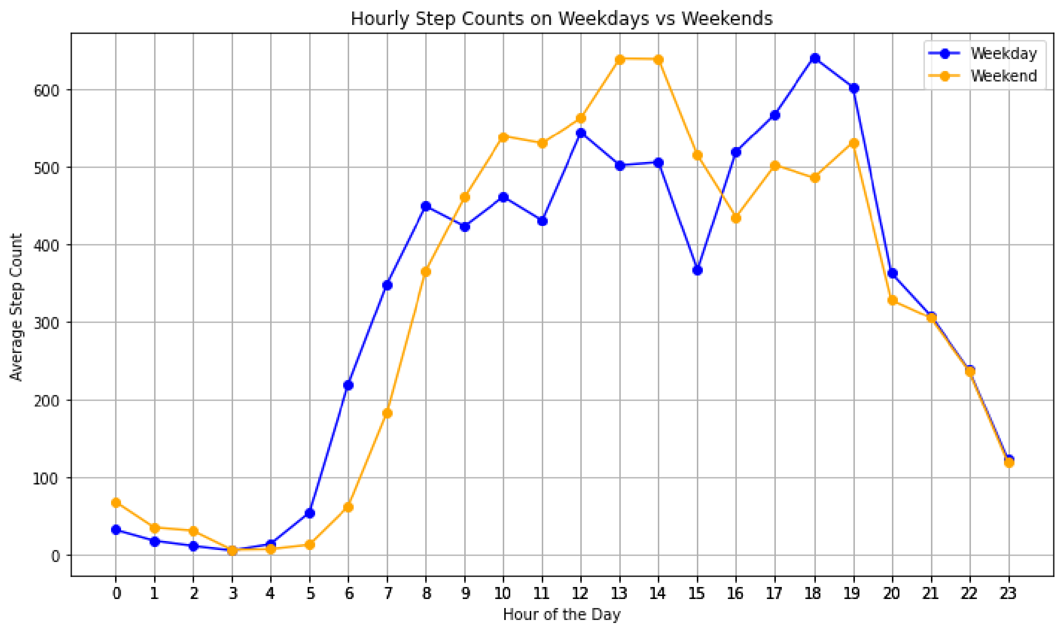
1. It can be seen that both the calorie expenditure and step count are starts increasing from morning 5 A.M to Evening 7 P.M and then starts decreasing. It is due to the fact that most of the people wakes up early and goes for running or do some exercise.



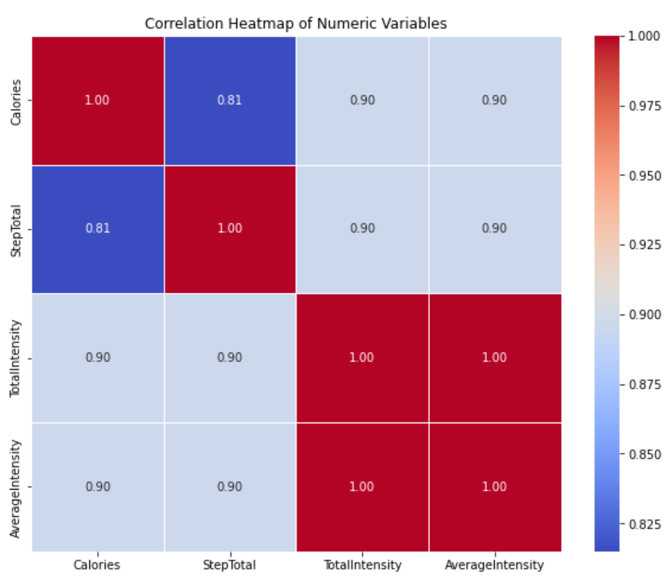
2. Analysis of the hourly data reveals that the intensity of the activities are starts increasing from morning 5 A.M to evening 7 P.M and then starts decreasing. This is due to the fact that most people engage in early morning workouts, running and other physical activities and then they leave for their workplace where they engage in work till 7 P.M which justifies the upside trend of intensity level.



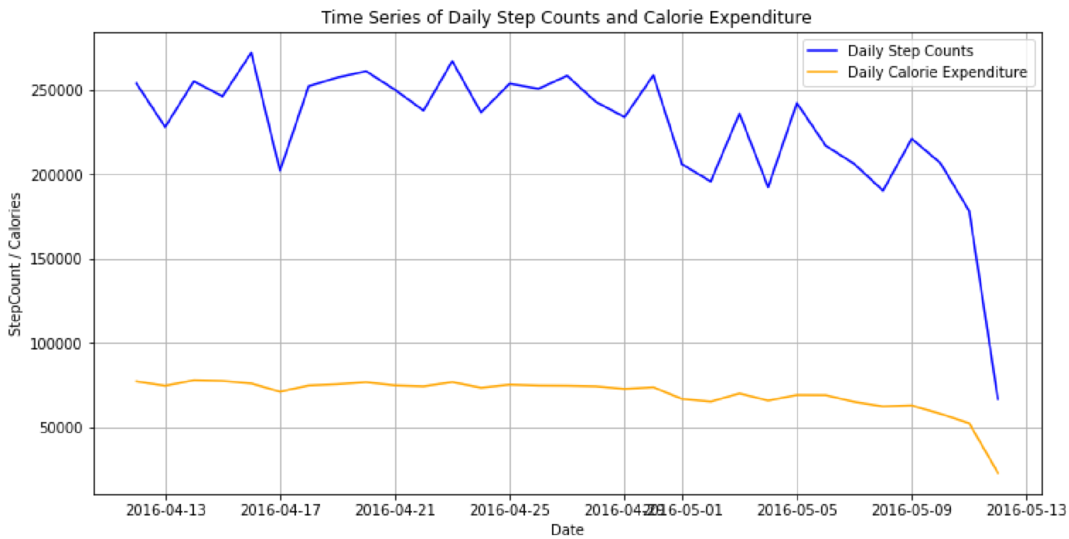
3. Step count of users on weekdays are more for some particular period of time as compared to weekends. Users on weekend evening takes less steps as compared to weekday evening. This is due to the fact that on weekend evening, users prefer to stay at home rather than going outside.



4. There is a very strong positive correlation between Sleep Duration and all the other sleep metrics (Restless Sleep, Deep Sleep, Light Sleep, Sleep Efficiency). This means that people who sleep for longer durations tend to also get more restful sleep, deep sleep, light sleep, and have better sleep efficiency. There is a moderate positive correlation between Sleep Duration and Weekend. This suggests that people tend to sleep for longer durations on weekends compared to weekdays.



5. If we observe the long term trend of step count and calories expenditure, we can clearly see that there are fluctuations in the daily total step count of all users and before the last day, the step count of all the users decrease drastically. While calories expenditure has remain almost constant with low fluctuations throughout the experiment and on the last day, it decreases.



MINUTE WISE ACTIVITY DATASET

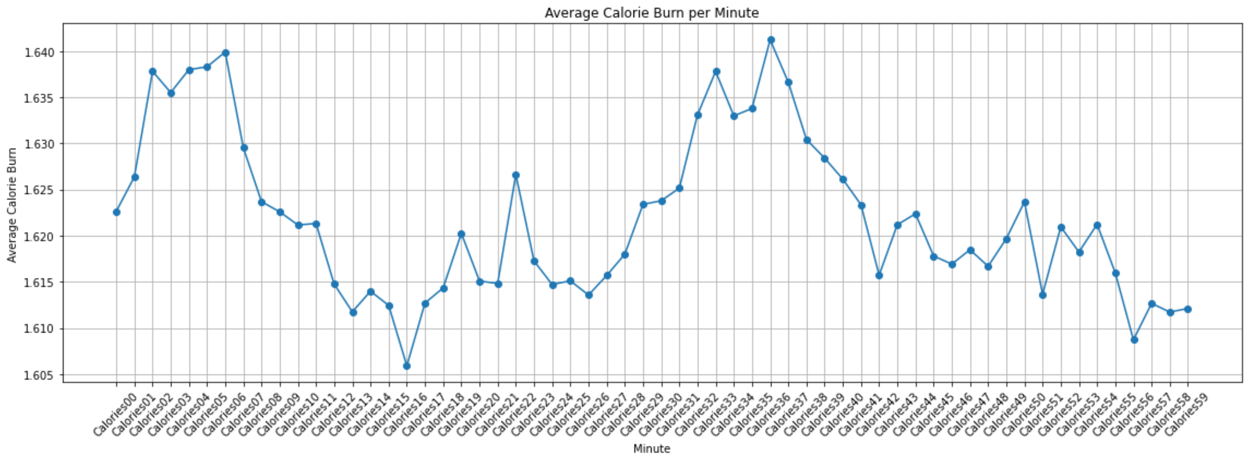
OVERVIEW OF THE MERGED DATASET

	Id	ActivityHour	Calories00	Calories01	Calories02	Calories03	Calories04	Calories05	Calories06	Calories07	...	Steps50	Steps51	Steps52	Step
0	1503960366	4/13/2016 12:00:00 AM	1.8876	2.2022	0.9438	0.9438	0.9438	2.0449	0.9438	2.2022	...	0	9	8	
1	1503960366	4/13/2016 1:00:00 AM	0.7865	0.7865	0.7865	0.7865	0.9438	0.9438	0.9438	0.7865	...	0	0	0	
2	1503960366	4/13/2016 2:00:00 AM	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	...	0	0	0	
3	1503960366	4/13/2016 3:00:00 AM	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	...	0	0	0	
4	1503960366	4/13/2016 4:00:00 AM	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	0.7865	...	0	0	0	

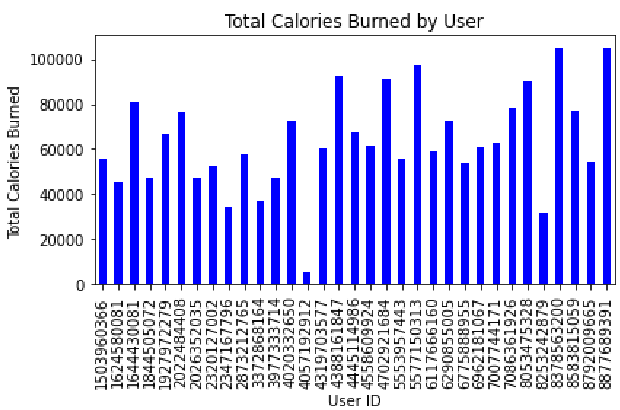
5 rows × 182 columns

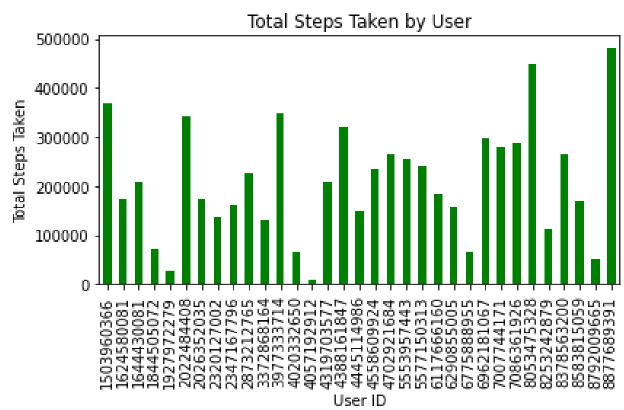
INSIGHTS

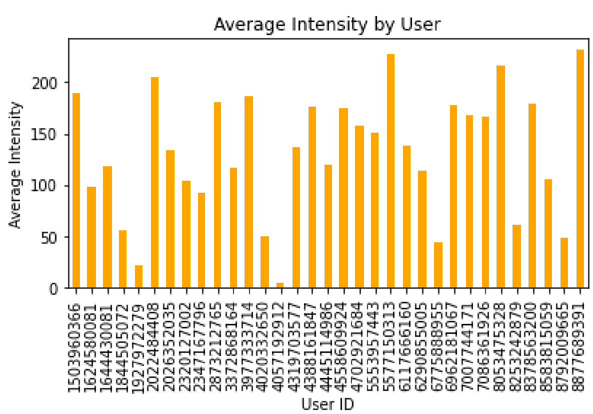
1. Below visualization has shown the average calorie expenditure per minute in an hour. We can clearly see that the calorie burn fluctuates throughout an hour. Users on an average burns more calories around their 36th minute of every hour in a day.



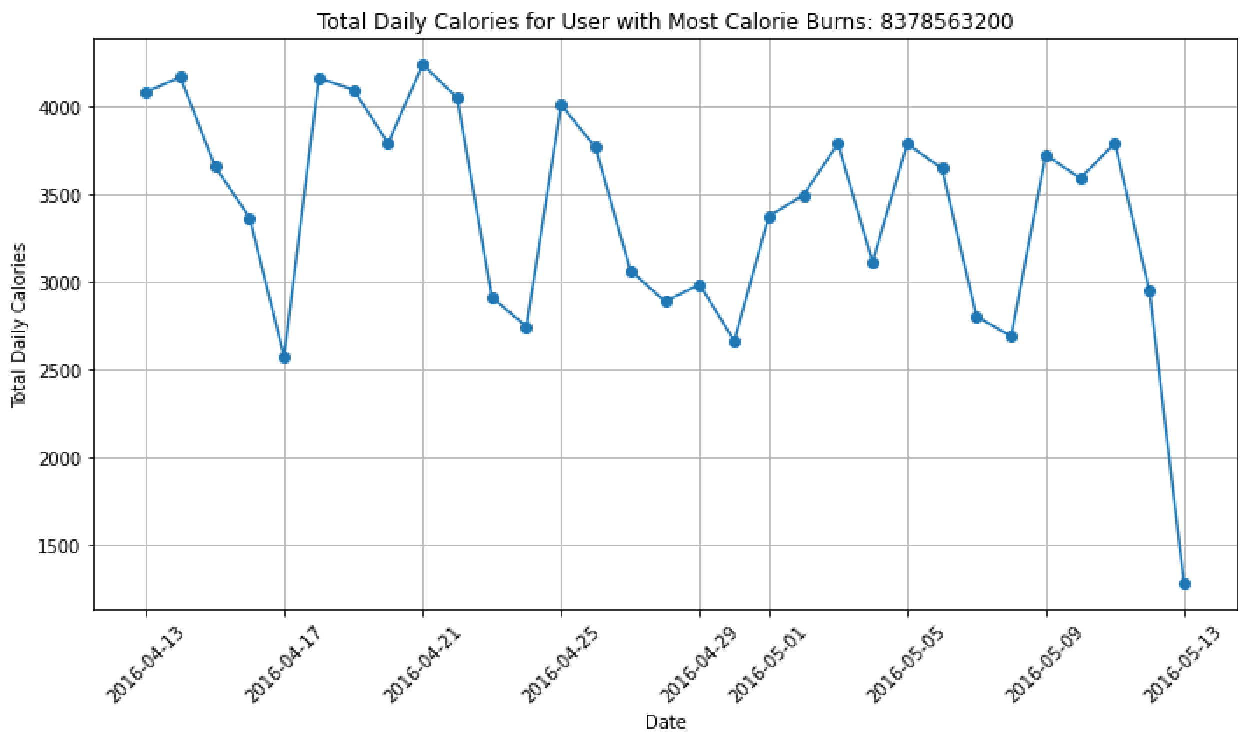
2. We have grouped the data of users to see how each user burn calories, takes how much steps and perform activities at what intensity for the whole time period of 13th April 2016 to 13th May 2016.

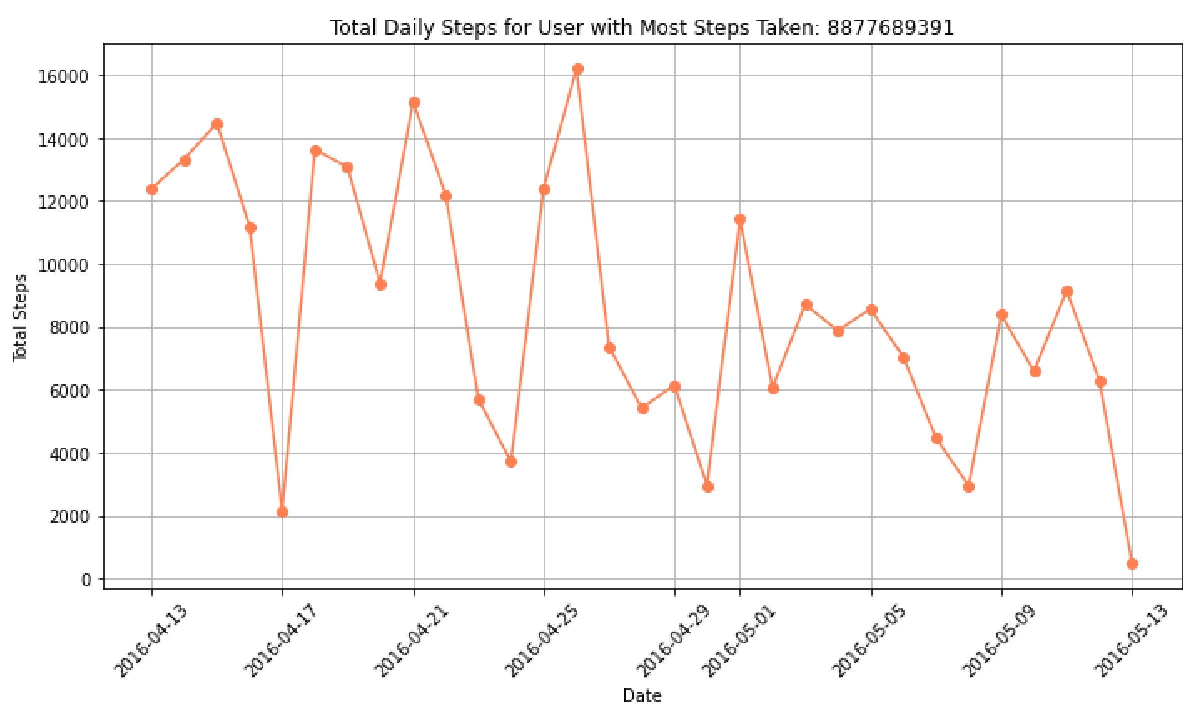


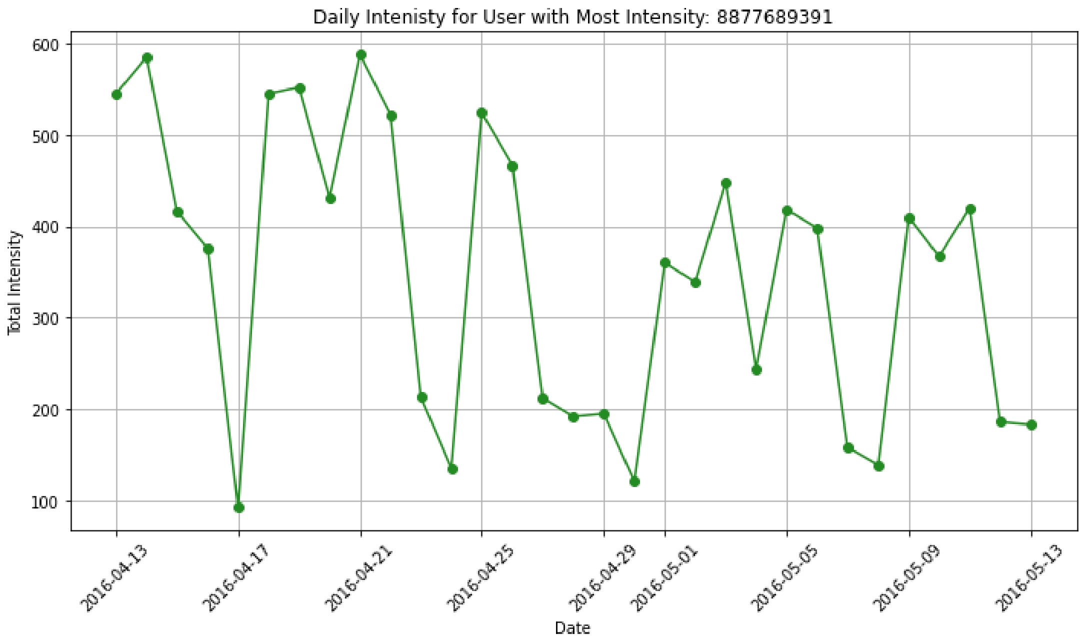




3. Amongst all the users, user with ID 8378563200 burn the most calories in total for the whole time period of 30 days. But the calorie expenditure decrease for this user after 10th may 2016. User with ID 8877689391 have the most step count and average intensity among all the other users (see 2nd and 3rd figure).







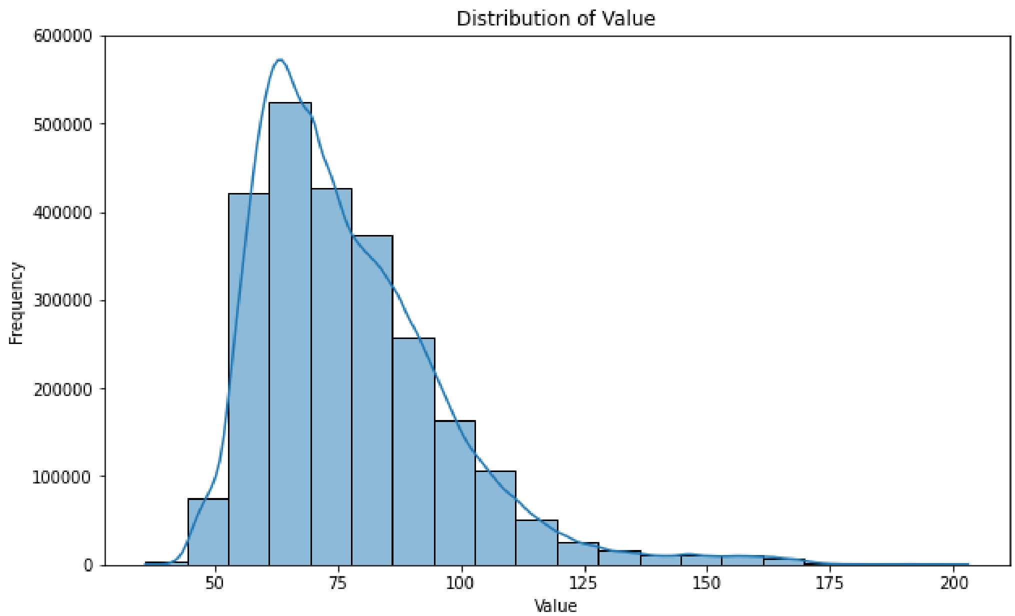
DAILY HEART RATE DATASET

OVERVIEW OF THE DATASET

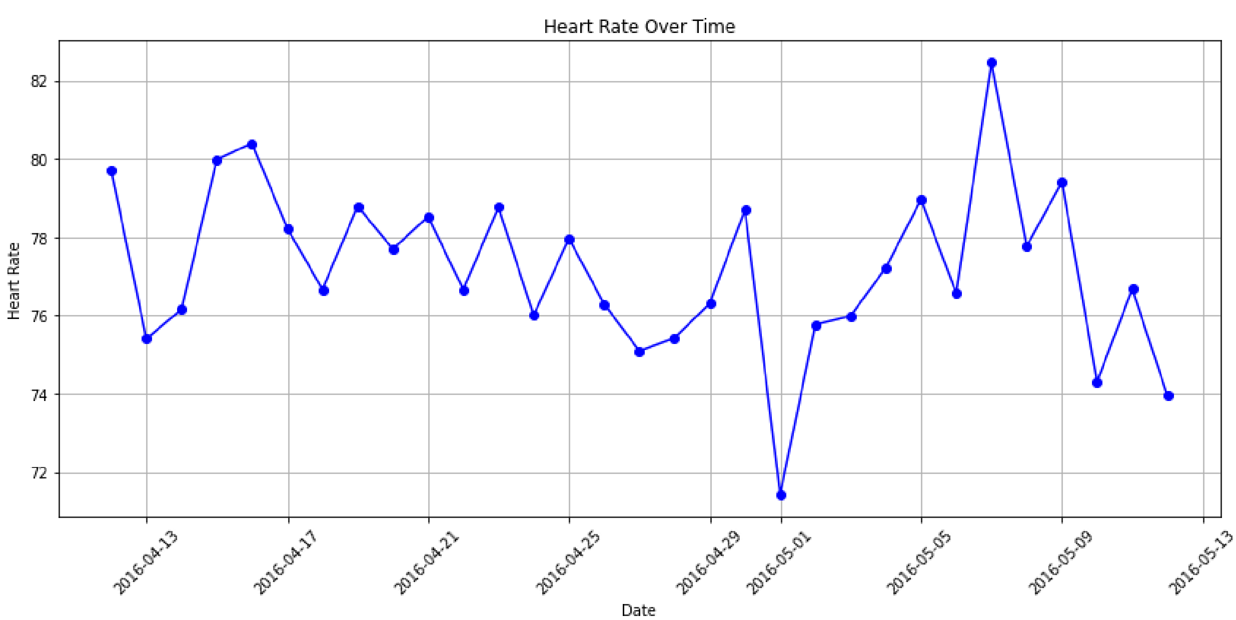
	Id	Time	Value
0	2022484408	4/12/2016 7:21:00 AM	97
1	2022484408	4/12/2016 7:21:05 AM	102
2	2022484408	4/12/2016 7:21:10 AM	105
3	2022484408	4/12/2016 7:21:20 AM	103
4	2022484408	4/12/2016 7:21:25 AM	101

INSIGHTS

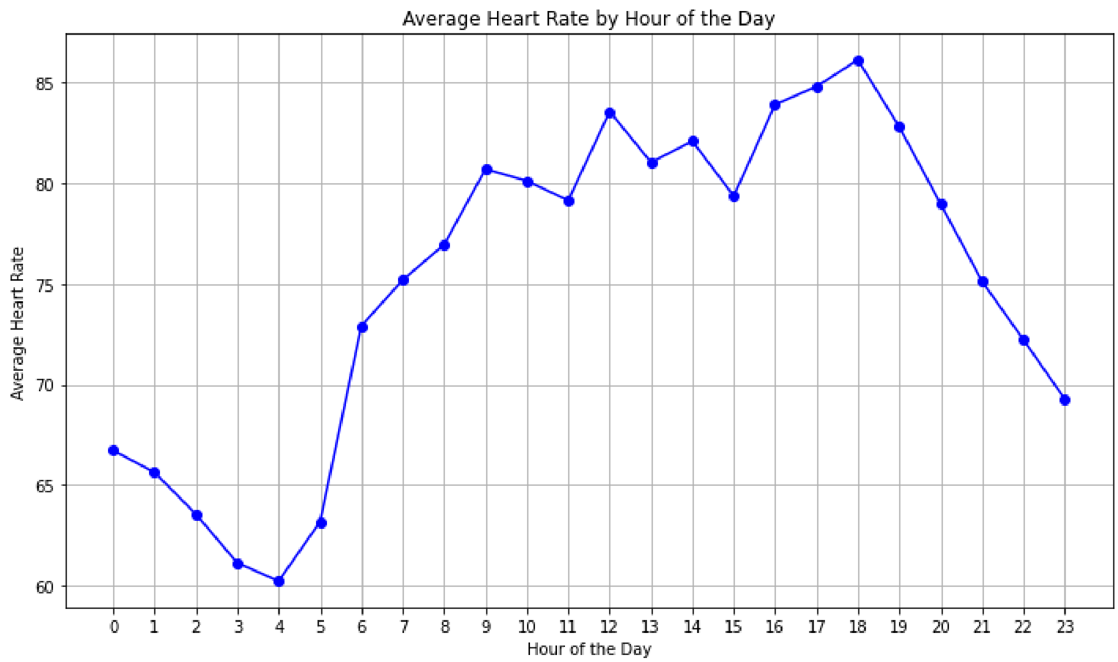
1. Most of the values in the dataset are comes in the range of 50 to 100. It means most of the users have average heart of 50 – 100 bpm.



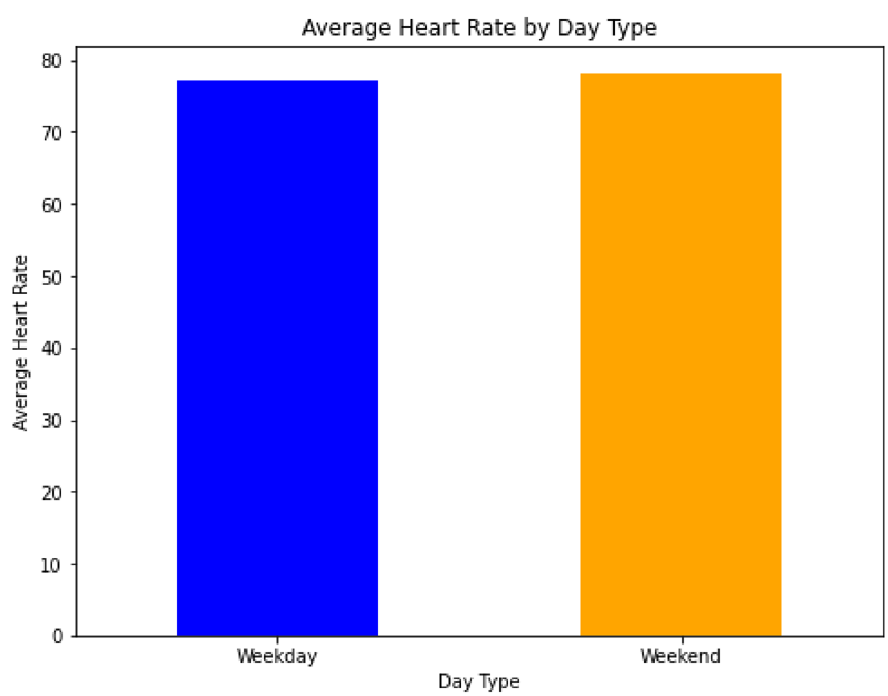
2. If we see the long-term trend of heart rate over the period of 30 days, we can see that there are sharp fluctuations in the trend lines which shows that heart rate changes from day to day depending on the activities users performed on that particular day.



3. Observing the hourly heart rate of users shows us that the heart rate starts increasing from morning 4 A.M as some users wakes up in early morning to do activities and daily chores. Heart rate continuously increasing till 7 P.M and after that it starts decreasing to normal.



4. Heart rate on weekends are surprisingly slightly more than that of weekdays. If we see the graph, we can clearly see that average heart rate on weekends are more than average heart rate on weekdays.



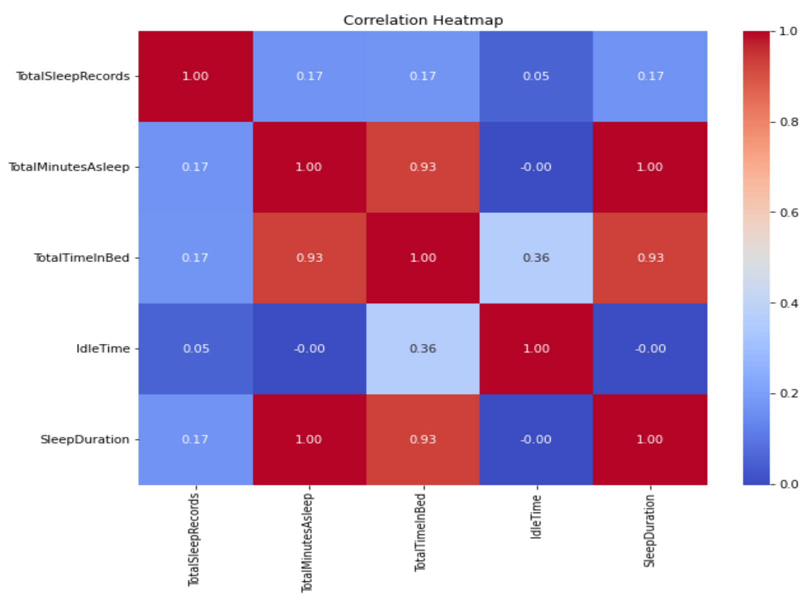
DAILY SLEEP DATASET

OVERVIEW OF THE DATASET

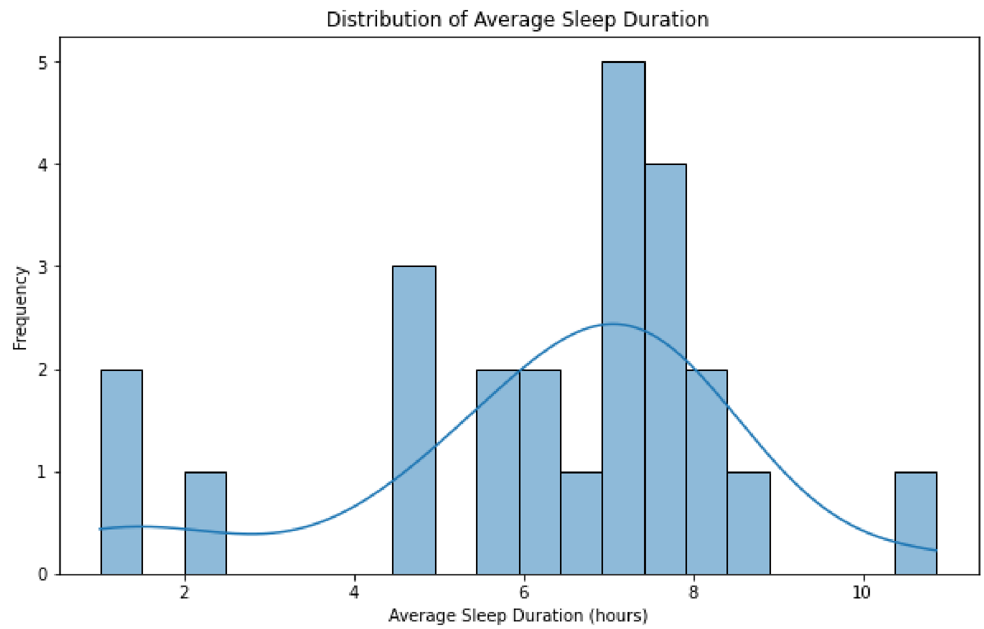
	Id	SleepDay	TotalSleepRecords	TotalMinutesAsleep	TotalTimeInBed
0	1503960366	4/12/2016 12:00:00 AM	1	327	346
1	1503960366	4/13/2016 12:00:00 AM	2	384	407
2	1503960366	4/15/2016 12:00:00 AM	1	412	442
3	1503960366	4/16/2016 12:00:00 AM	2	340	367
4	1503960366	4/17/2016 12:00:00 AM	1	700	712

INSIGHTS

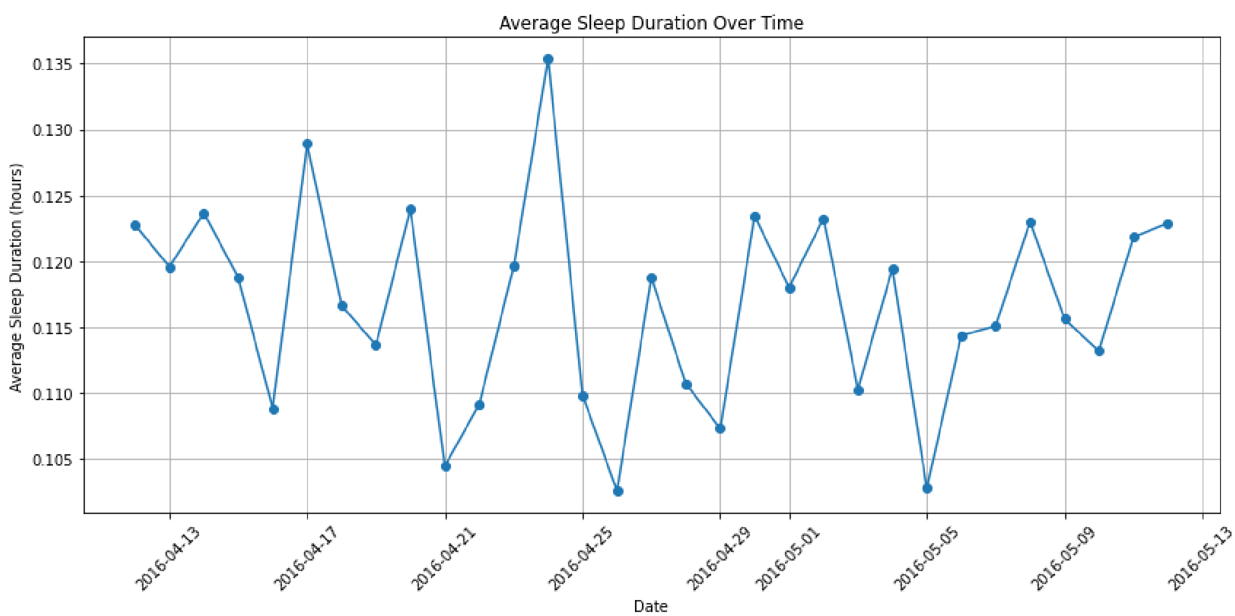
1. From the below heatmap, we can clearly see that the variables TotalMinutesAsleep & SleepDuration have a perfect correlation of 1.0, TotalMinutesAsleep & TotalTimeInBed have a correlation of 0.93 and SleepDuration & TotalTimeInBed have positive correlation of 0.93 as well.



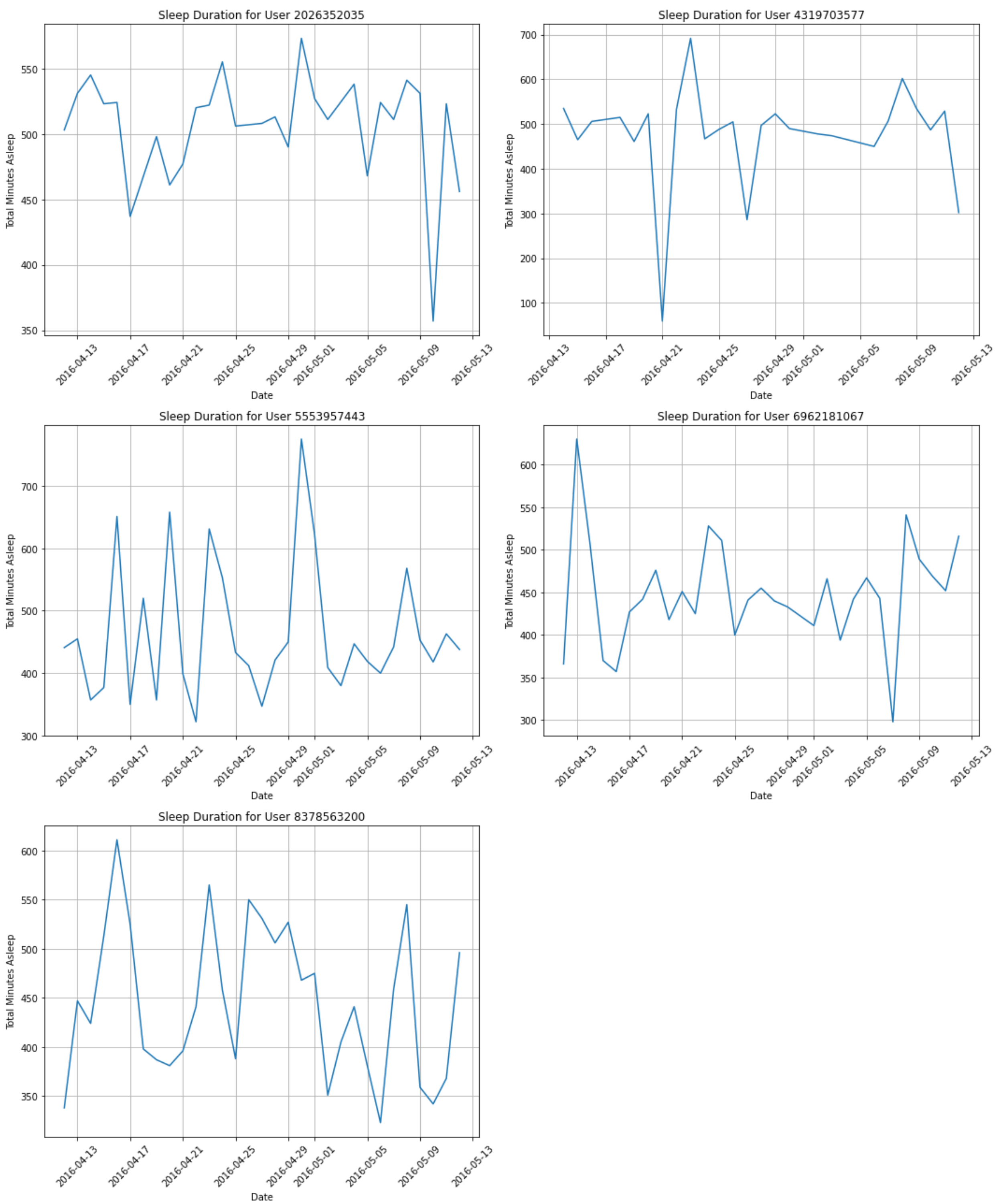
2. On an average, 50-60% of users sleep for 7-9 hours which is optimal time for the human body. From the graph we can clearly see that majority of the users sleep for 7-9 hours daily.



3. If we visualize the long-term trend of average sleep a user getting over a period of time, we can clearly observe that there are sharp fluctuations in the trend line which shows not every user gets same amount of sleep every day.



4. Top 5 Users with most hours of sleep in Total for a time period of 30 days.



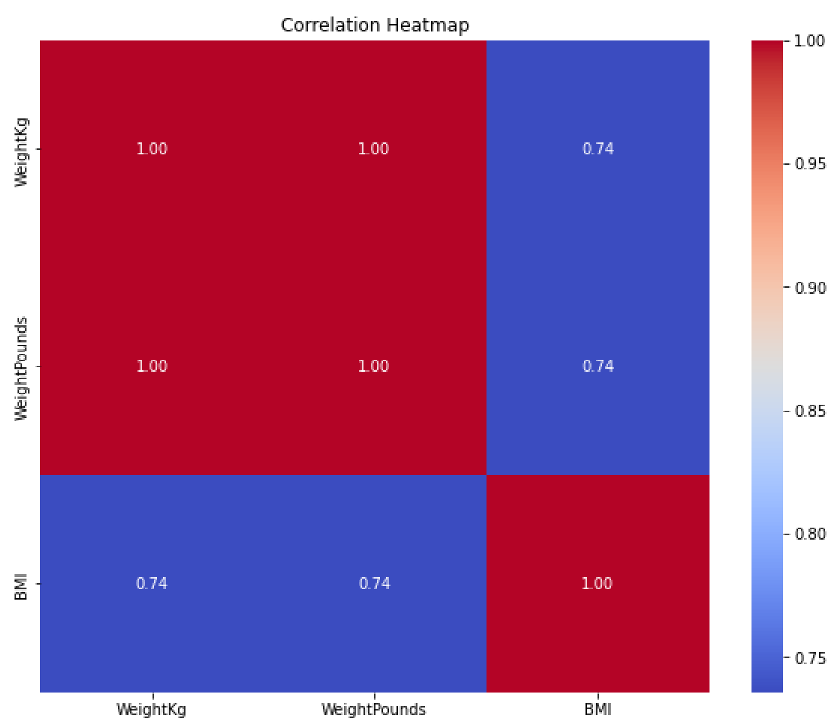
DAILY WEIGHT LOG DATA

OVERVIEW OF THE DATASET

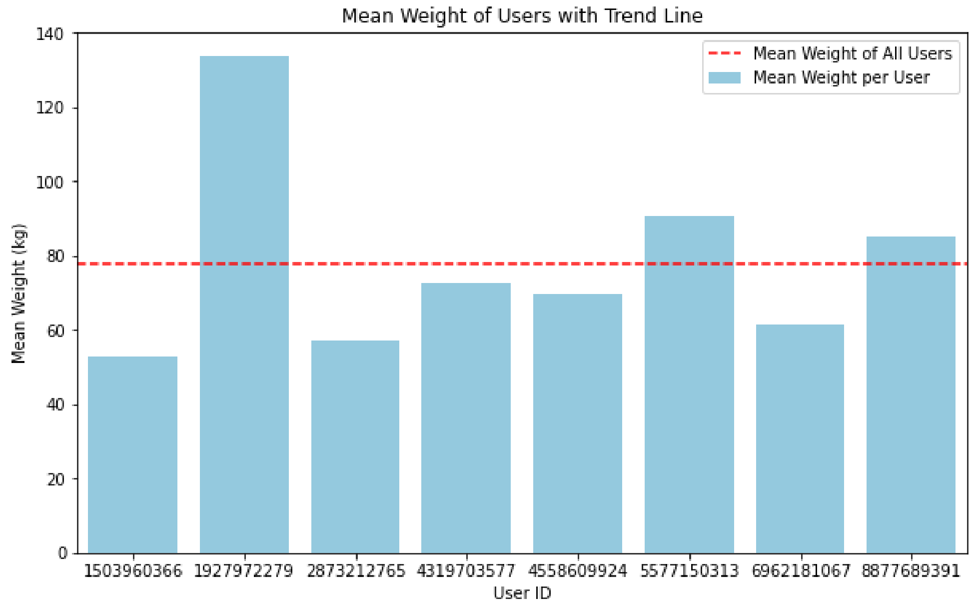
	Id	Date	WeightKg	WeightPounds	Fat	BMI	IsManualReport	LogId
0	1503960366	5/2/2016 11:59:59 PM	52.599998	115.963147	22.0	22.650000	True	1462233599000
1	1503960366	5/3/2016 11:59:59 PM	52.599998	115.963147	NaN	22.650000	True	1462319999000
2	1927972279	4/13/2016 1:08:52 AM	133.500000	294.317120	NaN	47.540001	False	1460509732000
3	2873212765	4/21/2016 11:59:59 PM	56.700001	125.002104	NaN	21.450001	True	1461283199000
4	2873212765	5/12/2016 11:59:59 PM	57.299999	126.324875	NaN	21.690001	True	1463097599000

INSIGHTS

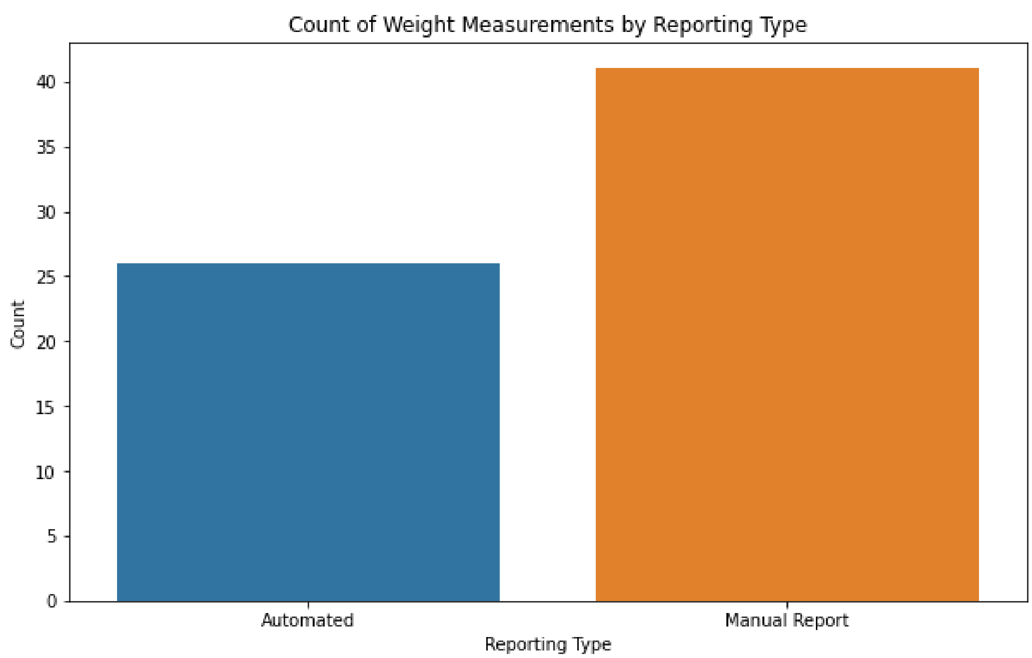
1. Below heatmap shows the positive correlation of 0.74 between BMI and WeightinKG variables. There is also a positive correlation of 0.74 between BMI and WeightInPounds variables. This means these variables are highly correlated to each other.



2. If we compare the average weight of each user with the average weight of all the users, we can observe that very few users have weight more than the average of all the users. Most of the users have weight less than the mean weight of all the users.



3. From the analysis we have found out that most of the users have manually reported their weight daily for the period of 30 days and less users opted for automated weight log.



RECOMMENDATIONS

Here are some recommendations for the marketing team based on the analysis of FitBit Fitness Tracker App data:

1. **Targeted Marketing Campaigns**:
    - Identify user segments with high engagement and tailor marketing campaigns to target these segments specifically.
    - Use personalized messaging to appeal to different user demographics, such as age groups, fitness goals, and activity levels.
2. **Promotion of Key Features**:
    - Highlight popular features of the FitBit app, such as step tracking, sleep monitoring, heart rate analysis, and fitness challenges, in marketing materials.
    - Educate users on how these features can help them achieve their fitness goals and improve overall health and well-being.
3. **User Engagement Strategies**:
    - Implement strategies to increase user engagement with the app, such as gamification, rewards programs, and social sharing features.
    - Encourage users to set goals, track progress, and participate in challenges to stay motivated and committed to their fitness journey.
4. **Health and Wellness Content**:
    - Create and promote content related to health, fitness, and wellness that aligns with the interests and preferences of FitBit app users.
    - Provide valuable tips, insights, and resources to help users lead healthier lifestyles and achieve their fitness objectives.
5. **Partnerships and Collaborations**:
    - Explore partnerships with fitness influencers, health professionals, and wellness brands to expand brand reach and credibility.
    - Collaborate on co-branded campaigns, sponsored content, and exclusive offers to attract new users and retain existing ones.
6. **Data-Driven Insights**:
    - Continuously analyze user data to identify emerging trends, user preferences, and areas for improvement.

- Use data-driven insights to refine marketing strategies, optimize user experiences, and innovate new features that resonate with users.

7. **Customer Feedback and Engagement**:
   - Encourage user feedback through surveys, reviews, and customer support channels to understand user needs and preferences better.
   - Actively engage with users on social media platforms, forums, and community groups to foster a sense of belonging and loyalty among FitBit app users.

CONCLUSION

In conclusion, the analysis of the FitBit Fitness Tracker App data revealed several key insights that can inform marketing strategies and improve customer engagement.

1. **Trends Identified**: We identified various trends in user behavior, including peak activity hours, sleep patterns, and weight fluctuations. Understanding these trends can help tailor marketing campaigns to target specific user segments more effectively.
2. **Customer Application**: These trends can be applied to enhance the customer experience by providing personalized recommendations, targeted promotions, and proactive support. By leveraging insights from the data, the FitBit app can become a more indispensable tool for users in achieving their fitness and wellness goals.
3. **Marketing Strategy**: The insights gained from the analysis can guide marketing strategies to focus on key features and functionalities that resonate most with users. For example, highlighting the benefits of tracking sleep quality or promoting social features that encourage competition and accountability among users can be effective strategies.
4. **Future Opportunities**: As the fitness tracking industry continues to evolve, there are opportunities to further leverage data analytics to innovate and differentiate the FitBit app. Exploring new features, partnerships, and integrations based on user preferences and behaviors can drive continued growth and engagement.