

# MEHMET YILDIRGAN

## CS 210 - Introduction to Data Science Term Project

### Abstract

In this project, personal health data is analyzed using data analysis techniques.

### Introduction

In this project, I analyze more than 220,000 lines of my own health data recorded by Mi Band 4 from August 2020 to May 2023. The dataset includes detailed measurements such as daily step count (day-to-day and minute-to-minute), sleep duration, and deep and shallow sleep time. This report contains the results of detailed analysis on the dataset. The dataset and associated python notebook are available from the link: [https://github.com/myildirgann/CS\\_210.git](https://github.com/myildirgann/CS_210.git)

### The Data Source

Mi Fit Health Tracker App & Mi Band 4

### Data Analysis Techniques

EDA (Exploratory Data Analysis) technique is used to explore the structure of the dataset. Data preprocessing techniques such as data imputation, are applied to ensure the quality and completeness of the dataset. Correlation analysis is employed to examine the relationships between activity and sleep parameters. Finally, visualization techniques, such as bar charts with gradient colors, are used to visually represent and interpret complex data parameters.

### Motivation & Aim

The aim of this project is to analyze personal health data and make inferences about the results using data analysis techniques such as Exploratory Data Analysis (EDA). The motivation for working on personal health data is to learn more about my sleep and activity patterns during the day. Because the dataset provides extensive data covering more than 900 days, it is easy to make robust and unbiased inferences about the results.

### Content Of Data

Activity:

1. day by day step count,
2. minute to minute activity,
3. total distance and running distance,
4. calories burned during activities

Sleep:

1. day by day sleep time,
2. deep sleep time,
3. shallow sleep time

## PART 1: HYPOTHESIS TESTING

### **Hypothesis 1:** The More Active Day Leads to the More Total Sleep Duration

- Null Hypothesis (H0): There is a significant correlation between the number of steps taken in a day and the total sleep duration.
- Alternative Hypothesis (H1): There is not a positive correlation between the number of steps taken in a day and the total sleep duration.

### **Hypothesis 2:** Active Days Result in More Calory Burning

- Null Hypothesis (H0): The level of activity (measured by steps) is associated with the calories burned during the activities.
- Alternative Hypothesis (H1): More active days (higher step count) are not associated with more calory burn.

### **Hypothesis 3:** Total Sleep Duration Affects the Sleep Quality

- Null Hypothesis (H0): There is a significant relationship between the total sleep duration and sleep quality (measured by deep sleep duration).
- Alternative Hypothesis (H1): Better sleep quality is not associated with a total sleep time.

## 1.1: CORRELATION HEATMAP

To reach a comprehensive intuition about the parameters, the correlation heatmap is used below (Figure 1).

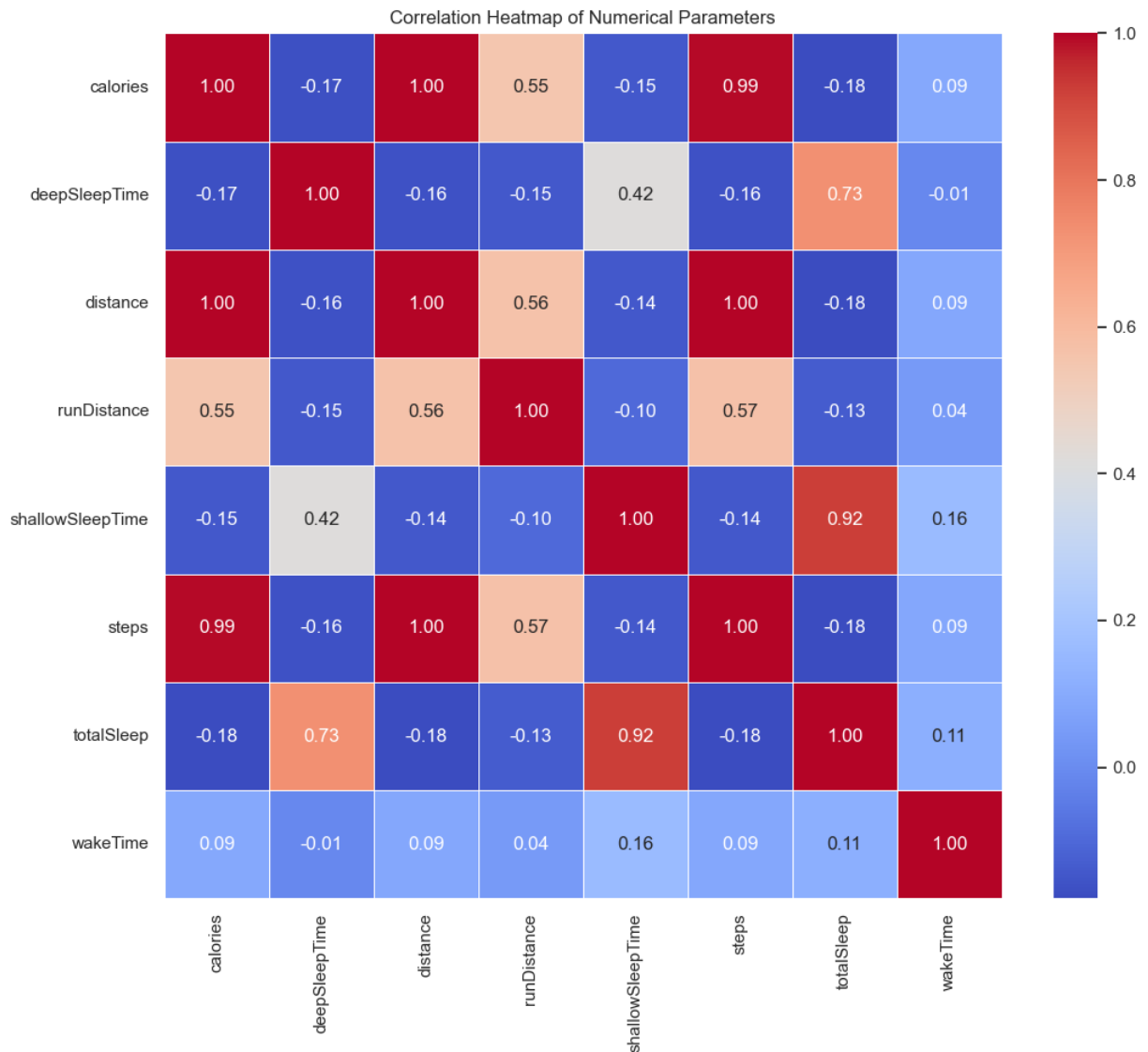


Figure 1: Correlation Heatmap of Numerical Parameters

From the correlation heatmap in Figure 1, we draw the following conclusion(s):

1. There is a strong correlation between "distance" and "calories".
  2. There is a strong correlation between "steps" and "distance".
  3. There is a relatively strong correlation between "total sleep time" and "shallow sleep time".
- Additionally, "deep sleep duration" is also correlated with "total sleep duration," although it is not as strongly correlated as "shallow sleep time".

## 1.2: SCATTER PLOTS

### 1.2.1 Hypothesis 1: The More Active Day Leads to the More Total Duration

- Null Hypothesis (H0): There is a significant correlation between the number of steps taken in a day and the total sleep duration.
- Alternative Hypothesis (H1): There is not a positive correlation between the number of steps taken in a day and the total sleep duration.

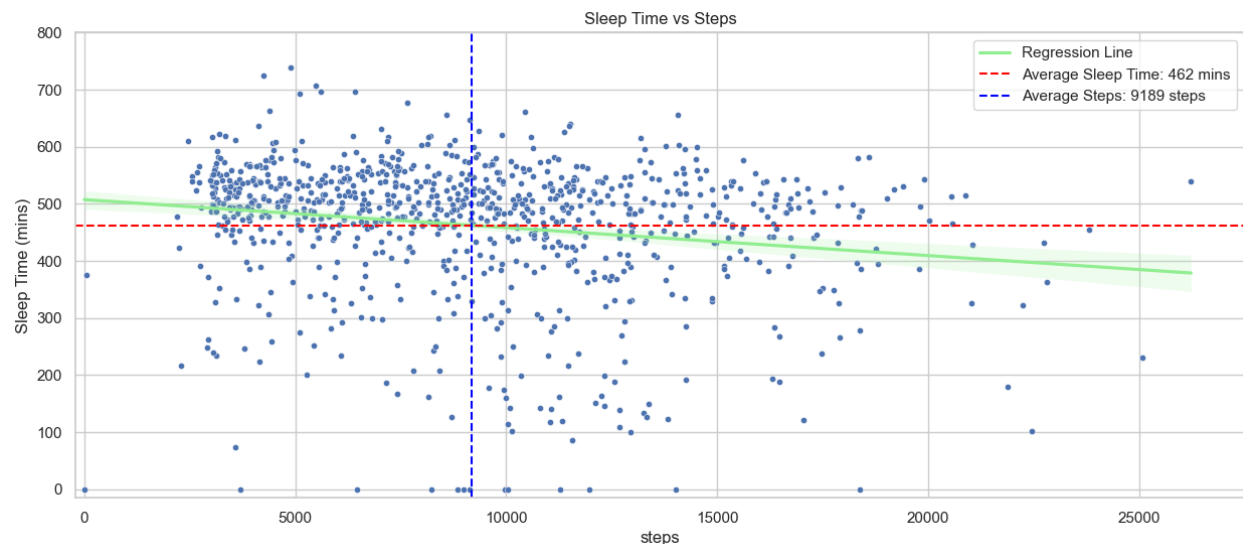


Figure 2: Scatter Plot showing the relationship between 'Steps' and 'Sleep Time'

Correlation Coefficient between 'Steps' and 'Sleep Time': -0.17706

From the scatter plot in Figure 2, we draw the following conclusion(s):

1. There is not a strong correlation between sleep time and steps taken during the day. In this case, we need to reject Null Hypothesis (H0).

#### Note:

Since my initial hypothesis and motivation to start this project is to investigate the relationship between the sleep and activity patterns, I put the scatter plot of "sleep" vs "steps", here. One remark is, since my hypothesis was "the more active I am in day, I sleep more", I used the following day's sleep data to compare with the activity of each day. However, the results show that there is not a strong correlation between sleep time and physical activity, directly. (The reason that drives me to produce the hypothesis is also explained in Part 2.2.2, page 14)

### 1.2.2 Hypothesis 2: Active Days Result in More Calory Burning

- Null Hypothesis (H0): The level of activity (measured by steps) is associated with the calories burned during the activities.
- Alternative Hypothesis (H1): More active days (higher step count) are not associated with more calory burn.

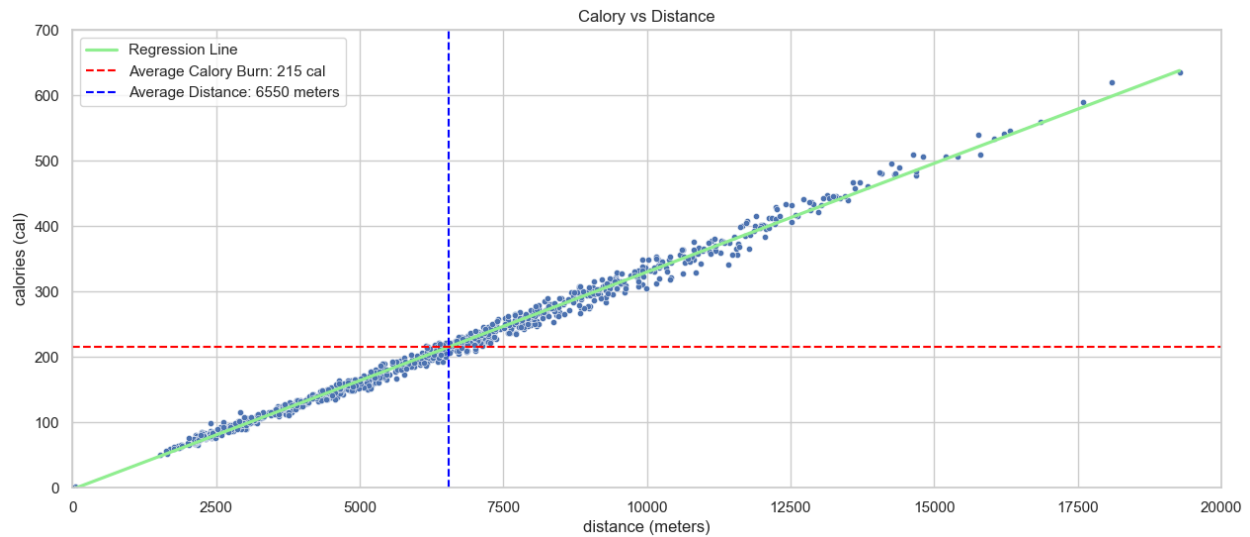


Figure 3: Scatter Plot showing the relationship between 'Calory' and 'Distance'

Correlation Coefficient between 'Calory' burned during activities and 'Distance' taken: 0.99692

From the scatter plot in Figure 3, we draw the following conclusion(s):

1. There is a strong correlation between "calories" and "distance".
2. Correlation coefficient between "calories" and "distance" is 0.997.

### 1.2.3 Hypothesis 3: Total Sleep Duration Affects the Sleep Quality

- Null Hypothesis (H0): There is a significant relationship between the total sleep duration and sleep quality (measured by deep sleep duration).
- Alternative Hypothesis (H1): Better sleep quality is not associated with a total sleep time.

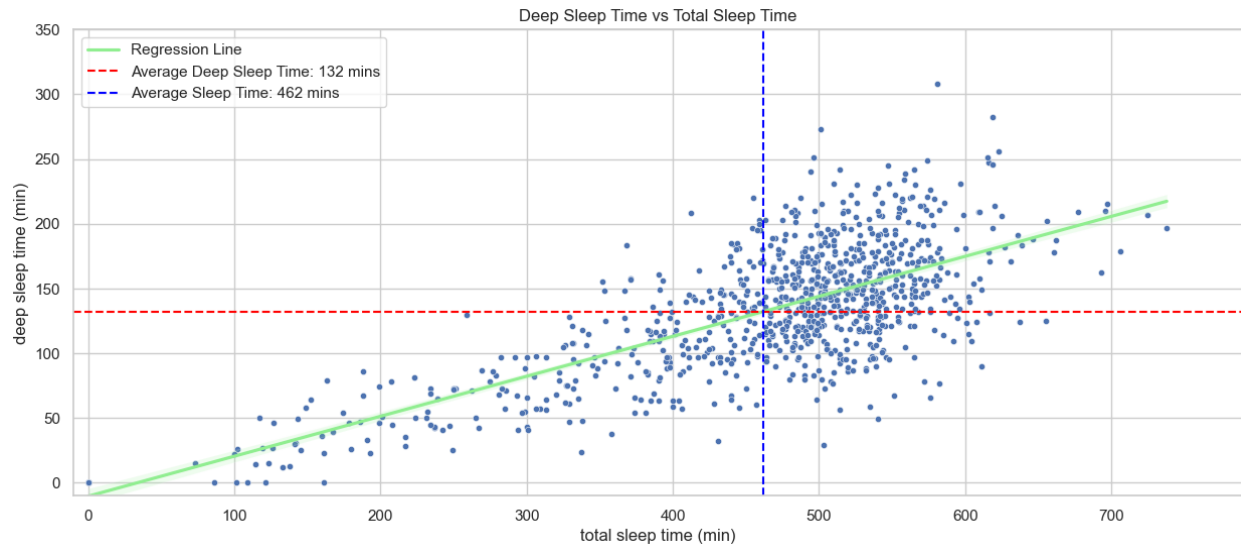


Figure 4: Scatter Plot showing the relationship between 'Deep Sleep Time' and Total 'Sleep Time'

Correlation Coefficient between Deep Sleep Time and Sleep Time: 0.73242

From the scatter plot in Figure 4, we draw the following conclusion(s):

1. There is a relatively strong correlation between "deep sleep time" and "total sleep time". In this case, we can carefully accept the Null Hypothesis (H0.)
2. Correlation coefficient between "deep sleep time" and "total sleep time" is 0.733.

## PART 2: EDA (Exploratory Data Analysis)

### 2.1: ACTIVITY DATA ANALYSIS

#### 2.1.1 Average Daily Steps Across the Months (2020 Aug-2023 May)

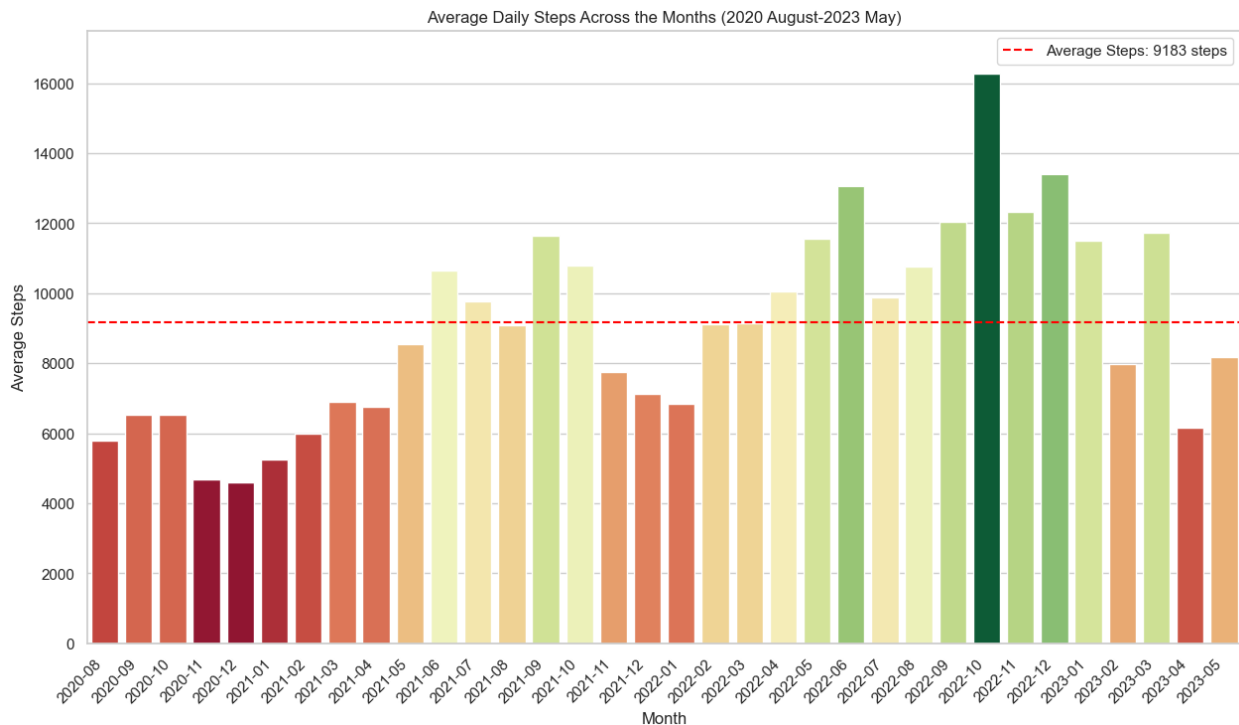


Figure 5: Average Daily Steps Across the Months (2020 August-2023 May)

From the bar chart in Figure 5, we draw the following conclusion(s):

1. Compared with post-pandemic term, i had taken less steps during pandemic.
2. Following 2020 November, there was a noticeable decrease in physical activity.
3. When controlled normalization began at 2021 March, my step count also increased.
4. After removing the Covid-19 precautions at 2021 June, there was an increase in step counts.
5. When face-to-face education started at 2021 September, September becomes the month I stepped the most in 2021.
6. Compared with 2021, I was more active in 2022.
7. 2022 was the month I stepped the most on average between 2020 August and 2023 May.

\*Resource: <https://www.aa.com.tr/tr/koronavirus/turkiyenin-kovid-19-salginiyla-mucadelesinin-2-yili/2529977>

## 2.1.2 Average Daily Steps Across the Months (All Data)

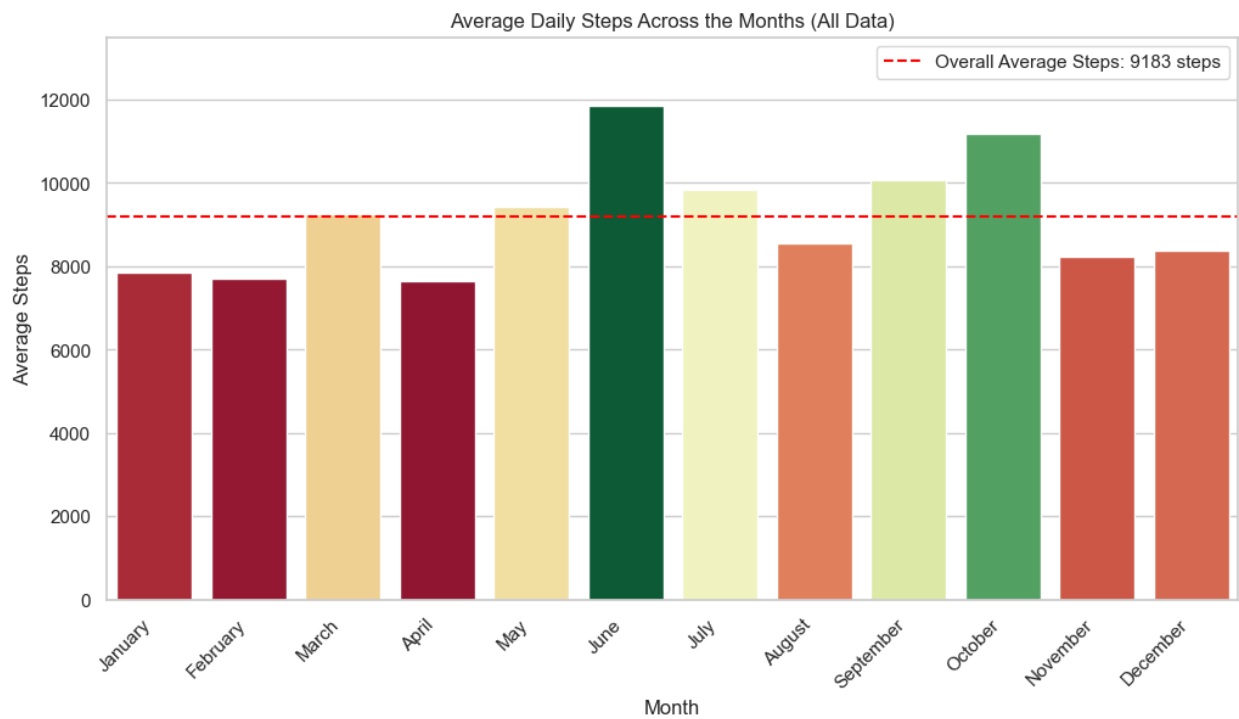


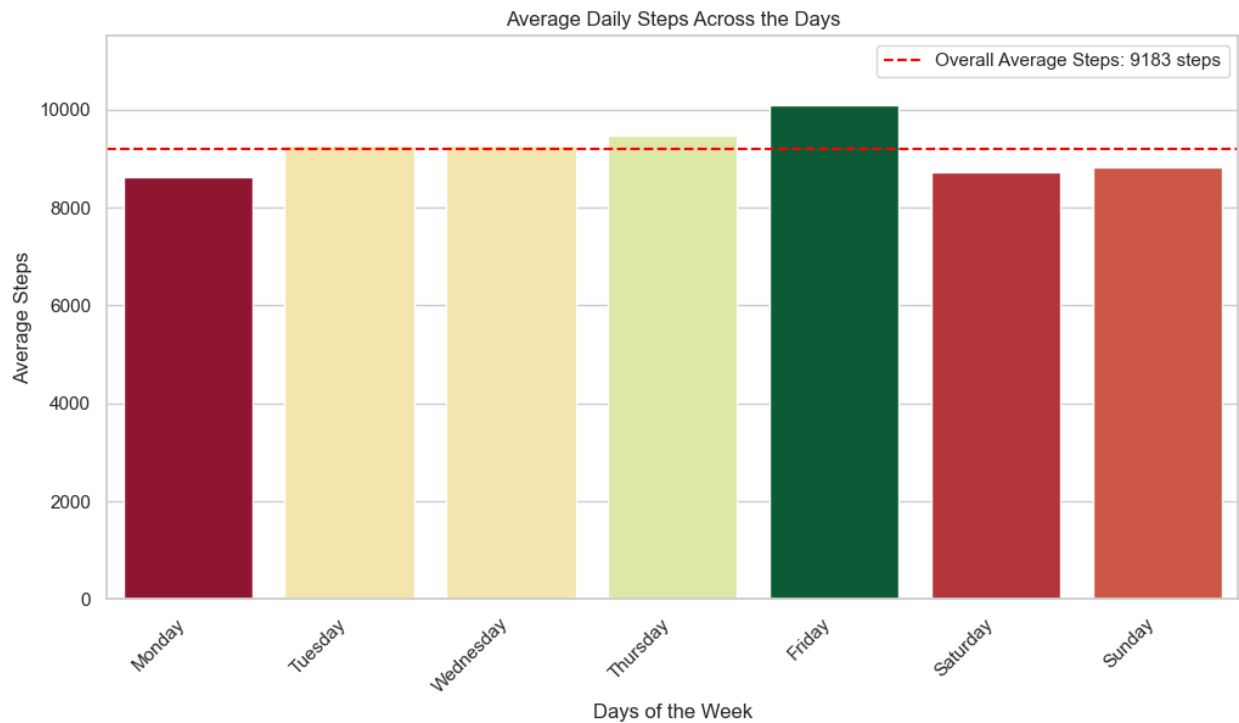
Figure 6: Average Daily Steps Across the Months (All Data)

From the bar chart in Figure 6, we draw the following conclusion(s):

1. On average, I take more than 9000 steps in a day.
2. I am most active in June, which is beginning of the summer.
3. 2nd most active month I am is October, which is coincident with schools open.
4. I am more active in summer than winter.



### 2.1.3 Average Daily Steps (Days of the Week)



*Figure 7: Average Daily Steps Across the Days of the Week*

*From the bar chart in Figure 7, we draw the following conclusion(s):*

1. There is a gradual increase in step count in weekdays.
2. In weekends, I am generally less active.
3. Friday is the day that I am physically most active.

## 2.1.4 Average Steps Across the Hours

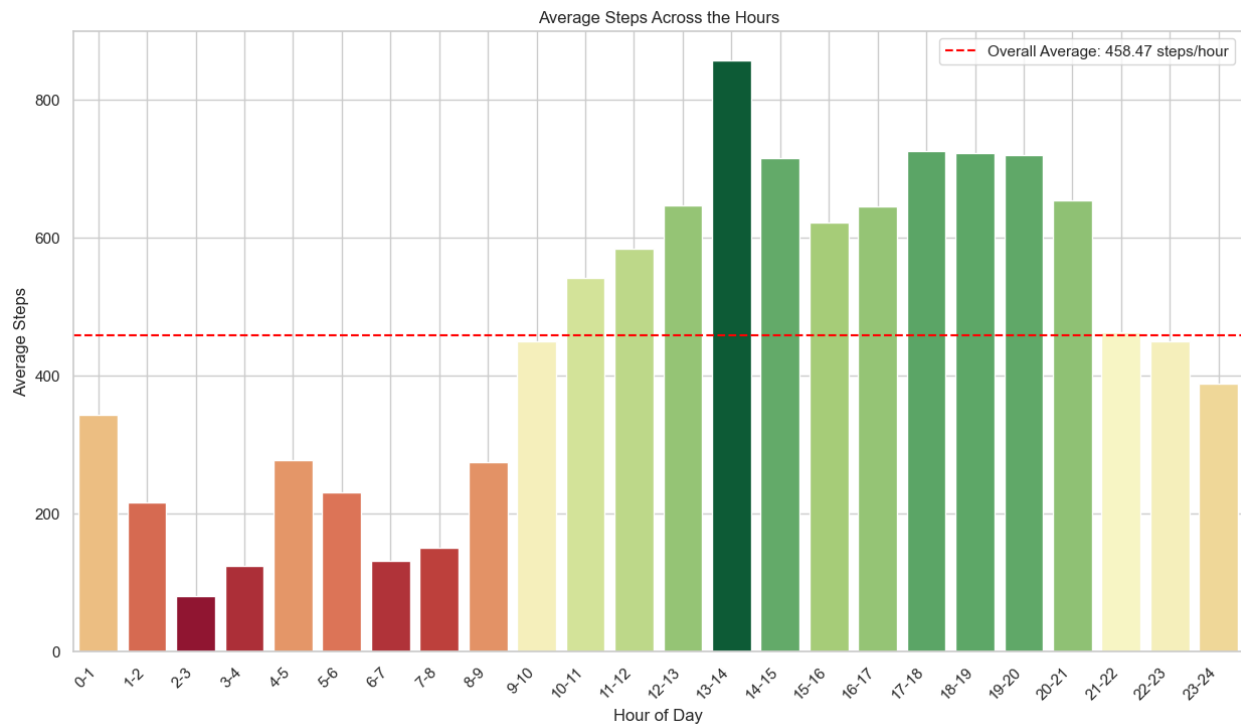
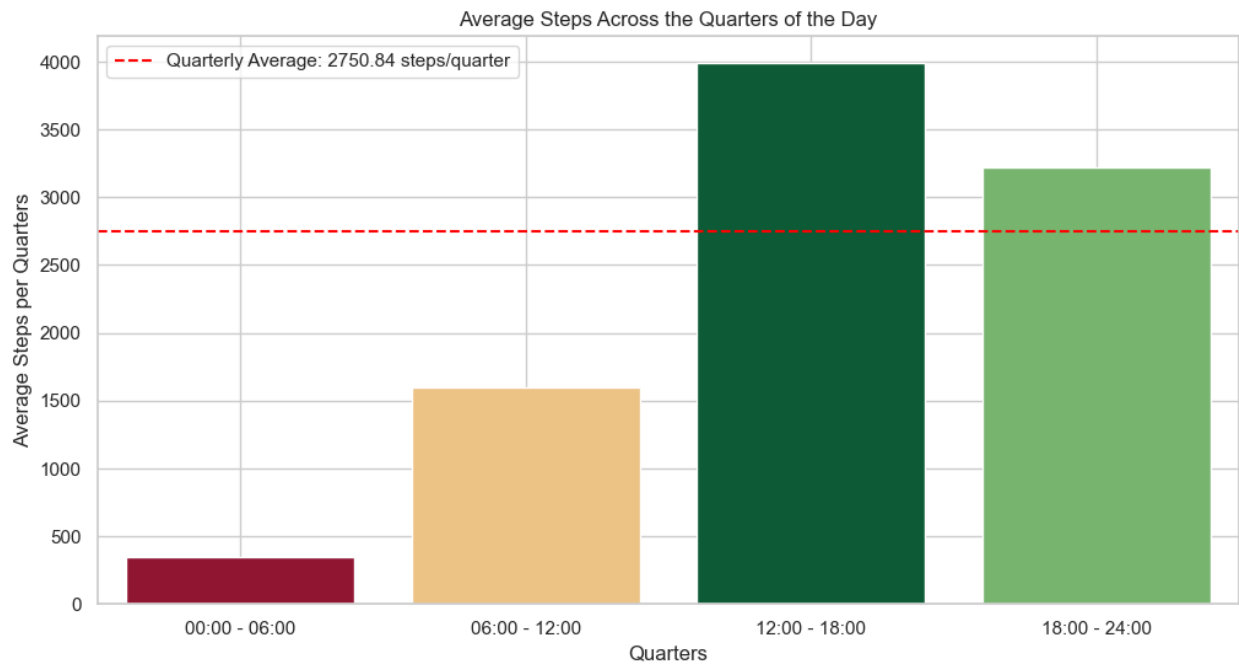


Figure 8: Average Steps Across the Hours

From the bar chart in Figure 8, we draw the following conclusion(s):

1. I am more active in days than nights.
2. I step most between 1pm-2pm.
3. There is a gradual decrease in physical activity during the day starting at 7pm.

## 2.1.5 Average Steps Across the Quarters of the Day



*Figure 9: Average Steps Across the Quarters of the Day*

*From the bar chart in Figure 9, we draw the following conclusion(s):*

1. 12:00 - 18:00 is the quarter that I am most active in a day.
2. 00:00 - 06:00 is the quarter that I am least active in a day.

## 2.1.6 Run Distance Ratio Across the Months (All Data)

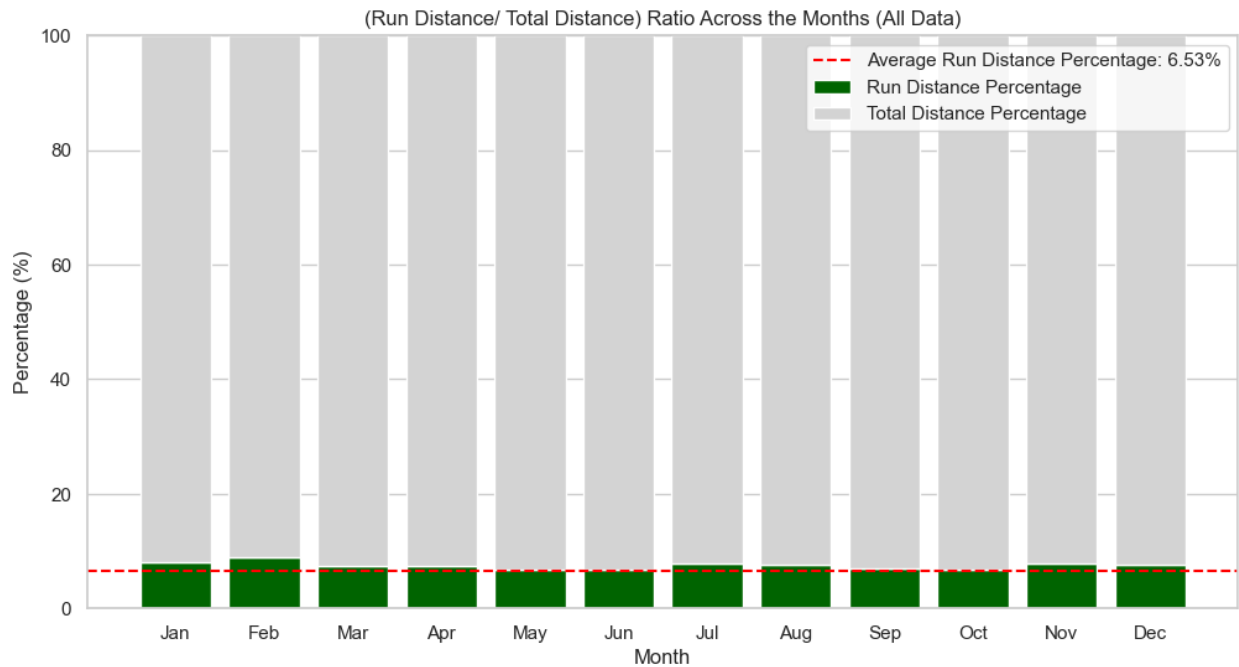


Figure 10: (Run Distance/ Total Distance) Ratio Across the Months (All Data)

From the bar chart in Figure 10, we draw the following conclusion(s):

1. In general I take 6.53% of my steps with running.
2. This shows that, most of the time (94%), I step on an average velocity. This result can also be further used to interpret my personal characteristics, as well :)
3. In Februarys, the average running/walking ratio increases.

## 2.2: SLEEP DATA ANALYSIS

### 2.2.1 Average Daily Sleep Time Across the Months (2020 Aug-2023 May)

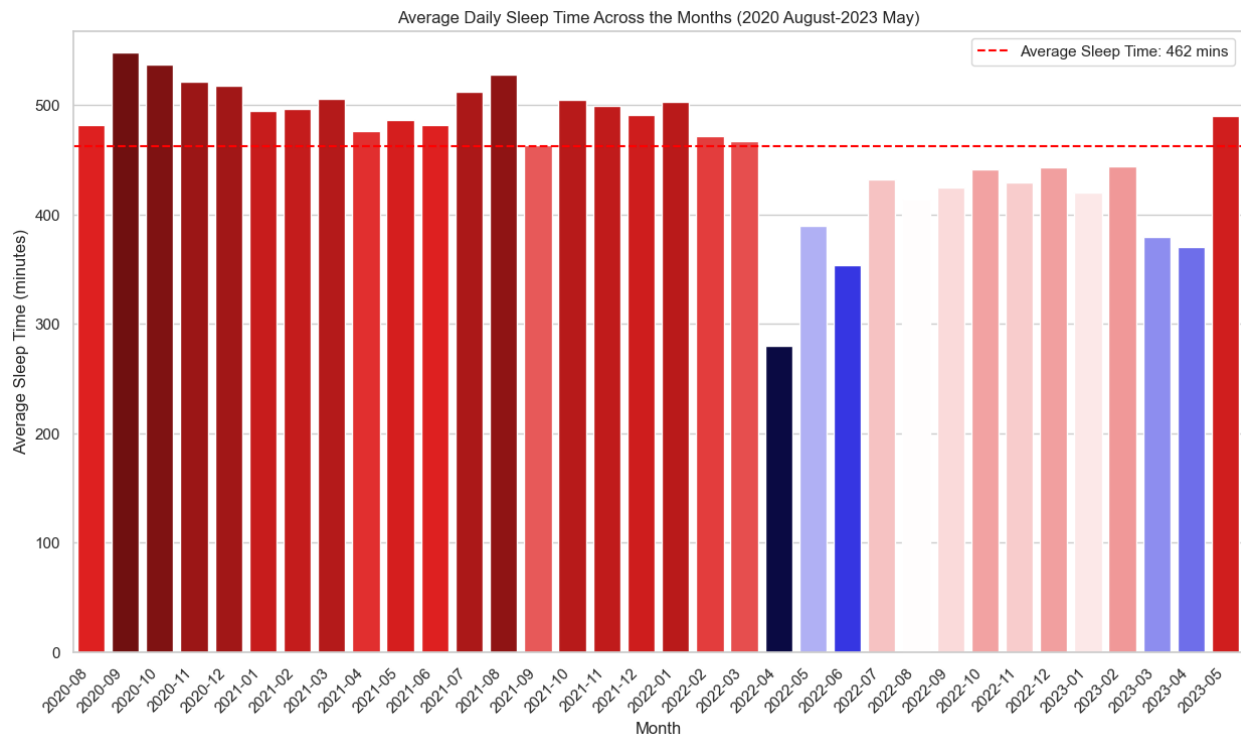


Figure 11: Average Daily Sleep Time Across the Months (2020 August-2023 May)

From the bar chart in Figure 11, we draw the following conclusion(s):

1. Between 2020 September and 2020 December my sleep time was comparably high. This term coincident with the quarantine precautions implied for the young people under 20s.
2. Following 2022 April, my average sleep time were under the overall average. This term is coincident with the rest of my academic career of the second term of the Freshman year. Hence, the reason for decrease in sleep time might be the academic challenges :) (Since the data covers only first few days of the 2023 May, increase in average sleep time on May cannot solely disprove the hypothesis :)

## 2.2.2 Average Daily Sleep Time Across the Months (All Data)

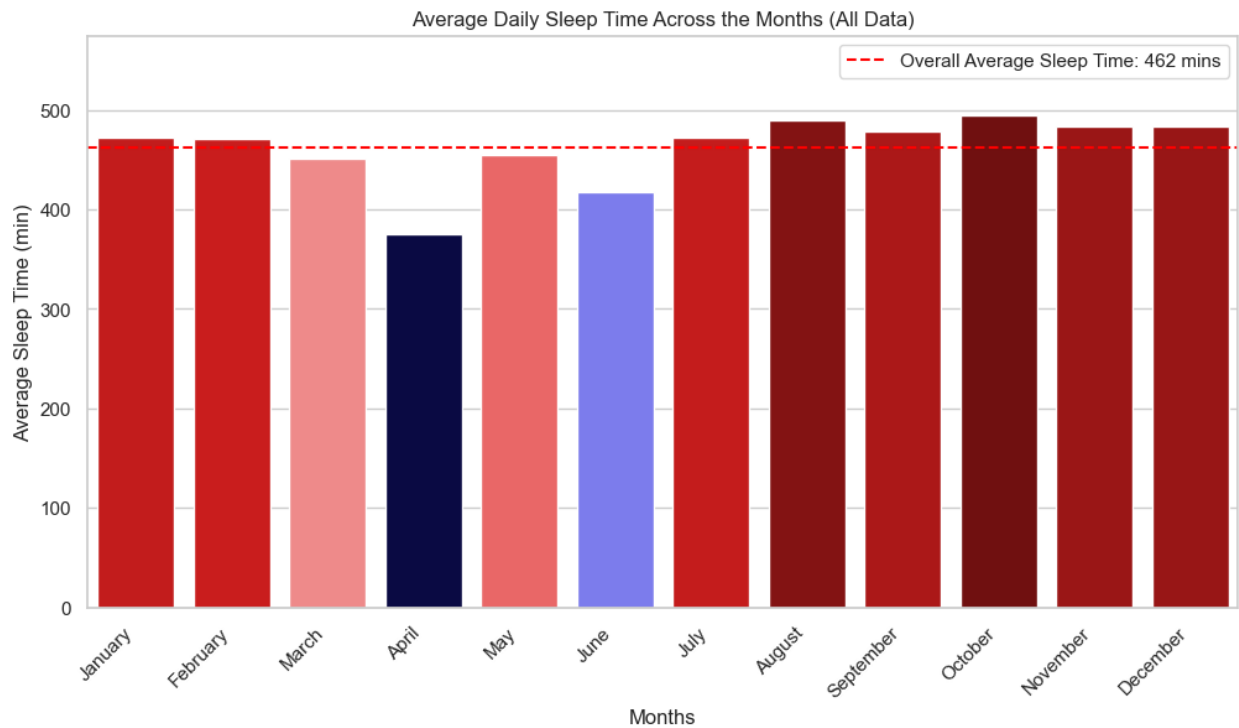


Figure 12: Average Daily Sleep Time Across the Months (All Data)

From the bar chart in Figure 12, we draw the following conclusion(s):

1. On average, I sleep 462 minutes (7 hours 42 mins) in a day.
2. I sleep most in Octobers, which is the 2nd month I am most active, in general.
3. In summer months I sleep less then winter months. This may be relevant with the difference between daytime in summers and winters.
4. I sleep least in Aprils, which is the month that I am least active, in general.
5. Article 2 and 4 led me to produce hypothesis, "the more active I am in day, I sleep more" (see Hypothesis 1, page 4). However, as can be seen from the correlation heatmap at the beginning, there is not a direct correlation between the total sleep time and physical activity.

### 2.2.3 Average Daily Sleep Time Across the Days of the Week

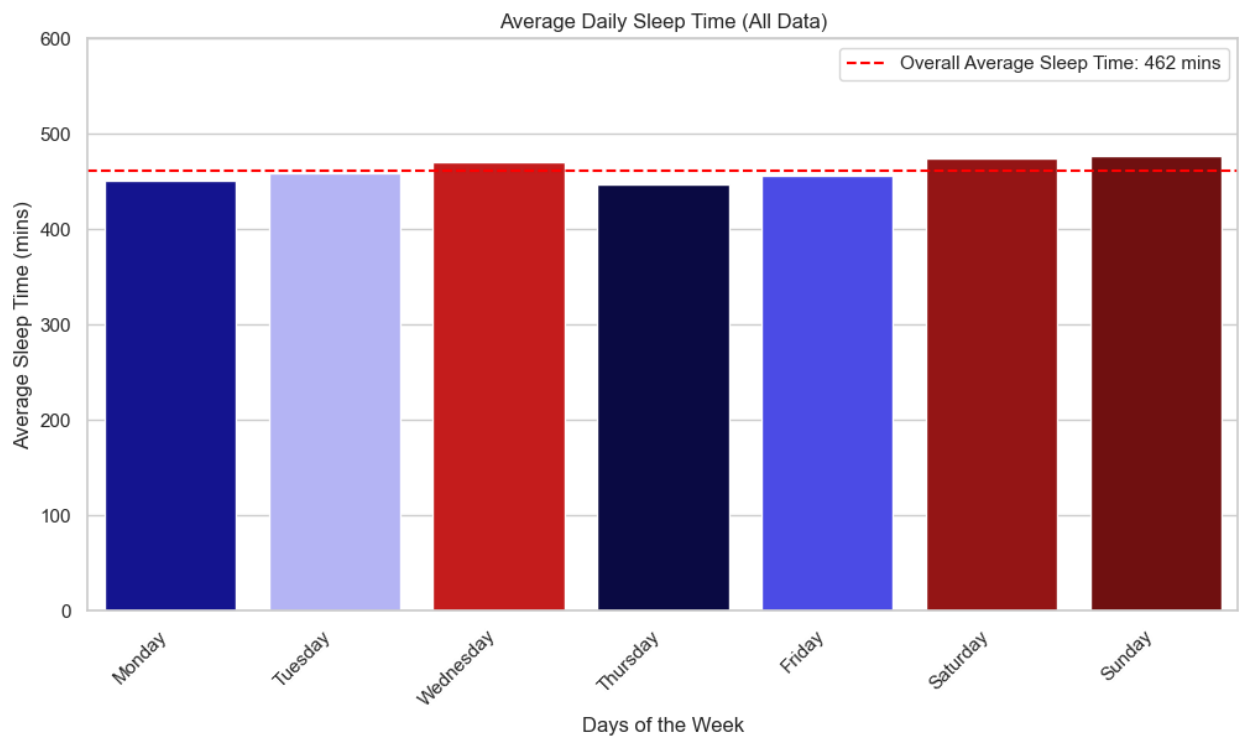


Figure 13: Average Daily Sleep Time Across the Days of the Week (All Data)

From the bar chart in Figure 13, we draw the following conclusion(s):

1. I sleep more on weekends than weekdays.
2. I sleep the most on Sunday.

## 2.2.4 Deep Sleep Time Ratio Across the Months (2020 Aug-2023 May)

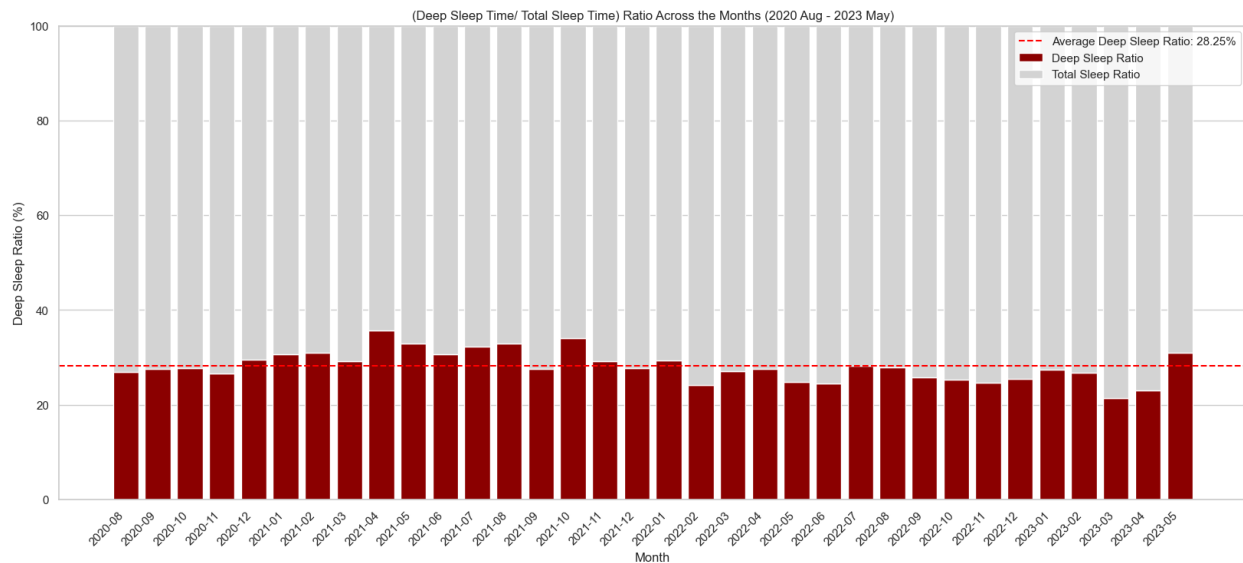


Figure 14: (Deep Sleep Time/ Total Sleep Time) Ratio Across the Months (2020 Aug - 2023 May)

From the bar chart in Figure 14, we draw the following conclusion(s):

1. On average, deep sleep time corresponds to the 28.25% of the total sleep time, which is about the one third.
2. Starting from 2022 February, deep sleep time ratio is below the average, in general.