SUMMARY/REPORT

Case Study : How Can a Wellness Technology Company Play It Smart?

**#About The Company**

“Bellabeat” , a high-tech manufacturer of health-focused products for women was founded in 2013 by Urška Sršen and Sando Mur.

Sršen developed and designed technology that informs and inspires women around the world by collecting data on activity, sleep, stress, and reproductive health which empower women with knowledge about their own health and habits. By 2016, Bellabeat had opened offices around the world and launched multiple products.

Key-highlights:

**#Products offered by Bellabeat:**

1) **Bellabeat app:** The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

2) **Leaf:** Bellabeat’s classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.

3) **Time:** This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

4) **Spring:** This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

5) **Bellabeat membership:** Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

**#Business Goal:**

Analysis of Bellabeat’s available consumer data to reveal more opportunities for growth.

**#Deliverables:**

1) A clear summary of the business task

2) A description of all data sources used

3) Documentation of any cleaning or manipulation of data

4) A summary of analysis

5) Supporting visualizations and key findings

6) Top high-level content recommendations based on analysis

**1)Business Task:**

To analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices, in this case “Fitbit” data.

* What are some trends in smart device usage?
* How could these trends apply to Bellabeat customers?
* How could these trends help influence Bellabeat marketing strategy?

**2) Data Sources:**

* Link to data: <https://www.kaggle.com/datasets/arashnic/fitbit>
* This is a Public Domain, dataset made available through Mobius: This Kaggle data set contains personal fitness tracker from 33 Fitbit users. 33 eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits.
* The dataset was originally generated by respondents through a distributed survey via Amazon Mechanical Turk.
* The survey spanned a 30 day period starting April 12th, 2016 to May 12th, 2016.

Note: This data set have some limitations, adding another data to it help address those limitations as we begin to work more with this data.

**Limitations:**

This data is not representative of all Fitbit users as only 33 people took part in the survey.

The largest table in the provided data is dailyActivity\_merged.

Roughly one fourth participant didn’t record sleep data, and almost three fourth participants didn’t record weight data.

Few records are missing as the participants were not asked to wear the Fitbit gadget for entirety of 30 day period.

This data is old as it was surveyd in 2016. The data comes from third party so there is no way to examine the data credibility and any potential biases.

**3) Documentation of cleaning, manipulation and analysis of data:**

I have done the data cleaning, manipulation and analysis for this case study in SQL server (Bigquery console).

Here is the link to SQL query/code script:

<https://console.cloud.google.com/bigquery?sq=1035602808525:c5b9079fb24e413eaf0e932edb94a07a>

The tables I have analyzed from this dataset are following:

dailyActivity\_merged, sleeplog\_merged and weightLogInfo\_merged, which were renamed and cleaned before performing analysis.

The reason to choose these 3 tables in this case study is, the insufficient data in other tables.

Participants in this study collected far more activity data than data on sleep or weight.

**Cleaning Steps:**

* Consistency: I identified the standard character length for each user ID (10). Across all tables, there were no deviations from this and/or incorrect inputs.
* Find unique participants by examining the Ids in all three tables. ( 33 unique IDs found in daily\_activity table, 24 unique IDs found in sleep\_log table, 8 unique IDs found in weight\_info table)
* Converted data type: converted Boolean values in IsManualReport column in weight\_info table to string "True" and "False"
* Duplicates: I found and deleted 3 duplicate rows total, all in the sleep log table
* Creating a new column: From ActivityDate column to get corresponding days of the week to identify if the date falls under weekdays or weekends category.
* Misleading records: 77 records in the dataset reported a total of zero daily steps. While technically possible that these records reflect days that participants were wholly inactive—most of the same records also claim 24 hours of
* sedentary activity—they’re more likely reflective of days the users didn’t wear their FitBits, making the values potentially misleading. As a result, these 77 records were deleted from the dataset.

**4) & 5) Summary Of Analysis and supporting Visualization :**

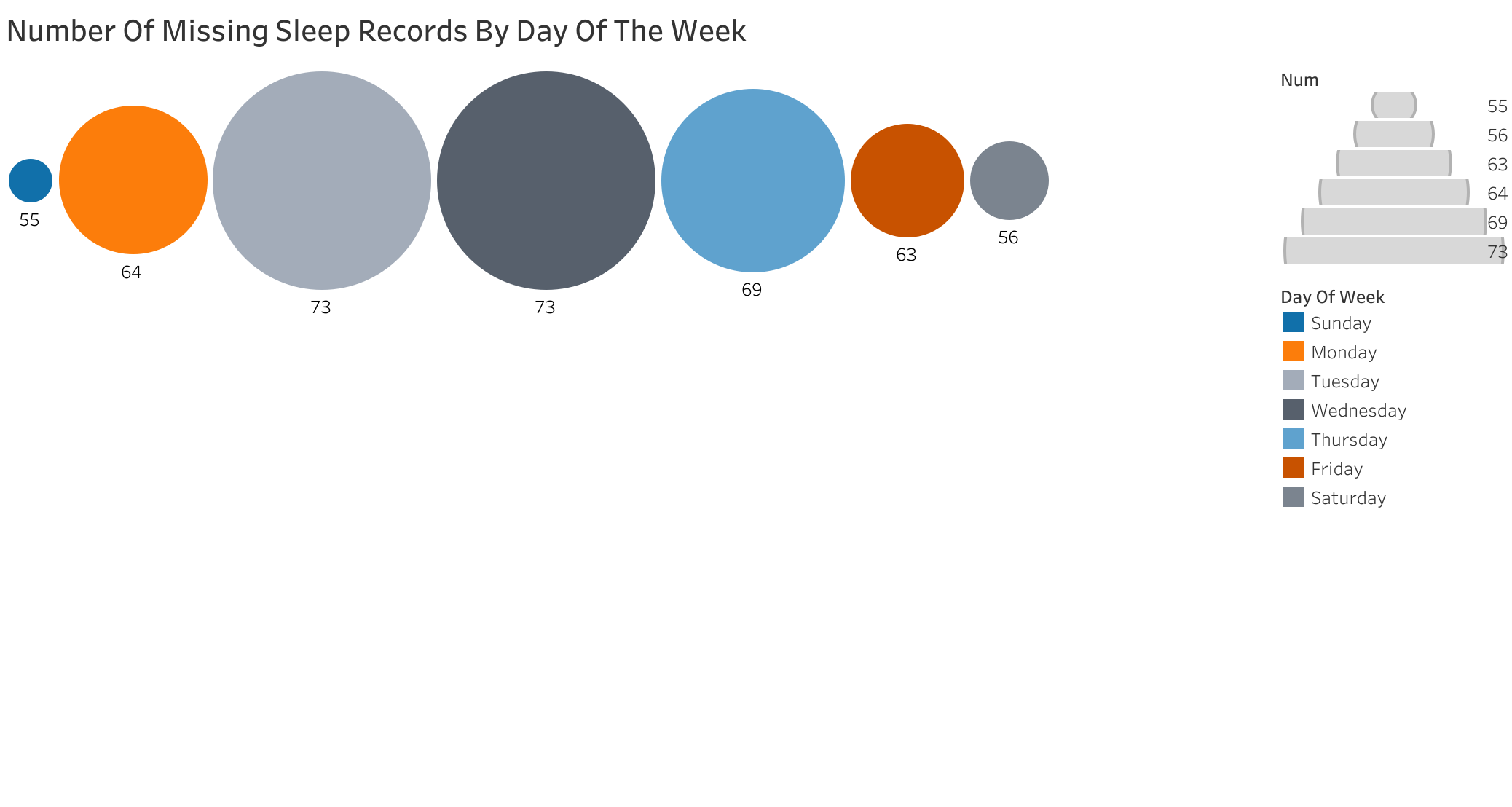
FitBit/Bellabeat users use this device data to improve their health, so it makes sense to analyze and identify the healthy and unhealthy tendencies shown in the data. Primarily, I’ll be looking at sleep, calories, and steps/activity data.

https://public.tableau.com/views/FitbitBellabeatCaseStudy/Sheet2?:language=en-US&:display\_count=n&:origin=viz\_share\_link

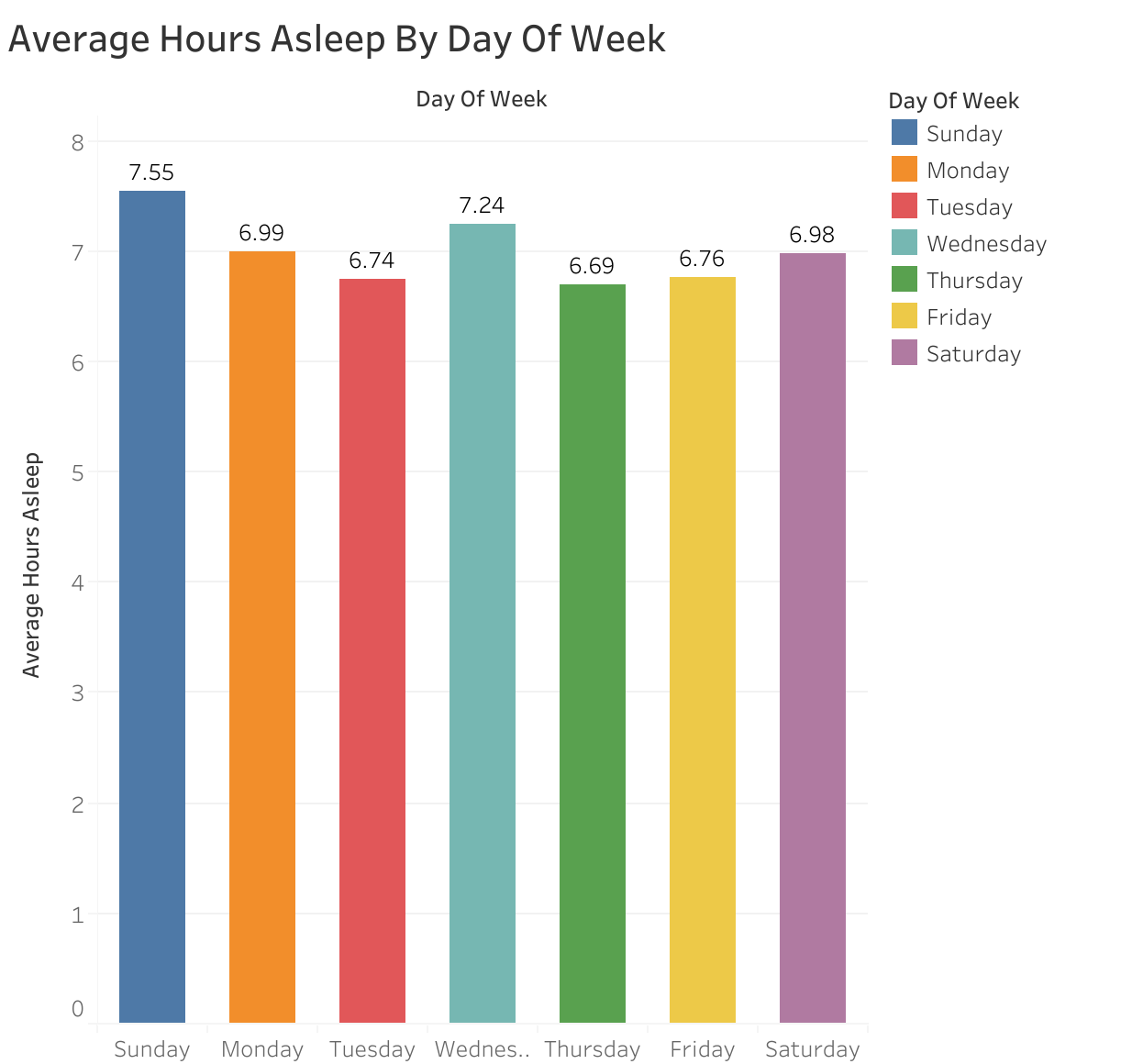
**A) Sleep Analysis:**

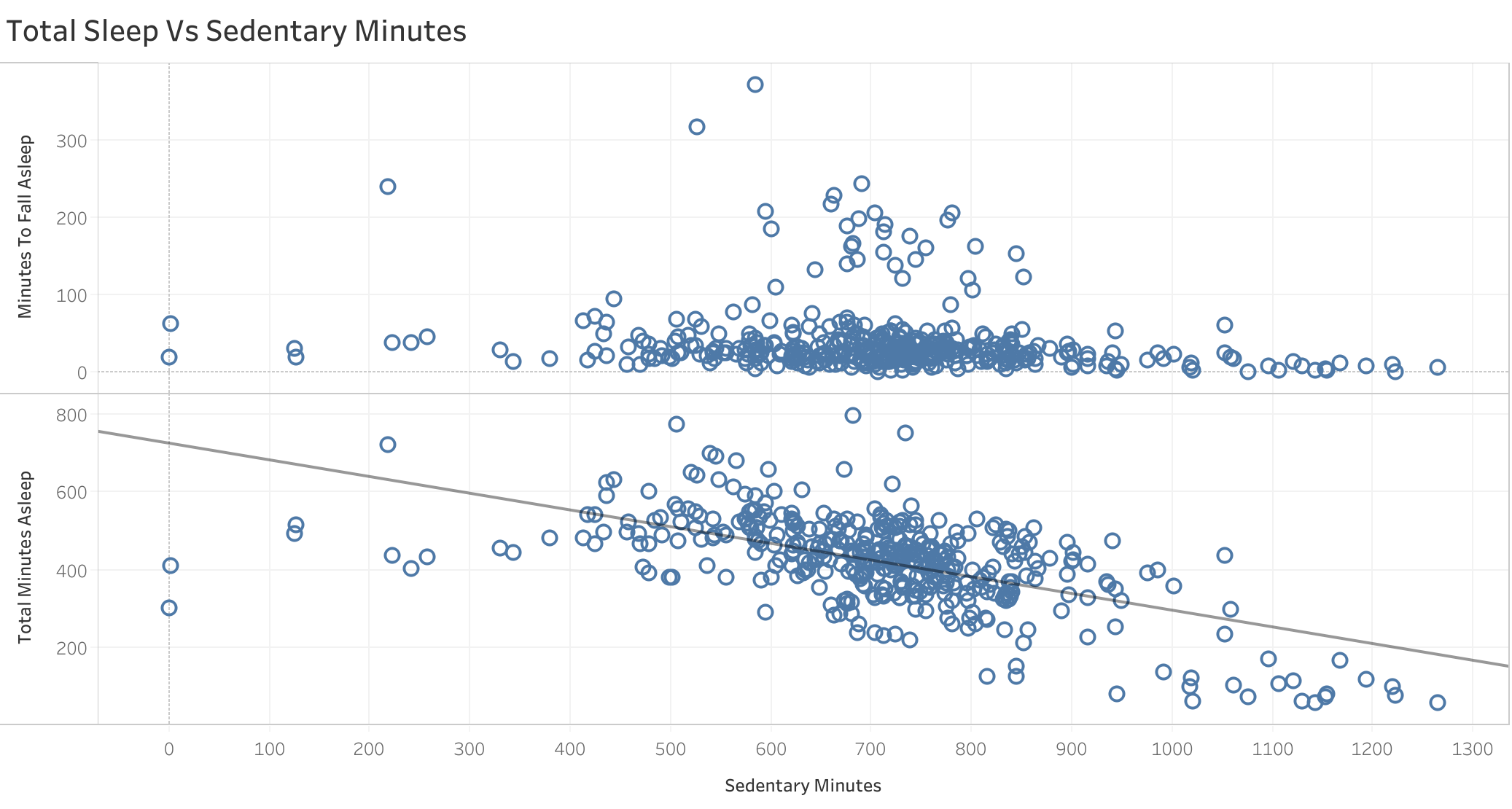
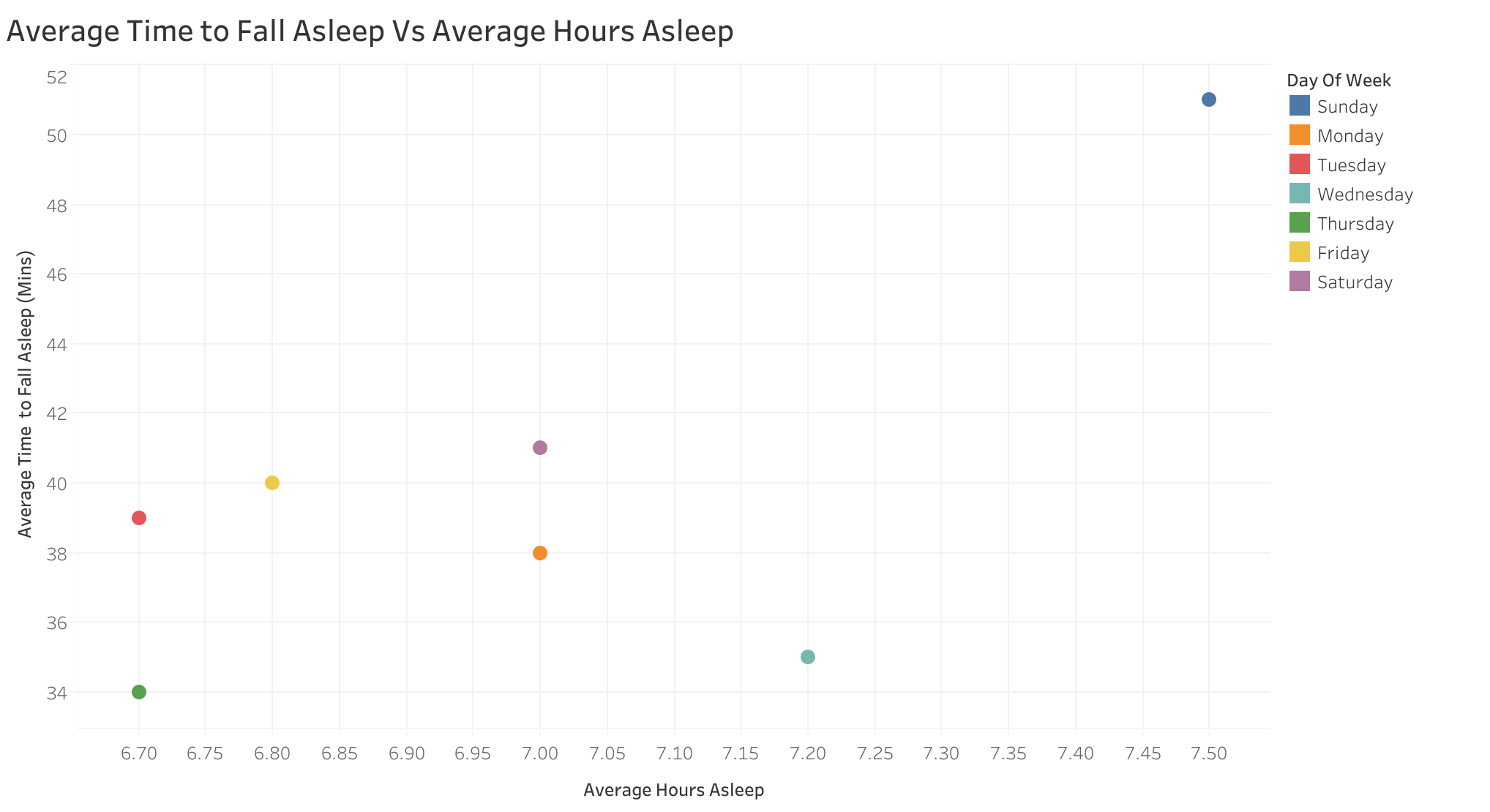
**1) Missing sleep records:** The data was aggregated by week of the day to understand the sleeping patterns on weekly basis. Its clear from the chart that majority of people forgot to record the sleep log on Tuesday and Wednesday.

2)

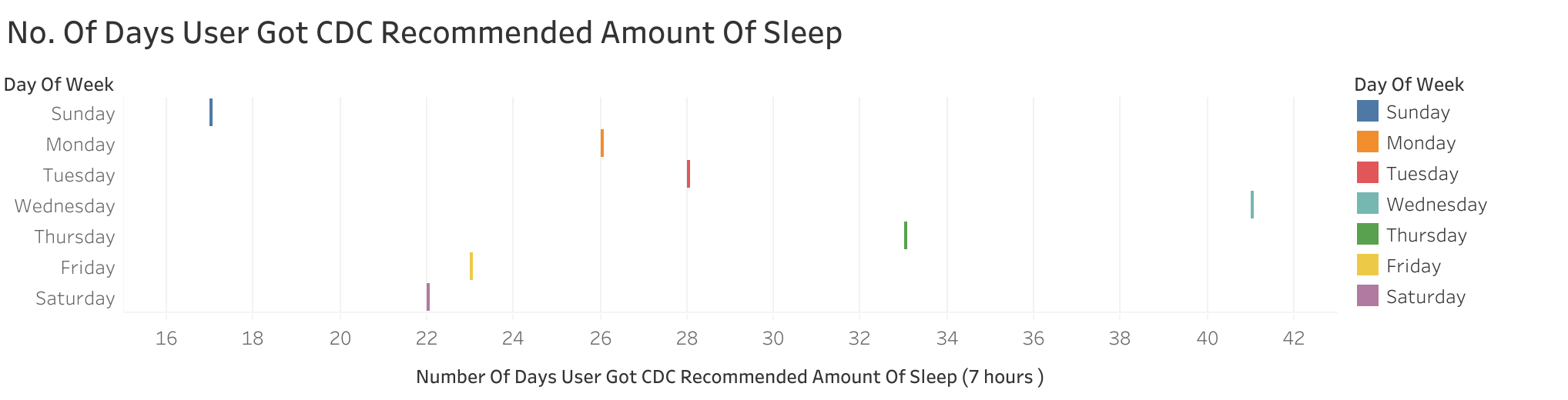


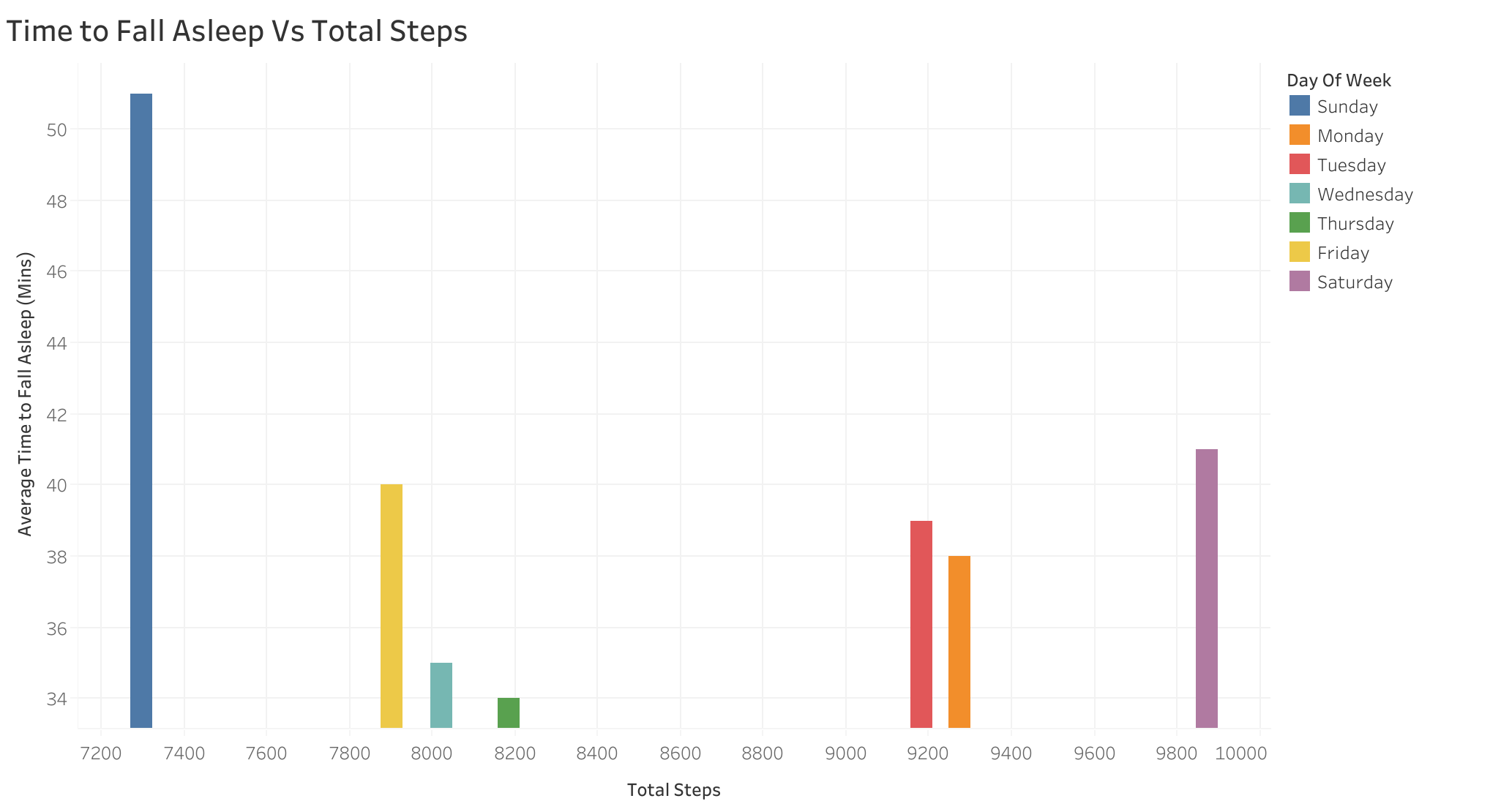
**2) Average hours asleep by day of the week:** The data indicates that people took maximum sleep on Sundays & Wednesdays.

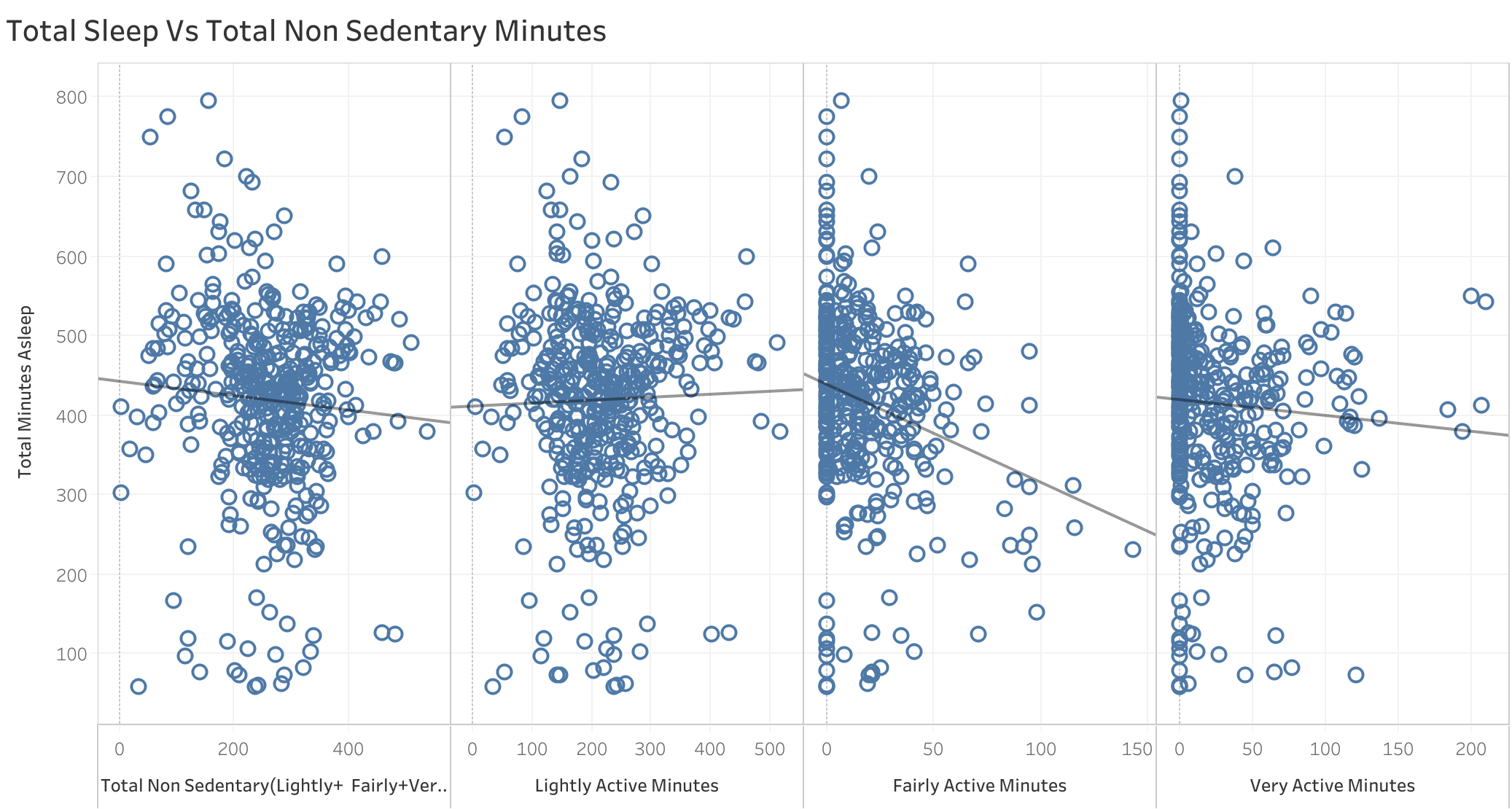
**3) Average time to fall asleep Vs average hours asleep:** The relationship between the time to fall asleep and hours asleep is positively correlated. The more rest people take and they will be less tired and would take more time to fall asleep

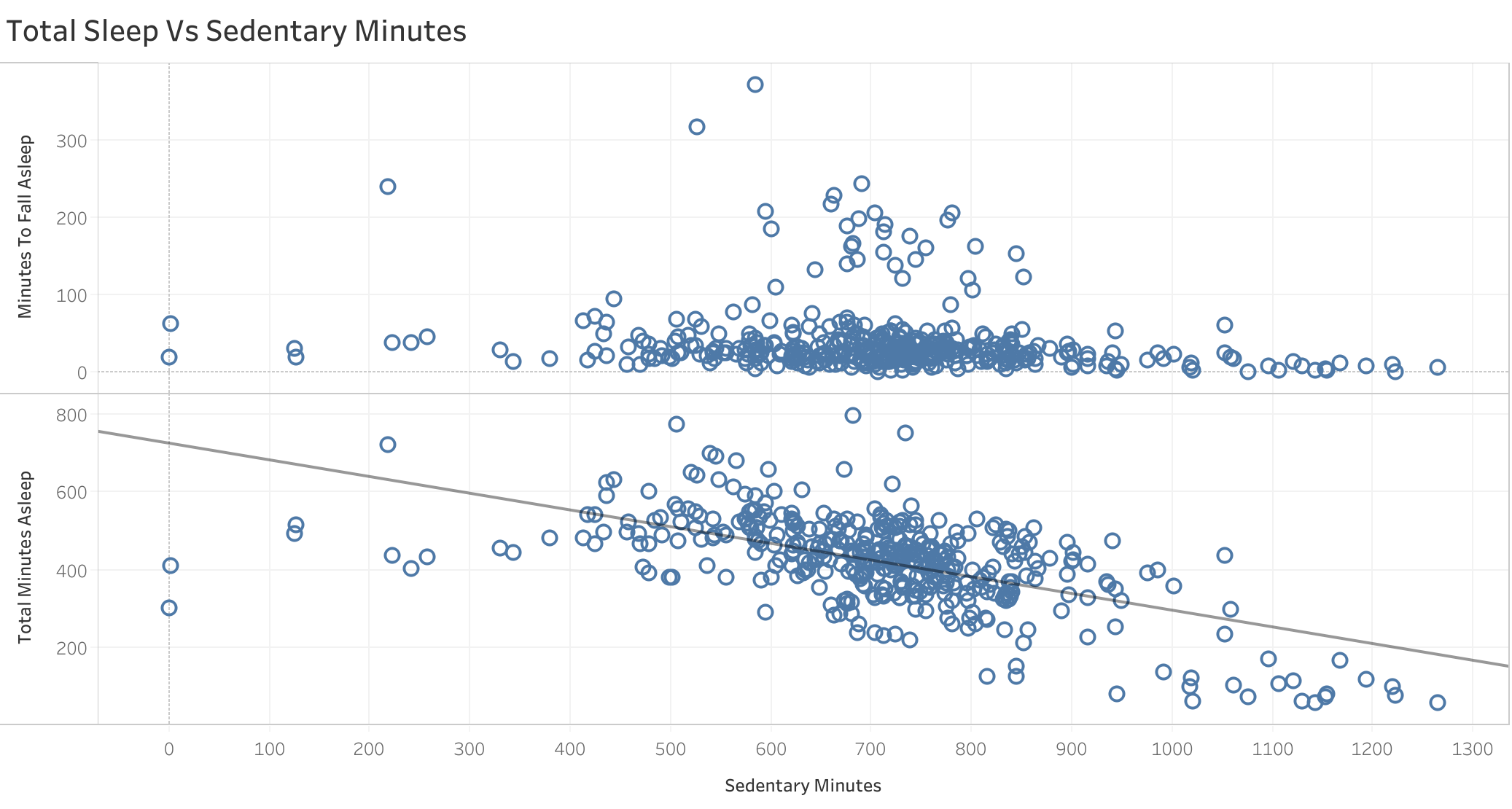


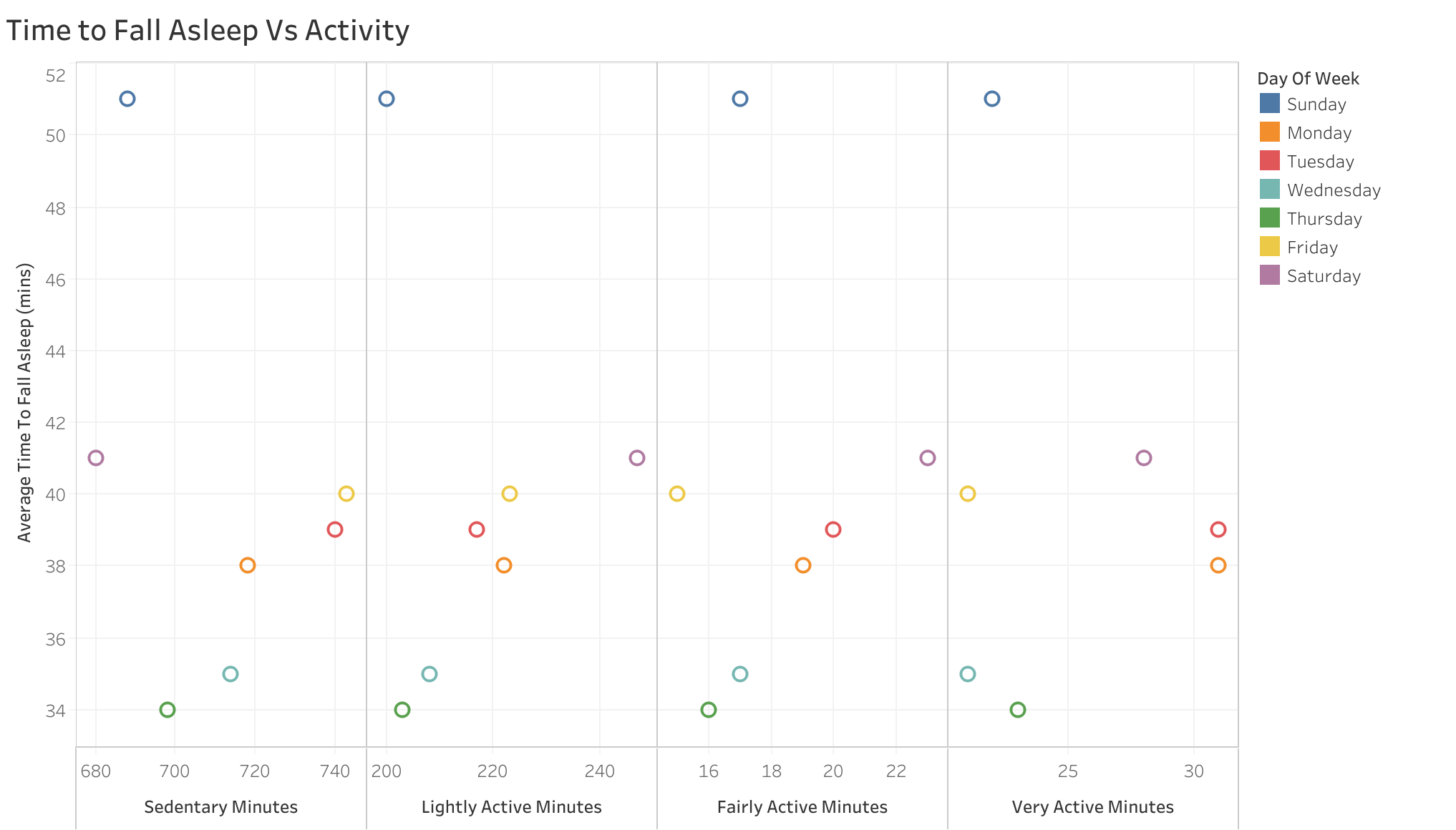
**4) Number of days user got recommended amount of sleep:** As per CDC guidelines, 7 hours is the recommended sleep a person needs in a day. As per the chart its clear that on mostly Mondays (26 days) and Tuesdays (28 days) people get 7 hours of sleep.



**5) Time to fall asleep Vs total steps taken:** When the total steps taken are less then it takes more time to fall asleep.

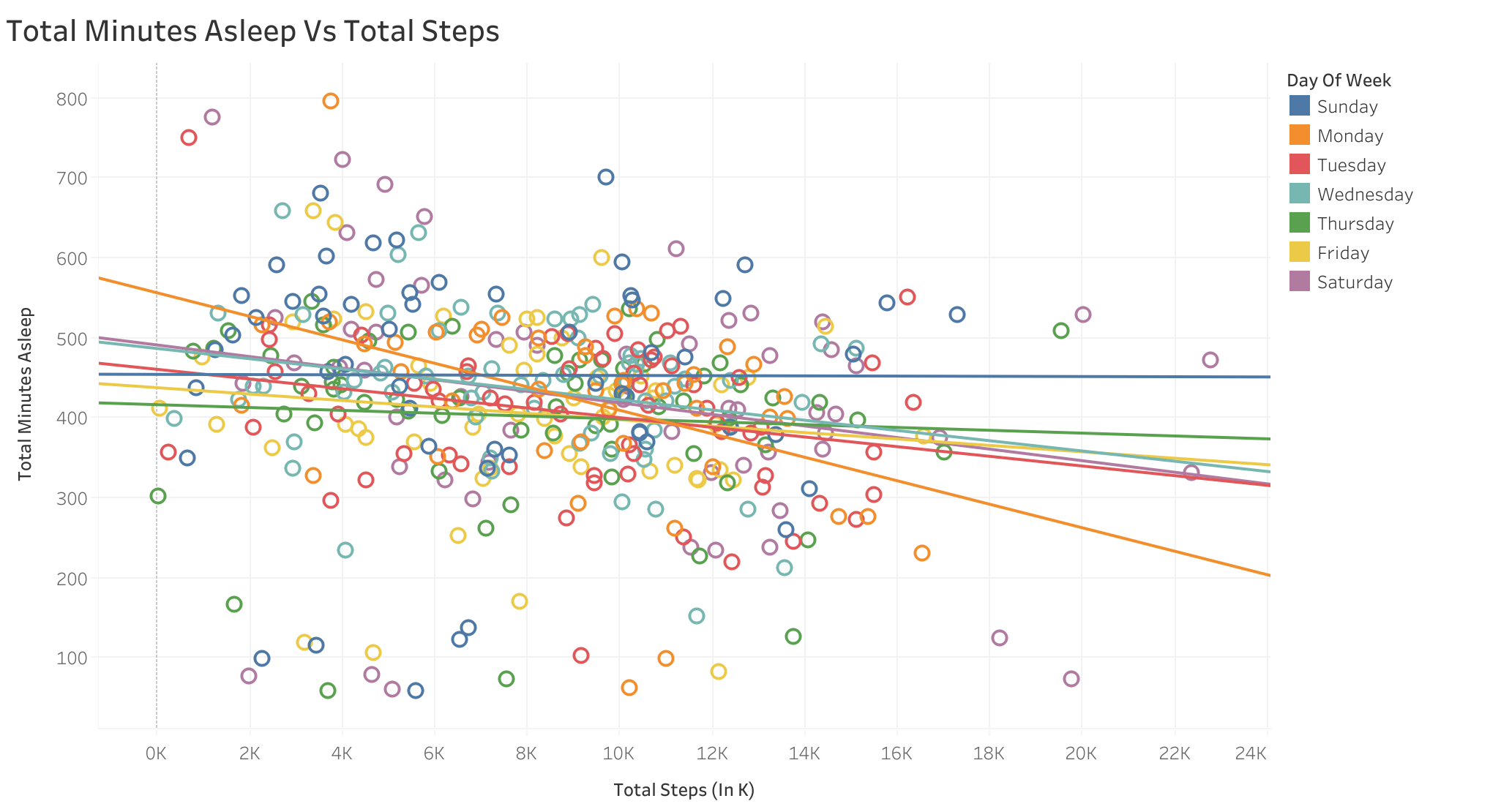
**6) Total sleep Vs total non sedentary minutes:**  In case of very active and fairly active minutes, the total minutes asleep is more.

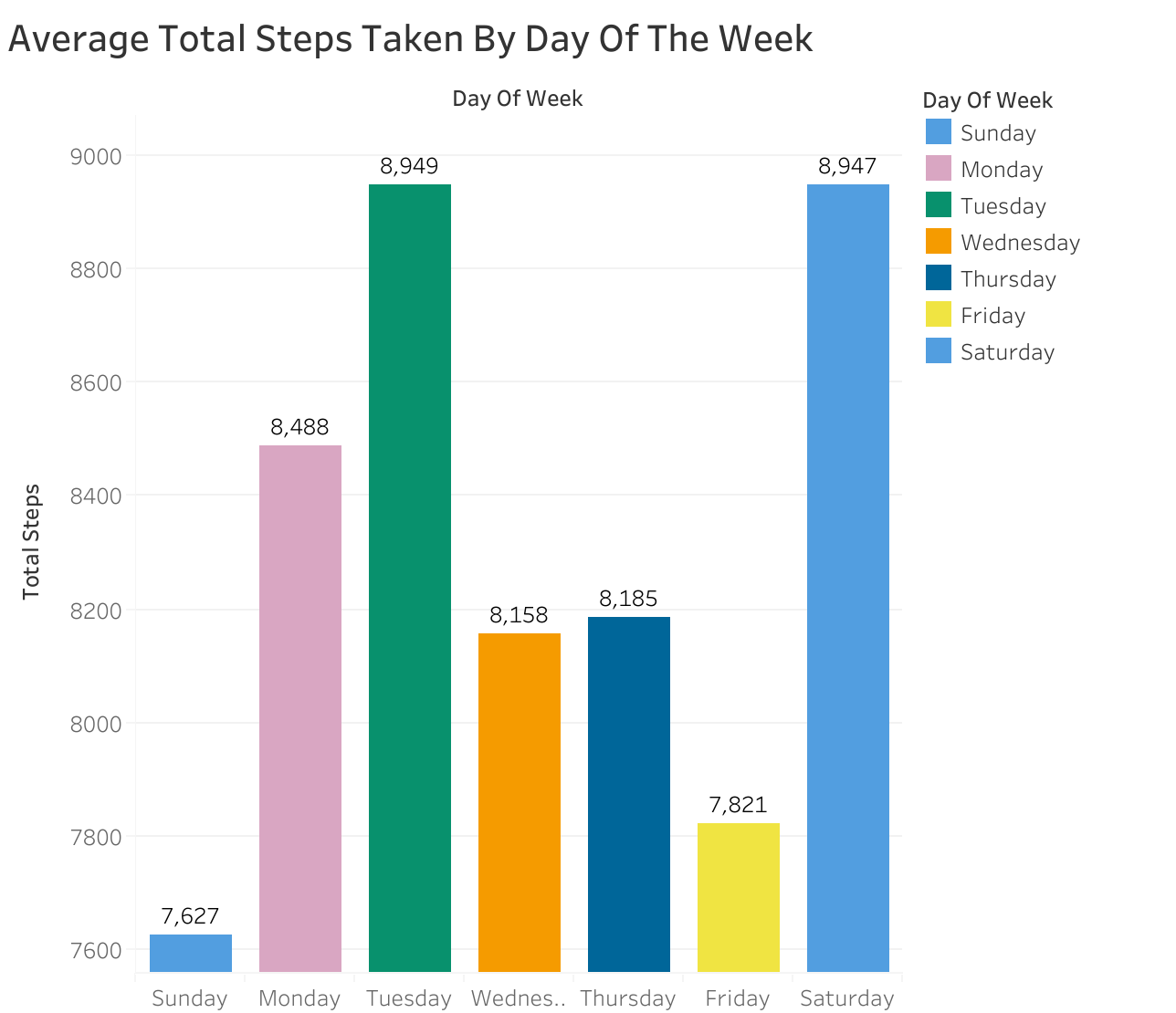
**7) Total sleep Vs total sedentary minutes:** More the sedentary minutes, lesser will be the time of sleep..

**8) Average time to fall asleep Vs activity minutes:** On an average it takes 34 minutes to 42 minutes to fall asleep depending on the activity level.

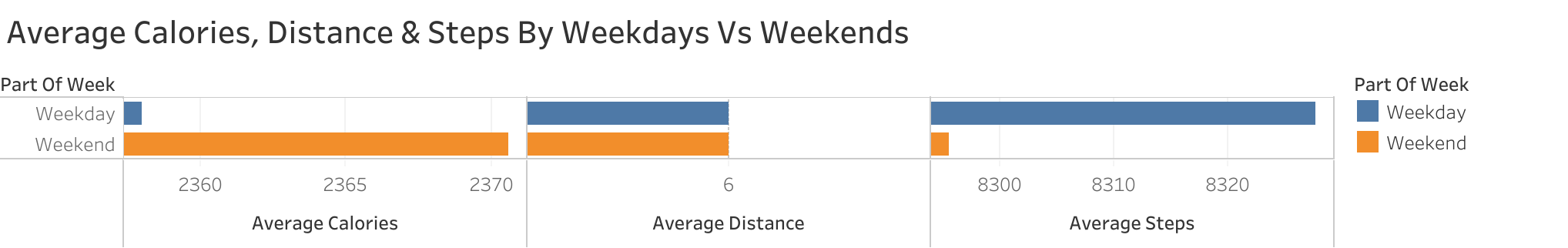
**9) Total sleep Vs total steps:** People are more active and take more steps on Tuesdays and Saturdays, and the amount of sleep

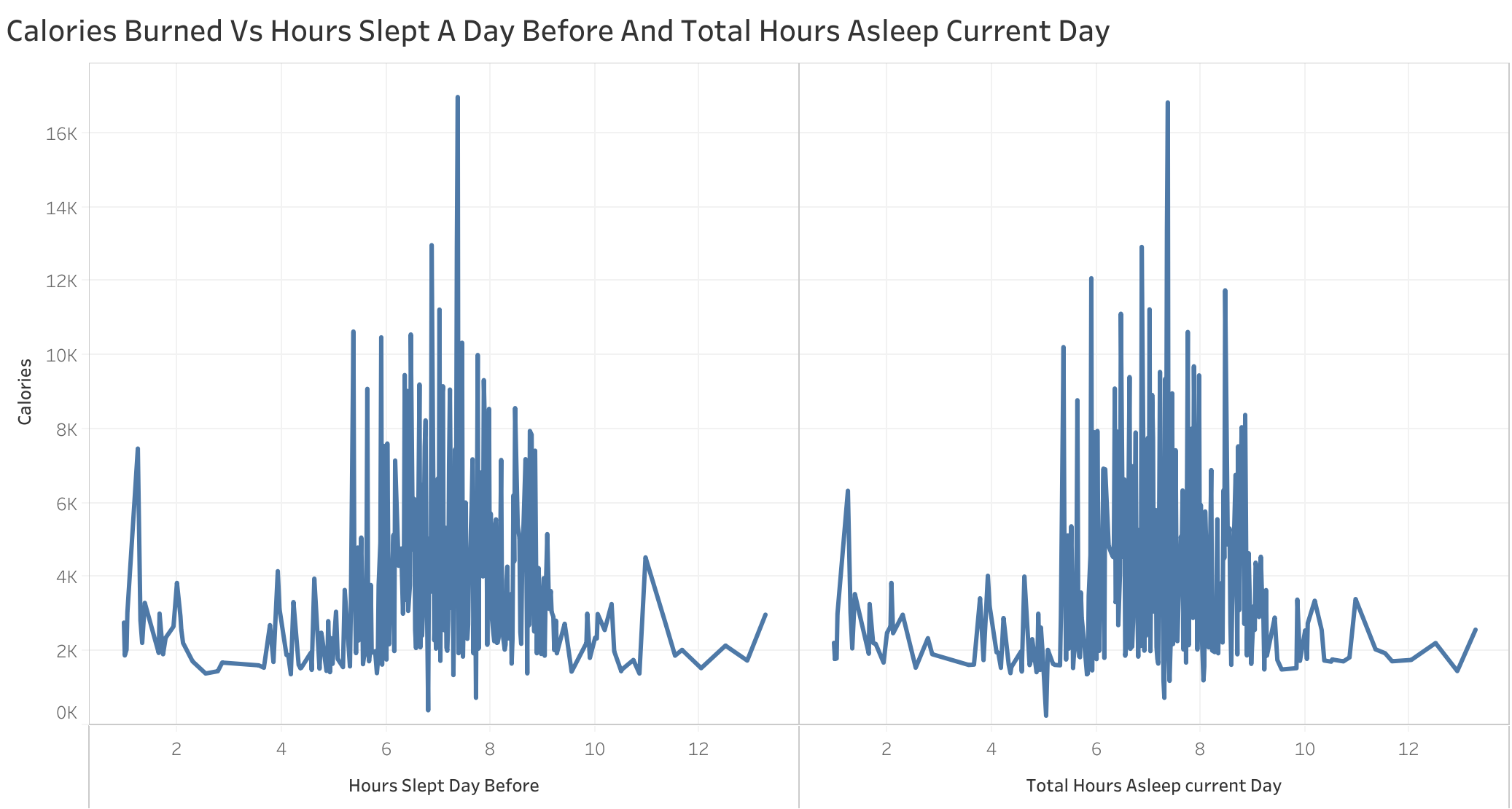
is accordingly good around 7 hours or more.

**10) Average steps taken by day of the week:** Fridays (May be because of the movie nights) and Sundays (People might prefer going for outing on Saturdays and would like to take rest before the week starts ) are the least active days as per the observation.

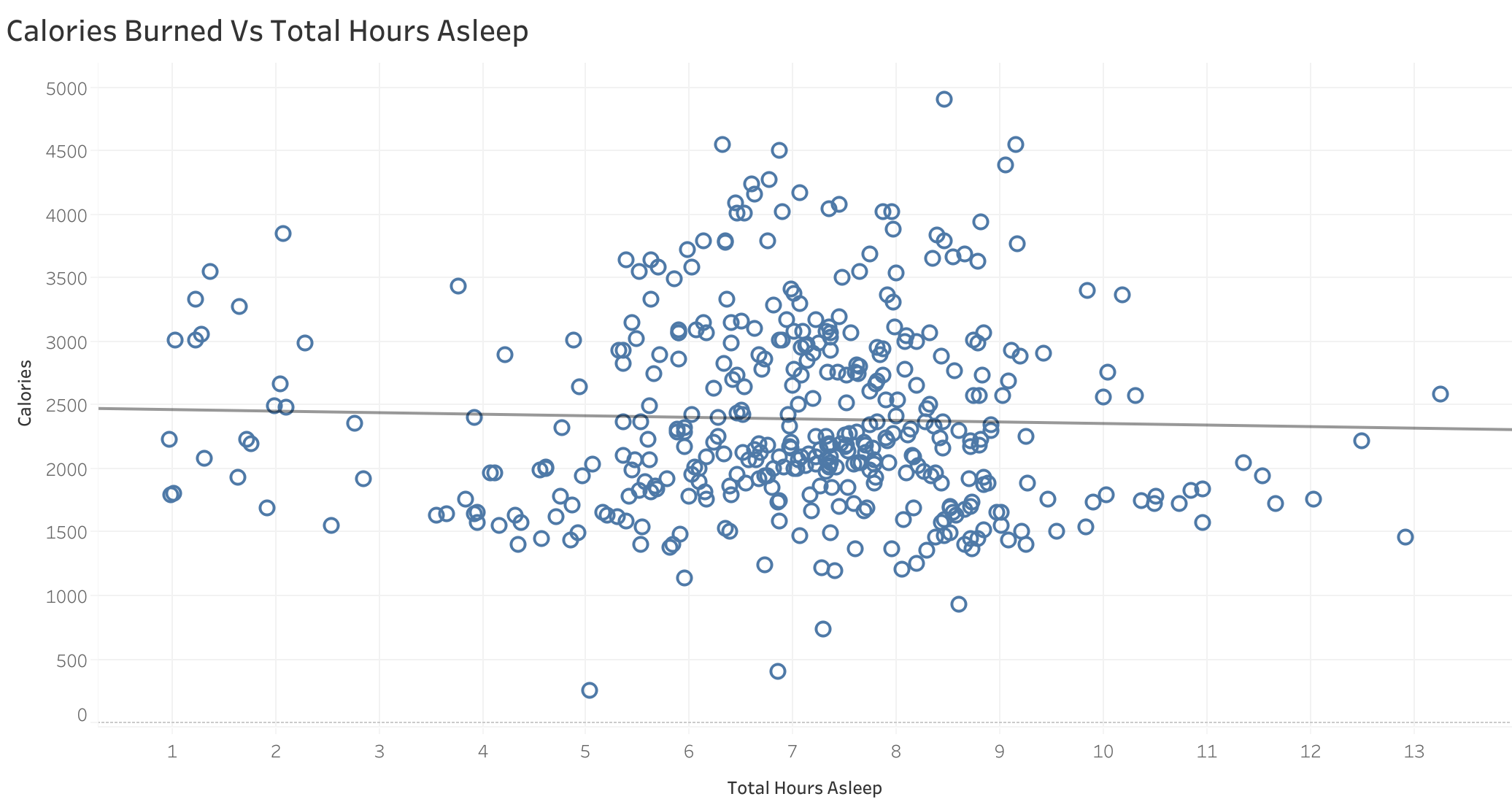


**B) Calories Analysis:**

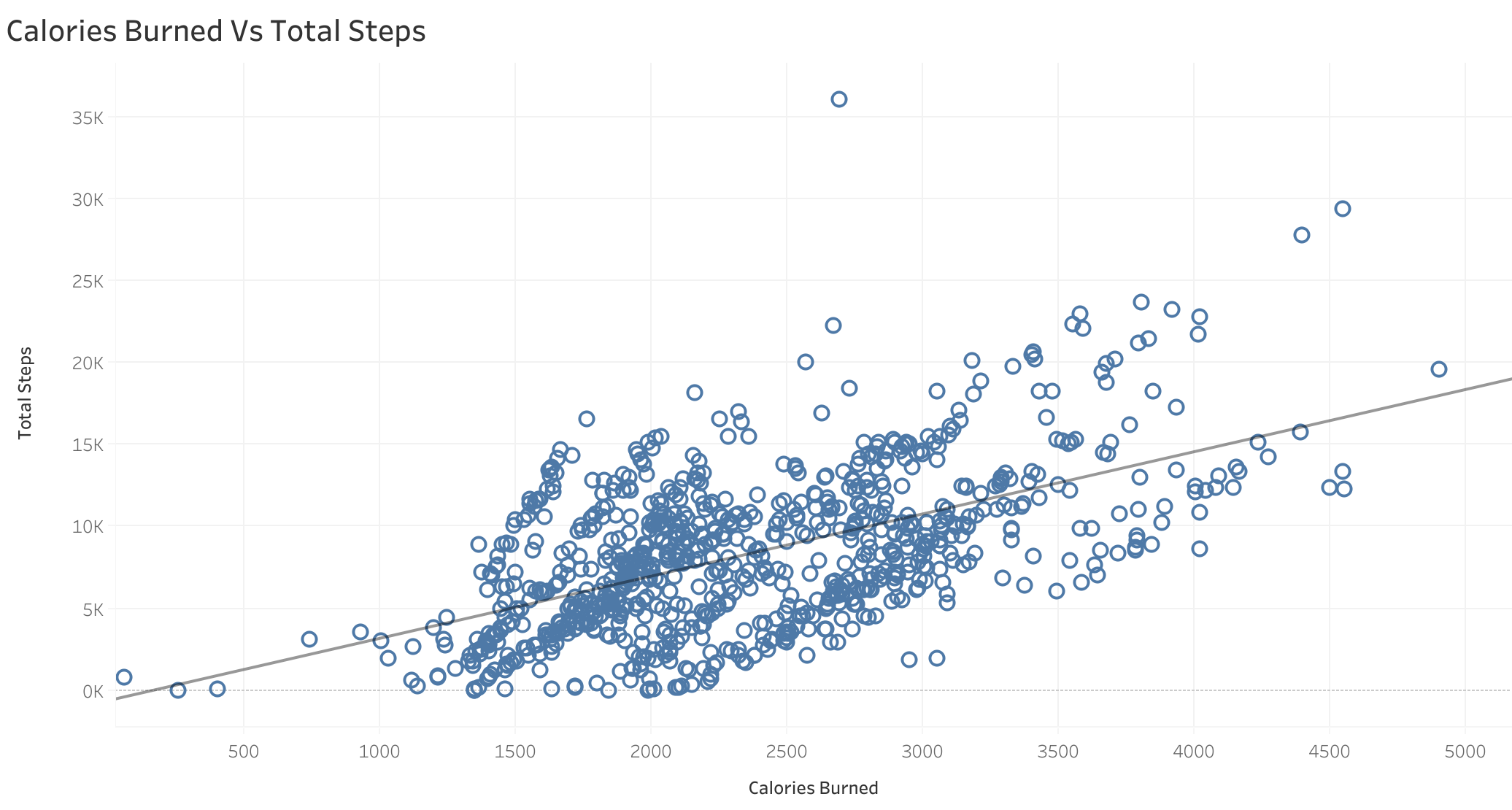
**1) Calories Vs Hours slept a day before:** Hours slept a day before is positively correlated with the amount of calories burned. When the body was well rested, the amount of calories burned on the other day is nearly higher as compared to the least rest taken on previous day.



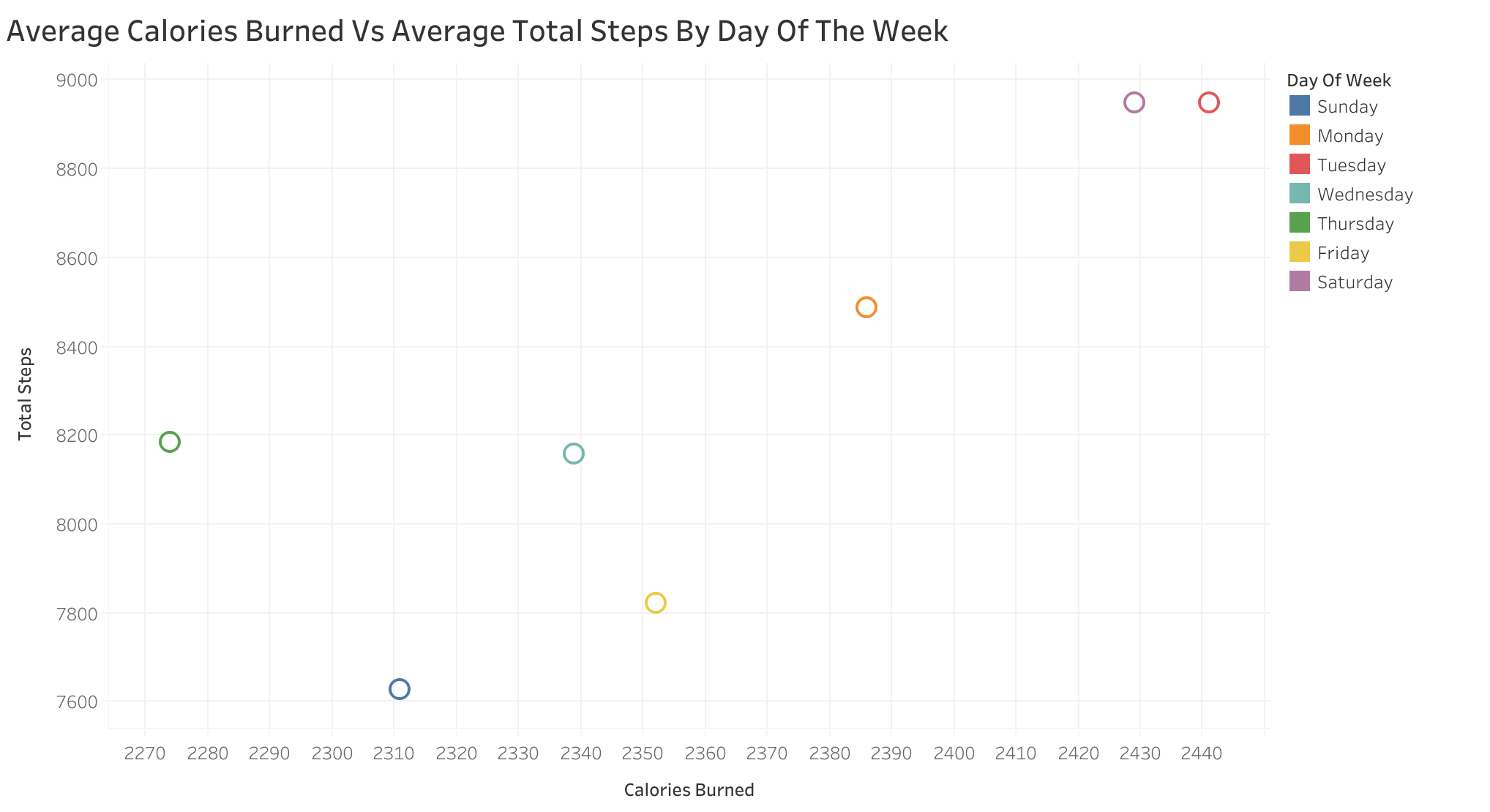
**2) Calories burned Vs total hours asleep:** The Calories burned on an average is maximum when the sleep hours are between 6 to 8.



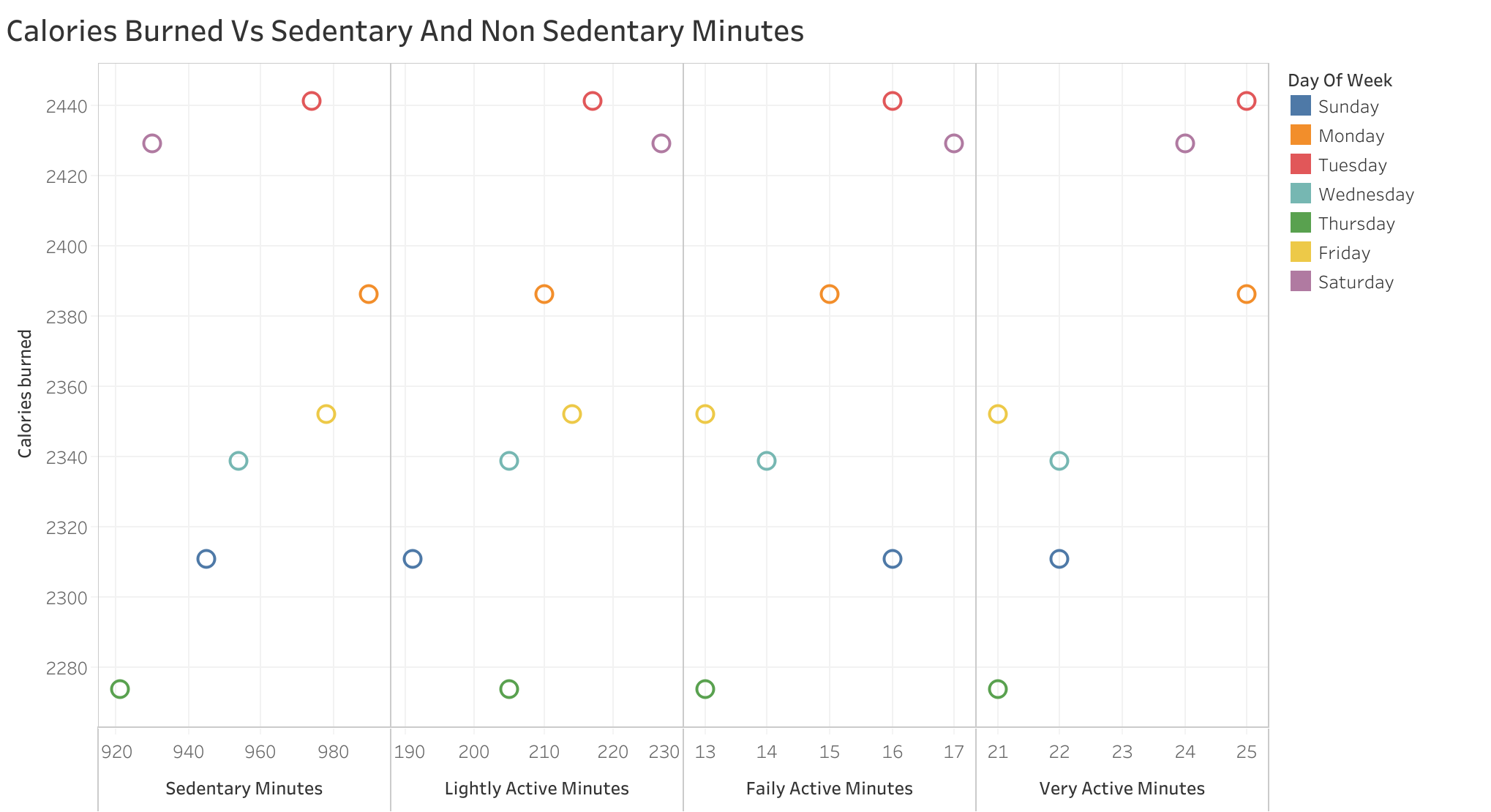
**3) Calories burned Vs total steps:** There is a linear correlation between total steps taken and calories burned.

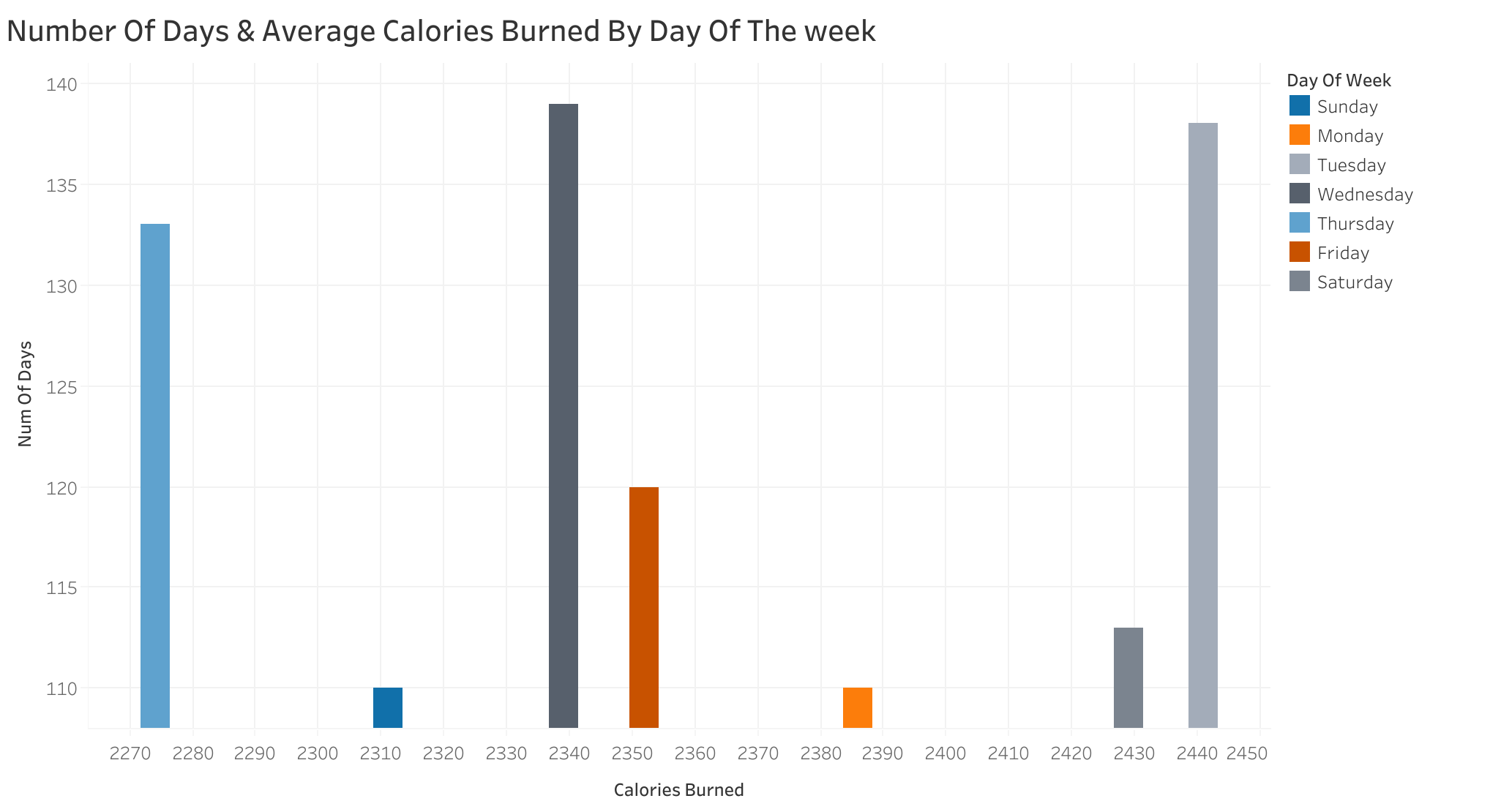


**4) Average calories burned Vs average total steps taken by day of the week:** Tuesdays and Saturday are the more active days as a result they burn more calories.

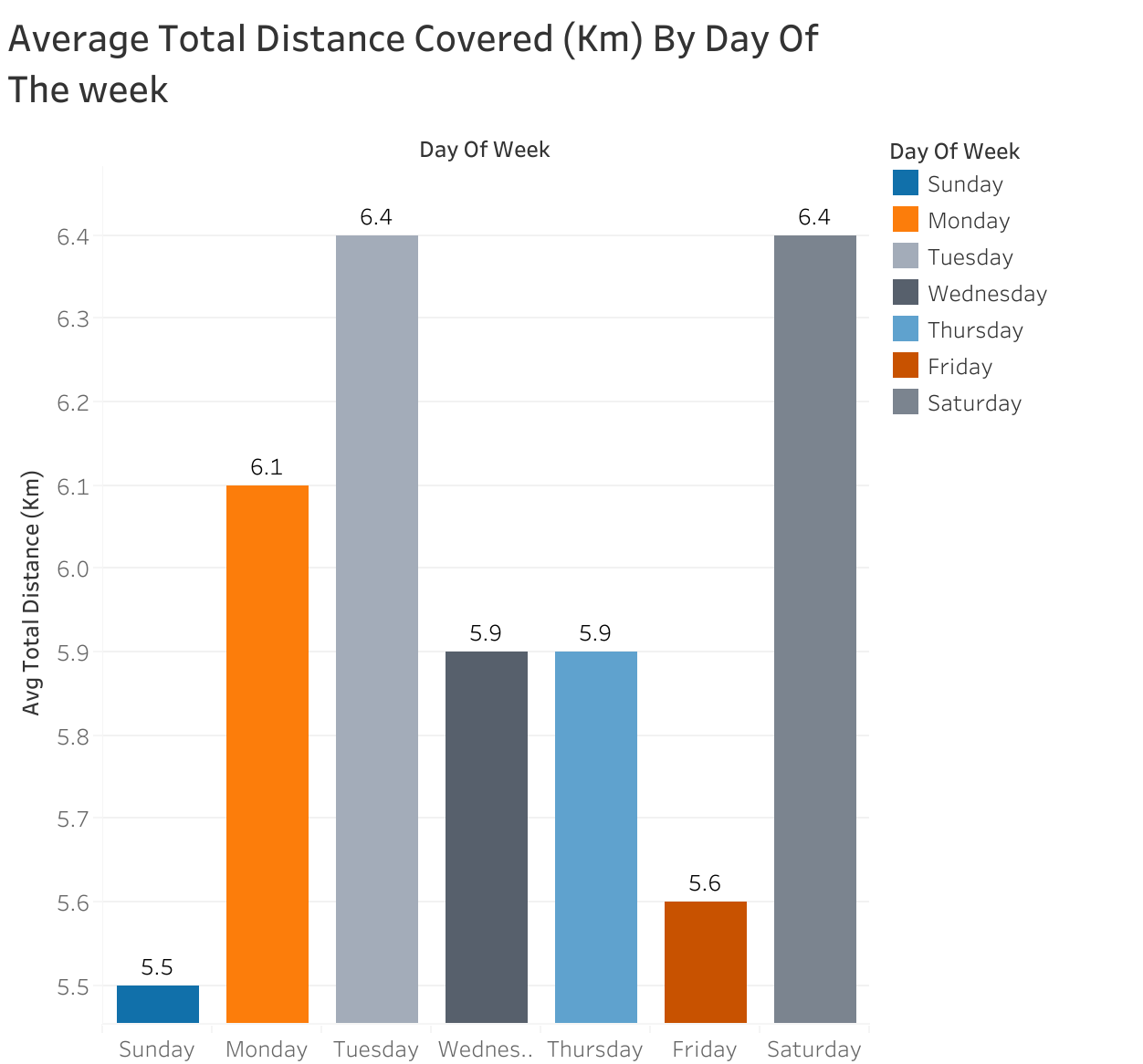


**5) Calories burned Vs sedentary and non sedentary minutes:** Its clear from the chart that more is the activity level, more calories are burned as a result of it.

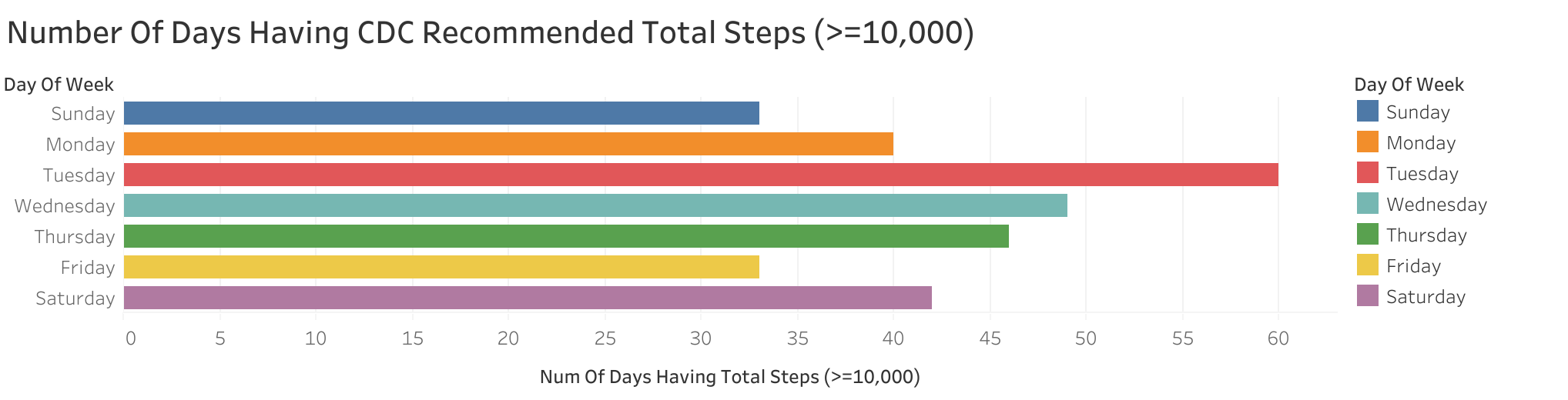


**6) Number of days and average calories burned by day of t****he week:** People burned maximum calories on Saturdays.

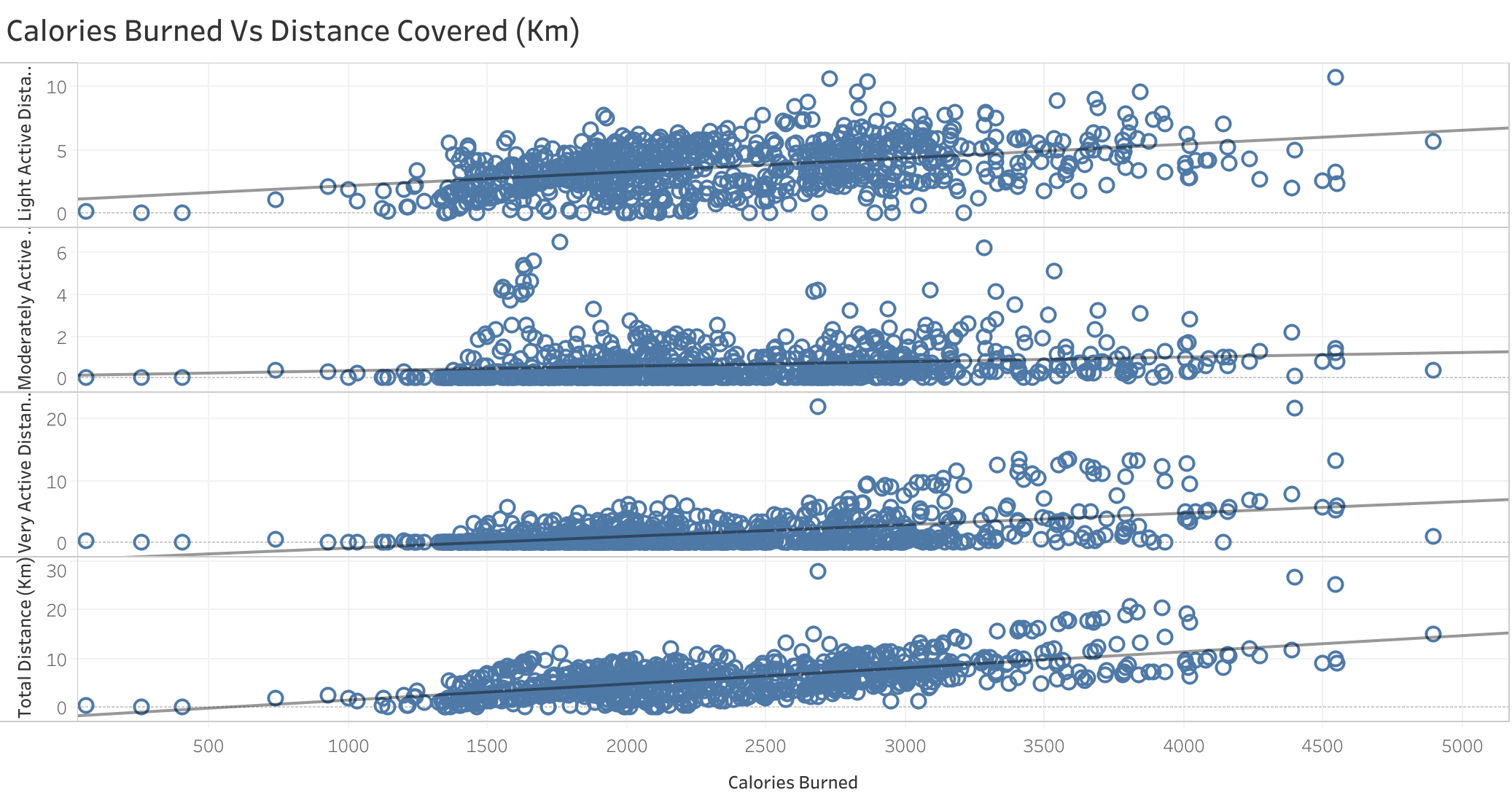
**C) Activity/Distance covered Analysis:**

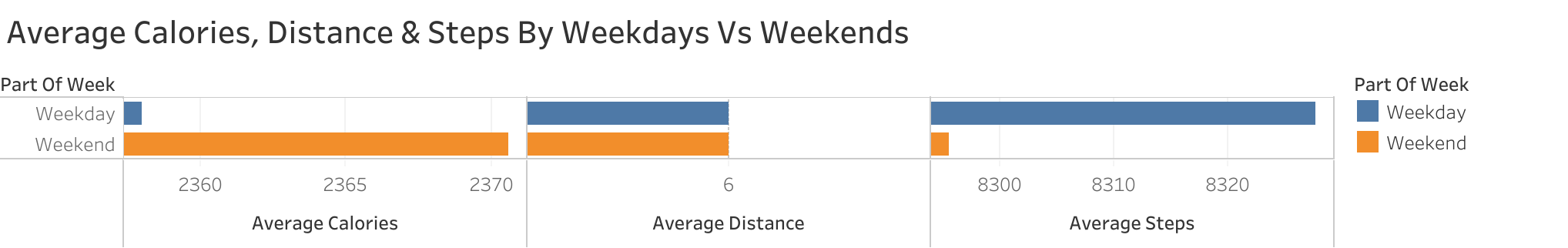
**1) Average total distance covered by day of the week:** As per the findings, the maximum distance of 6.4 KM is covered on Saturdays and Tuesdays. Which ultimately indicates that users were most active on these 2 days of the week.

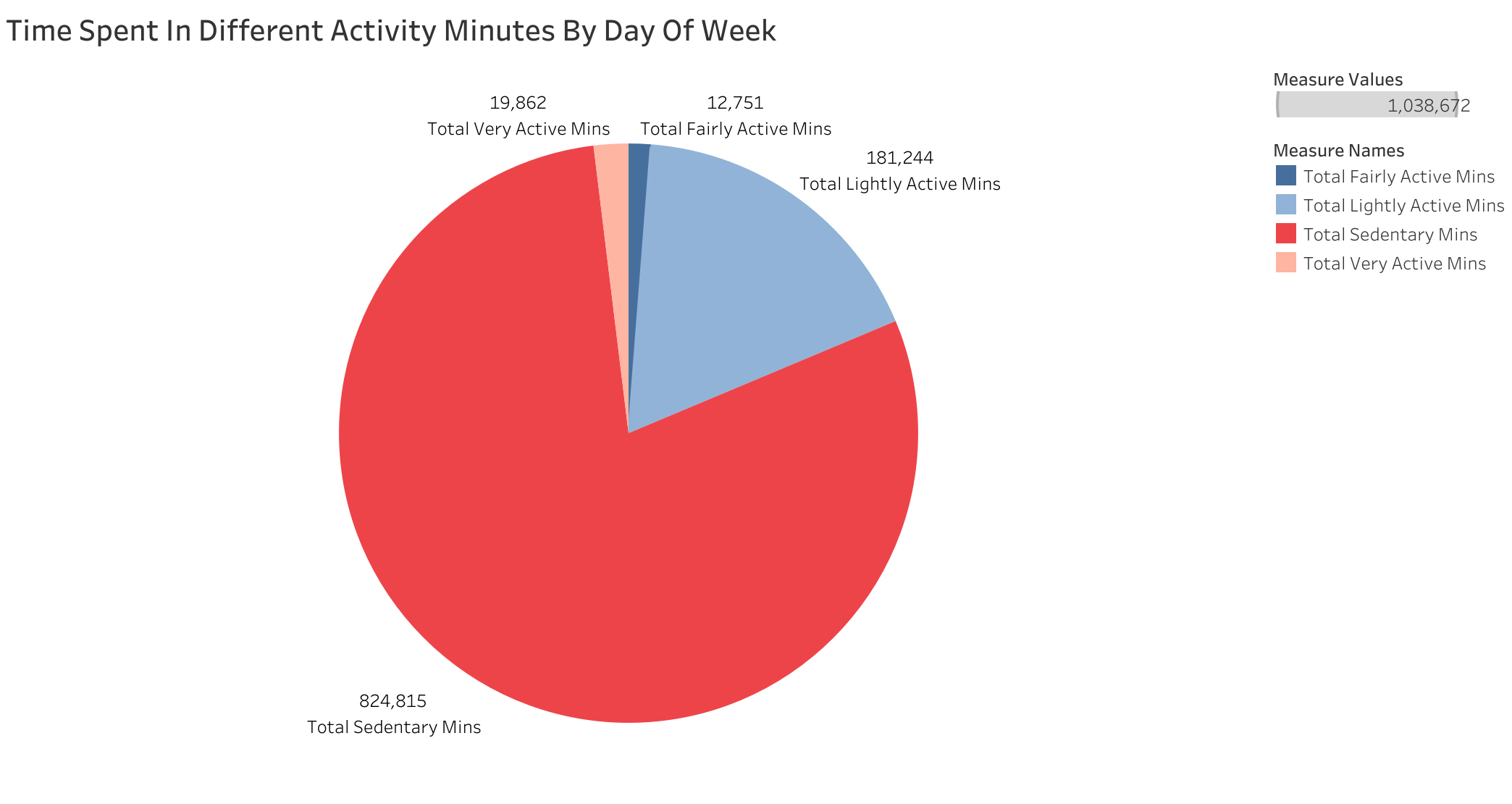
**2) Number of days having CDC recommended total steps ( >= 10,000) :** Its clear from the chart, Tuesdays and Saturdays are the highest recorded days for completing CDC recommended steps.

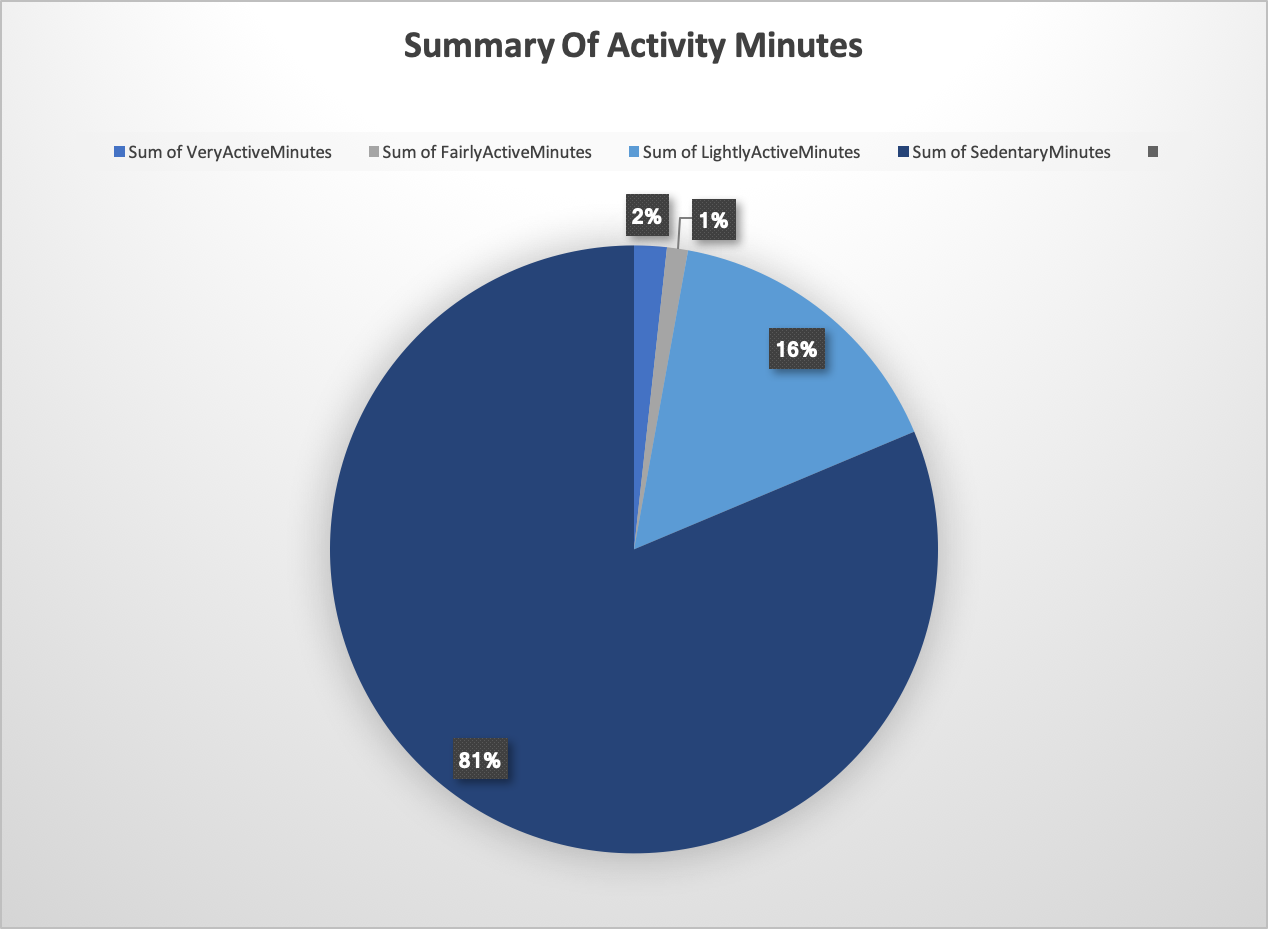


**3) Calories burned Vs distance covered:** A slight positive correlation is found between total distance covered Vs calories burned.



**4) Average calories, distance and steps by weekdays Vs weekends:** The summary indicates that the people prefer taking rest on weekends.

**D) Summary of time spend in different activity level:** Majority of time people are not active. The fairly and very active minutes are very less as compared to the sedentary minutes.



**6) Top high-level content recommendations based on analysis:**

Below are some recommendations to the marketing team for curating their marketing strategies based on my understanding after the data analysis.

**Observations and Trends:**

* More steps means more calories burned.
* Most users walk about 5K to 10K steps on average in a day.
* Sundays are preferred for sleeping/taking rest having least activity.
* Saturdays are preferred for the longest walks or most number of steps taken while also catching up on sleep.
* People on average spend more than half of their time as sedentary time, probably because they are seated for their desk jobs.
* To improve quality and duration of sleep, they should reduce their sedentary time and be more physically active.
* People tend to go for highest intensity runs/walks and workouts on tuesdays and that keeps reducing till weekends as people get busy with their work throughout the week. Saturdays see a little rise while Sundays being the least as it's preferred as a rest day.

**Recommendations:**

* People in long sedentary/seated positions can be reminded of taking a walk around every hour with pop-ups/reminders in their smart devices.
* Send reminders/notifications on device everyday for users to stick to sleep routine and sleep at least 6-8 hrs everyday.
* Highlight the benefits of a proper daily sleep with tips, small reads and articles. Direct and inspire people to use more of their sleep tracker in the devices.
* Create personalized challenges/goals like taking 5K steps, then 8K steps, sleeping 8 hrs, on time, walking everyday and such.
* Marketing Campaigns and Tie ups with gyms and fitness clubs to focus on the right user groups, who are interested in getting fit. The memberships can be clubbed with device costs which will also provide huge sets of data from different types of users, unfit to fit, over periods of time, for further exploration.