Fitbit Tracker Dataset Analysis

Group 3 - Team Members:

Gifty Appiagyei
Travis Jones
Enimienlen Lawani
Yunnan Li
Yao Xiao

Outline

- Overview

- Analysis (visualizations)

- Findings/Observations

- Summary/Conclusion

- Recommendations

Overview

- Why we choose this topic, what interested us?

The data broke down the user activities into the minutes and hours throughout the day,
With this data we were able to see how many calories they burned during different periods throughout the day and where they could do better.

- The data we use for analysis (what, where)

We found the data on Kaggle.com.
The data show a 2 month period where they recorded 33 different users activities, from how many calories they burned to how many steps they took in a day even to how much time they spent in bed both awake and asleep.

The Data Exploration

- ## Limitations

* There are limited users. The sample size may not be representative of the population.
* There are only a few categorical variables to categorize the data for analysis.
* Each dataset has 30 days of data from 4/12/2016 to 5/12/2016, but the dataset description

- ## Data Summary

The data shows the activity rate throughout the day allowing us to compare different elements of the daily activities.

Analysis

- What is the average daily step count for users?

- How does physical activities vary across different days of the week?

- What time of day do users typically reach their peak activity levels?

- What's the intensity level among all users and how it varies within a day?

- Do the activities during the day affect individuals sleep quality? How?

Q1

- What is the average daily step count for users?

•Importance of staying active.
•Insights from data.

Q1

Q1
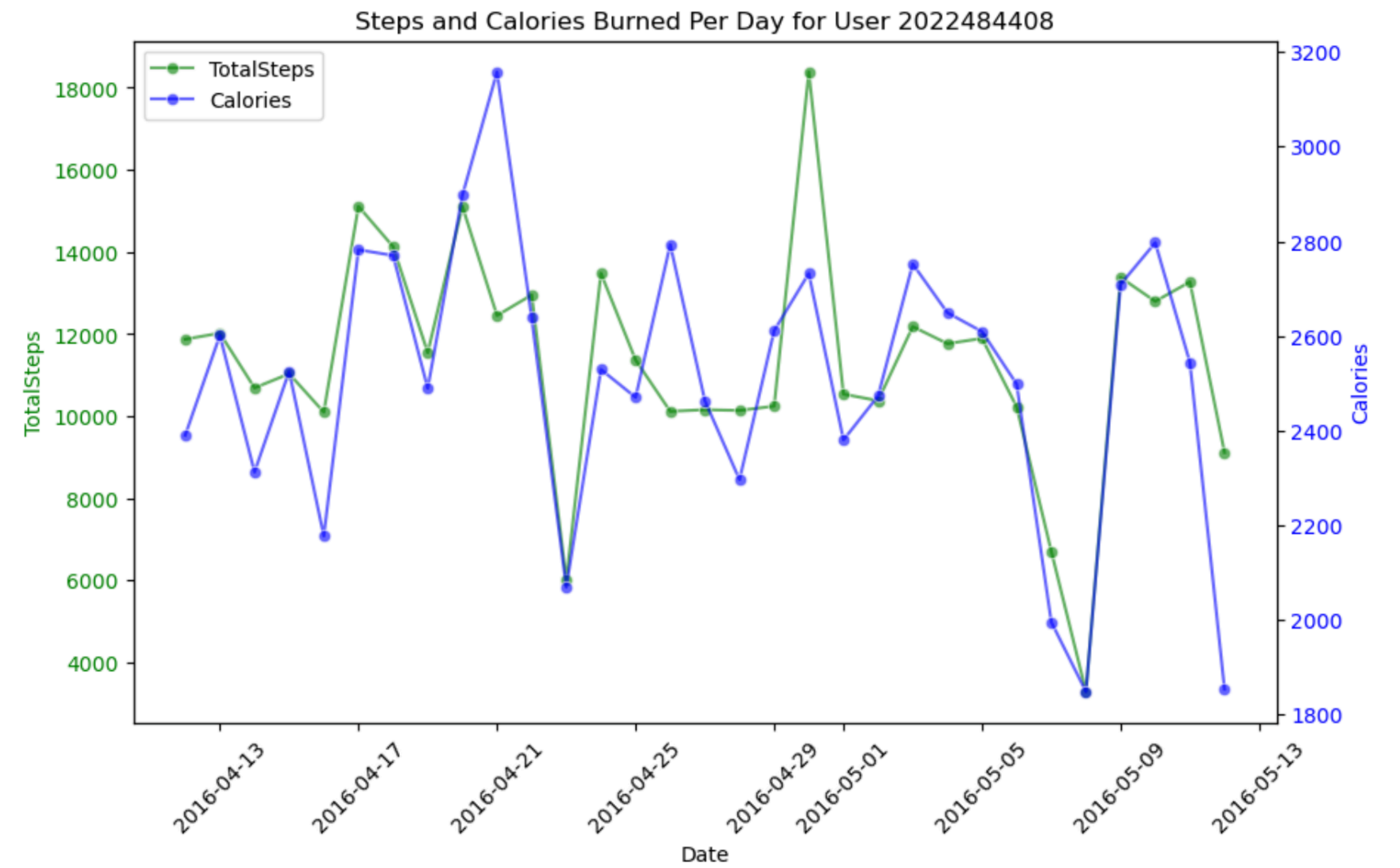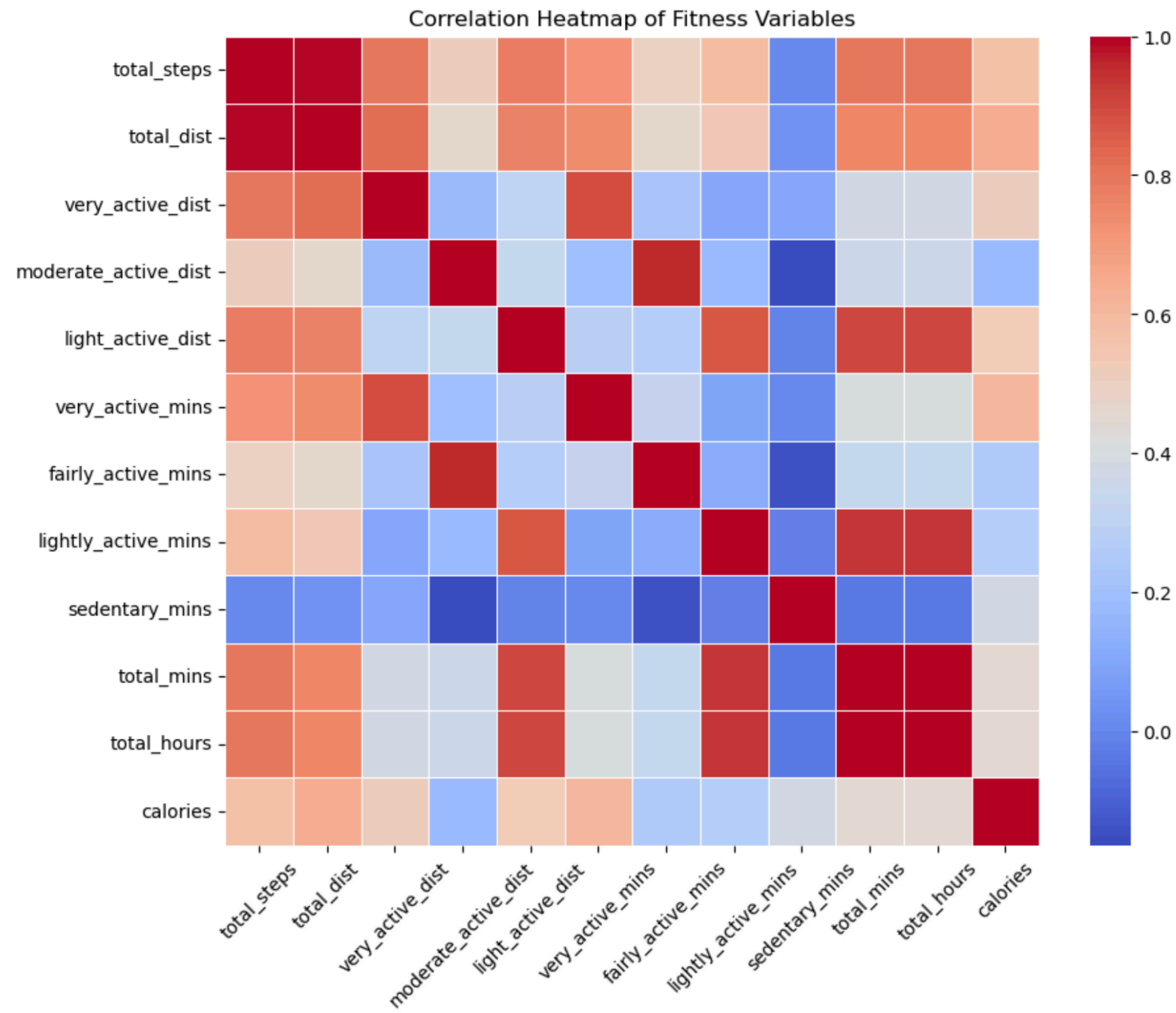
```python
# Generate lineplot of the total steps and total calories burnt for a single user with time.
user_id = 2022484408
user_activity = df_activity[df_activity["id"] == user_id]


fig, ax1 = plt.subplots(figsize=(10, 6))
plt.xticks(rotation= 45)

sns.lineplot(data=user_activity, x="date", y="total_steps", marker="o", ax=ax1, color="g", label="TotalSteps", alpha=0.6, legend=False)
ax1.set_ylabel("TotalSteps", color="g")
ax1.tick_params(axis="y", labelcolor="g")

ax2 = ax1.twinx()
sns.lineplot(data=user_activity, x="date", y="calories", marker="o", ax=ax2, color="b", label="Calories", alpha=0.6, legend=False)
ax2.set_ylabel("Calories", color="b")
ax2.tick_params(axis="y", labelcolor="b")

plt.title(f"Steps and Calories Burned Per Day for User {user_id}")
ax1.set_xlabel("Date")

lines, labels = ax1.get_legend_handles_labels()
lines2, labels2 = ax2.get_legend_handles_labels()
ax1.legend(lines + lines2, labels + labels2, loc="upper left")
plt.show()
```
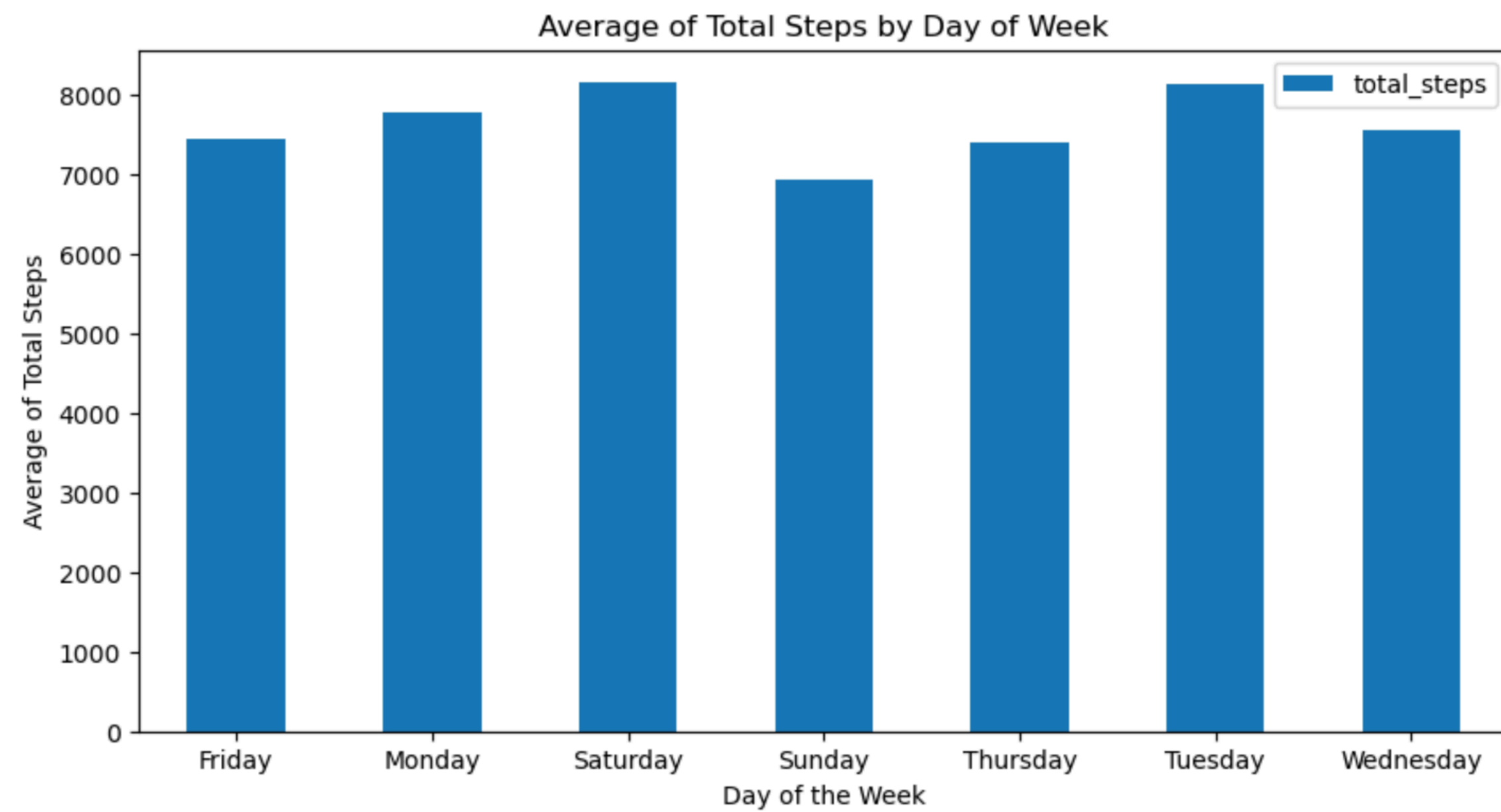
Q2

- How does physical activities vary across different days of the week?

The average of total daily steps by the users is 7,638 steps.According to a study conducted in 2011 by BMC/BioMed Central, taking 10,000 steps a day is a reasonable target for healthy adults, helping reduce certain health conditions, such as high blood pressure and heart disease.

Q3

- What time of day do users typically reach their peak activity levels?

What are the most active hours of the users' day on average? Are they occurring during the day or at night?
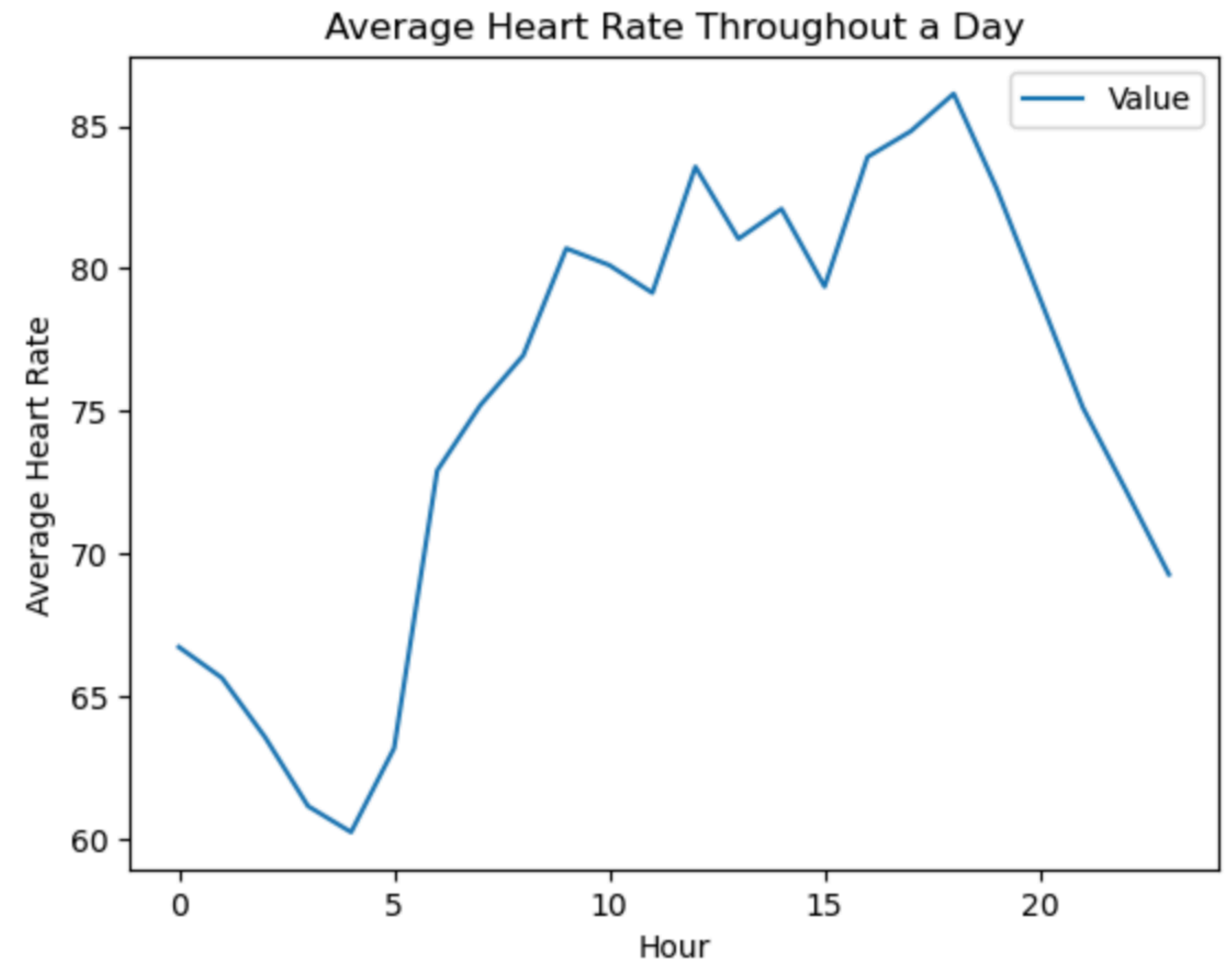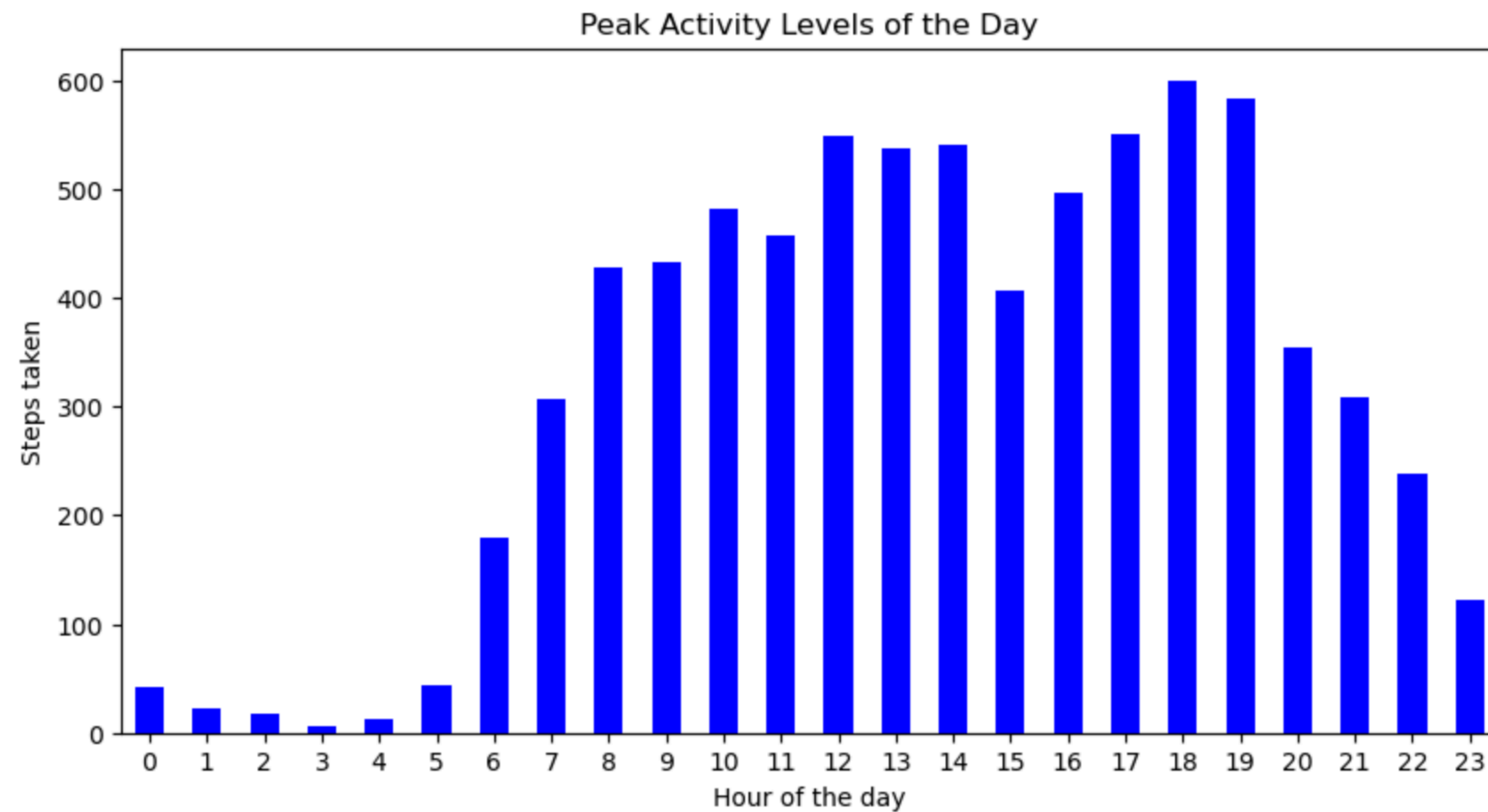
Q4

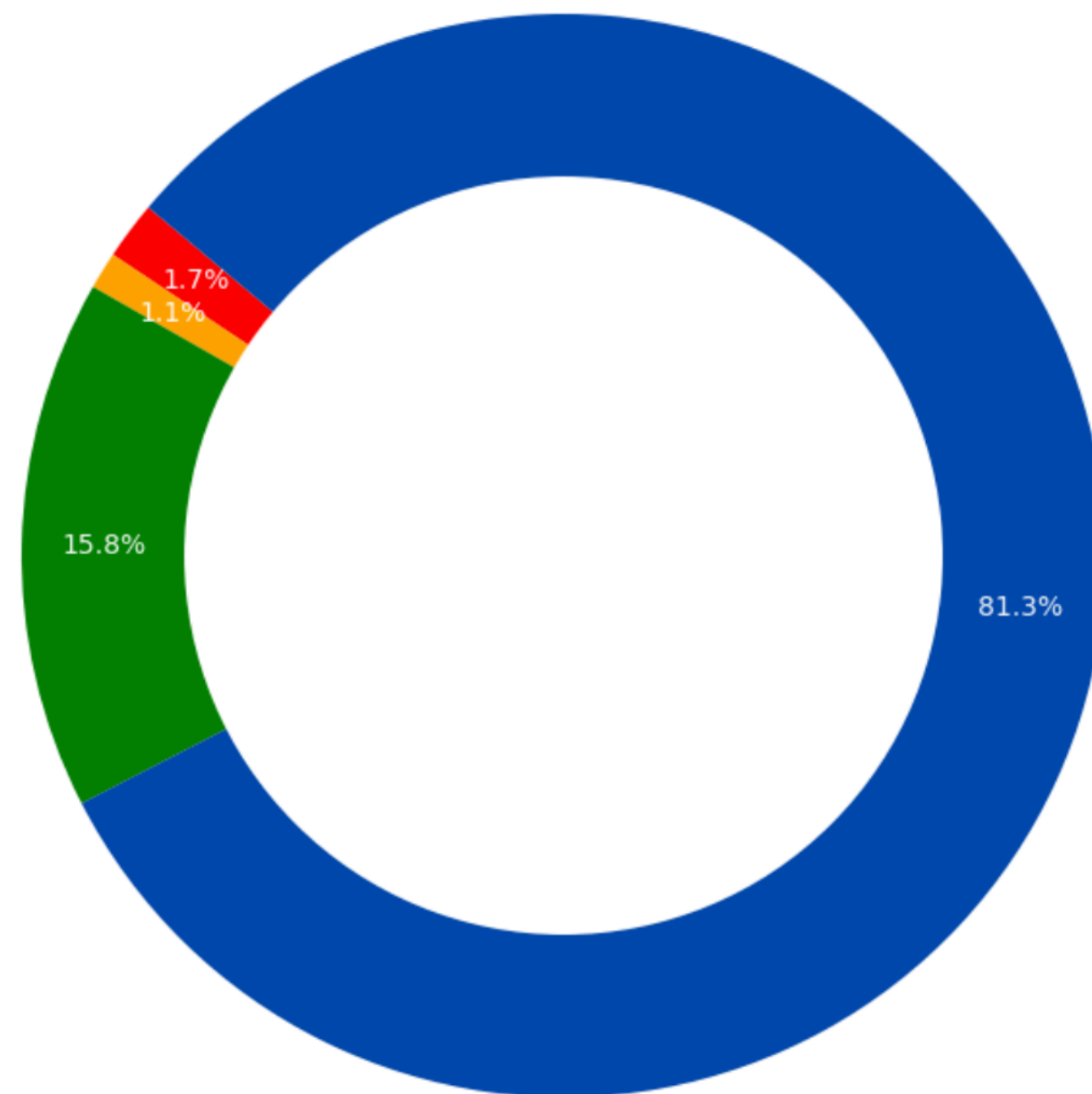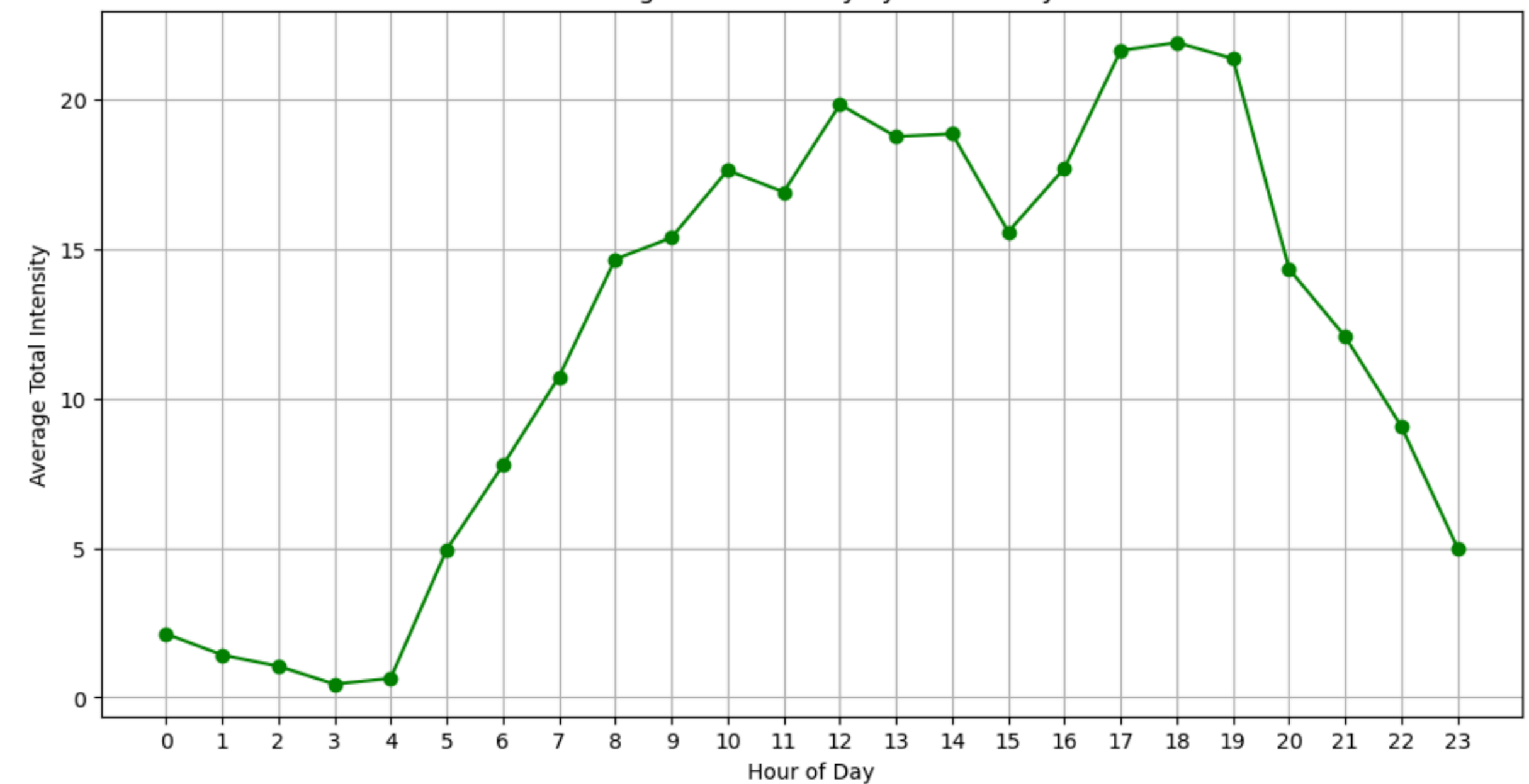- What's the intensity level among all users and how it varies within a day?

Distribution of Intensity Levels

Intensity Levels
- Very Active
- Fairly Active
- Lightly Active
- Sedentary

81.3%
15.8%
1.7%
1.1%

Average Total Intensity by Hour of Day

Q5

- Do the activities during the day affect individuals sleep quality? How?

	Id	ActivityDay	Calories
0	1503960366	4/12/2016	1985
1	1503960366	4/13/2016	1797
2	1503960366	4/14/2016	1776
3	1503960366	4/15/2016	1745
4	1503960366	4/16/2016	1863
...	...	...	...
935	8877689391	5/8/2016	
936	8877689391	5/9/2016	
937	8877689391	5/10/2016	
938	8877689391	5/11/2016	
939	8877689391	5/12/2016	

940 rows × 3 columns

	Id	ActivityDay	SedentaryMinutes	LightlyActiveMinutes	FairlyActiveMinutes	VeryActiveMinutes
0	1503960366	4/12/2016	728	328	13	25
1	1503960366	4/13/2016	776	217	19	21
2	1503960366	4/14/2016	1218	181	11	30
3	1503960366	4/15/2016	726	209	34	29
4	1503960366	4/16/2016	773	221	10	36
	...	...	...	...	...	...
	174			245	4	17
	131			217	19	73
	187			224	11	18
	127			213	12	88
	770			137	1	23

	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
...	...	...	...	...	...
408	8792009665	4/30/2016 12:00:00 AM	1	343	360
409	8792009665	5/1/2016 12:00:00 AM	1	503	527
410	8792009665	5/2/2016 12:00:00 AM	1	415	423
411	8792009665	5/3/2016 12:00:00 AM	1	516	545
412	8792009665	5/4/2016 12:00:00 AM	1	439	463

413 rows × 5 columns

Q5

The r-value is: 0.8657510739504785

y = 0.87x +22.12
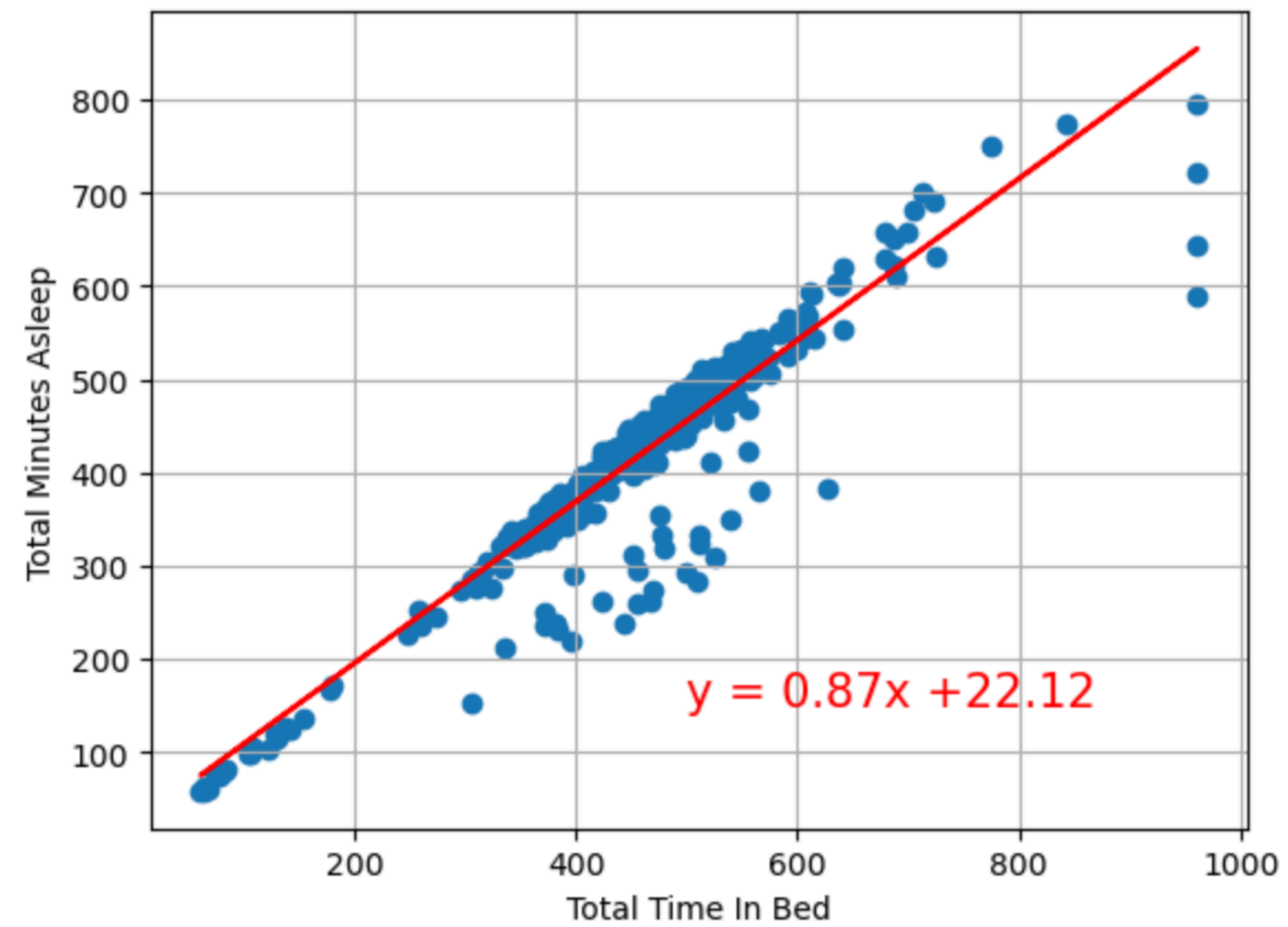
AwakeTimeInBed
19.0
23.0
30.0
27.0
12.0
...
17.0
24.0
8.0
29.0
24.0

Q5

```python
fitbit_127_drop.loc[(fitbit_127_drop['TotalMinutesAsleep'] >= 300) & (fitbit_127_drop['TotalMinutesAsleep'] <= 510)]
```

```python
fitbit_127_drop_filter.loc[(fitbit_127_drop_filter['AwakeTimeInBed'] >= 60) | (fitbit_127_drop_filter['AwakeTimeInBed'] <= 10)]
```

```python
awaketime_bins = [0, 10, 60, 227]
labels = ["<10min", "10-60min", ">60min"]
```

```python
df_cal = fitbit_127_drop_filter_awake_combined_copy.groupby(["Awake Time In Bed"])["Calories"].median()
df_very = fitbit_127_drop_filter_awake_combined_copy.groupby(["Awake Time In Bed"])["VeryActiveMinutes"].median()
df_fairly = fitbit_127_drop_filter_awake_combined_copy.groupby(["Awake Time In Bed"])["FairlyActiveMinutes"].median()
df_lightly = fitbit_127_drop_filter_awake_combined_copy.groupby(["Awake Time In Bed"])["LightlyActiveMinutes"].median()
df_seden = fitbit_127_drop_filter_awake_combined_copy.groupby(["Awake Time In Bed"])["SedentaryMinutes"].median()
```

Q5

	Calories (median)	VeryActiveMinutes (median)	FairlyActiveMinutes (median)	LightlyActiveMinutes (median)	SedentaryMinutes (median)
Awake Time In Bed					
<10min	2731.5	34.0	14.0	215.0	736.0
>60min	1624.0	11.0	22.0	191.5	676.0

Summary

- 1. Importance of staying active.

- 2. The resulting plot visually represent the average total steps taken on each day of the week, showing that Saturday and Tuesday have the highest total steps across different days.

- 3. The highest peak activity level is between 6:00pm - 7:00pm.

- 4. The pattern in question 4 can provide insights for designing interventions to increase daily physical activity, encouraging users to incorporate more moderate to vigorous activities into their daily routines could help balance the high proportion of sedentary and light activities. For example, targeted reminders or fitness programs during peak low-activity times.

- 5. The more calories burnt during the day, the less time you will take to fall asleep. If you have sleep issues, we would recommend users to try be more active during the daytime.

Thank you!