Introduction

The Bellabeat data analysis project focuses on analyzing user data to gain insights and inform marketing strategies for Bellabeat, a high-tech manufacturer of health-focused products for women. Through my analysis, I will extensively explore relationships and patterns within historical health tech data. My aim is to use my findings, along with research papers and industry reports, to develop a new data-backed marketing strategy for Bellabeat's stakeholders.

Documentation

SQL Cleaning & Manipulation Markdown
R Manipulation, Analysis & Visualizations Markdown
Correlation & Regression Markdown

Data Source

- The dataset is sourced from Kaggle and contains personal tracker data from 30 Fitbit users who consented to share their data.
- The data includes minute-level output for physical activity, heart rate, and sleep monitoring.
- The dataset spans from April 12, 2016, to May 12, 2016, providing a single month of data.
- For this project I used 10 out of 18 CSV files.

Data Storage

- The datasets have been stored locally in a folder named bellabeat_export_4.12.16-5.12.16 & duplicated for retention of the original data prior to cleaning and manipulation.
- Files were uploaded to BigQuery with new table names recorded within documentation.

Challenges

Data Limitations and Considerations

The dataset used in this analysis spans from April 12, 2016, to May 12, 2016, providing only one month of data. This limited timeframe may restrict the scope and generalizability of the findings. Additionally, the 30 user sample size is small, and the methods of participant selection through Amazon Mechanical Turk are not detailed on Kaggle, which introduces potential bias into the analysis. It is also important to note that the data is from 2016 and may not reflect the most recent trends or user behaviors. Furthermore, the lack of demographic information limits the depth and context of the insights that can be derived from this analysis.

Consistency

Although the Fitbit Fitness Tracker Dataset description claims that thirty eligible Fitbit users consented to the use of their data, 33 distinct IDs were found within the records during exploration.

The weightlog table was the most explicit example of this dataset's challenges and was subsequently dropped:

1. Incomplete data: Only 8 unique IDs, which may not provide a comprehensive view of user behavior and patterns.
2. Missing values: High number of null values in the Fat column, suggesting unreliable data capture or reporting.
3. Relevance: Weight data was not considered a major variable for the analysis.

Datetime Formatting

While uploading the CSVs into BigQuery there was an error uploading hourlycalories, hourlyintensities, sleepday, weightlog due to issues with datetime formatting. This was resolved within google sheets. I renamed each changed file with v2.

Key Findings

- There is a strong positive correlation between total steps and calories burned for both weekdays and weekends. As total steps increase, calories burned also tends to increase.
- There is a strong positive correlation between very active minutes and calories burned for both weekdays and weekends. As very active minutes increase, calories burned also tends to increase.
- There is a significant negative correlation between sedentary minutes and total time in bed for both weekdays and weekends. As total time in bed increases, sedentary minutes tend to decrease.
- On average, people exercise with the greatest intensity during afternoon workouts, followed by mornings, with the lowest intensity in the evenings.
- On average, people take more steps during the afternoon, followed by mornings, with the lowest step count in the evenings.

Target Market

Bellabeat's target market comprises health-conscious, tech-savvy women who prioritize a holistic approach to wellness. They value integrating health into their daily lives, aligning their routines with their menstrual cycles, and being part of a supportive community. These women appreciate innovative, data-driven solutions and stylish, functional products, spanning a diverse age range from young adults to older women.

Recommendations

- Leverage the strong positive correlation between steps and calories burned by incorporating gamification elements into Bellabeat's apps and products. Develop personalized step challenges and rewards that align with each user's fitness level and goals, taking into account their cycle phases and other biometric data. This approach can motivate users to increase their physical activity and improve overall health outcomes while catering to the unique needs of women.

- Implementing a social network element to bellabeats app could leverage our target market, social connections and encourage current users to invite family and friends, possibly creating new memberships.

- A social hub could be a place to manage promotional events, challenges and possible rewards via partnerships. According to "Incentive-Based Interventions for Health-Related Behaviors" Implementing finely crafted financial or non-financial incentives based on users' performance can further motivate behavior change. For example, providing discounts or coupons for reaching activity milestones could be an effective strategy.

- Create dynamic challenges based on continuous data analysis. For example, since the case study revealed that the highest exercise intensity is in the afternoon, a challenge involving early morning activities could be introduced for a week, offering a new and engaging challenge for the community.

- Conduct a larger-scale study with a diverse, representative sample of Bellabeat's target audience to gain deeper insights into their needs, preferences, and behaviors. Use this data to refine product features, marketing strategies, and content offerings.

Tables

Summary Statistics Tables

Steps summary:

Min Steps	Max Steps	Mean Steps	Median Steps
0	36019	7637.910638	7405.5

Distance summary:

Min Distance	Max Distance	Mean Distance	Median Distance
0	28.03000069	5.489702122	5.244999886

Average Steps by Time of Day

Day Period	Avg Steps
Afternoon	508.4705083
Morning	380.5684588
Night	239.9626947

Average activity by day type

DayType	Avg Steps	Avg Calories	Avg Distance
Weekday	7668.699281	2301.516547	5.505107907
Weekend	7550.571429	2309.546939	5.445999997

Visualizations

General:

- There is a strong statistically significant positive correlation between Total Steps and Calories Burned for both weekdays and weekends. As Total Steps increase, Calories Burned also tends to increase.
- The data points form a dense cluster, suggesting that the relationship between Total Steps and Calories Burned is relatively consistent across users.

Weekdays:

- The weekday data points show a steady increase in Calories Burned as Total Steps increase, with a few high-calorie outliers at various step counts.

Weekends:

- Weekend data points (blue) are more spread out and extend further to the right on the Total Steps axis, indicating that users generally have higher step counts on weekends.

General:

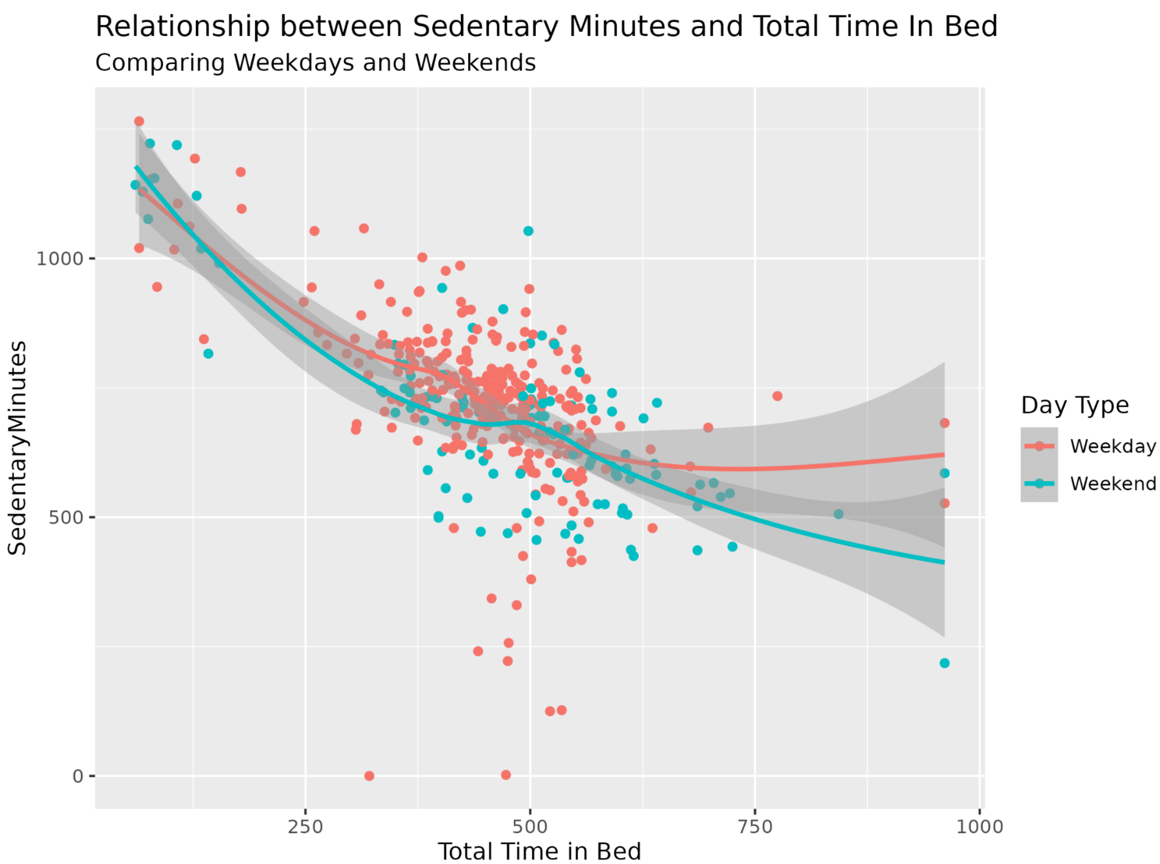
- There is a statistically significant positive correlation between Very Active Minutes and Calories Burned for both weekdays and weekends across multiple users. As Very Active Minutes increase, Calories Burned also tends to increase.

Weekdays:

- For a given number of Very Active Minutes, users tend to burn slightly more Calories on weekdays compared to weekends.

Weekends:

- Users generally have more Very Active Minutes on weekends compared to weekdays, as the weekend data points extend further to the right on the x-axis.
- The weekend data points show a steeper increase in Calories Burned as Very Active Minutes increase, suggesting that users may engage in more high-intensity activities or have longer durations of active minutes on weekends.



General:

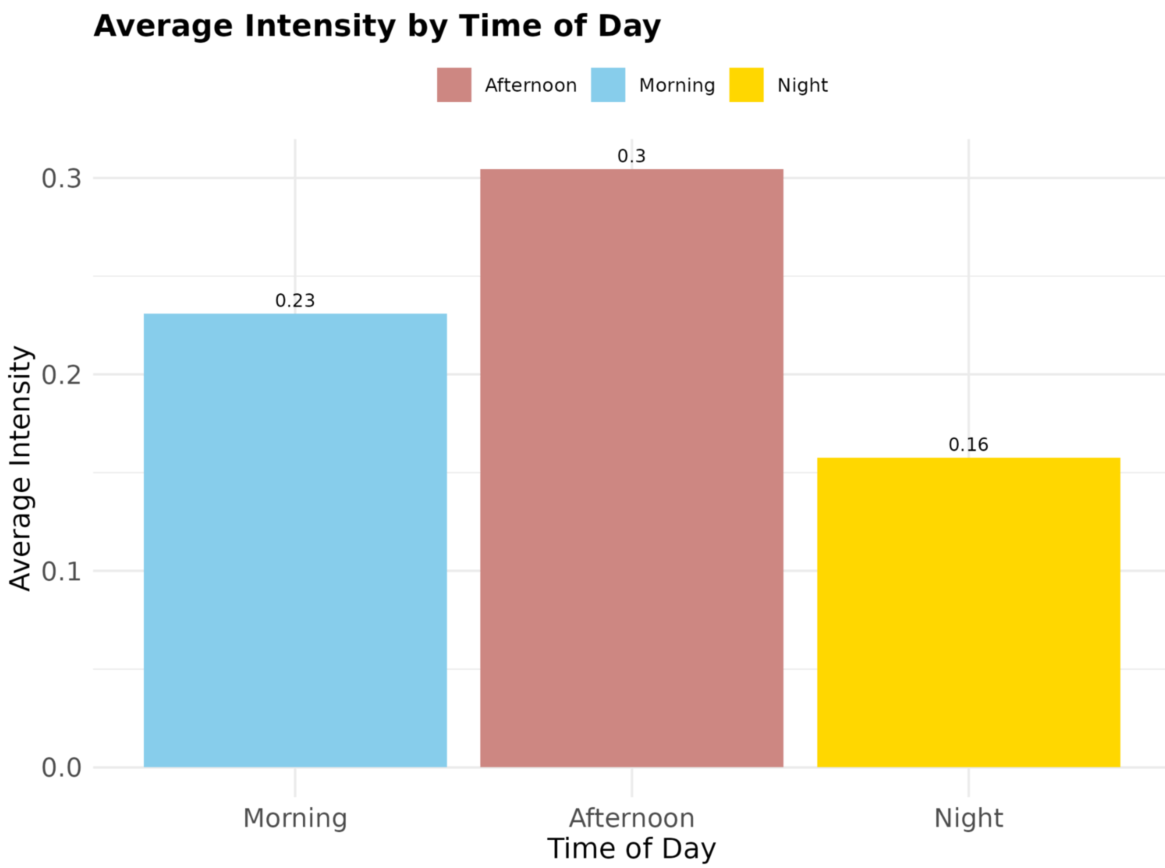
- There is a statistically significant negative correlation between Sedentary Minutes and Total Time in Bed for both weekdays and weekends. As Total Time in Bed increases, Sedentary Minutes tend to decrease.
- The data points are concentrated in the middle of the plot, suggesting that most users have a similar range of Total Time in Bed and Sedentary Minutes.

Weekdays:

- Weekday data points (pink) are more tightly clustered compared to weekend data points, indicating less variability in the relationship between Sedentary Minutes and Total Time in Bed on weekdays.
- For a given Total Time in Bed, weekday data points tend to have higher Sedentary Minutes compared to weekend data points.

Weekends:

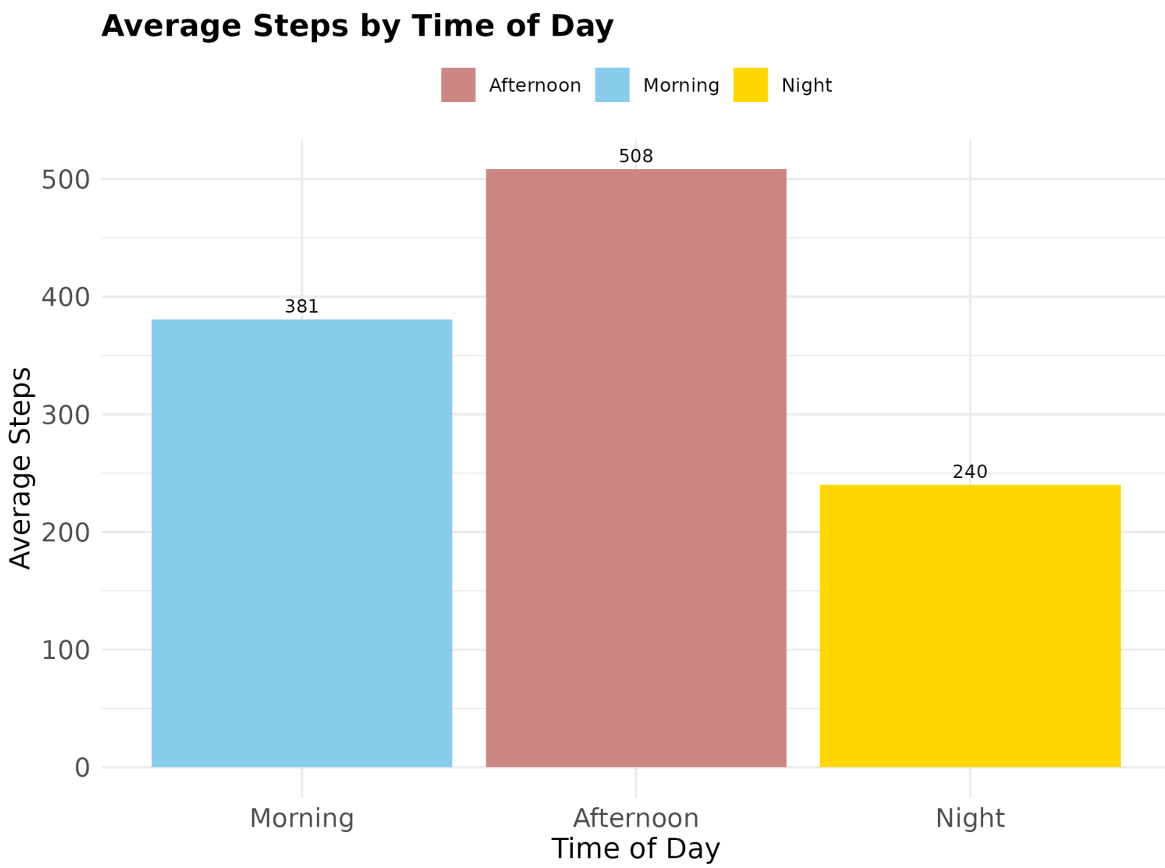
- For a given Total Time in Bed, weekend data points tend to have lower Sedentary Minutes compared to weekday data points.



General:

This data suggests that on average, people are exercising with the greatest intensity during afternoon workouts, followed by mornings, with the lowest intensity in the evening.

Morning: .23 **Afternoon:** .3 **Night:** .16



General:
This data suggests that on average, people are taking more steps during the afternoon, followed by mornings, with the lowest count in the evening.
Morning: 381 **Afternoon:** 508 **Night:** 240

Relevant Studies

"Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps" (BMJ Open)

Summary:
This systematic review evaluates the effectiveness of gamification elements in smartphone health apps for promoting healthy behaviors. It identifies key gamification strategies such as badges, leaderboards, competitions, rewards, and avatars that have been shown to motivate users and sustain engagement. The study emphasizes the use of behavior change techniques like goal setting, feedback, and social connectivity, which align with established health behavior change theories.

“Changing health behaviors using financial incentives: a review from behavioral economics" (BMC Public Health)

Summary:
This study reviews the effectiveness of financial incentive schemes in promoting health-related behaviors. It emphasizes the importance of well-designed incentive programs, highlighting behavioral economics principles like loss aversion and immediate rewards. Key findings include that incentives tied to immediate outcomes and framed as losses can be more motivating than those framed as gains.

Citations:

- Study: "Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps." *BMJ Open*, 2023. Available at: BMJ Open.
- Study: "Changing health behaviors using financial incentives: a review from behavioral economics." *BMC Public Health*, 2023. Available at: BMC Public Health.

User Average Activity

Id	AvgDailySteps	AvgDailyCalories	AvgDailyActiveMinutes
8877689391	16040.03226	3420.258065	310.7096774
8053475328	14763.29032	2945.806452	245.7096774
1503960366	12116.74194	1816.419355	277.8064516
2022484408	11370.64516	2509.967742	313.0967742
7007744171	11323.42308	2544	328.0384615
3977333714	10984.56667	1513.666667	254.9333333
4388161847	10813.93548	3093.870968	272.8709677
6962181067	9794.806452	1982.032258	287.1290323
2347167796	9519.666667	2043.444444	286.5555556
7086361926	9371.774194	2566.354839	211.7741935
8378563200	8717.709677	3436.580645	225.0322581
5553957443	8612.580645	1875.677419	242.6129032
4702921684	8572.064516	2965.548387	268.6451613
5577150313	8304.433333	3359.633333	265.1
4558609924	7685.129032	2033.258065	309.0645161
2873212765	7555.774194	1916.967742	328.2258065
1644430081	7282.966667	2811.3	209.4
4319703577	7268.83871	2037.677419	244.6774194
8583815059	7198.516129	2732.032258	170.1612903
6117666160	7046.714286	2261.142857	291.9642857
3372868164	6861.65	1933.1	341.15
8253242879	6482.157895	1788	151.7368421
1624580081	5743.903226	1483.354839	167.9677419
6290855005	5649.551724	2599.62069	234
2026352035	5566.870968	1540.645161	257
4445114986	4796.548387	2186.193548	217.4516129
2320127002	4716.870968	1724.16129	202.1290323
4057192912	3838	1973.75	105.25
1844505072	2580.064516	1573.483871	116.8709677
6775888955	2519.692308	2131.769231	65.96153846
4020332650	2267.225806	2385.806452	87.48387097
8792009665	1853.724138	1962.310345	96.79310345
1927972279	916.1290323	2172.806452	40.67741935

Statistical Analysis

Correlation & Regression RMD

Statistical significance should be interpreted cautiously given the small sample size of 30 users. I recommend replicating the findings with larger, more representative samples to increase confidence in this case studies findings.

Relationship between Very Active Minutes and Calories Burned:

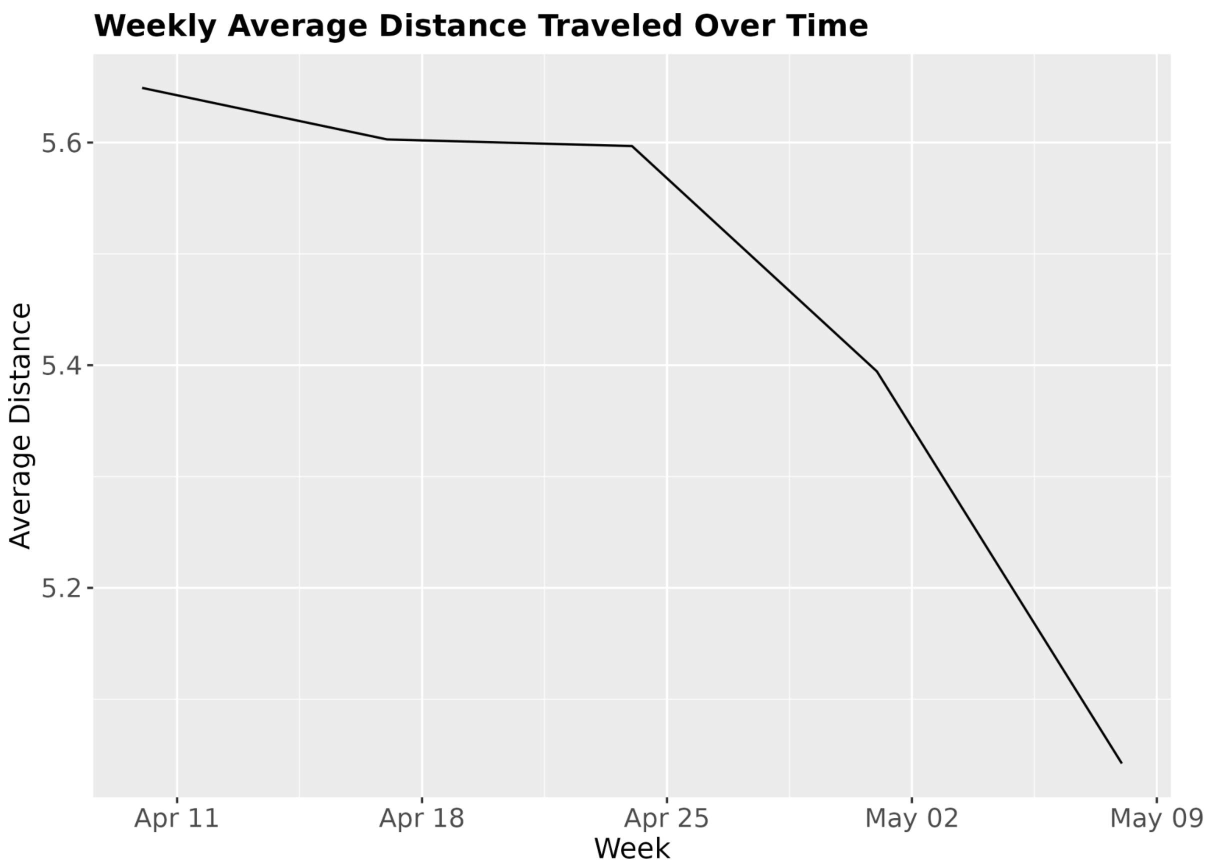
- The correlation coefficient (0.6111983) indicates **a strong positive correlation between Very Active Minutes and Calories Burned.**
- The p-value (< 2.2e-16) is highly significant, suggesting that the correlation is **statistically significant.**
- The linear regression model shows a significant positive relationship between Very Active Minutes and Calories Burned (coefficient estimate: 12.7989, p-value: < 2e-16).
- The adjusted R-squared (0.372) indicates that 37.2% of the variance in Calories Burned can be explained by Very Active Minutes.

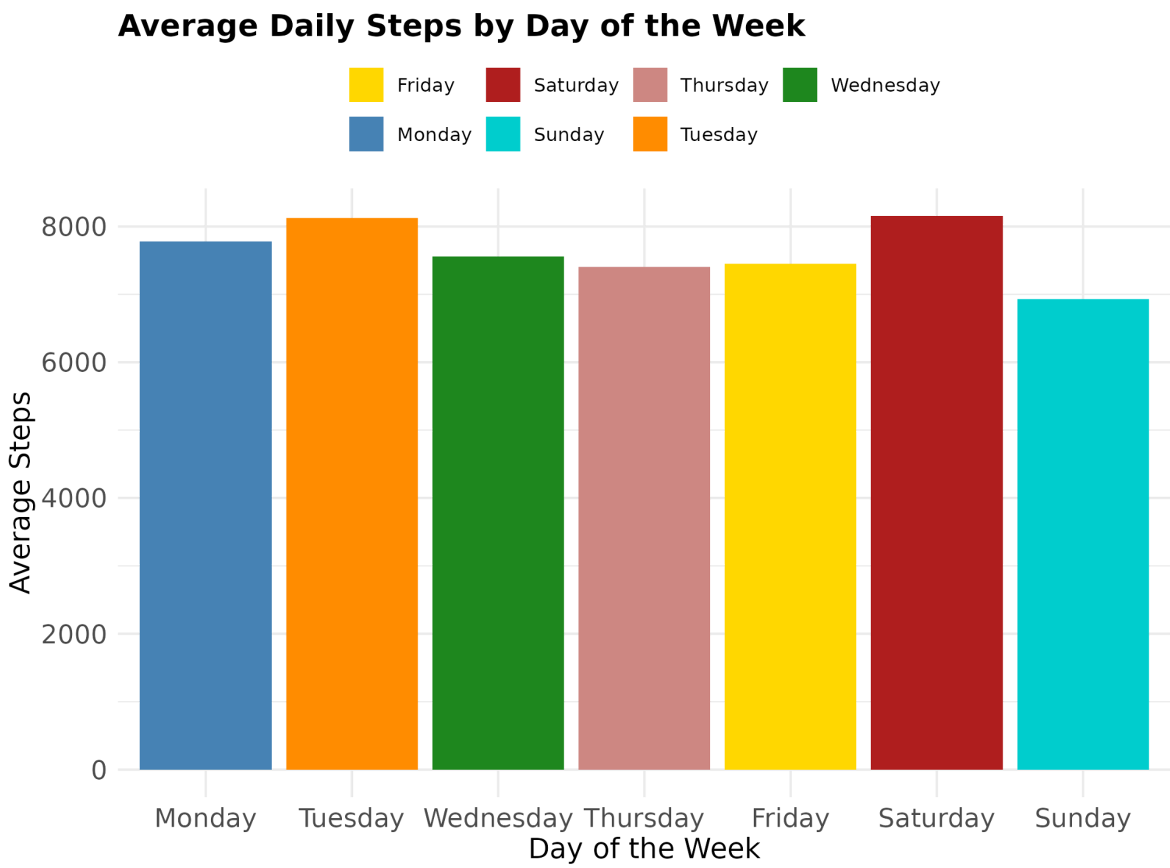
Relationship between Total Steps and Calories Burned:

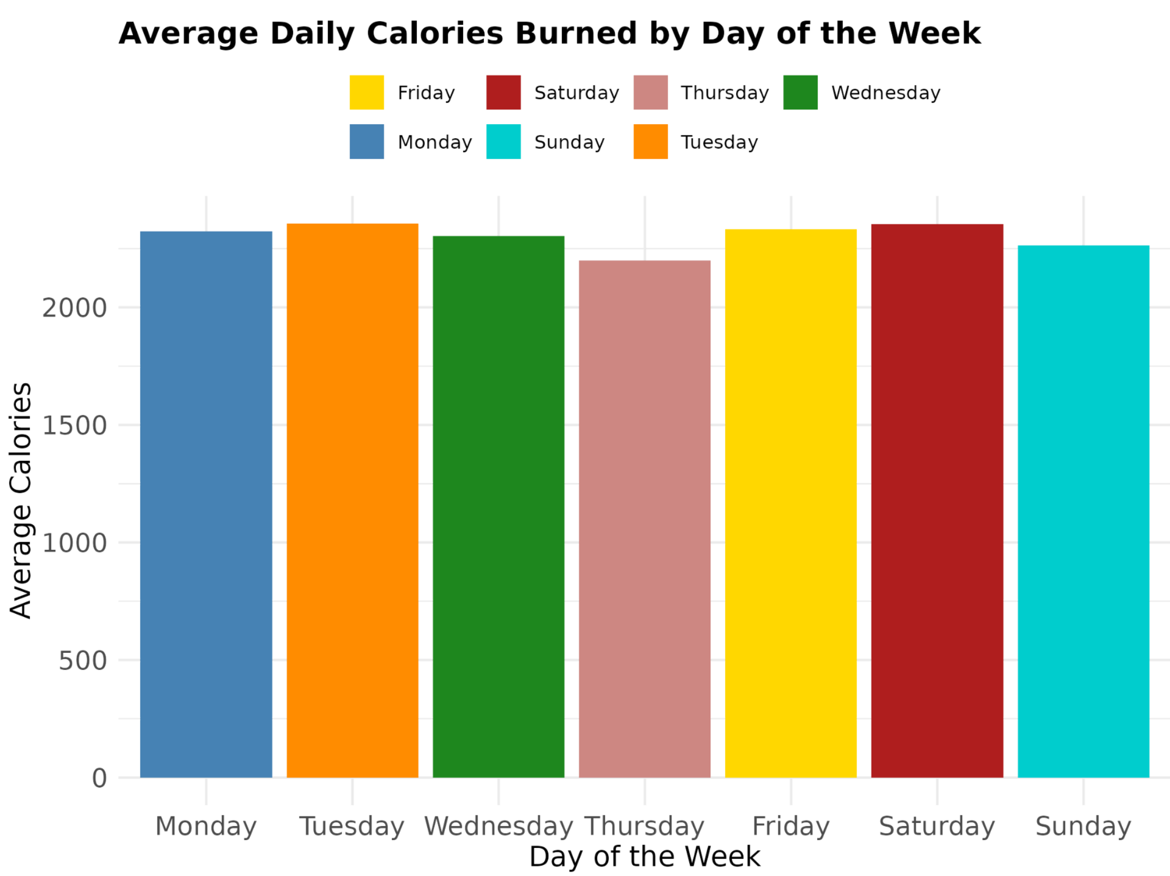
- The correlation coefficient (0.4063007) indicates a moderate positive correlation between Total Steps and Calories Burned.
- The p-value (< 2.2e-16) is highly significant, suggesting that the correlation is **statistically significant.**
- The linear regression model shows a significant positive relationship between Total Steps and Calories Burned (coefficient estimate: 0.07412, p-value: < 2e-16).
- The adjusted R-squared (0.163) indicates that 16.3% of the variance in Calories Burned can be explained by Total Steps.

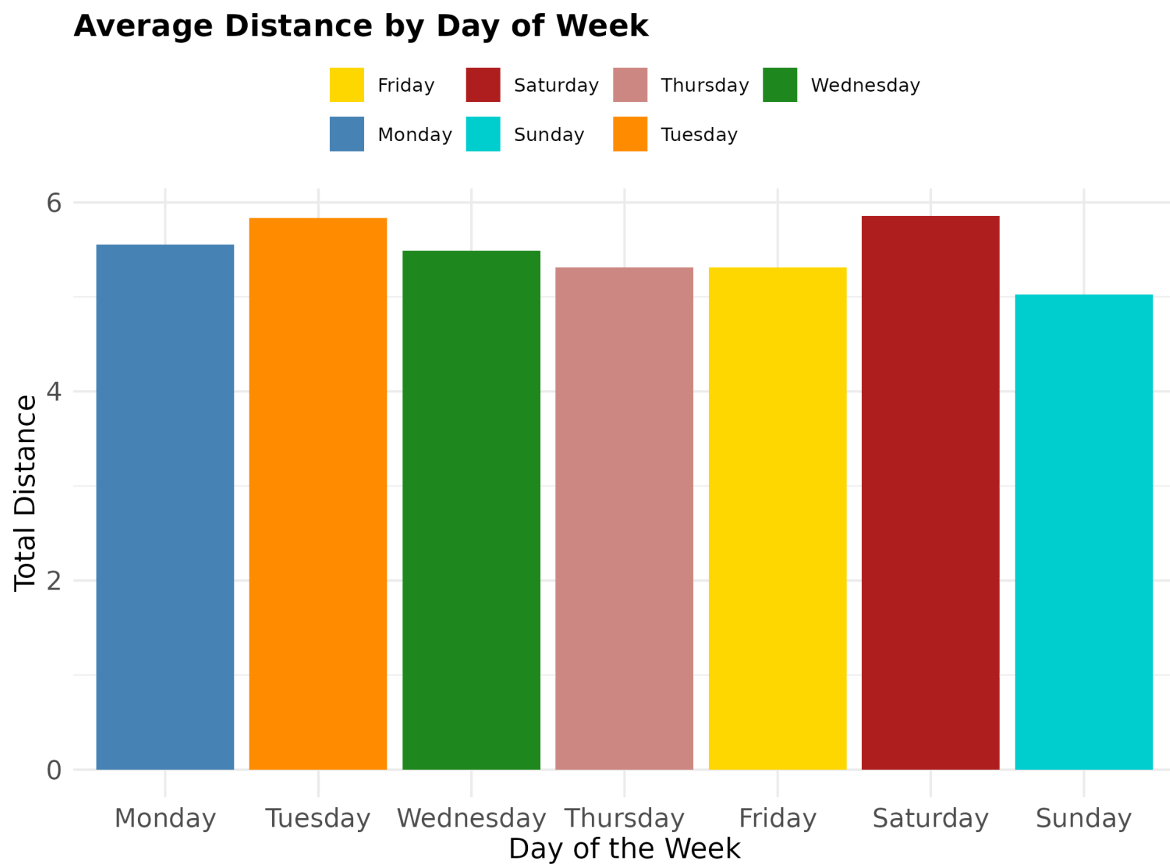
Relationship between Sedentary Minutes and Total Time in Bed:

- The correlation coefficient (-0.6202804) indicates a **strong negative correlation between Sedentary Minutes and Total Time in Bed.**
- The p-value (< 2.2e-16) is highly significant, suggesting that the correlation is **statistically significant.**
- The linear regression model shows a significant negative relationship between Sedentary Minutes and Total Time in Bed (coefficient estimate: -0.47574, p-value: < 2e-16).
- The adjusted R-squared (0.3832) indicates that 38.32% of the variance in Total Time in Bed can be explained by Sedentary Minutes.

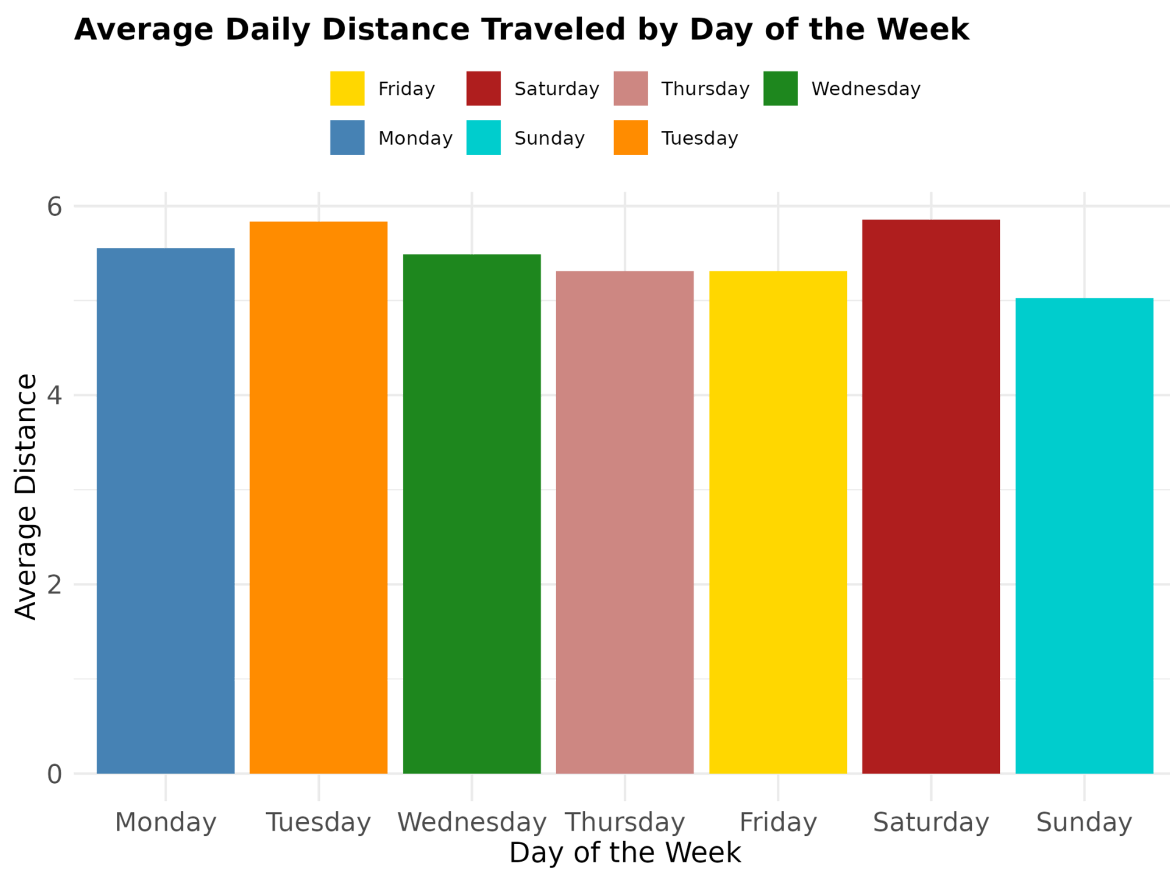

Weekly Average Distance Traveled Over Time
Average Distance
5.6
5.4
5.2
Apr 11
Apr 18
Apr 25
May 02
May 09
Week


Average Daily Steps by Day of the Week
Friday
Saturday
Thursday
Wednesday
Monday
Sunday
Tuesday
Average Steps
8000
6000
4000
2000
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week

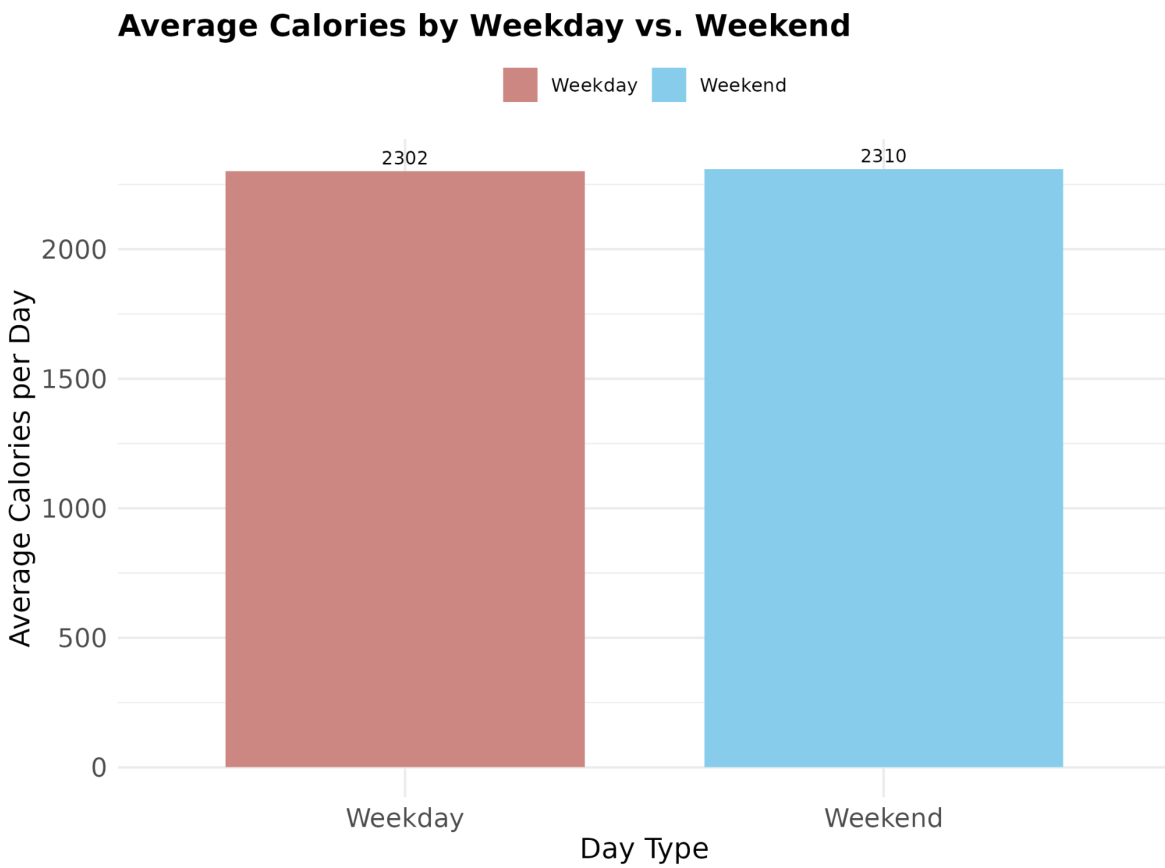

Average Daily Calories Burned by Day of the Week
Friday
Saturday
Thursday
Wednesday
Monday
Sunday
Tuesday
Average Calories
2000
1500
1000
500
0
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week

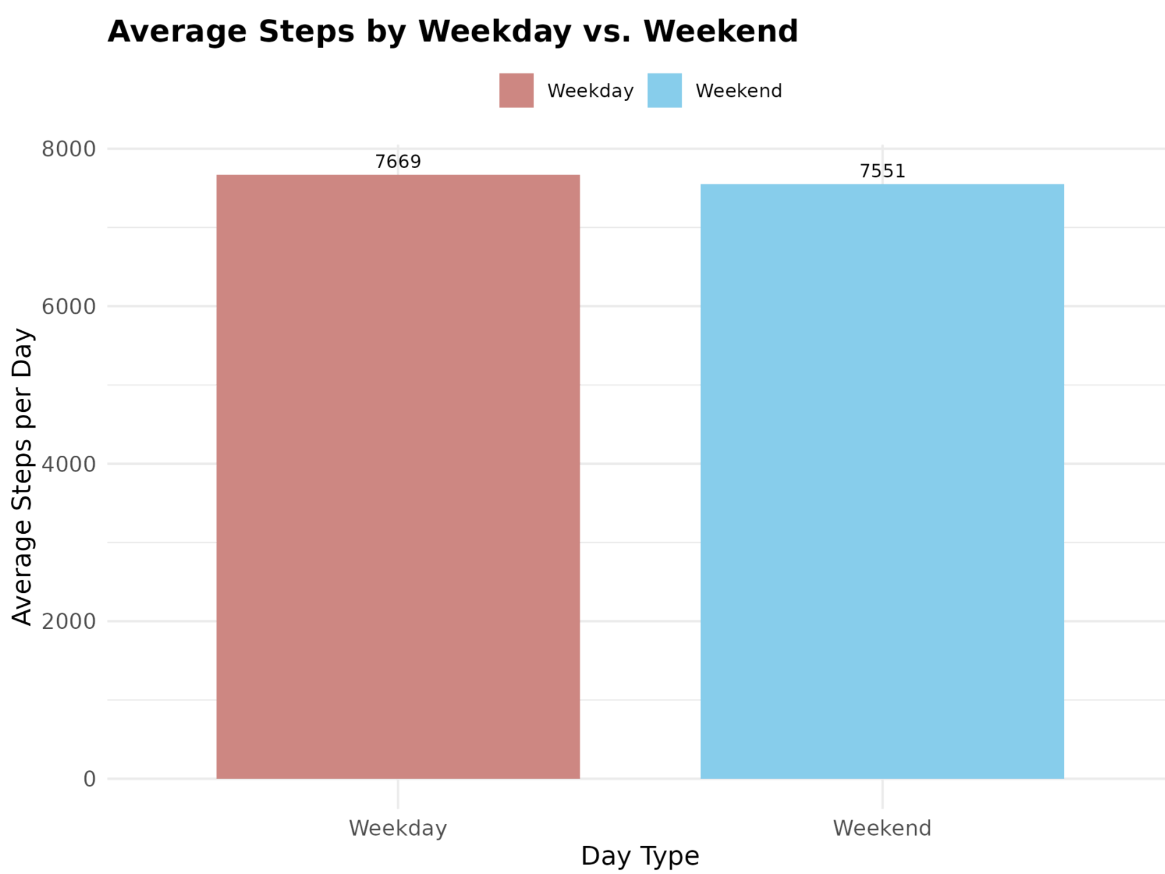


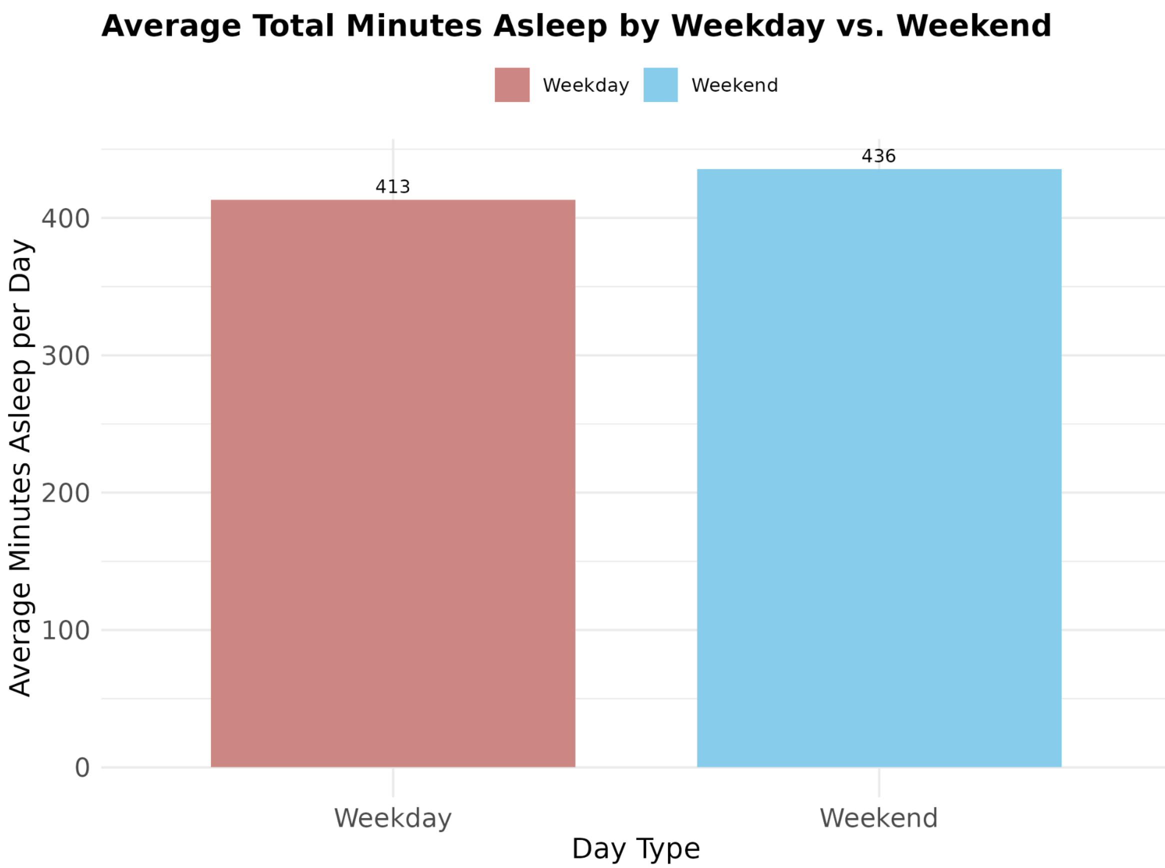
Average Distance by Day of Week
Friday
Saturday
Thursday
Wednesday
Monday
Sunday
Tuesday
6
4
2
0
Total Distance
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week

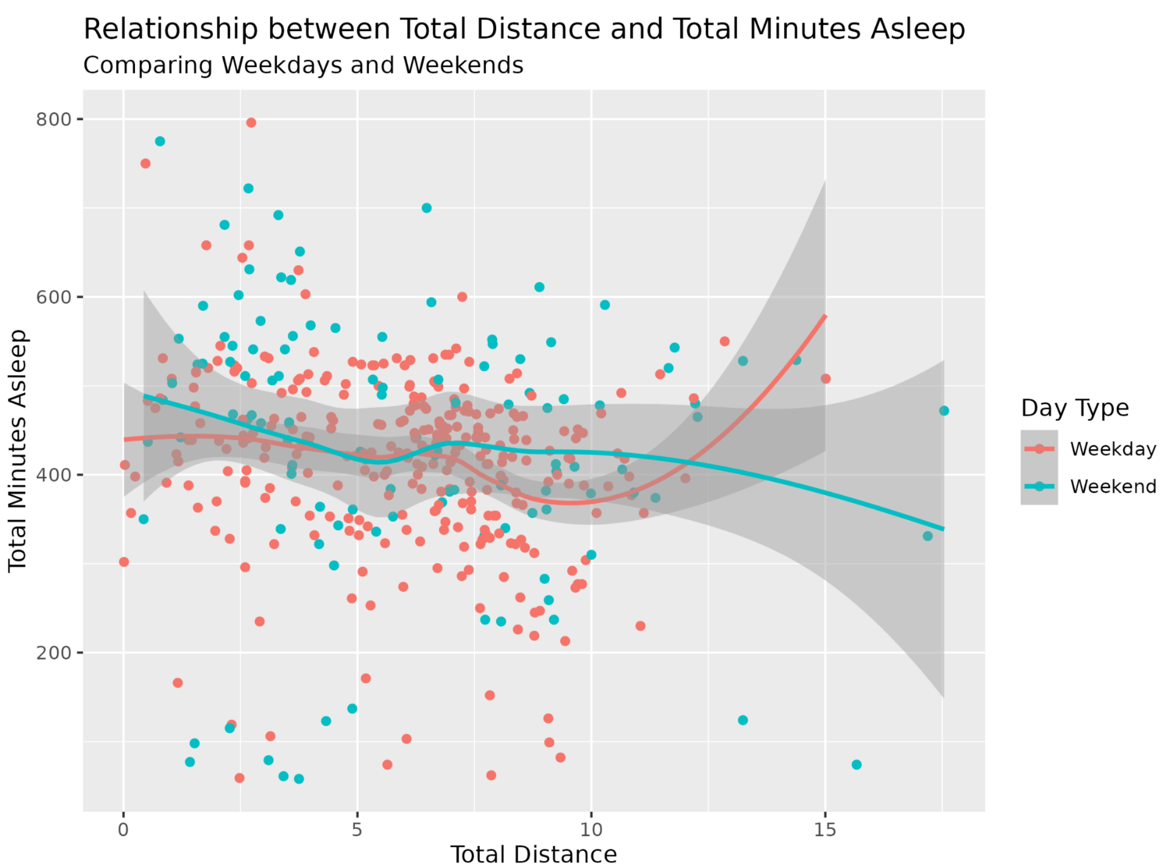


Average Daily Distance Traveled by Day of the Week
Friday
Saturday
Thursday
Wednesday
Monday
Sunday
Tuesday
6
4
2
0
Average Distance
Monday
Tuesday
Wednesday
Thursday
Friday
Saturday
Sunday
Day of the Week


Average Calories by Weekday vs. Weekend
Weekday
Weekend
2302
2310
2000
1500
1000
500
0
Average Calories per Day
Weekday
Weekend
Day Type


Average Steps by Weekday vs. Weekend
Weekday
Weekend
8000
7669
7551
6000
4000
2000
0
Average Steps per Day
Weekday
Weekend
Day Type


Average Total Minutes Asleep by Weekday vs. Weekend
Weekday
Weekend
413
436
Average Minutes Asleep per Day
0
100
200
300
400
Weekday
Weekend
Day Type


Relationship between Total Distance and Total Minutes Asleep
Comparing Weekdays and Weekends
Total Minutes Asleep
800
600
400
200
0
5
10
15
Total Distance
Day Type
Weekday
Weekend