

GOOGLE DATA ANALYTICS CAPSTONE PROJECT

Fadhil Umar Al Farouq
10/10/2023



TRACK 1.

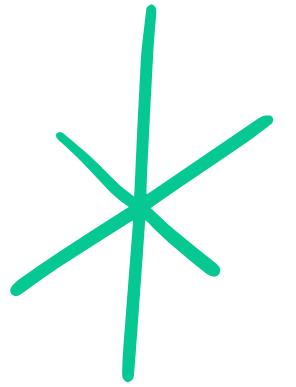
CASE STUDY 2

**How Can a Wellness Technology Company
Play It Smart?**

A high-tech manufacturer of health-focused
smart products for women.

In this case study, we consider the company
Bellabeat.

Company website : <https://bellabeat.com/>



FRAMEWORK OVERVIEW

1. Ask

- Understand the problem
- Ask the right question

2. Prepare

- Collecting Data
- Verify Data

3. Process

- Create clean data
- Recheck data

4. Analyze

- Identify trends and relationships
- Organize and format data

5. Share

- Create Visualization
- Present findings

6. Act

- Share next steps
- Evaluate

1. ASK

Our objective is to understand **consumer behaviour** to improve **marketing strategy**. In this particular case, we consider a **smart device data** as our main source of information. We want to know if the data can help gain insight on how costumer use non-Bellabeat smart devices.

Some questions we want to answer :

1. What are some **trends** in smart device usage?
2. Are there other external **behaviour** or **bias**?
3. Are there some hidden **relationships** we can utilize?
4. How to **use** this information to Bellabeat customers?
5. How could we **influence** marketing strategy using this information?

We can then generate problem statement as

Capturing customer behaviour of non-Bellabeat smart devices and create a marketing strategy to increase sales and customer satisfaction



2. PREPARE

To realize our problem statement, we consider the [FitBit Fitness Tracker Data](#). Which is a **public dataset** on minute-level FitBit user data.

Facts about the data :

1. The data comprises of **thirty consented fitbit users**.
2. Data tracked is on **minute-level** scale and includes physical activity, heart rate, and sleep monitoring.
3. the data have been rated **10.00** in usability in *Kaggle* platform. The score is based on completeness, credibility, and compatibility. Assuring the data abides to **ROCCC**.
4. The data comprises of **18 different files** for each measurement.
5. The data does not give identity about **customer gender** which might not reflect Bellabeat customer distribution.
6. The **metadata** of the dataset is not provided, but a [Kaggle discussion](#) finds a reliable metadata file for the dataset through the fitabase archive.
7. We will import the data to **Spreadsheets** to gain initial analysis.

3. PROCESS

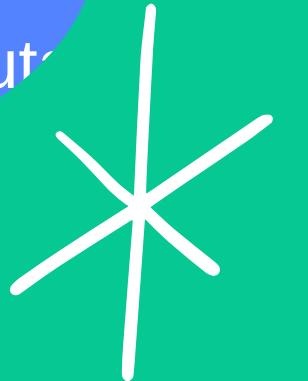
After Importing each files into spreadsheet we can conclude some things :

- The data are taken from 4/12/2016 to 5/12/2016 which means **31 days** for each user.
- After verifying, the total **number of users are 33**.
- The data have a minute, hourly, and daily level of measurement. We will only use daily and hourly data to make it **consistent**.

Data we will use includes :

1. **DailyActivity** data for daily trends
2. **HourlyCalories**, **HourlyIntensities**, and **HourlySteps** for daily patterns
3. **sleepDay** data for sleep data
4. **weightLog** to identify relationship with weight measurement

We will use R as data exploration tool and reporting. R is known for its reputation in statistical computing and high quality graphics.



3. PROCESS

Cleaning Log :

1. **Import** Data.
2. **Merge** daily data.
3. **Format date** as date and separate datetime into two columns, date and time.
4. **Add** dayname and intensity calculations.
5. **Merge** daily data.
6. **Clean** weight log outliers.

For more log detail refer to the [notebook](#)

4. ANALYZE

Analysis are done using R and can be found in this [link](#).

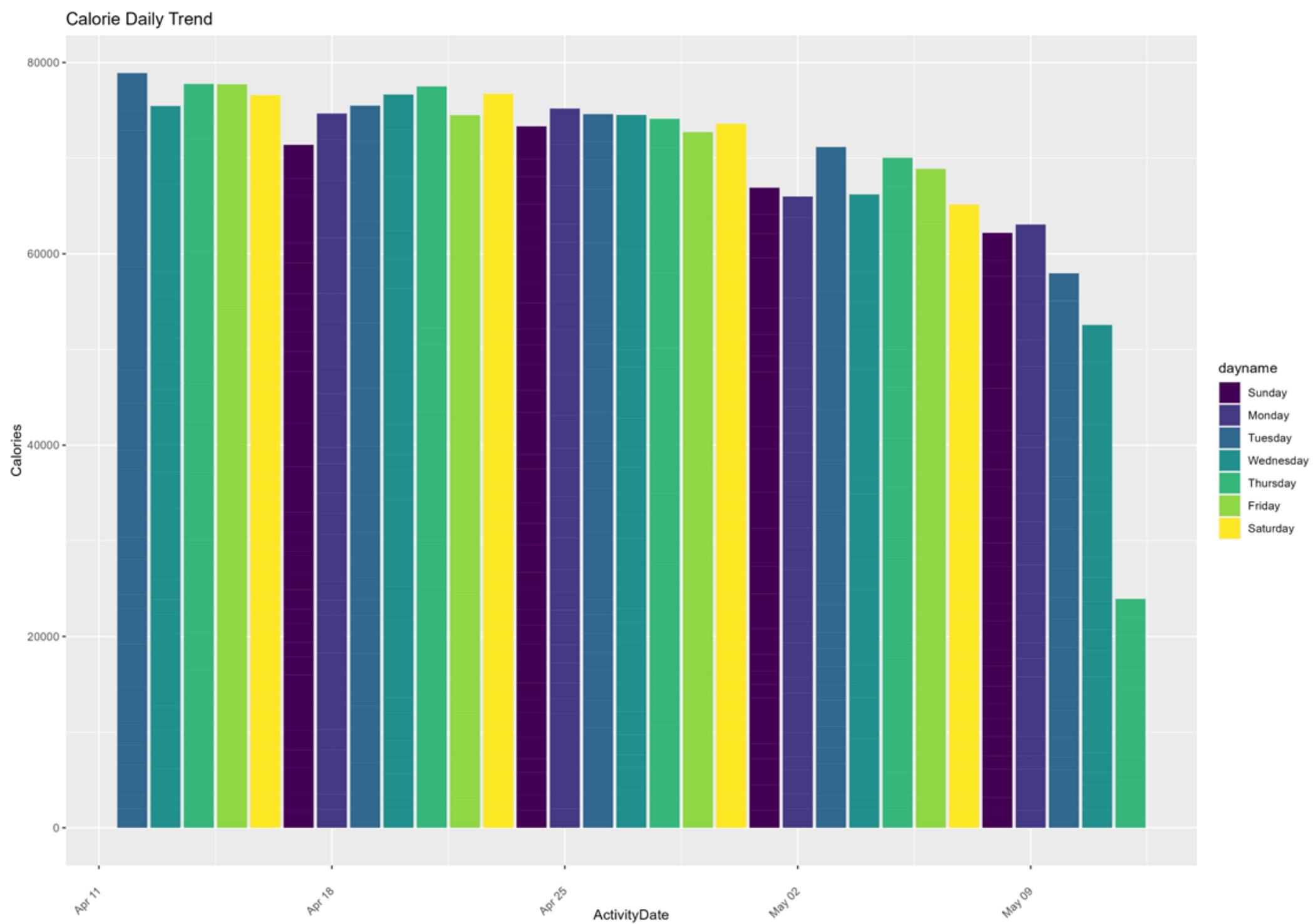
Some notable findings are :

- 1. Temporal Scope:** The dataset spans a period of **31 days**, from **4/12/2016** to **5/12/2016**, offering a comprehensive snapshot of user activity during this timeframe.
- 2. User Demographics:** It's established that there are **33 distinct users** contributing to this dataset. Notably, gender information is absent. Assuming a uniform distribution of genders might slightly skew interpretations, given Bellabeat's female-centric focus.



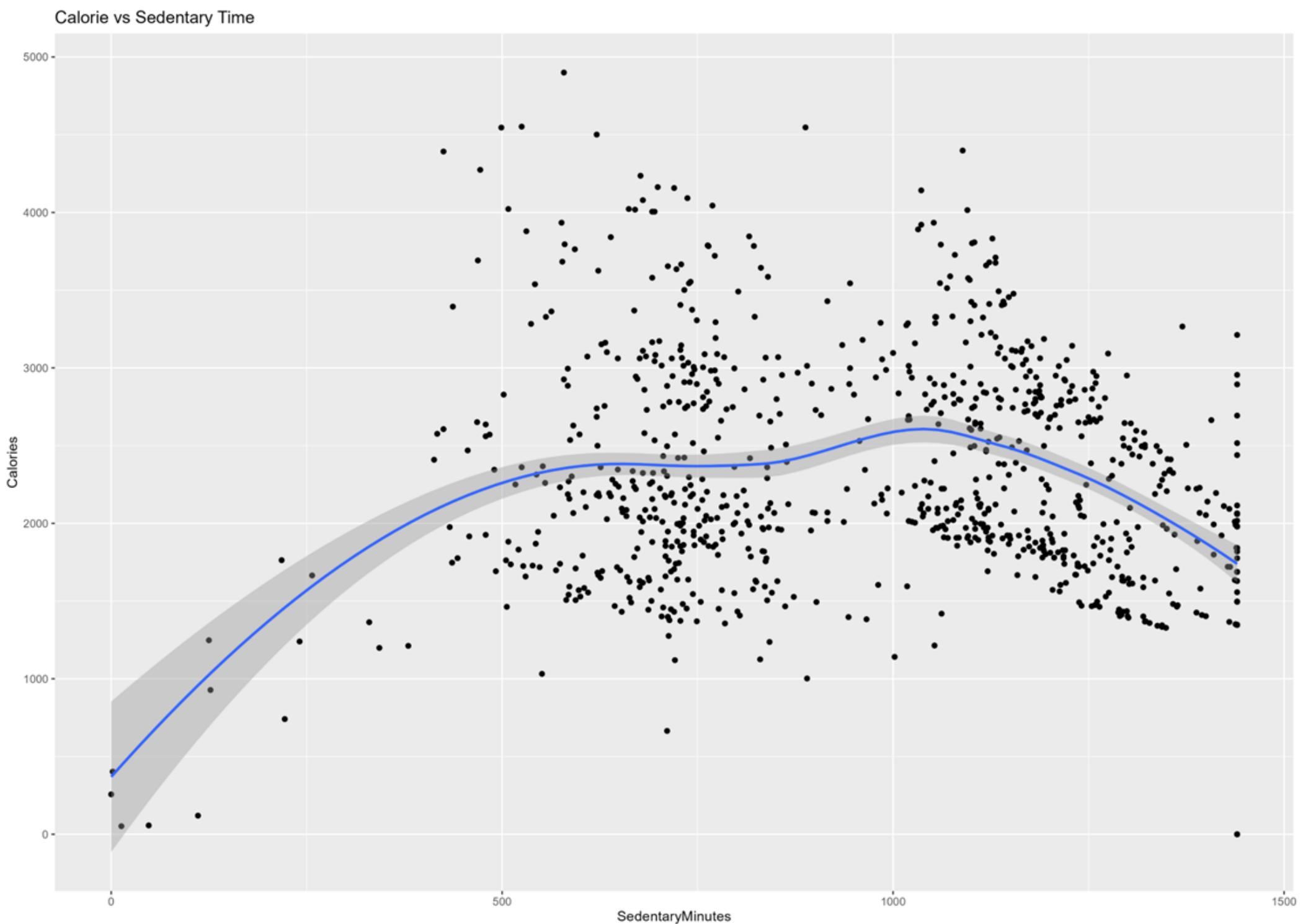
4. ANALYZE

3. Caloric Trends : A noticeable trend reveals a gradual decrease in daily calorie expenditure over the observed period. Additionally, Sundays stand out with lower calorie burn compared to weekdays, suggesting potential variations in activity routines during weekends.



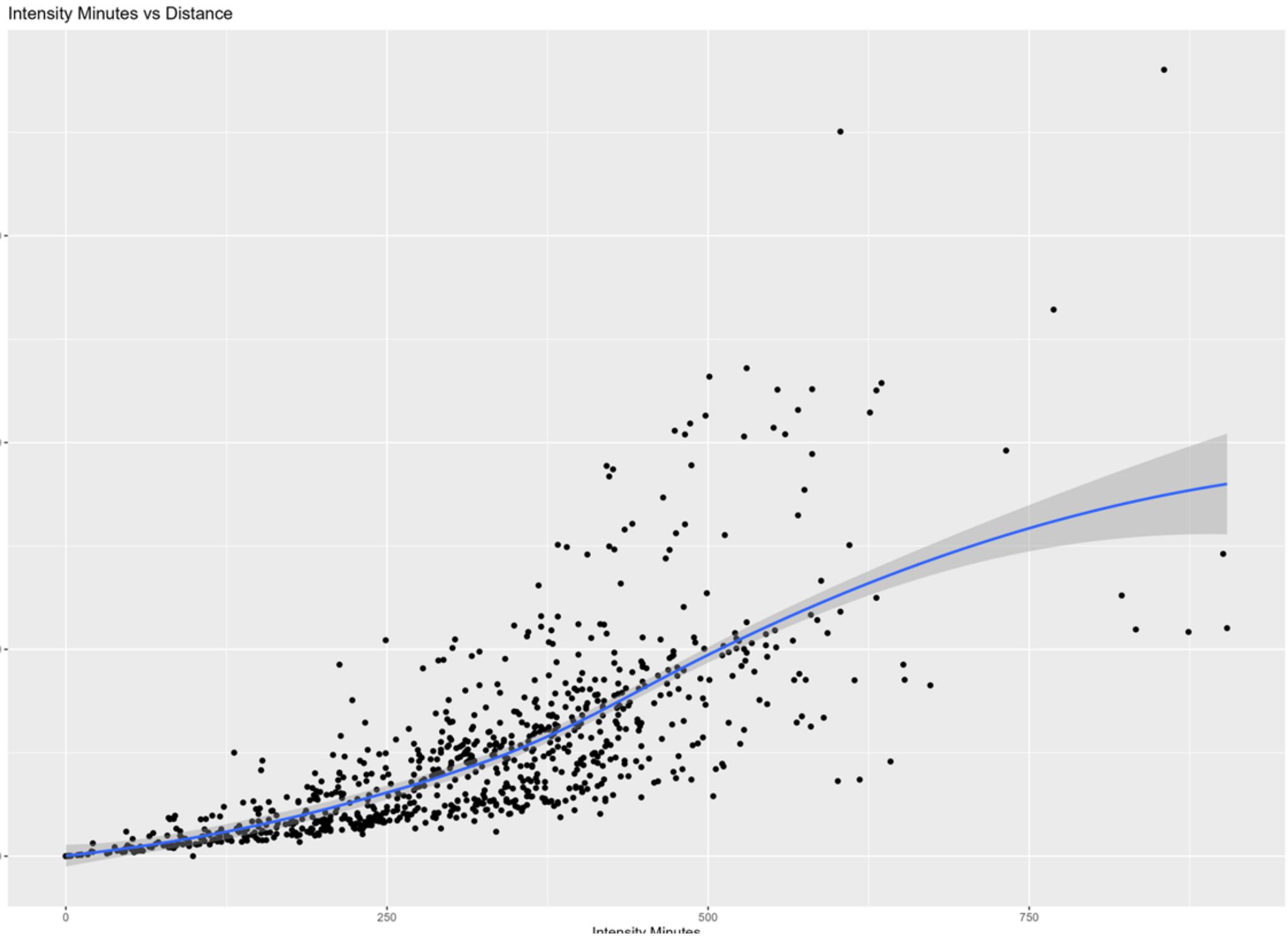
4. ANALYZE

4. Caloric Expenditure vs. Sedentary Behavior: Surprisingly, there's a lack of a strong linear correlation between calories burned and sedentary minutes. This suggests that factors beyond sedentary behavior contribute significantly to caloric expenditure.



4. ANALYZE

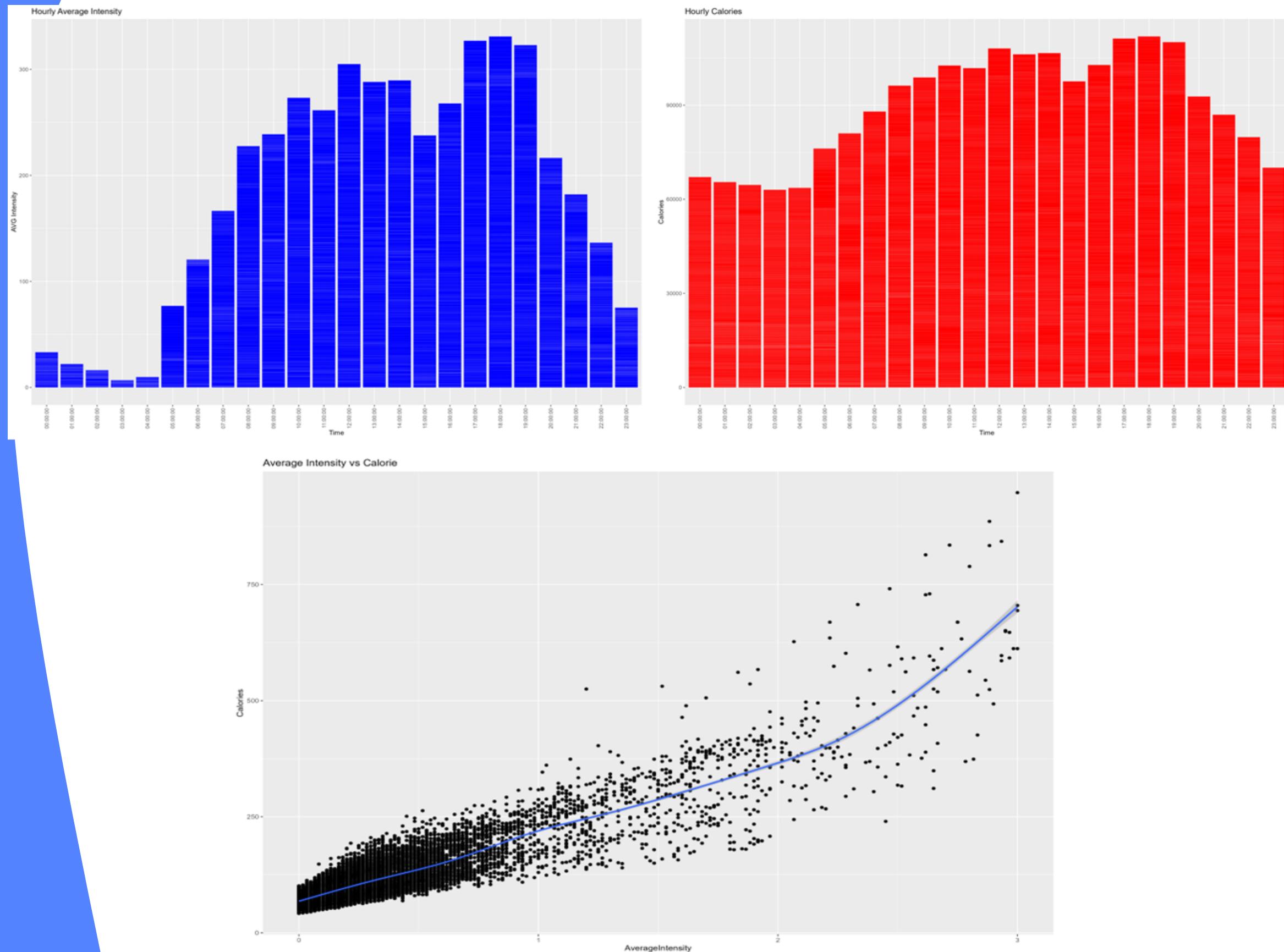
5. Intensity Score Significance: The calculated intensity score exhibits a robust relationship with both intensity minutes and distance covered, reinforcing its relevance in gauging activity levels. Both also shows strong correlation.



4. ANALYZE

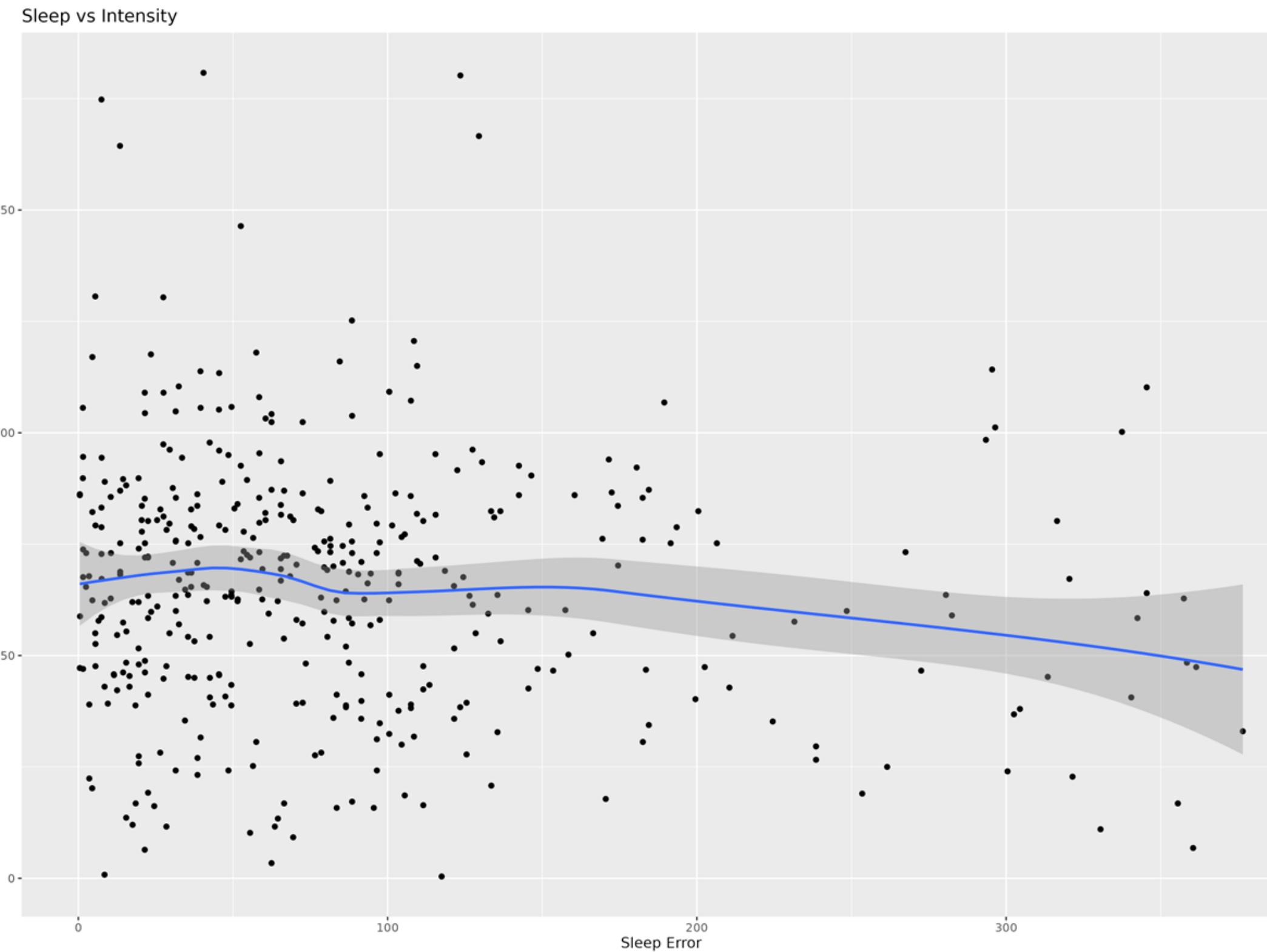
6. Hourly Activity Peaks: Hourly trends indicate heightened activity levels and calorie expenditure between 17:00 and 20:00, potentially indicating a popular timeframe for physical activity.

7. Intensity vs. Caloric Expenditure: Hourly data underscores a strong connection between intensity levels and calorie expenditure, further emphasizing the influence of activity intensity on caloric burn.



4. ANALYZE

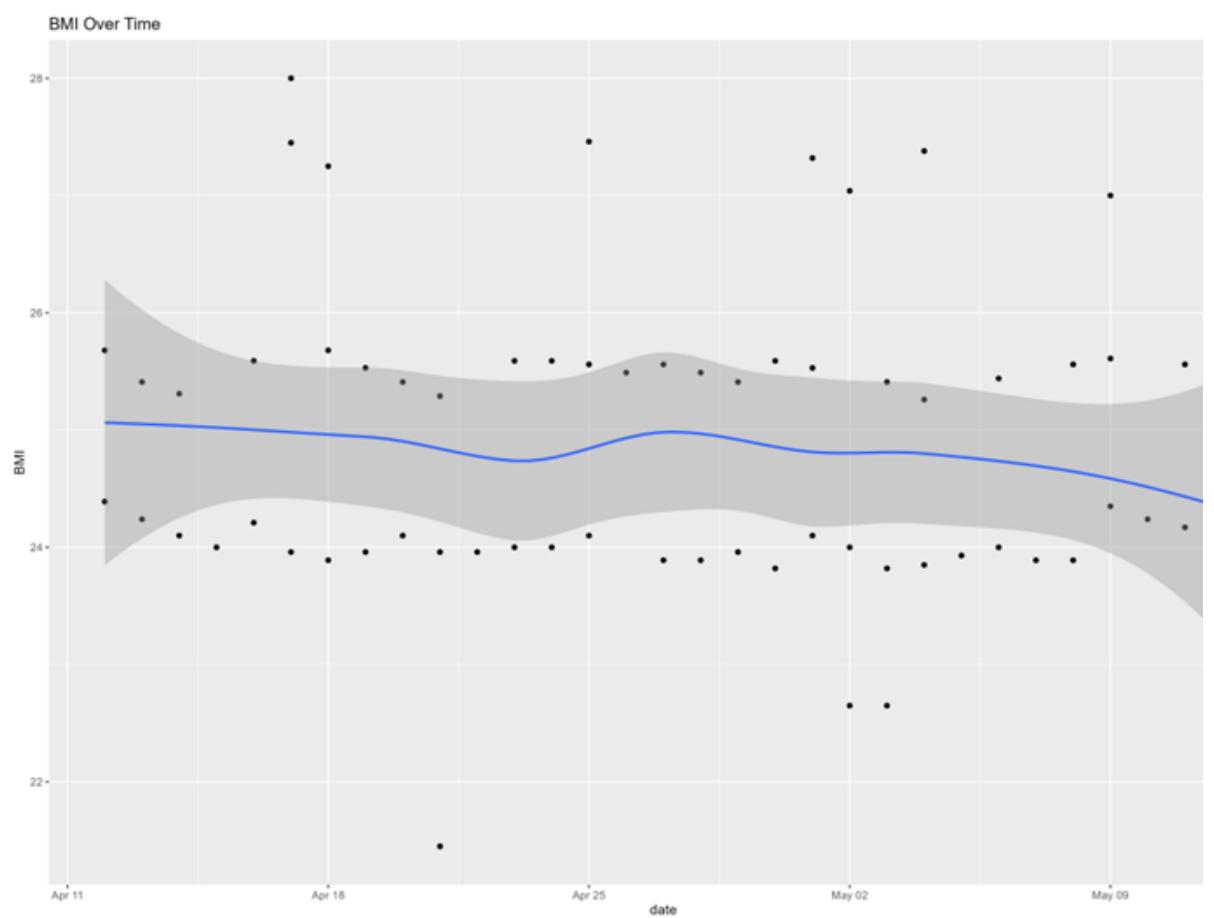
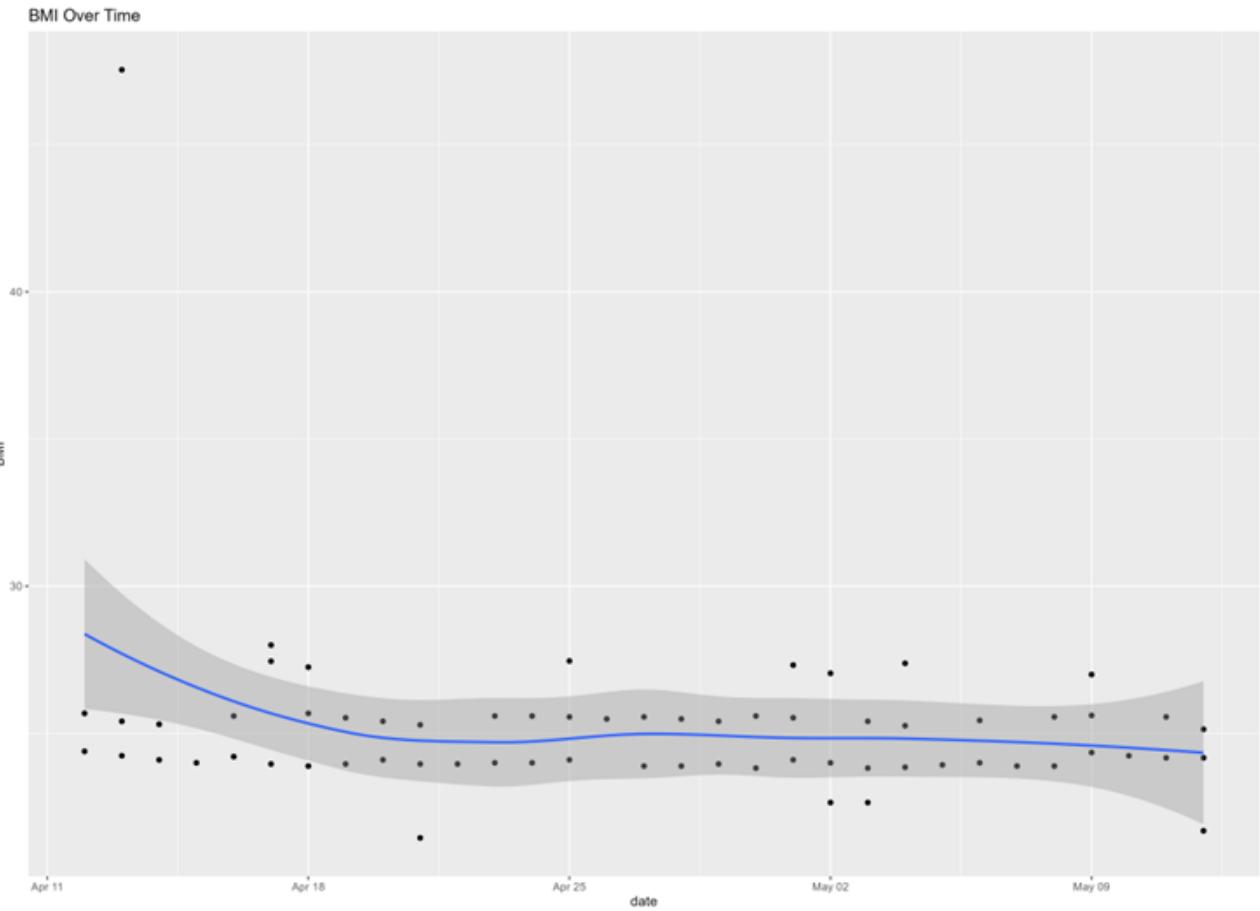
8. Sleeping Error Metric: Defined as the absolute difference between actual sleep time and its mean, the sleeping error metric introduces an interesting dimension. However, it shows no strong linear correlation with intensity minutes, implying that activity levels may not directly impact sleep quality.



4. ANALYZE

9. Outlier Handling in Weight Log:

The decision to remove a single user from the weight log data is noteworthy. This action addresses a potential outlier that might otherwise skew trends, resulting in a more accurate representation of BMI trends.



5. SHARE

Based on our thorough analysis, we will go back to our original question and try to answer them.

1. What are some **trends** in smart device usage? (calorie burnt shows gradual decrease over the span of one months)
2. Are there other external **behaviour** or **bias**? (the data used does not have information about user gender which is different from bellabeat's target customer)
3. Are there some hidden **relationships** we can utilize? (calorie expenditure and intensity peaked at 17:00 to 20:00 indicating a common time for workout)
4. How to **use** this information to Bellabeat **customers**? (the use of smart devices for health tracking usually to maintain good health not necessarily have a set goal)
5. How could we **influence** marketing strategy using this information? (by verifying the behaviour of customers, we can create personalized plan program or milestone events)

6. ACT

Simplifying our analysis here are our key points.

- We can conclude from our analysis that the main customer behaviour is that some customer have **long sitting time** that is caused by working hour (**verified by hourly data analysis**).
- Customer essentially use the device to track their health activity but **not necessarily have a set target to achieve** (**verified by weight log and calorie trends**).
- One thing missing from the analysis is that calorie expenditure does not have strong correlation with slack time which imply that **user diet must have been another factor** (**verified by calorie and sedentary time plot**).

We propose a **personalized health tracking plan** with additional features such as **diet plan**. This solution will account for the **amount of user slack time**. Furthermore, we can set a reminder to **keep moving** when in a static position for too long, this can help reduce back pain on sitting for a long time and remind to take a break. Finally, we can create an **afterwork workout program** followed by a guide to have a **healthy sleep** after workout and a long day.

THANK YOU!

All questions are welcome!

