



# GOOGLE DATA ANALYTICS — CAPSTONE PROJECT

Bellabeat — Jesslyn Jane

Kaggle — Bellabeat

# INTRODUCTION

I am wearing the hat of a **junior data analyst**, working in the marketing analyst team at **Bellabeat**, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market.

Urška Sršen, cofounder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. They offer different smart devices that collect data on activity, sleep, stress, and reproductive health to empower women with knowledge about their own health and habits. The main objective of this case is to analyze smart devices fitness data and determine how it could help unlock new growth opportunities for Bellabeat. We will focus on one of Bellabeat's products: **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.

# METHOD APPROACH

In this project, I will use 6 steps to ensure its completion:

1. Ask

2. Prepare

3. Process

4. Analyze

5. Share

6. Act

# STEP 1. ASK

**Business task:** Identify some trends in how consumers use the Bellabeat devices and how these trends can help improve new opportunities growth for Bellabeat as well as marketing strategy.

## **Key Stakeholders:**

- 1.Urška Sršen** – Bellabeat's cofounder and Chief Creative Officer
- 2.Sando Mur** – Mathematician and Bellabeat's cofounder
- 3.Bellabeat's marketing analytics team** – a team of data analytics

## STEP 2. PREPARE

In this project, I will be using datasets from FitBit Fitness Tracker Data, which has been published by Möbius to Kaggle.com under the CC0: Public Domain Creative Common License.

I downloaded zip files provided, then extracted to csv files. There are 18 datasets, but I only used 6 datasets for this study.

I used excel to take a glimpse of the data. The data collected was only during a 31-day period (04-12-2016 to 05-12-2016), so it was quite outdated as fitness trackers matured a lot since then. No demographic information (gender, location, age) collected makes the data bias even higher.

## 3. PROCESS

To begin processing the data, I used SQL in Bigquery as one of the data analytics tools, to import the dataset, do the process of cleaning and organizing.

The cleaning process included adjusting data type formats, removing duplicates and null data.

I extracted the clean data to new csv and stored it.

I documented the whole process of cleaning.

## 4. ANALYZE

I organized and formatted the data.

I performed some calculations and identified trends as well as relationships between each variable.

### **Data:**

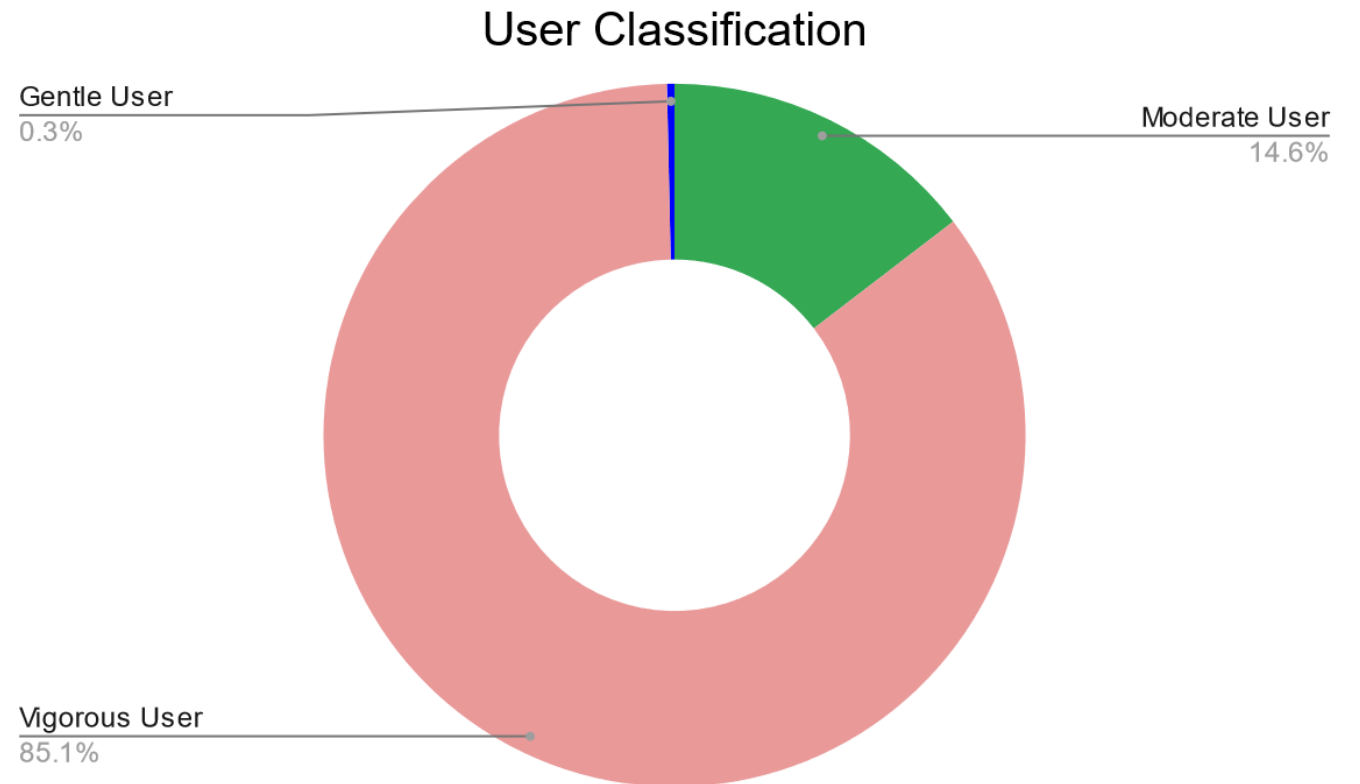
- User type distribution by tracker-wear
- Look Deep into Activity
  - Average activity minutes by day of week
  - User type distribution by total steps per day
  - Average Steps, distances, calories by day of week
  - Average total steps vs calories
- Look Deep into Sleep
  - Average sleep time and awake time by day of week
  - Average total minutes asleep, steps, calories
- Look Deep into Weight
  - Average weight vs non-sedentary minutes

# USER TYPE DISTRIBUTION BY TRACKER-WEAR

- Vigorous User → wear their tracker for 21 – 31 days
- Moderate User → 11 – 20 days
- Gentle User → 1 -10 days

## Key Takeaway:

**More than 80%** of the users used Fitbit Fitness Tracker frequently – between 21 to 31 days (Vigorous User)



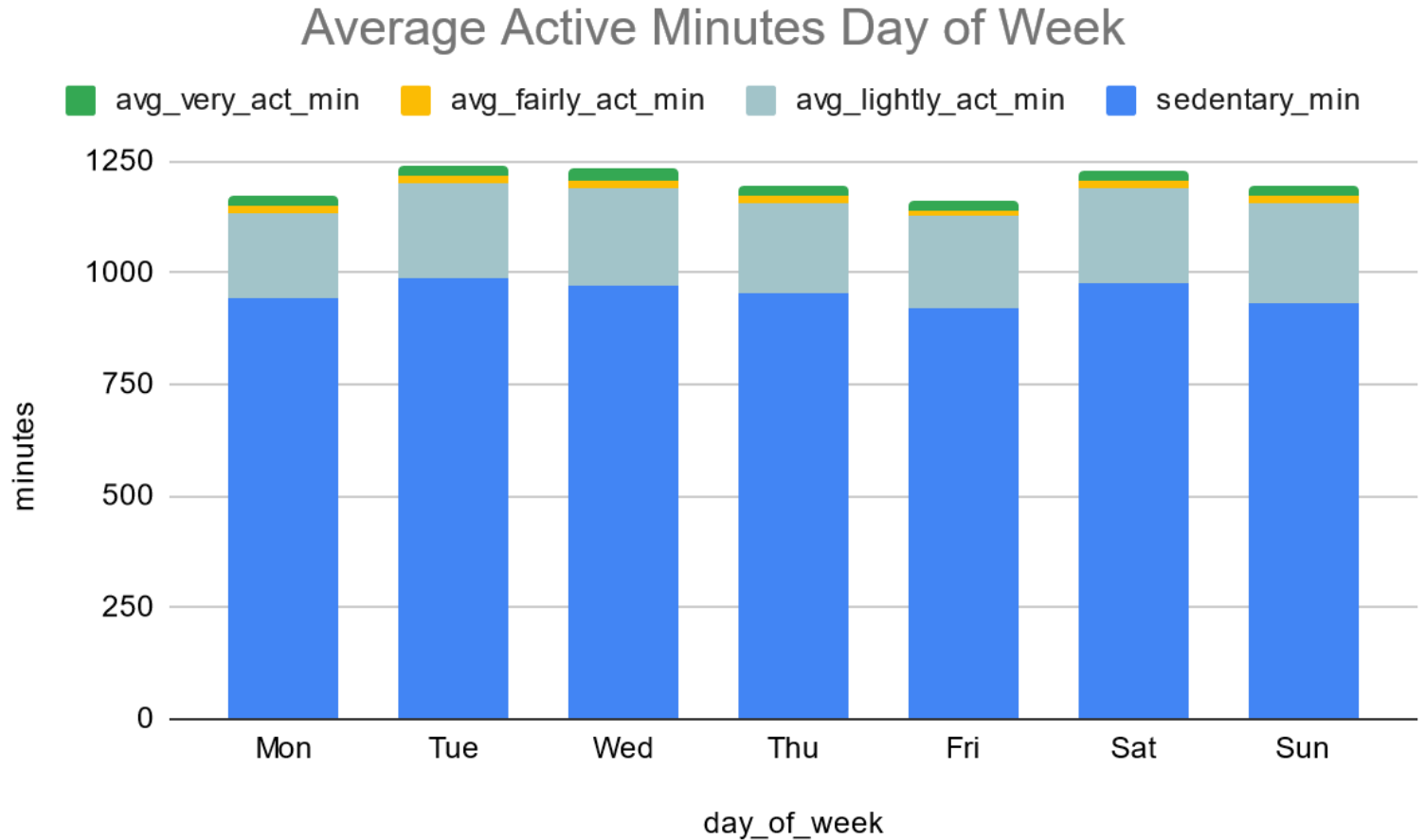


# AVERAGE ACTIVE MINUTES BY DAY OF WEEK

## Key Takeaway:

Total average active minutes shows **slightly difference** throughout the week

**Sedentary Minutes** are the highest type of active minutes



# USER TYPE DISTRIBUTION BY TOTAL STEPS PER DAY

- Sedentary Lifestyle → total steps per day <5000
- Physically Inactive → 5000 - 7499
- Moderately Active → 7500 - 9999
- Physically Active → 10000 – 12499
- Physically Active → >= 12500

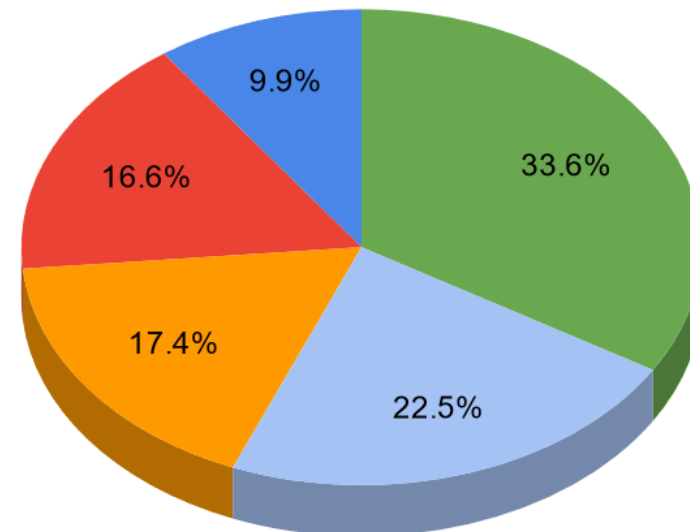
*Tudor-Locke and Bassett (2004)*

Key Takeaway:

**More than a third** of the users are considered moderately active

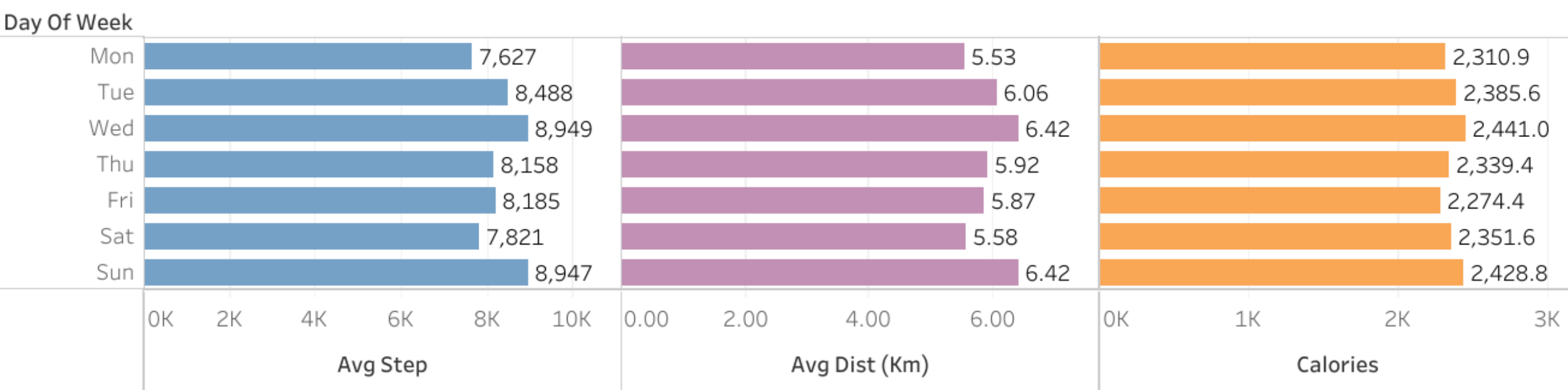
User Type Distribution By Total Steps

● Moderately Active   ● Physically Inactive   ● Physically Active   ● Very Active  
● Sedentary Lifestyle



# AVERAGE STEPS, DISTANCE, CALORIES BY DAY OF WEEK

Average Steps, Distances, Calories Per Day of Week



## Key Takeaway:

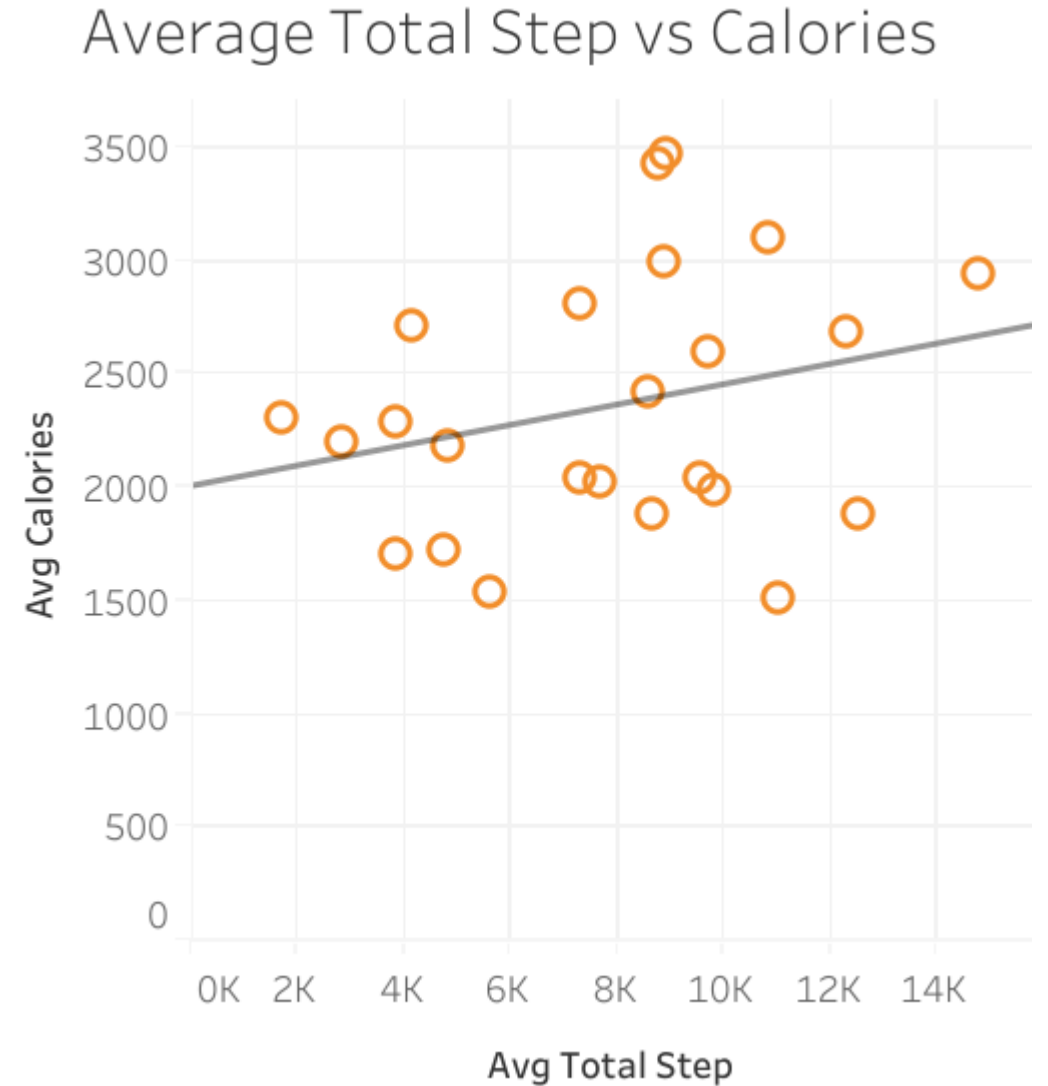
The highest average step and distance per day was on **Sunday** and **Wednesday** with almost **9 thousand steps** and **6 km distance**.

Average calories shows little difference throughout the week.

# AVERAGE TOTAL STEP VS CALORIES

Key Takeaway:

There is **high correlation** between average total step and calories



# AVERAGE SLEEP TIME VS AWAKE TIME BY DAY OF WEEK

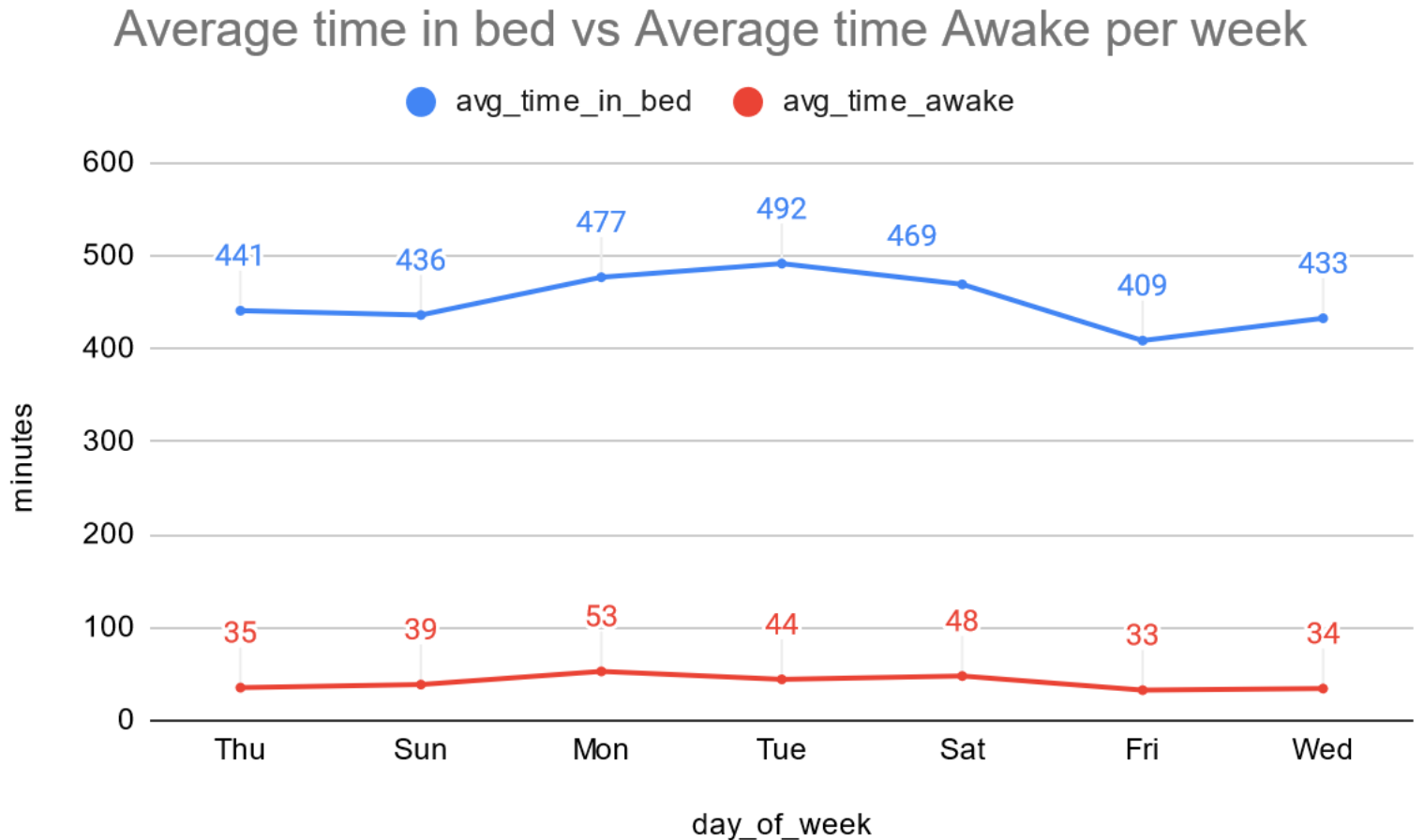
## Key Takeaway:

There isn't a whole lot of difference between each day in terms of average time in bed and time awake.

The highest average time in bed within a week falls on **Tuesday**.

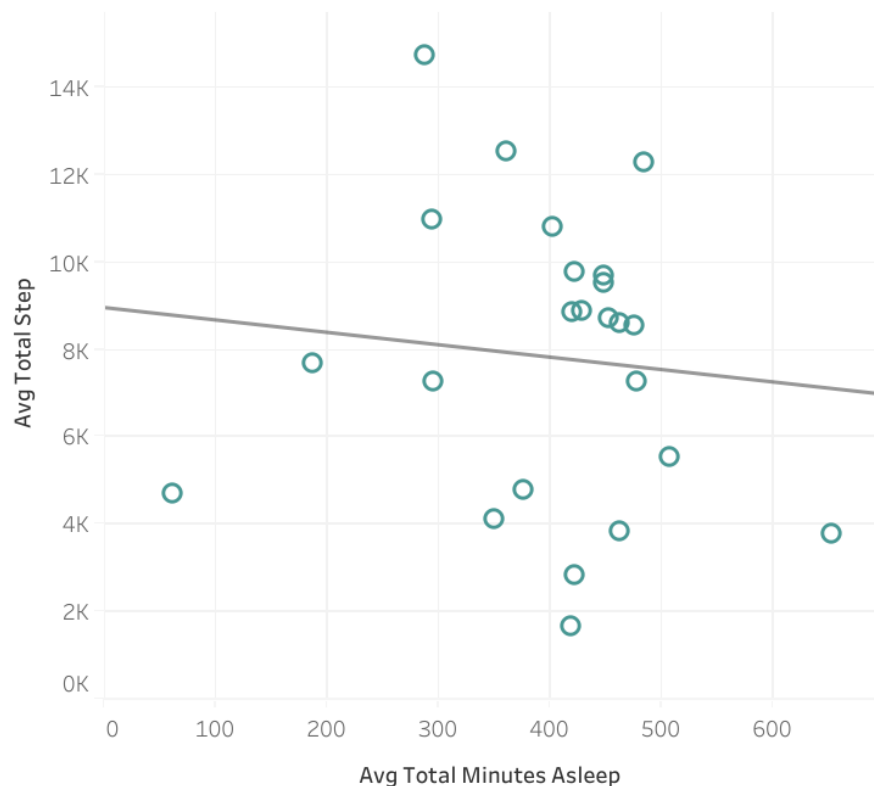
## Further analysis:

Check on whether this pattern was the same every month

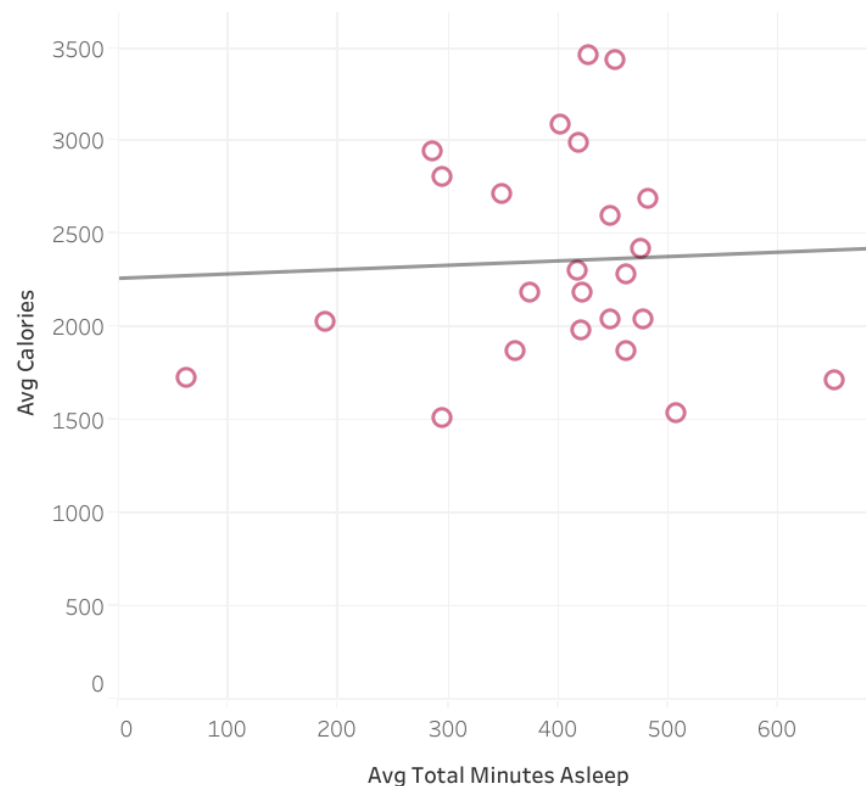


# AVERAGE TOTAL MINUTES ASLEEP, STEPS, CALORIES

Average Total Minutes Asleep vs Total Steps



Average Total Minutes Asleep vs Calories



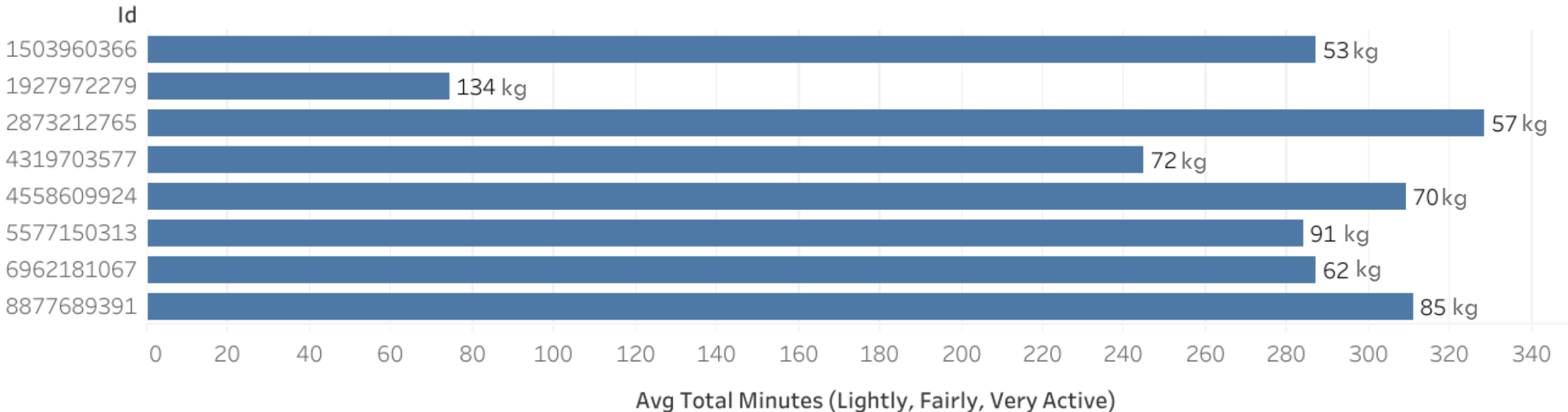
## Key Takeaway:

There is **no correlation** between average total steps and average amount of minutes users sleep.

**No correlation** is found between average total minutes of sleep and average calories.

# AVERAGE TOTAL ACTIVE MINUTES VS WEIGHT

Average Total Active Minutes vs Weight



## Key Takeaway:

User with **low exercise** shows **overweighted** (134 kg).

Users with weight of <70 kg had been more active.

## 5. SHARE

The share phase is done by presenting this powerpoint.

Some conclusions we can draw from the data:

- In a period of a month (April to May 2016), there were 83% of users in Vigorous category when it came to how frequently they wore their tracker.
- Most of the users were not active, as sedentary minutes show the highest value.
- 73,5% users were not reaching 10.000 steps a day.
- Wednesday and Sunday made up to the most active day.
- Higher steps spent means that higher calories was burnt.
- The highest average time in bed within a week falls on Tuesday.
- There is no significant correlation between steps, calories, and total minutes asleep.
- The more intense users worked out, the better for them to reach their weight goal.



## 6. ACT

The act phase would be done by the production and marketing team of the company. The main takeaway will be the top three recommendations:

- 1. Marketing team can showcase about the **importance of doing sport activities daily**, by highlighting the strong correlation between total active minutes with healthy weight, so that users might build a self-conscious about using the product regularly.
- 2. Production team can improve the **notification feature** of the tracker app as reminders for the users to achieve their goal and increase **total steps** each day.
- 3. Production team can provide **reward system**, based on total amount of steps reaching daily, weekly, monthly, to boost the energy and motivation of workout.



**THE END**

Thank you.