Bellabeat Fitness Data Analysis

Date : October 21, 2022

Introduction:

Bellabeat is a high-tech manufacturer of health-focused smart products for women. Bellabeat's app and multiple smart devices collect data on activity, sleep, stress, hydration levels, and reproductive health to empower women with an understanding of their own health and habits. The company was founded in 2013 by Urška Sršen and Sando Mur and has expanded quickly since, now with the possibility to become a greater player in the global smart device market.

Bellabeat's product line is made up of the Bellabeat app, which allows users insight into their health by providing data on their activity, sleep, stress, menstrual cycle, and mindfulness habits. The Bellabeat app also connects to the company's line of smart device products. Leaf is Bellabeat's classic wellness tracker that can be worn as a bracelet, necklace, or clip. Leaf tracks the user's activity, sleep, and stress and connects to the Bellabeat app. Time is a wellness smart watch that also tracks the user's activity, sleep, and stress and connects to the Bellabeat app. Spring is a smart water bottle that tracks the daily water intake of its user to ensure proper hydration levels are maintained throughout the day. Spring also connects to the Bellabeat app to track this data. Bellabeat membership is a subscription-based membership program that provides users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health, beauty, and mindfulness based on their lifestyle and goals.

Business task:

Analyze consumers use of an existing competitor to identify potential opportunities for growth and recommendations for the Bellabeat marketing strategy
Primary stakeholders: Urška Sršen and Sando Mur, executive team members.
Secondary stakeholders: Bellabeat marketing analytics team.

Questions for the analysis:

1. What are some trends in smart device usage?
2. How could these trends apply to Bellabeat customers?

Prepare:

- The data is from 30 FitBit users who consented to the submission of personal tracker data and generated by from a distributed survey via Amazon Mechanical Turk.
- Data minute-level output for physical activity, heart rate, and sleep monitoring. While the data tracks many factors in the user activity and sleep, but the sample size is small and most data is recorded during certain days of the week.

  The dataset has limitations:

- Only 30 user data is available. The central limit theorem general rule of n≥30 applies and we can use the t test for statstic reference. However, a larger sample size is preferred for the analysis.
- Most data is recorded from Tuesday to Thursday, which may not be comprehensive enough to form an accurate analysis.

Process :

Convert ActivityDate into date format and add a column for day of the week:

```
daily_activity <- daily_activity %>% mutate( Weekday =
weekdays(as.Date(ActivityDate, "%m/%d/%Y")))
```

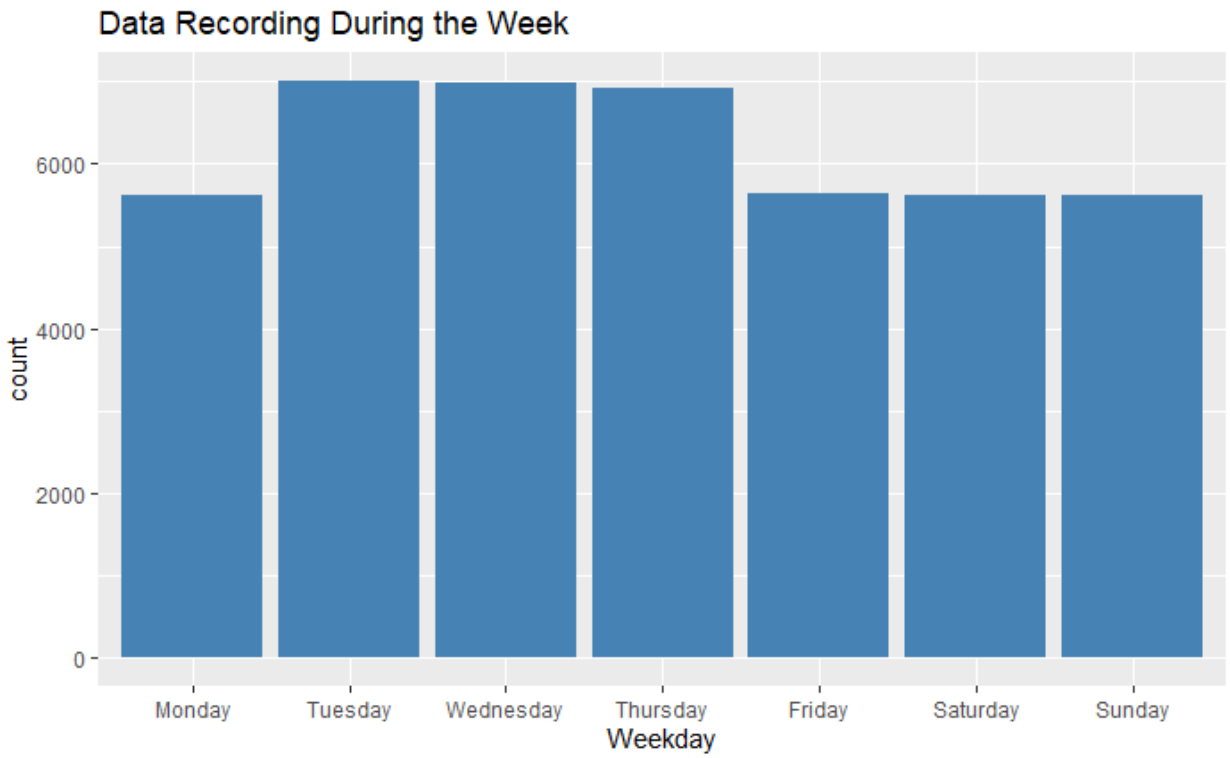
Check to see if we have 30 users using n_distinct(). The dataset has 33 user data from daily activity, 24 from sleep and only 8 from weight. If there is a discrepency such as in the weight table, check to see how the data are recorded. The way the user record the data may give you insight on why there is missing data.

```
weight %>%
  filter(IsManualReport == "True") %>%
  group_by(Id) %>%
```

```
  summarise("Manual Weight Report"=n()) %>%
  distinct()
```

Additional insight to be awared of is how often user record their data. We can see from the ggplot() bar graph that the data are greatest from Tuesday to Thursday. We need to investigate the data recording distribution. Monday and Friday are both weekdays, why isn't the data recordings as much as the other weekdays?

```
ggplot(data=merged_data, aes(x=Weekday))+
  geom_bar(fill="steelblue")
```



Merge the three tables:

```
merged_data <- merge(merged_activity_sleep, weight, by = c("Id"), all=TRUE)
```

Analyze :

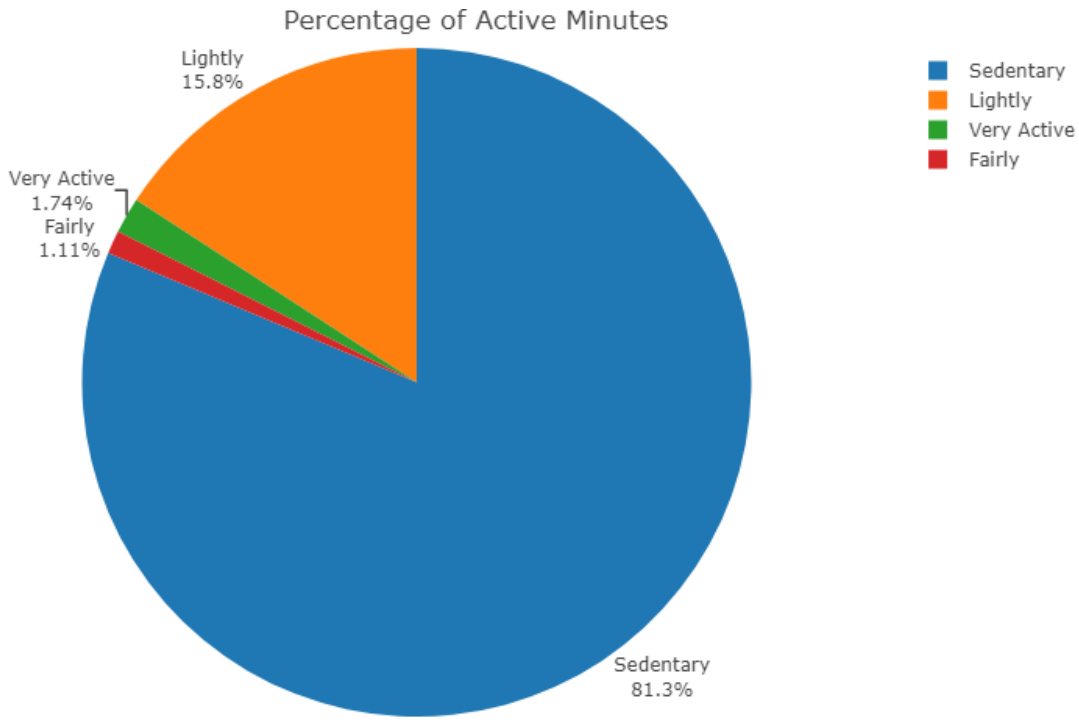
Check min, max, mean, median and any outliers. Avg weight is 135 pounds with BMI of 24 and burn 2050 calories. Avg steps is 10200, max is almost triple that 36000 steps. Users spend on avg 12 hours a day in sedentary minutes, 4 hours lightly active, only half hour in fairly+very active! Users also gets about 7 hour of sleep.

```
merged_data %>%
  dplyr::select(Weekday,
         TotalSteps,
         TotalDistance,
         VeryActiveMinutes,
         FairlyActiveMinutes,
         LightlyActiveMinutes,
         SedentaryMinutes,
         Calories,
         TotalMinutesAsleep,
         TotalTimeInBed,
         WeightPounds,
         BMI
         ) %>%
```

Active Minutes:

Percentage of active minutes in the four categories: very active, fairly active, lightly active and sedentary

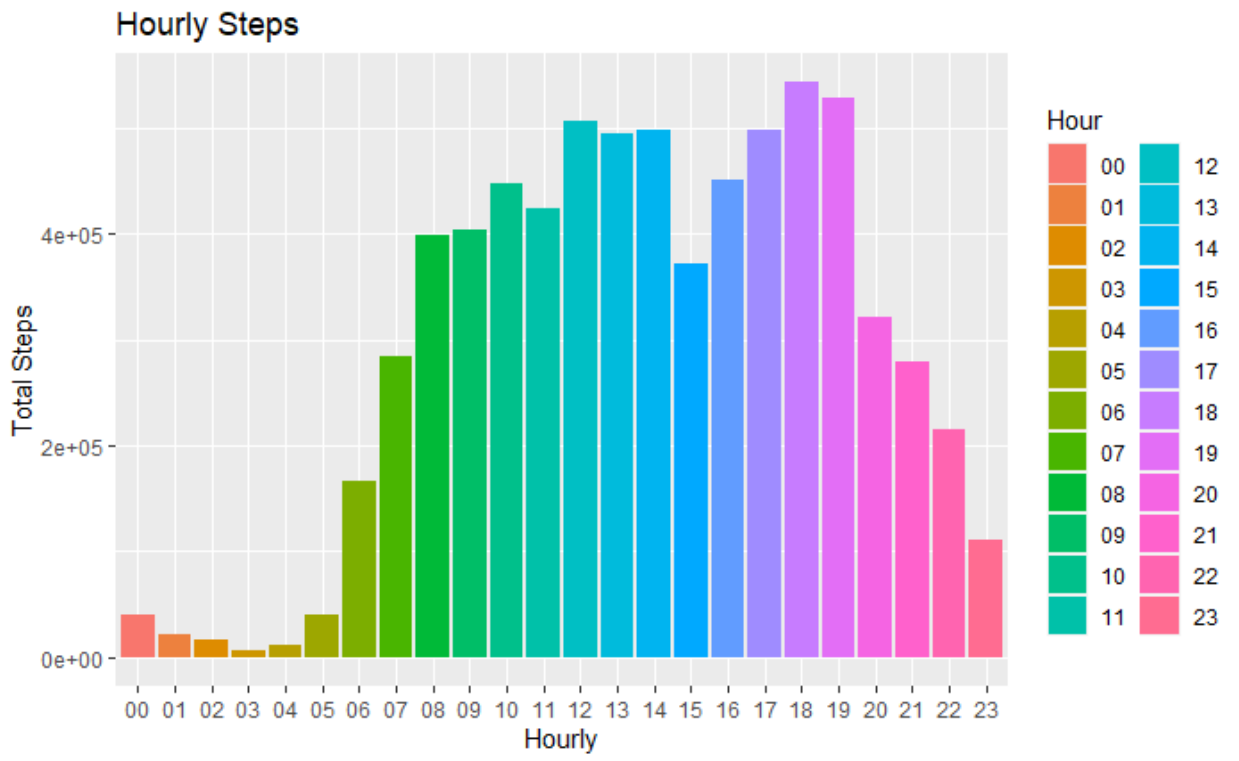
```
percentage <- data.frame(
  level=c("Sedentary", "Lightly", "Fairly", "Very Active"),

minutes=c(sedentary_percentage,lightly_percentage,fairly_percentage,a
ctive_percentage)
)

plot_ly(percentage, labels = ~level, values = ~minutes, type =
'pie',textposition = 'outside',textinfo = 'label+percent') %>%
  layout(title = 'Activity Level Minutes',
         xaxis = list(showgrid = FALSE, zeroline = FALSE,
showticklabels = FALSE),
         yaxis = list(showgrid = FALSE, zeroline = FALSE,
showticklabels = FALSE))
```



Total Steps:

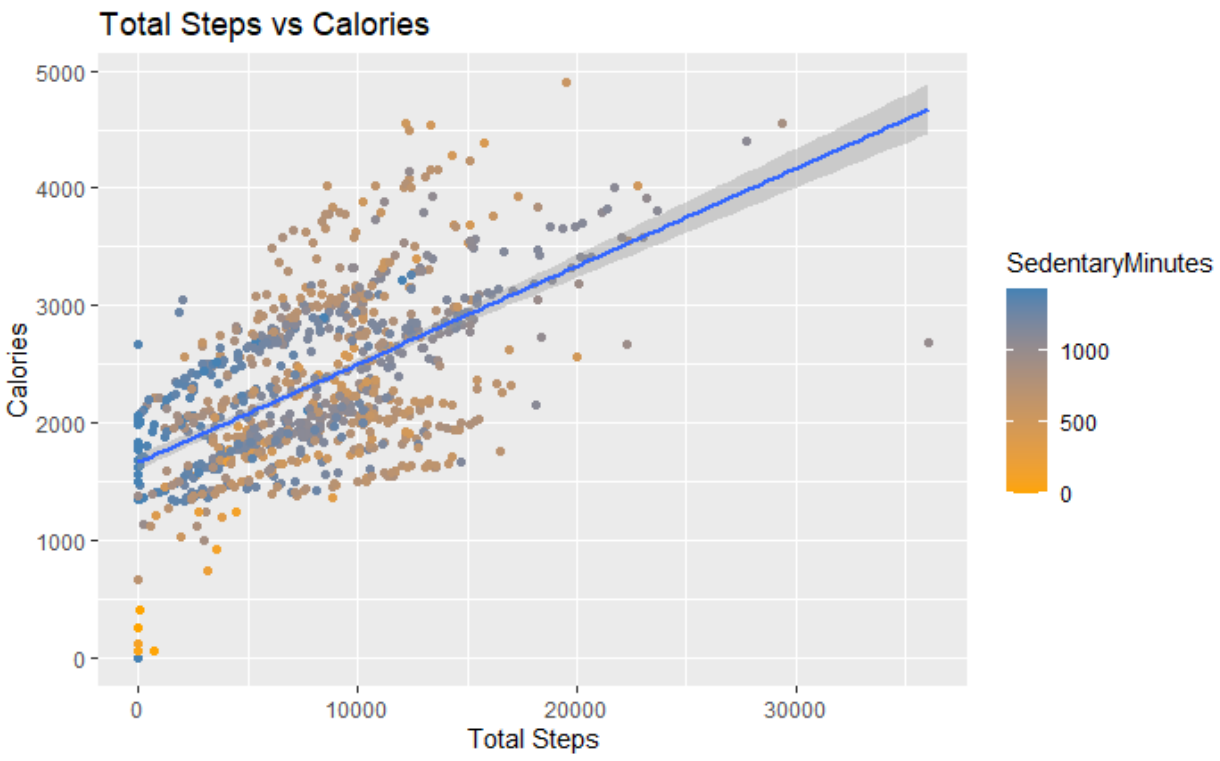
Let's look at how active the users are per hourly in total steps. From 5PM to 7PM the users take the most steps.

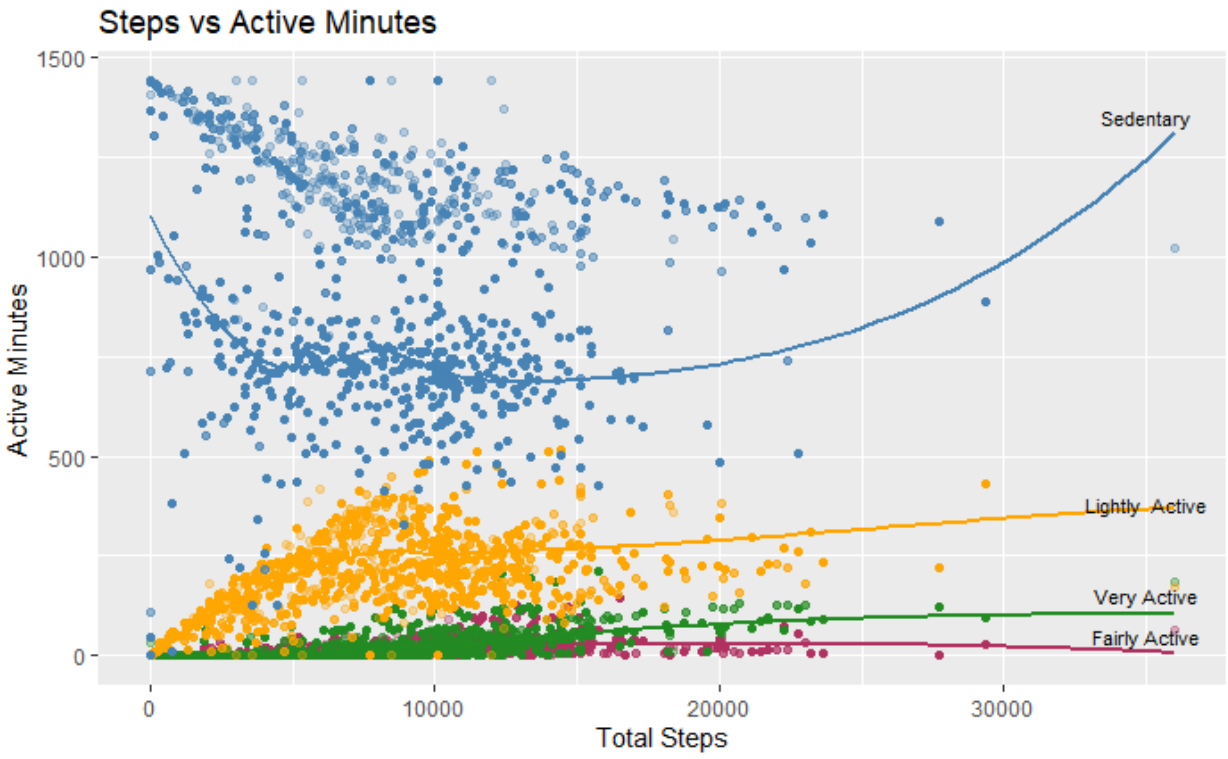
```
ggplot(data=hourly_step, aes(x=Hour, y=StepTotal, fill=Hour))+
  geom_bar(stat="identity")+
  labs(title="Hourly Steps")
```



The more active that you're, the more steps you take, and the more calories you will burn. This is an obvious fact, but we can still look into the data to find any interesting. Here we see that some users who are sedentary, take minimal steps, but still able to burn over 1500 to 2500 calories compare to users who are more active, take more steps, but still burn similar calories.

```
ggplot(data=daily_activity, aes(x=TotalSteps, y = Calories,
color=SedentaryMinutes))+
  geom_point()+
  stat_smooth(method=lm)+
  scale_color_gradient(low="steelblue", high="orange")
```


Total Steps vs Calories
Calories
Total Steps
SedentaryMinutes
1000
500
0
5000
4000
3000
2000
1000
0
0
10000
20000
30000


Steps vs Active Minutes
Active Minutes
Total Steps
Sedentary
Lightly Active
Very Active
Fairly Active
1500
1000
500
0
0
10000
20000
30000

Comparing the four active levels to the total steps, we see most data is concentrated on users who take about 5000 to 15000 steps a day. These users spent an average between 8 to 13 hours in sedentary, 5 hours in lightly active, and 1 to 2 hour for fairly and very active.

Act :

- We see the most change on Saturday: users take more steps, burn more calories, and spend less time sedentary. Sunday is the most "lazy" day for users.
- Users takes the most steps from 5 PM to 7 PM Users who are sedentary take minimal steps and burn 1500 to 2500 calories compared to users who are more active, take more steps, but still burn similar calories.

Recommendations:

- Heavily market Spring as Fitbit does not track hydration levels.
- Offer a bundle deal for the Spring and Leaf together.

RESOURCES:

FitBit Fitness Tracker Data

Bellabeat