Sean Hughes

Bellabeat Case Study

August 4th, 2023

Introduction

Bellabeat is a tech-driven wellness company for women that was founded by Urška Sršen and Sando Mur in 2013. Through the collection of data on sleep, stress, activity, and reproductive health Bellabeat has given women the knowledge and power of their health. Bellabeat's product line includes:

- **Bellabeat App** (a hub for all data tracked by Bellabeat products)
- **Leaf** (can be worn as a bracelet, necklace, or clip that tracks activity, sleep, and stress)
- **Time** (a smartwatch that tracks activity, sleep, and stress)
- **Spring** (a water bottle that tracks your hydration levels and connects to the Bellabeat App)
- **Bellabeat Membership** (gives 24/7 access to personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness)

Sršen believes there is untapped potential with Bellabeat and by analyzing smart device fitness data, better business and marketing decisions can be made. She asks the marketing analytics team to determine how consumers use their smart devices based on the data, then apply the findings to a Bellabeat product.

Ask

Business Task

The business task is to analyze consumer smart device data and apply such findings to the Bellabeat 'Time' smartwatch. These insights will guide stakeholders in their quest of creating smart devices that cater to the needs of their customers.

Stakeholders

The key stakeholders are Bellabeat's cofounders Urška Sršen (Chief Creative Officer) and Sando Mur (Mathematician). The Bellabeat marketing analytics team makes up the secondary stakeholders.

Prepare

Data Sources

The data comes from [thirty Fitbit users] who consented to share personal information like the output of physical activity, heart rate, steps, and sleep monitoring. The dataset is separated into 18 .csv files.

ROCCC Analysis

Reliable - Low, a sample size of 30 is small and could be misleading.

Original - Low, originally collected by Amazon Mechanical Turk and distributed by a third party.

Comprehensive - Low, the demographics of participants are missing. Information like gender, age, health, etc. Data could not be randomized, leading to further bias.

Current - Medium, this data is from 2016. Smartwatch technology and subsequent consumer behavior may not be representative of today.

Cited - Unknown, there is skepticism about the credibility of the Amazon Mechanical Turk source.

Process

Tools

Google Sheets, Google BigQuery, RStudio Cloud, Tableau

Data Cleaning

In BigQuery, the number of unique ‘Id’ values in each spreadsheet was counted, dailyActivity_merged (33), sleepDay_merged (24), and weightLogInfo_merged (8). The syntax is as follows:

```
SELECT DISTINCT
int64_field_0
FROM `data_activity_merged.all_data`

```

Then, using Google Sheets, each unique 'Id' was conditional formatted in order to color code every row like such:

A1 | fx Id

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Id	ActivityDate	TotalSteps	TotalDistance	TrackerDistance	LoggedActivitiesDistance	VeryActiveDistance	ModeratelyActiveDistance	LightActiveDistance	SedentaryActiveDistance	VeryActiveMinutes	FairlyActiveMinutes	LightlyActiveMinutes	SedentaryMi
16	1503960366	4/26/2016	13755	8.789999962	8.789999962	0	2.329999924	0.9200000167	5.539999962	0	31	23	279	
17	1503960366	4/27/2016	18134	12.21000004	12.21000004	0	6.400000095	0.4099999964	5.409999847	0	78	11	243	
18	1503960366	4/28/2016	13154	8.529999733	8.529999733	0	3.539999962	1.159999967	3.789999962	0	48	28	189	
19	1503960366	4/29/2016	11181	7.150000095	7.150000095	0	1.059999943	0.5	5.579999924	0	16	12	243	
20	1503960366	4/30/2016	14673	9.25	9.25	0	3.559999943	1.419999957	4.269999981	0	52	34	217	
21	1503960366	5/1/2016	10602	6.809999943	6.809999943	0	2.289999962	1.600000024	2.920000076	0	33	35	246	
22	1503960366	5/2/2016	14727	9.710000038	9.710000038	0	3.210000038	0.5699999928	5.920000076	0	41	15	277	
23	1503960366	5/3/2016	15103	9.659999847	9.659999847	0	3.730000019	1.049999952	4.880000114	0	50	24	254	
24	1503960366	5/4/2016	11100	7.150000095	7.150000095	0	2.460000038	0.8700000048	3.819999933	0	36	22	203	
25	1503960366	5/5/2016	14070	8.899999619	8.899999619	0	2.920000076	1.080000043	4.880000114	0	45	24	250	
26	1503960366	5/6/2016	12159	8.029999733	8.029999733	0	1.970000029	0.25	5.809999943	0	24	6	289	
27	1503960366	5/7/2016	11992	7.710000038	7.710000038	0	2.460000038	2.119999886	3.130000114	0	37	46	175	
28	1503960366	5/8/2016	10060	6.579999924	6.579999924	0	3.529999971	0.3199999928	2.730000019	0	44	8	203	
29	1503960366	5/9/2016	12022	7.71999979	7.71999979	0	3.450000048	0.5299999714	3.74000001	0	46	11	206	
30	1503960366	5/10/2016	12207	7.769999981	7.769999981	0	3.349999905	1.159999967	3.25999999	0	46	31	214	
31	1503960366	5/11/2016	12770	8.130000114	8.130000114	0	2.559999943	1.00999999	4.550000191	0	36	23	251	
32	1503960366	5/12/2016	0	0	0	0	0	0	0	0	0	0	0	
33	1624580081	4/12/2016	8163	5.309999943	5.309999943	0	0	0	5.309999943	0	0	0	146	
34	1624580081	4/13/2016	7007	4.550000191	4.550000191	0	0	0	4.550000191	0	0	0	148	
35	1624580081	4/14/2016	9107	5.920000076	5.920000076	0	0	0	5.909999847	0.009999999776	0	0	236	
36	1624580081	4/15/2016	1510	0.9800000191	0.9800000191	0	0	0	0.9700000286	0	0	0	96	
37	1624580081	4/16/2016	5370	3.49000001	3.49000001	0	0	0	3.49000001	0	0	0	176	
38	1624580081	4/17/2016	6175	4.059999943	4.059999943	0	1.029999971	1.519999981	1.49000001	0.009999999776	15	22	127	
39	1624580081	4/18/2016	10536	7.409999847	7.409999847	0	2.150000095	0.6200000048	4.619999886	0.009999999776	17	7	202	
40	1624580081	4/19/2016	2916	1.899999976	1.899999976	0	0	0	1.899999976	0	0	0	141	
41	1624580081	4/20/2016	4974	3.230000019	3.230000019	0	0	0	3.230000019	0	0	0	151	
42	1624580081	4/21/2016	6349	4.130000114	4.130000114	0	0	0	4.110000134	0.01999999955	0	0	186	
43	1624580081	4/22/2016	4026	2.619999886	2.619999886	0	0	0	2.599999905	0	0	0	199	
44	1624580081	4/23/2016	8538	5.550000191	5.550000191	0	0	0	5.539999962	0.009999999776	0	0	227	
45	1624580081	4/24/2016	6076	3.950000048	3.950000048	0	1.149999976	0.9100000262	1.889999986	0	16	18	185	
46	1624580081	4/25/2016	6497	4.21999979	4.21999979	0	0	0	4.199999809	0.01999999955	0	0	202	
47	1624580081	4/26/2016	2826	1.840000033	1.840000033	0	0	0	1.830000043	0.009999999776	0	0	140	
48	1624580081	4/27/2016	8367	5.440000057	5.440000057	0	1.110000014	1.870000005	2.460000038	0	17	36	154	

dailyActivity_merged

Explore

In RStudio Cloud, the "tidyverse" "ggplot2" and "dplyr" packages were installed and the following data was imported:

```
>  weight_log <- read.csv("weightLogInfo_merged.csv")
> sleep_data <- read.csv("sleepDay_merged.csv")
> daily_activity <- read.csv("dailyActivity_merged.csv")
```

Data cleaning in RStudio Cloud began with the clean function. Every row that contained missing values was removed. The following syntax was applied to 2 of the datasets. The weight log dataset contained a column with only two values, so that row, 'Fat', was removed:

```
> cleaned_daily_activity <- na.omit(daily_activity)
```

The unique function, which was applied to all 3 datasets, was used to eliminate any duplicate value or row:

```
> unique_daily_activity <- unique(cleaned_daily_activity)
```

Analyze

To get an initial understanding of the datasets, the summary function was used.

Average Weight (lbs)

```
> average_weight_per_id <- unc_unique_weight_log %>%
+     group_by(Id) %>%
+     summarise(average_weight = mean(WeightPounds),
+               number_of_values = n())
```

	Id	average_weight	number_of_values
1	1503960366	115.9631	2
2	1927972279	294.3171	1
3	2873212765	125.6635	2
4	4319703577	159.5045	2
5	4558609924	153.5299	5
6	5577150313	199.9593	1
7	6962181067	135.7019	30
8	8877689391	187.7144	24

As confirmed from early data cleaning, only 8 participants submitted weight logs. On top of that, 5 of the 8 participants submitted only 1 or 2 weight entries.

Total Steps

```
> summary(unique_daily_activity$TotalSteps)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0    3790    7406    7638   10727   36019 
```

Typically, a healthy adult should aim for approximately 10,000 steps per day. However, the average number of steps in this dataset is 7,638. 1 in 4 participants averaged 3,790 steps or less per day, significantly lower than the recommended 10,000 steps.

Calories

```
> summary(unique_daily_activity$Calories)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
      0    1828    2134    2304    2793    4900 
```

The average number of calories burned per day for all participants is 2,304. This is actually more than the suggested amount of calories burned daily for women (~2,000) but slightly behind the number of calories men should burn daily (~2,700). Unfortunately, there are many limitations of this dataset, this being one of them.

Very Active Minutes

```
> summary(unique_daily_activity$VeryActiveMinutes)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    0.00    4.00   21.16   32.00  210.00
```

Half of all participants averaged 4 daily 'very active minutes', quite below the recommended 30 minutes of activity per day.

Very Active Minutes vs Calories Burned

To produce a scatterplot with the 'very active minutes' data in relation to calories burned, the following code was entered:

```
> ggplot(unique_daily_activity, aes(x = VeryActiveMinutes, y = Calories)) +
+     geom_point(color = "orange") +
+     geom_smooth(method = "lm", se = FALSE, color = "blue") +
+     labs(title = "Active Minutes vs Calories Burned",
+          x = "Very Active Minutes",
+          y = "Calories")
```

The scatterplot illustrates a direct relationship between active minutes and calories burned. It also points to the fact that most data in the 'very active minutes' axis is near or at zero, which confirms an earlier summary table.

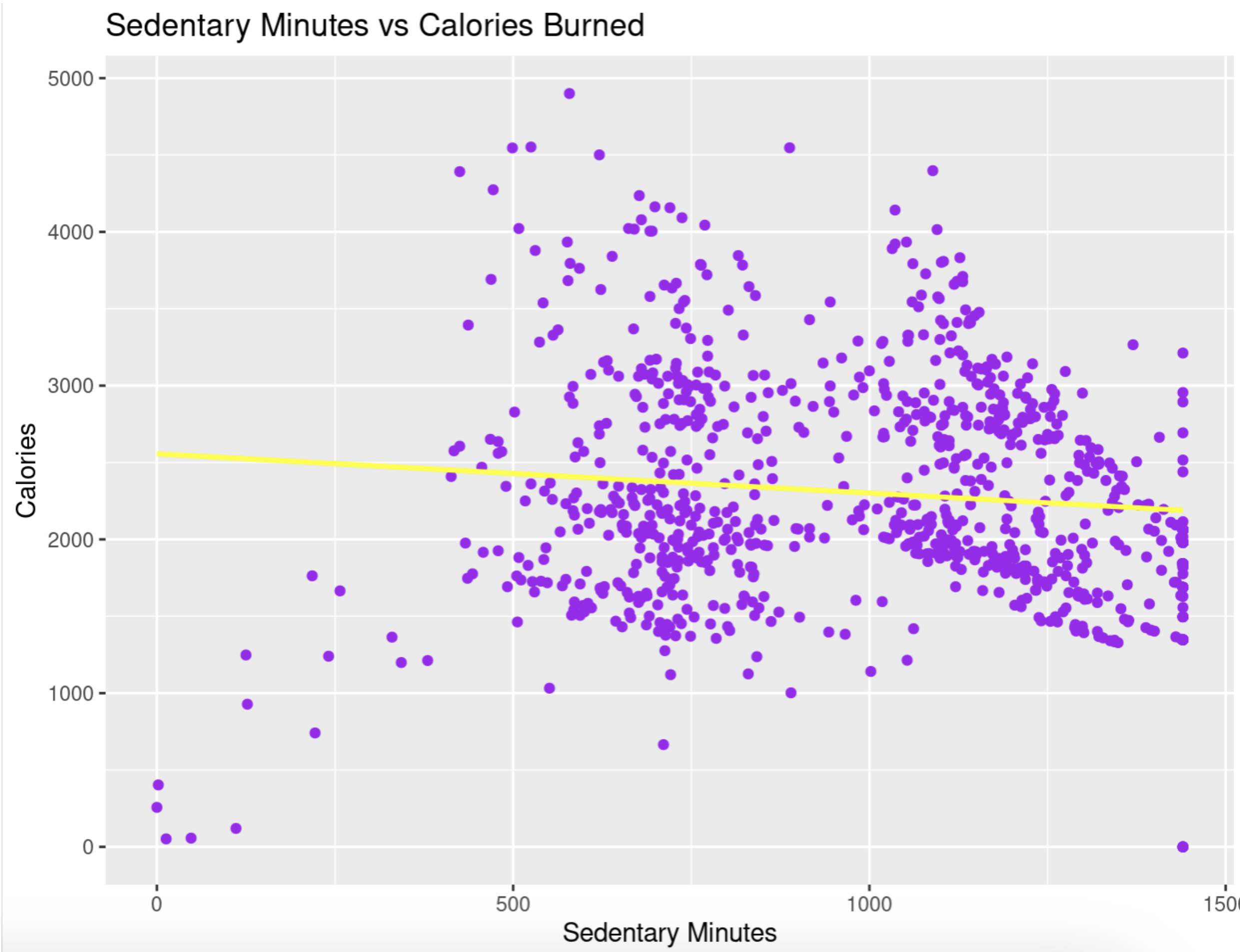
Total Steps vs Calories Burned

```
> ggplot(unique_daily_activity, aes(x = TotalSteps, y = Calories)) +
+     geom_point(color = "green") +
+     geom_smooth(method = "lm", se = FALSE, color = "red") +
+     labs(title = "Total Steps vs Calories Burned",
+          x = "Total Steps",
+          y = "Calories")
```

Again, we see a direct relationship between total steps and calories burned.

Sedentary Minutes vs Calories Burned

```
> ggplot(unique_daily_activity, aes(x = SedentaryMinutes, y = Calories)) +
+     geom_point(color = "purple") +
+     geom_smooth(method = "lm", se = FALSE, color = "yellow") +
+     labs(title = "Sedentary Minutes vs Calories Burned",
+          x = "Sedentary Minutes",
+          y = "Calories")
```



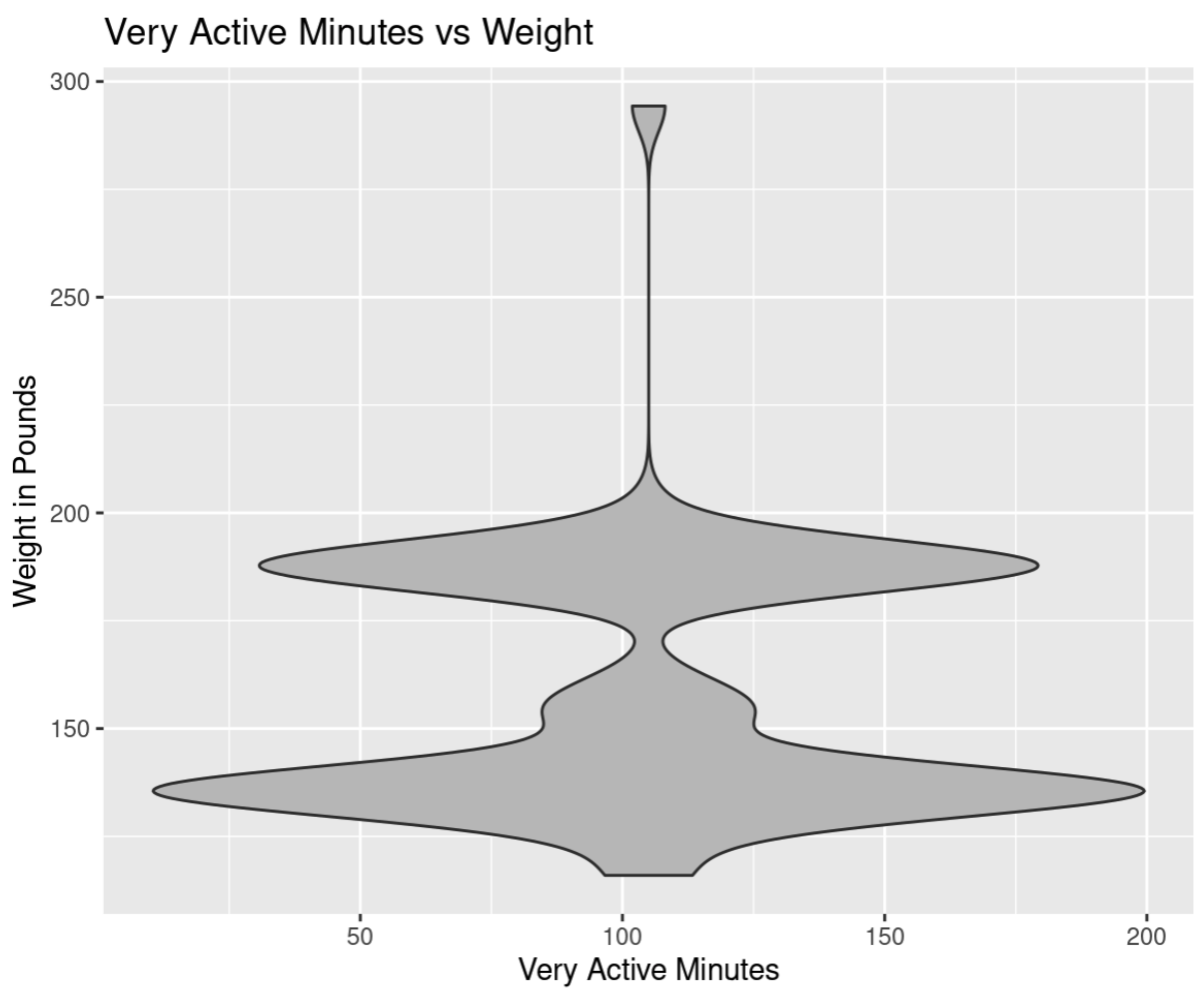
The resulting scatterplot shows an indirect relationship but is less pronounced than the direct relationships from the graphs above. This suggests that calories can still be burned even with large amounts of one's day being inactive.

Very Active Minutes vs Weight (lbs)

A violin plot was created to illustrate which weight groups are most active:

```
> merged_data <- merge(unc_unique_weight_log, unique_daily_activity, by = "Id")
```

```
> ggplot(merged_data, aes(x = VeryActiveMinutes, y = WeightPounds)) +
+     geom_violin(fill = "grey") +
+     labs(title = "Very Active Minutes vs Weight",
+          x = "Very Active Minutes",
+          y = "Weight in Pounds")
```

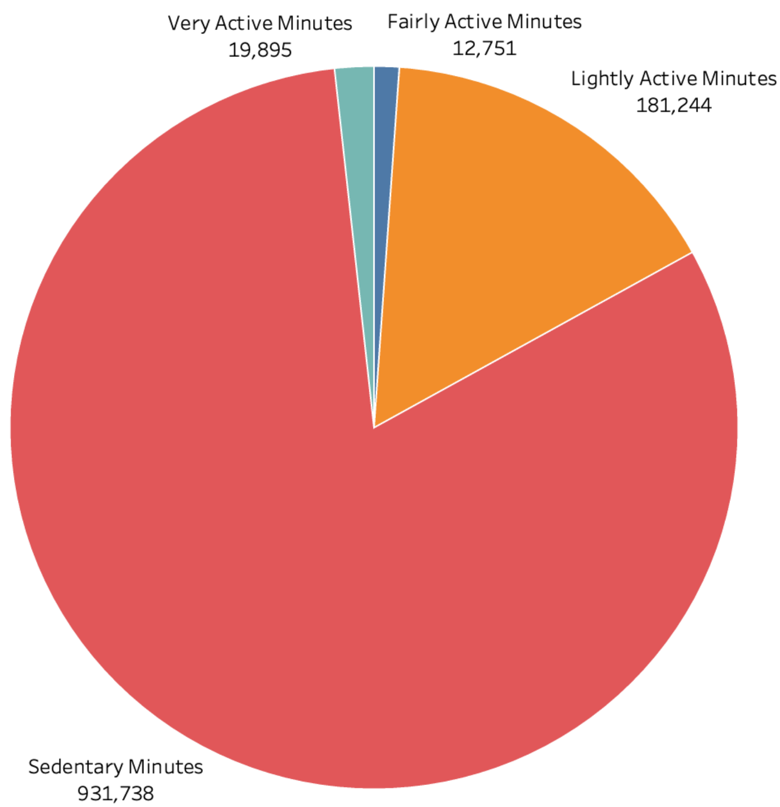


Participants between 125 to 175 pounds were the most active.

Share

Activity Level Pie Chart

A pie chart was created in Tableau, which was supplemented with each activity level's percentage. 81% of daily activity among all people was spent inactive, the remaining 19% was spent in varying degrees of activity. Over 97% of daily activity was spent either lightly active or completely inactive. 2.86% of a day amounts to about 40 minutes of high-intensity or fairly high-intensity activity.



	Activity Level				
	Very Active	Fairly Active	Lightly Active	Sedentary	Total
Percentage	1.74%	1.12%	15.83%	81.33%	100%

Average Hours of Sleep by Day

The 'SleepDay' column contained both the date and time, 'sleep_day' was created, separating the date from the time:

```
> unique_sleep_data$sleep_day <- substr(unique_sleep_data$SleepDay, 1, 9)
```

The ‘TotalMinutesAsleep’ column was converted into hours in a new column called ‘TotalHoursAsleep’:

```
> unique_sleep_data <- unique_sleep_data %>%
+     mutate(TotalHoursAsleep = TotalMinutesAsleep / 60)
```

Then, the column, ‘sleep_day’, was formatted as a date:

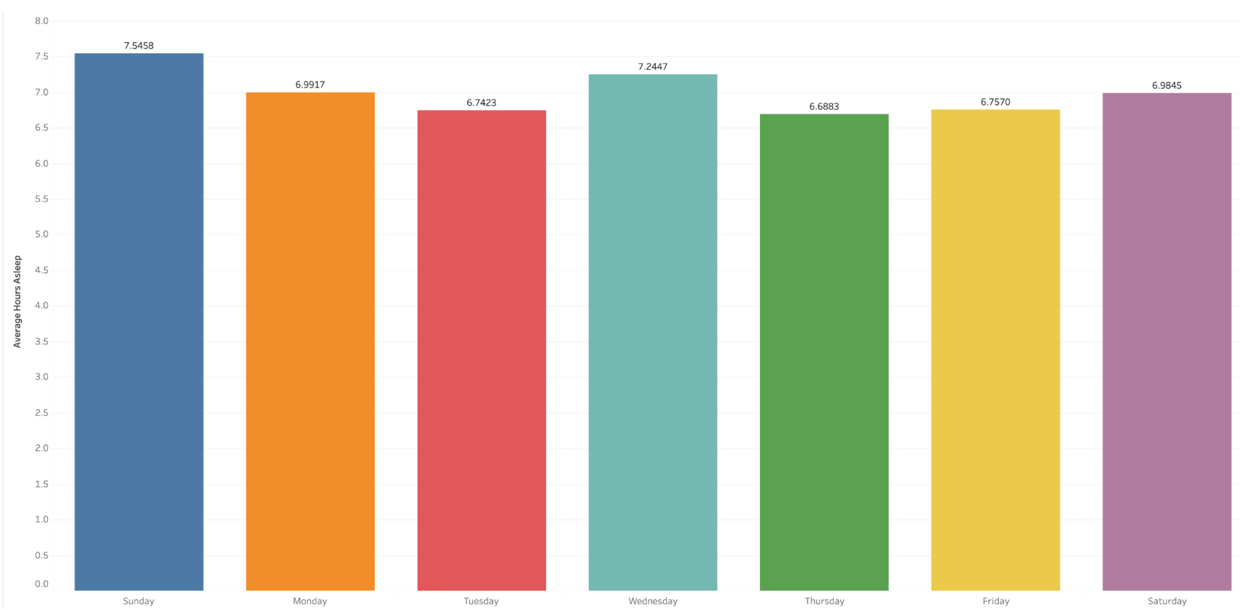
```
> unique_sleep_data$sleep_day <- as.Date(unique_sleep_data$sleep_day, format = "%
m/%d/%Y")
```

A new column, ‘day_of_week’, transformed the dates from ‘sleep_day’ to days of the week:

```
> unique_sleep_data$day_of_week <- weekdays(unique_sleep_data$sleep_day)
```

The average amount of sleep in hours for each day of the week was calculated, yielding this bar graph:

```
> average_sleep_per_day <- unique_sleep_data %>%
+     group_by(day_of_week) %>%
+     summarise(average_hours_asleep = mean(TotalHoursAsleep, na.rm = TRUE))
```



I was expecting significantly more sleep on Friday and Saturday compared to the middle of the work week, but that is evidently not the case. In fact, when comparing the amount of sleep on Monday to Saturday and Tuesday to Friday, you’ll find almost no differences. More context could be given if the data specified the time at which each sleep log began or ended. The fact that a sample of 24 participants slept nearly the same amount regardless of the day of the week suggests that parts of life, like work or going out on the weekends, did not impact sleep length. Presumably, the body has an internal clock; going to bed earlier, like during the work week, equates to waking up earlier. Similarly, going to bed later, like during the weekend, equates to waking up proportionally later.

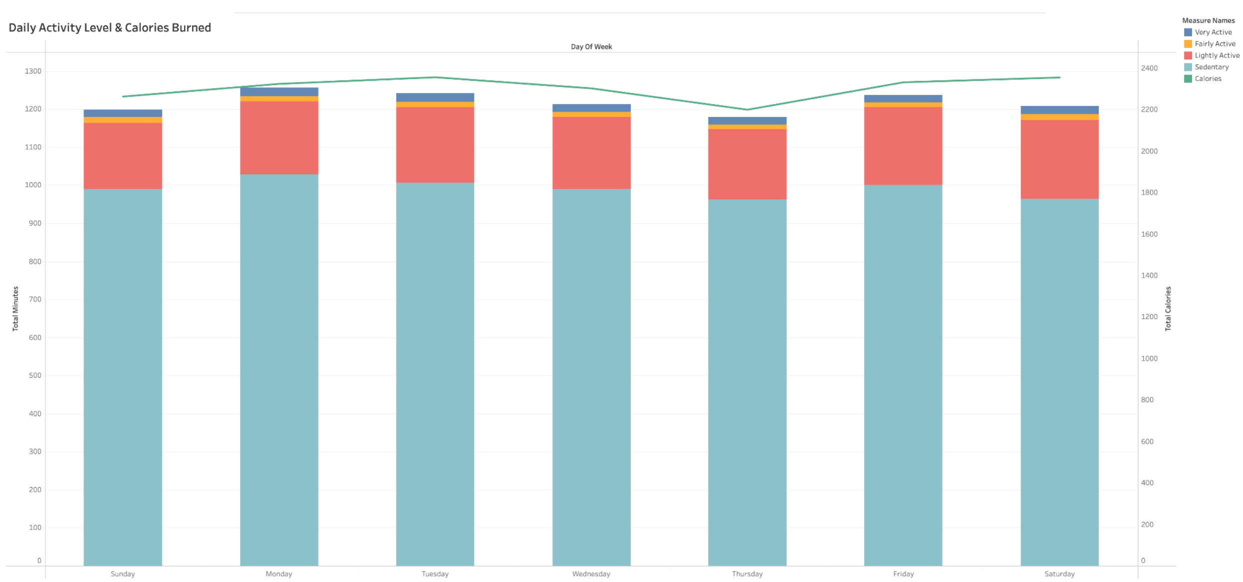
Activity Level & Burned Calories by Day

To examine how burned calories and activity level change based on the day of the week, the syntaxes below were respectively entered:

```
> average_calories_per_day <- unique_daily_activity %>%
+     group_by(day_of_week) %>%
+     summarise(average_calories = mean(Calories, na.rm = TRUE))
```

```
> average_activity_per_day <- unique_daily_activity %>%
+     group_by(day_of_week) %>%
+     summarise(
+         average_VeryActiveMinutes = mean(VeryActiveMinutes, na.rm = TRUE),
+         average_FairlyActiveMinutes = mean(FairlyActiveMinutes, na.rm = TRUE),
+         average_LightlyActiveMinutes = mean(LightlyActiveMinutes, na.rm = TRUE),
+         average_SedentaryMinutes = mean(SedentaryMinutes, na.rm = TRUE)
+     )
```

Both tables produced were imported to Tableau to create a combination chart.



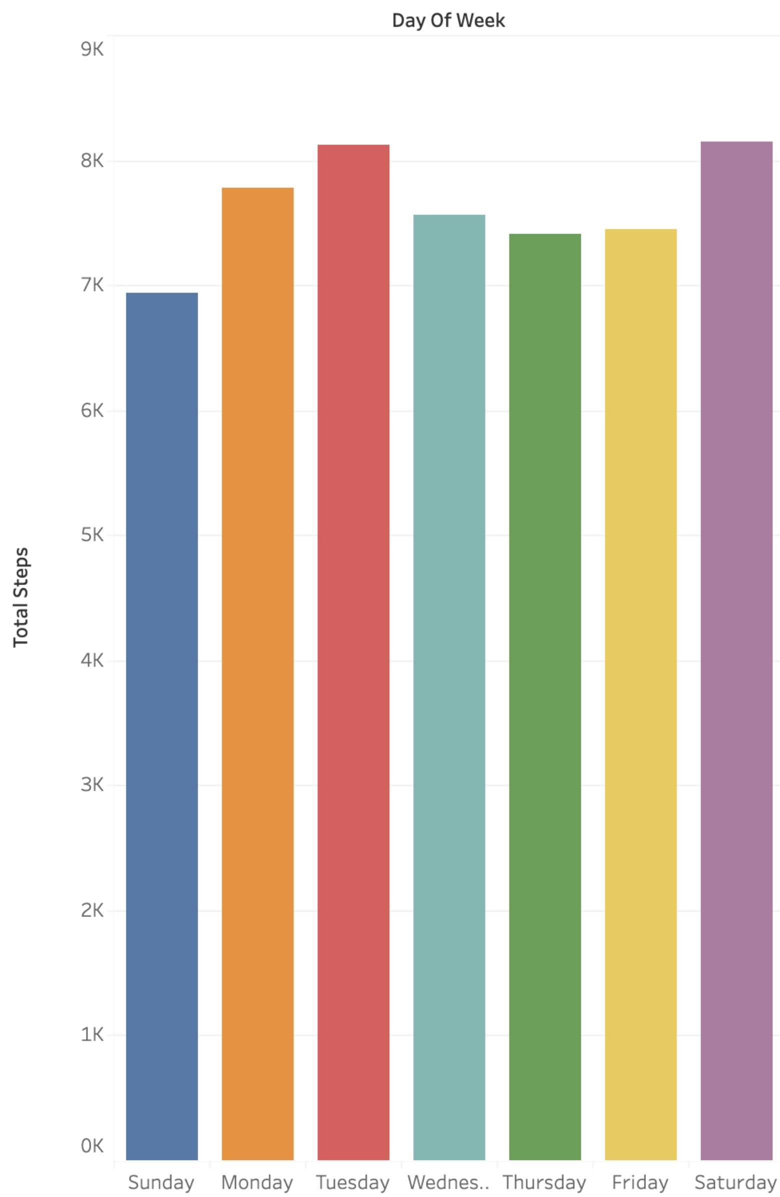
There is a clear relationship between calories burned and activity level. When activity level suffers, so does the amount of calories burned.

Average Steps by Day

The average number of steps taken per weekday was calculated with this syntax:

```
> average_steps_per_day <- unique_daily_activity %>%
+     group_by(day_of_week) %>%
+     summarise(average_steps = mean(TotalSteps, na.rm = TRUE))
```

Tableau summarized that data with a bar graph:



This graph is consistent with the combination chart from above and the idea that activity is cyclical - steps peak early in the work week, before dropping from Wednesday to Friday, and then going back up on Saturday. Sunday, often referred to as a day of rest, averaged the least amount of steps.

Act

Bellabeat's founders, Urška Sršen and Sando Mur, knew the company was capable of more, so they looked toward the power of data and the insights it could bring. Analyzing data about similar products on the market ultimately facilitated recommendations for future Bellabeat products and their subsequent marketing strategies.

The business task was to apply the findings of the data analysis to a single Bellabeat product. Bellabeat offers a multitude of products but one of the most essential parts of all of them is the Bellabeat App. Without a well-designed app, all of Bellabeat's products are inherently less insightful, which is ultimately why my recommendations are specific to the Bellabeat App.

Providing consumers with useful app features will immediately improve a range of Bellabeat products. Consumers in this market may vary in their pursuit of physical fitness, which is why Bellabeat+ should exist. Bellabeat+, $8.99/month, is geared towards very active consumers who want to optimize their current health standards: sleep, diet, and activity level. Bellabeat+ is powered by an ai assistant - Bellabot, which gives the app character and performs quintessential parts of the premium subscription. On the other hand, the free version of the Bellabeat App is designed for consumers looking to *become* more physically active.

Bellabeat+

Sleep Tracking

- Bellabot will learn how long you sleep and what time you typically wake up on specific days. It will notify you what time to be asleep to achieve what your body considers healthy sleep. Bellabot will also notify you about any trends in your sleep, including failure to meet healthy sleep standards.

Nutrition Tracking

- Bellabeat+ allows users to input all of the foods and drinks they consume either via search or barcode scan. Simply scan the food or drink packaging you consume and all major nutritional facts will automatically be counted towards your daily total. If your food or drink has no barcode, you can search for the item in the food/drink database.   Bellabot will recommend daily caloric, protein, fat, and carbohydrate intake based on your current weight and targeted weight. Bellabot also suggests an appropriate target date for meeting said desired weight, encouraging healthy eating habits and sustainable weight changes.

Cardio Progression System

- Users will be prompted to enter how many times a week they aim to run. Next, they'll enter their current run distance and time and set a run distance and time goal. The goal also comes with a deadline decided by the user themselves or Bellabot can recommend a healthy goal deadline. By tracking runs with a Bellabeat smartwatch, Bellabot will let users know how they are progressing toward their running goals.

Weight Lifting Progression System

- Bellabeat+ members will be prompted to enter their weekly workout plan by selecting exercises from a database. They also have the option to let Bellabot generate a workout plan. After, they'll enter their current reps and weight of any workout they want to track. Then, a reps and weight goal for said workouts will be set along with a deadline for achieving the goal. Again, Bellabot will recommend a healthy target date, which helps prevent overexertion. How frequently a user must update their current exercise reps and weight is dependent on when their goal deadline is.

The Free Version of the Bellabeat App

There are a few big differences between the free version of the Bellabeat App and Bellabeat+. First, smart bedtime notifications are not included and users can only view their last week of sleep logs. Second, in regard to nutrition tracking, there is no barcode tracking in the free version of the Bellabeat App. Lastly, the Cardio & Weight Lifting Progression System is exclusive to Bellabeat+, but purchasing a Bellabeat product comes with a 3-month free trial.

References

Edward R. Laskowski, M. D. (2023, July 26). *How much exercise do you really need?*. Mayo Clinic. https://www.mayoclinic.org/healthy-lifestyle/fitness/expert-answers/exercise/faq-20057916#:~:text=As%20a%20general%20goal%2C%20aim,sitting%20time%20is%20important%2C%20too.

MediLexicon International. (n.d.-a). *Average steps per day by age, sex, and occupation*. Medical News Today. https://www.medicalnewstoday.com/articles/average-steps-per-day#summary

MediLexicon International. (n.d.-b). *Calories burned in a day: Calculation, factors, exercise, weight loss*. Medical News Today. https://www.medicalnewstoday.com/articles/319731