Case Study:
How can a wellness company play it smart?
By Estelle Martin
5/28/2022

I. Introduction:

The purpose of this case study is to help the technology-driven company Bellabeats make better informed decisions about their future products within their target audience: women and their health habits. The goal is to identify trends in smart device usage by users to be able to inform and inspire people around the world to care about their health. In this study I will perform calculations and create meaningful visualizations of smart device utilization to gain insights on growth opportunities to advise Bellabeats's future marketing strategy.

II. Phase 1 - Ask:

- *Stakeholders:*

- *Urška Sršen*: Bellabeat's cofounder and Chief Creative Officer
- *Sando Mur*: Mathematician and Bellabeat's cofounder

- *Business Task:*

What are people's smart devices utilization habits? How can we improve the marketing strategy of Bellabeats to cater to those needs and reveal opportunities for growth?

- *What is the problem I am trying to solve?*

Analyzing company data in order to draw insights that will promote Bellabeats' growth as a market leader in technology for women relating to their health habits.

- *What product is the analytics team focusing on?*

The Bellabeat App will be the center of this analysis and where insights will be applied to at the end in the form of recommended actions. Device usage data will be drawn from 'FitBit' datasets and will serve to study trends in four key areas: daily activity, intensities, sleep, and weight logs.

- *How can my insights drive business decisions?*

Based on my findings, Bellabeats will be able to decide which direction to take based on understanding smart devices' utilization by women. The company needs to know more about their target audience in order to cater to them by making products they will relate to and love.

III. Phase 2 - Prepare:

- ***Where is your data stored?***

Data has been pulled from the following Kaggle dataset: “Fitbit Fitness Tracker Data,” public domain data by Mobius. I have downloaded the csv files and opened them in Google Sheets for data cleaning, then later in R for further analysis.

- ***How is the data organized? Is it in long or wide format?***

The data is organized in long format. In Sheets, each column A of all four datasets represents the IDs of all the Fitbit users who consented to using this data for analysis or who were willing to log data based on their interests. Each row repeats each ID multiple times as the data was collected over multiple days in a two-month period. For example, in the Daily Activity dataset row 1 displays the variables collected on any given day such as ActivityDate, TotalSteps, TotalDistance, Calories etc..

- ***Are there issues with bias or credibility in this data? Does your data ROCCC?***

Stakeholder Sršen has mentioned that the data might contain limitations. Some potential limitations include:

- Stakeholders would like an analysis of Bellbeats products, however the data available is from popular consumer electronics and fitness company ‘FitBit’.
- How do we define the term ‘eligible Fitbit users’ stated in the dataset summary. E.g.: Do they identify as men, women, or non-binary? Do they have a disability that could affect the data? Which Fitbit item were they using during this test? What is the age range?
- Ongoing random life events in these people’s lives could have affected the data.
- The data only comes from one source, is outdated, and the sample can be less than 30.
- Sampling bias: How do we go around the fact that we are looking for data from ‘women’ who are the target of this analysis, yet the ‘Fitbit users’ aren’t defined as women only?

Nonetheless, the data is reliable, original, comprehensive, and cited which means that it still ROCCC. The dataset was last updated a year ago. For the purpose of this fictional analysis we will ignore the relevance of the dates when the data was pulled, which are from 03.12.2016 to 05.12.2016. It is important to note that if this were a current real life analysis to predict future strategies, this data would be outdated and of no use to draw insights from in 2022.

- ***How are you addressing licensing, privacy, security, and accessibility?***

The data was pulled from a trusted Kaggle user and is licensed under CC0 Public Domain. It is rated to be uncopyrighted. The Creative Commons states: “You can copy, modify, distribute and perform the work, even for commercial purposes, all without asking permission.” This means that the FitBit users from this dataset have waived their rights to the data to the public domain. It has a score of 10.0 for usability and remains accessible to all.

- *How did you verify the data's integrity?*

To verify the data's integrity I need to verify that the data types match the columns and that enough data is available to complete the study. After taking a glance at the dataset I can tell that there are 33 total IDs and that some data is missing. No user has all 60 possible entries (one for each day of the 2-month period) which indicates that many users probably didn't wear the smart device every day of the month. It is also worth mentioning that the study states it occurred between 03/12/2016 and 05/12/2016 however all data entries date from 4/12/2016 and up to 5/12/2016.
There is also a concern of sampling bias. Since the term 'FitBit users' is undefined, I cannot say for sure if the data is representative of the population as a whole. Moreover, Bellabeats' model and business goals focus on women's smart device usage, which is more than likely not every user in this sample. This is an important limitation to data integrity to keep in mind.

- *How does it help you answer your question?*

The preparation phase of the analysis helps me brainstorm the business task by allowing me to ponder multiple questions. This way, I can redirect my analysis towards areas that cause conflict or prove to be opportunities for growth. By sorting and filtering the data, I can efficiently identify outliers and null values that need to be addressed. In order to ensure the validity and integrity of my analysis I will make sure to account for all limitations to the dataset.

- *Are there any problems with the data?*

Other than a few missing entries and obvious limitations, there does not seem to be any more problems with the data at this stage.

IV. Phase 3 - Process:

- *What tools are you choosing and why?*

To begin taking a look at the data I chose to use Google Sheets in order to get a general idea of what the dataset looks like. This way, I can determine early on if there are gaps in the data. I also wanted the challenge of analyzing a large data set in Google Sheets to be able to see its limitations. Although there are many rows to go through I still found it useful to visualize the dataset in this format.
I then decided that I will be using R for the rest of the analysis due to the size of the dataset and the opportunities to create impactful visualizations.

- *What steps have you taken to ensure that your data is clean?*

  - *With Google Sheets:*

- <u>Data cleaning the dailyActivity merged file:</u>

- I text wrapped the header row and color-filled it for visibility.
- I deleted the extra empty rows at the bottom of the set to avoid confusion while sorting and filtering.
- I used conditional formatting to check for blank cells and found 1:
  ```
  =IF(ISBLANK(N4),IFERROR(0/0),"not blank")
  ```
- I removed duplicates and none were found.
- I increased decimal places for all distance columns to increase legibility.
- I color-coded each ID to reveal patterns.
- I formatted the “ActivityDate” column to show all dates in MM/DD/YYYY format.
- I checked to see if each ID had the same amount of entries for the enteriety of the trial and they do not. I then decided to delete rows with “0” data all across the board assuming that the smart device wasn’t worn that day. At this time I am only focusing on the data collected rather than taking into consideration “off” days. I used the “TotalSteps” column and filtered it to only show “0” values then deleted the rows.

I then repeated the cleaning process for the sleepDay_merged, dailyIntensities_merged, and weightLogInfo_merged. I renamed the files to ‘edited’ before I processed them into R.

<u>Note:</u> Although this is more time consuming than using R, I wanted to sharpen my Excel skills for my first case study and took the opportunity to practice.

➢ *<u>With R Studio:</u>*

- <u>Extra data cleaning steps in R for daily_activity dataset:</u>

```
### Renaming 'ActivityDate' column to 'Date'
colnames(daily_activity)[colnames(daily_activity) == "ActivityDate"]
<- "Date"

### Changing 'Date' data type from character to date type
daily_activity$Date <-
as.Date(daily_activity$Date,format='%m/%d/%Y')

### Changing 'SedentaryMinutes' data type from character to numeric
daily_activity$SedentaryMinutes <-
suppressWarnings(as.numeric(daily_activity$SedentaryMinutes))

### Adding a 'Weekdays' column for future analysis
daily_activity$Weekdays<- weekdays(daily_activity$Date, abbr = FALSE)
```

- <u>Extra data cleaning steps in R for sleep_day dataset:</u>

```
### Renaming 'SleepDay' column to 'Date'
colnames(sleep_day)[colnames(sleep_day) == "SleepDay"] <- "Date"

### Changing 'Date' data type from character to date type
sleep_day$Date <-  as.Date(sleep_day$Date,format='%m/%d/%Y')

### Adding a 'Weekdays' column for future analysis
sleep_day$Weekdays<- weekdays(sleep_day$Date, abbr = FALSE)
```

- <u>Extra data cleaning steps in R for weight_log dataset:</u>

```
### Changing 'Date' data type from character to date type
weight_log$Date <-  as.Date(weight_log$Date,format='%m/%d/%Y')

### Adding a 'Weekdays' column for future analysis
weight_log$Weekdays<- weekdays(weight_log$Date, abbr = FALSE)
```

- <u>Extra data cleaning steps in R for daily_intensities dataset:</u>

```
### Renaming 'ActivityDay' column to 'Date'
colnames(daily_intensities)[colnames(daily_intensities) ==
"ActivityDay"] <- "Date"

### Changing 'Date' data type from character to date type
daily_intensities$Date <-
as.Date(daily_intensities$Date,format='%m/%d/%Y')

### Adding a 'Weekdays' column for future analysis
daily_intensities$Weekdays<- weekdays(daily_intensities$Date, abbr =
FALSE)
```

- <u>Calculating the number of unique IDs in each dataset using Google Sheets functions:</u>

I inserted a column to the right and labeled it “Unique IDs.” I then calculated a list of unique IDs in column A using the `UNIQUE()` function: `=UNIQUE(A2:A863)`. To total the number of IDs I used the `COUNT()` function: `=COUNT(P2:P34)`.

I repeated these steps for each data set:

dailyActivity_entries = 33
dailyIntensities_entries = 33
weightlogInfo_entries = 8
sleepDay_entries = 24

These results are interesting, it tells me that out of 33 people all of them logged information about their daily activities and intensities, while only 8 logged weight info and 24 logged sleep entries. This early information will be relevant in the rest of the analysis.

- *<u>How can you verify that your data is clean and ready to analyze?</u>*

To verify that my data is cleaned and ready to analyze I double checked the list of commands I performed and loaded the new edited datasets into R.
Since the datasets aren't too long, I am able to glance over them and troubleshoot for errors.

- *<u>Have you documented your cleaning process so you can review and share those results?</u>*

Yes, I am using this Google Doc to report my case study analysis and document my cleaning process so I can share it with others as needed. I am laying out all my thoughts and processes in this document so I can refer back to it later. I will also be creating an R Markdown file to document all the analysis work performed in R.

V. Phase 4 - Analyze Data with R

- *<u>How should you organize your data to perform analysis on it?</u>*

At this point, I want to refer back to the business task to evaluate the approach I want to take without stirring away from business goals. Now that my data is clean, I decide to import the datasets into R for further analysis.

❖ ***Business Task**: What are people's smart device utilization habits?*

In order to answer this question I chose to create data frames in R to easily access data such as daily activity, daily sleep, daily intensities, and weight log:

```
daily_activity <- read_csv("dailyActivity_merged.csv")

sleep_day <- read_csv("sleepDay_merged.csv")

daily_intensities <- read_csv("dailyIntensities_merged.csv")

weight_log <- read_csv("weightLogInfo_merged.csv")
```

- *Comparing datasets:*

I start by comparing the number of rows in the daily activity set versus the data in the daily sleep set to see if they match or if I can identify a pattern:

```
daily_activity_nrow <- nrow(daily_activity)
sleep_day_nrow <- nrow(sleep_day)
daily_intensities_nrow <- nrow(daily_intensities)
weight_log_nrow <- nrow(weight_log)

data.frame(daily_activity_nrow, sleep_day_nrow, daily_intensities_nrow,
weight_log_nrow)

daily_activity_nrow sleep_day_nrow daily_intensities_nrow weight_log_nrow
[1]             862            410                    939              67
```

This is also achievable by looking at the data in the R environment:

daily_activity	862 obs. of 15 variables
daily_intensities	939 obs. of 12 variables
sleep_day	410 obs. of 7 variables
weight_log	67 obs. of 10 variables

This data tells me that there aren't the same amount of entries for each dataset. This more than likely means that people are most interested in common daily activity trends rather than using smart devices to keep track of their weight for example. At this stage we could assume that the data collected in the daily_Activity dataset automatically gets loaded when wearing the device while most people probably don't wear a smart device to sleep.

- *Creating summary tables to get better insights into our data:*

```
### Summary of daily_activity dataset:

daily_activity %>%
  select(TotalSteps,
         TotalDistance,
         Calories,
         VeryActiveMinutes,
         FairlyActiveMinutes,
         LightlyActiveMinutes) %>%
  summary()
```

OUTPUT:

```
 TotalSteps    TotalDistance       Calories    VeryActiveMinutes
FairlyActiveMinutes
 Min.   :    8   Min.   : 0.010   Min.   :  52   Min.   :  0.00    Min.
:  0.00
 1st Qu.: 4927   1st Qu.: 3.373   1st Qu.:1857   1st Qu.:  0.00    1st
Qu.:  0.00
 Median : 8054   Median : 5.590   Median :2220   Median :  7.00    Median
:  8.00
 Mean   : 8329   Mean   : 5.986   Mean   :2362   Mean   : 23.04    Mean
: 14.79
 3rd Qu.:11096   3rd Qu.: 7.905   3rd Qu.:2832   3rd Qu.: 35.00    3rd
Qu.: 21.00
 Max.   :36019   Max.   :28.030   Max.   :4900   Max.   :210.00    Max.
:143.00


 LightlyActiveMinutes
 Min.   :  0.0
 1st Qu.:147.0
 Median :208.5
 Mean   :210.3
 3rd Qu.:272.0
 Max.   :518.0
```

Summary of sleep_day dataset:

```
sleep_day %>%
  select(TotalSleepRecords,
         TotalMinutesAsleep,
         TotalTimeInBed) %>%
  summary()
```

OUTPUT:

```
TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
 Min.   :1.00      Min.   : 58.0      Min.   : 61.0
 1st Qu.:1.00      1st Qu.:361.0      1st Qu.:403.8
 Median :1.00      Median :432.5      Median :463.0
 Mean   :1.12      Mean   :419.2      Mean   :458.5
 3rd Qu.:1.00      3rd Qu.:490.0      3rd Qu.:526.0
 Max.   :3.00      Max.   :796.0      Max.   :961.0
```

Summary of daily_intensities dataset:

```
daily_intensities %>%
  select(Date,
         SedentaryMinutes,
         LightlyActiveMinutes,
         FairlyActiveMinutes,
         VeryActiveMinutes,
         SedentaryActiveDistance,
         LightActiveDistance,
         ModeratelyActiveDistance,
         VeryActiveDistance) %>%
  summary()
```

OUTPUT:

```
Date            SedentaryMinutes LightlyActiveMinutes FairlyActiveMinutes
VeryActiveMinutes
 Min.   :2016-04-12   Min.   :   2.0   Min.   :  0          Min.   :  0.00
Min.   :  0.00
 1st Qu.:2016-04-19   1st Qu.: 730.0   1st Qu.:127          1st Qu.:  0.00
1st Qu.:  0.00
 Median :2016-04-26   Median :1058.0   Median :199          Median :  6.00
Median :  4.00
 Mean   :2016-04-26   Mean   : 992.3   Mean   :193          Mean   : 13.58
Mean   : 21.19
 3rd Qu.:2016-05-04   3rd Qu.:1230.0   3rd Qu.:264          3rd Qu.: 19.00
3rd Qu.: 32.00
 Max.   :2016-05-12   Max.   :1440.0   Max.   :518          Max.   :143.00
Max.   :210.00

 SedentaryActiveDistance LightActiveDistance ModeratelyActiveDistance
VeryActiveDistance
 Min.   :0.000000        Min.   : 0.000      Min.   :0.0000           Min.
: 0.000
 1st Qu.:0.000000        1st Qu.: 1.950      1st Qu.:0.0000           1st
Qu.: 0.000
 Median :0.000000        Median : 3.370      Median :0.2400
Median : 0.210
 Mean   :0.001608        Mean   : 3.344      Mean   :0.5681           Mean
: 1.504
 3rd Qu.:0.000000        3rd Qu.: 4.785      3rd Qu.:0.8000           3rd
Qu.: 2.065
 Max.   :0.110000        Max.   :10.710      Max.   :6.4800           Max.
:21.920
```

```
### Summary of weight_log dataset:

weight_log %>%
  select(Date,
         WeightKg,
         WeightPounds,
         Fat,
         BMI,
         IsManualReport) %>%
  summary()
```

OUTPUT:

```
Date                  WeightKg        WeightPounds        Fat              BMI          IsManualReport
 Min.   :2016-04-12   Min.   : 52.60   Min.   :116.0   Min.   :22.00   Min.   :21.45   Mode :logical
 1st Qu.:2016-04-19   1st Qu.: 61.40   1st Qu.:135.4   1st Qu.:22.75   1st Qu.:23.96   FALSE:26
 Median :2016-04-27   Median : 62.50   Median :137.8   Median :23.50   Median :24.39   TRUE :41
 Mean   :2016-04-26   Mean   : 72.04   Mean   :158.8   Mean   :23.50   Mean   :25.19
 3rd Qu.:2016-05-04   3rd Qu.: 85.05   3rd Qu.:187.5   3rd Qu.:24.25   3rd Qu.:25.56
 Max.   :2016-05-12   Max.   :133.50   Max.   :294.3   Max.   :25.00   Max.   :47.54
                                                       NA's   :65
```

VI. Phase 5 - Share Visualizations

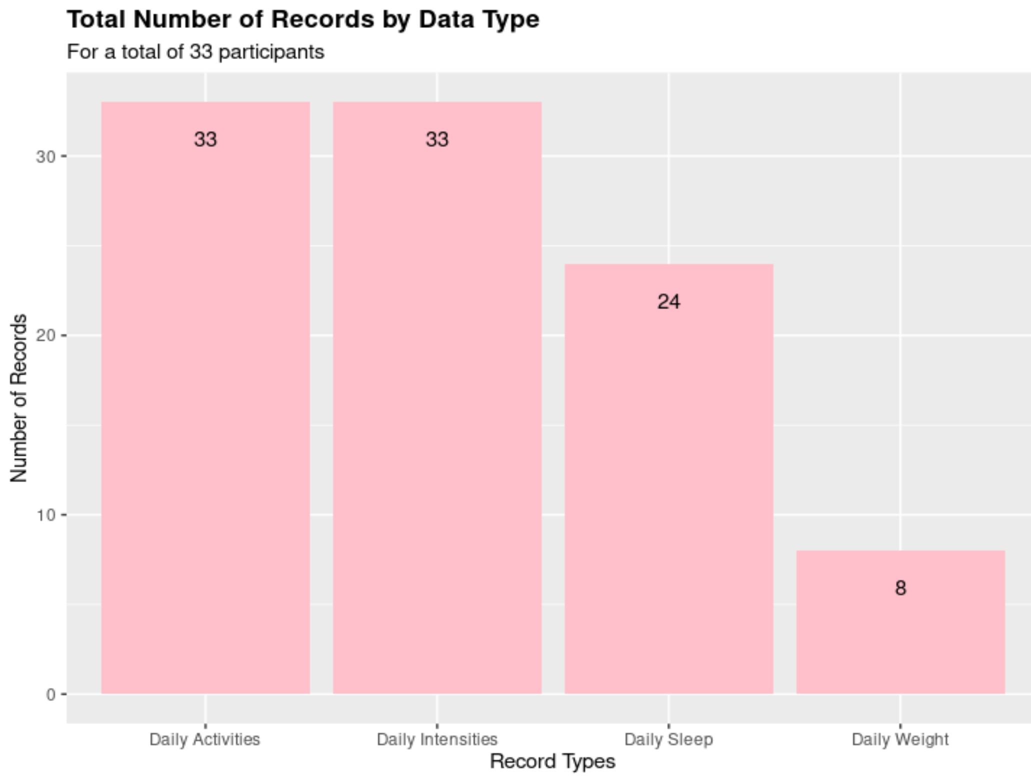
- *Establishing relationships in our datasets:*

1. Relationship between number of entries and data types:

```
record_type <- c("Daily Activities","Daily Intensities","Daily
Sleep","Daily Weight")
record_unique_IDs <-
c(n_distinct(daily_activity$Id),n_distinct(daily_intensities$Id
),n_distinct(sleep_day$Id), n_distinct(weight_log$Id))
record_number <- data.frame(record_type,record_unique_IDs)

ggplot(record_number, aes(record_type,record_unique_IDs)) +
  geom_bar(fill="pink", stat="identity") +
  geom_text(aes(label=record_unique_IDs), vjust=3,
element_text(face="bold"))+
  labs(title="Total Number of Records by Data Type",
       subtitle="For a total of 33 participants",
       x="Record Types", y="Number of Records") +
  theme(plot.title = element_text(face="bold"))
```

OUTPUT:

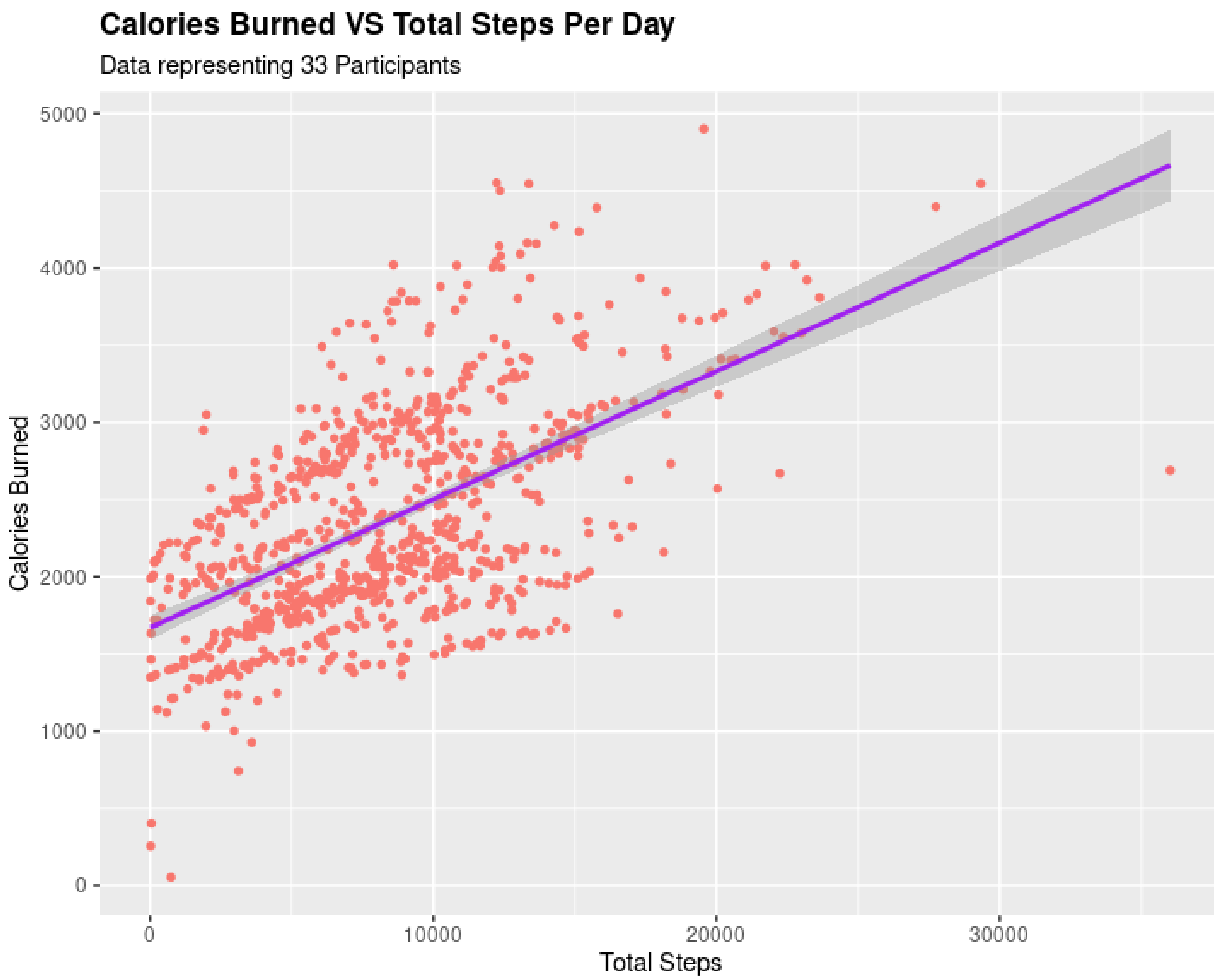


Conclusion: The data shows that not everyone has entries for each type of data that the Bellabeats App has to offer. At this point we can assume that people prefer to look at their daily activities and intensities stats rather than logging in their weight or sleep. We do need to keep in mind that those stats might be automated while logging weight and sleep might not.

2. Relationship between total steps and calories burned:

```
ggplot(daily_activity, aes(x=TotalSteps, y=Calories, color =
"orange")) +
  geom_point(size=1.25) +
  geom_smooth(method=lm, color="purple") +
  labs(title="Calories Burned VS Total Steps Per Day", subtitle="Data
representing 33 Participants", x="Total Steps", y="Calories Burned")+
  theme(legend.position="none")
```

OUTPUT:



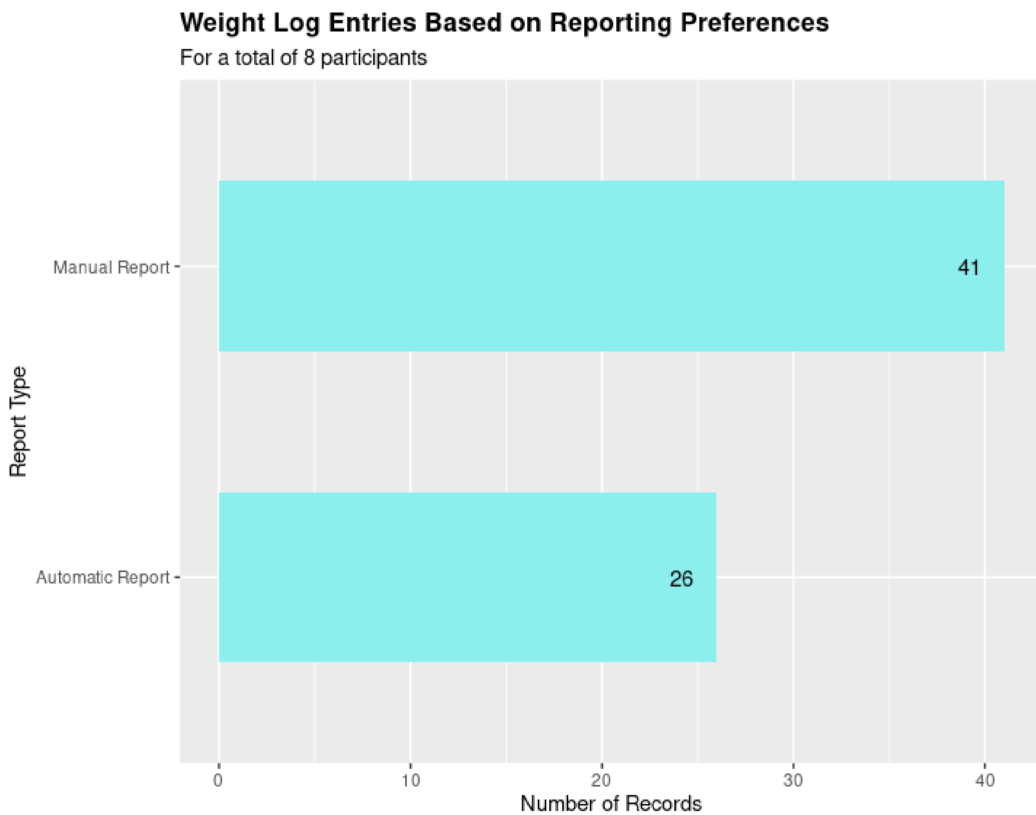
Conclusion: The data shows an upward trend meaning that the more steps people take the more calories they burn. Since all 33 participants have logged data about total steps and calories burned we can assume that these are data points that are of interest to them.

3. Graph representing people's preferences regarding logging in their weights:

```
report_type <- c("Manual Report","Automatic Report")
report_method <- c(length(which(weight_log$IsManualReport =="TRUE")),
                   length(which(weight_log$IsManualReport =="FALSE")))
report_preference <- data.frame(report_type,report_method)

ggplot(report_preference, aes(report_type,report_method)) +
  geom_bar(fill="darkslategray2", stat="identity", width = 0.55) +  coord_flip() +
  geom_text(aes(label=report_method), vjust=0.55, hjust=2) +
  labs(title="Weight Log Entries Based on Reporting Preferences",
       subtitle="For a total of 8 participants",
       x="Report Type", y="Number of Records") +
  theme(plot.title = element_text(face="bold"))
```

<u>OUTPUT:</u>



Conclusion: Out of 8 participants, most of the time people chose to manually report their weight rather than automatically let the smart device do it for them. This could be due to the fact that people found manual entry to be more accurate than what the AI suggests. This could be a potential opportunity for Bellabeats to improve their product.

4. Determining people's sleeping patterns based on weekdays

Creating a tibble to only see weekdays and average minutes asleep:

```
sleep_schedule <- sleep_day %>%
  group_by(Weekdays) %>%
  summarize(mean_TotalMinutesAsleep = mean(TotalMinutesAsleep))
head(sleep_schedule)
```

<u>OUTPUT:</u>

```
# A tibble: 6 × 2
  Weekdays         mean_TotalMinutesAsleep
  <chr>                       <dbl>
1 Friday                       405.
2 Monday                       420.
3 Saturday                     419.
4 Sunday                       453.
5 Thursday                     401.
6 Tuesday                      405.
```

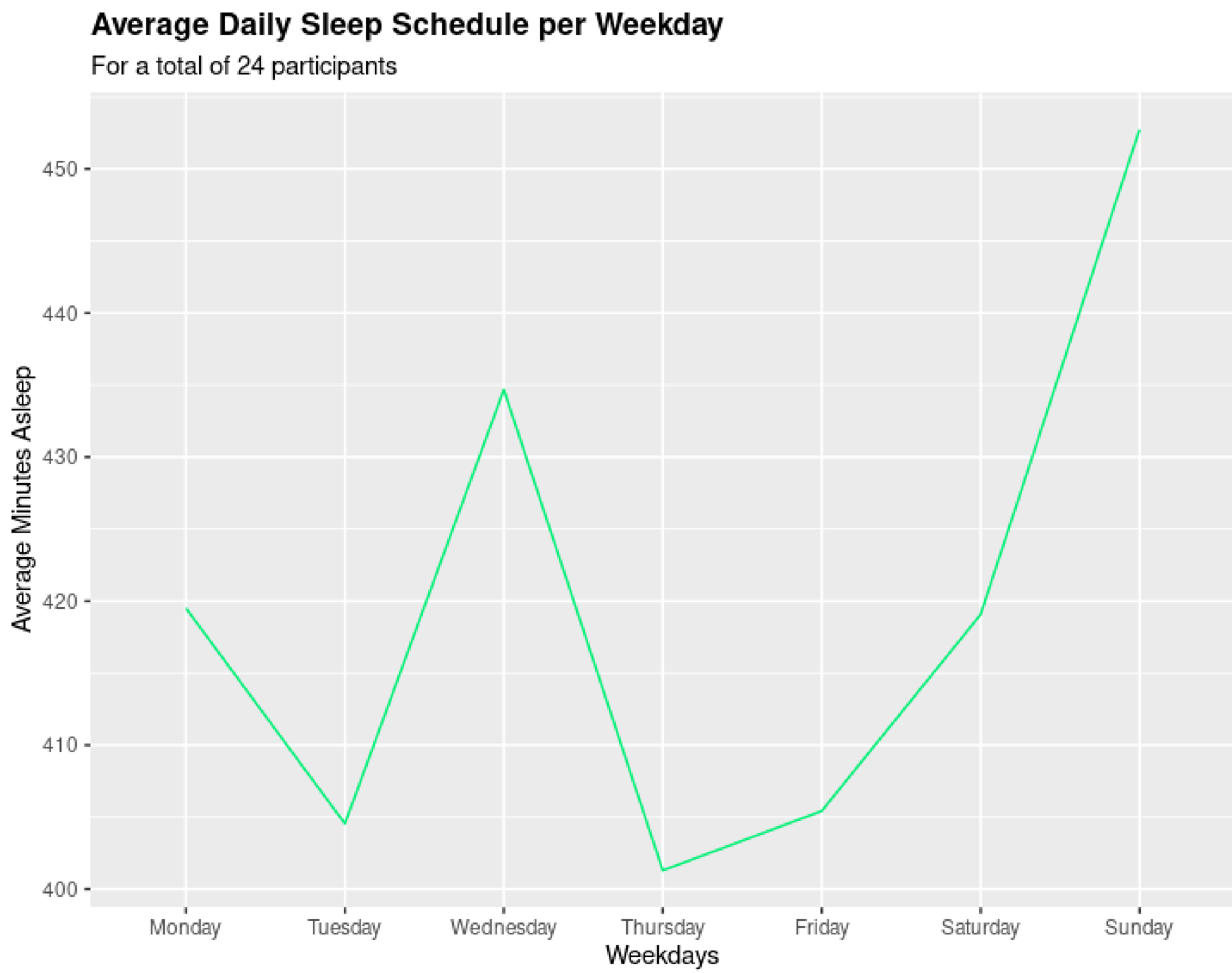
Ordering the weekdays in chronological order for better visibility:

```
sleep_schedule$Weekdays <- factor(sleep_schedule$Weekdays,
                                       levels = c("Monday",
"Tuesday", "Wednesday", "Thursday", "Friday", "Saturday", "Sunday"))
```

Creating a visualization of the tibble:

```
ggplot(sleep_schedule, aes(x=Weekdays,y=mean_TotalMinutesAsleep,
group=1))+
  geom_line(color = "springgreen2")+
  labs(title="Average Daily Sleep Schedule per Weekday",
       subtitle="For a total of 24 participants",
       x= "Weekdays", y="Average Minutes Asleep") +
  theme(plot.title = element_text(face="bold"))
```

OUTPUT:



Conclusion: Based on this chart it seems that on average participants sleep the most on Wednesdays and Sundays. It also appears that the participants on average get less than 7 hours of sleep per night.

5. Calculating the difference between minutes asleep VS total time in bed:

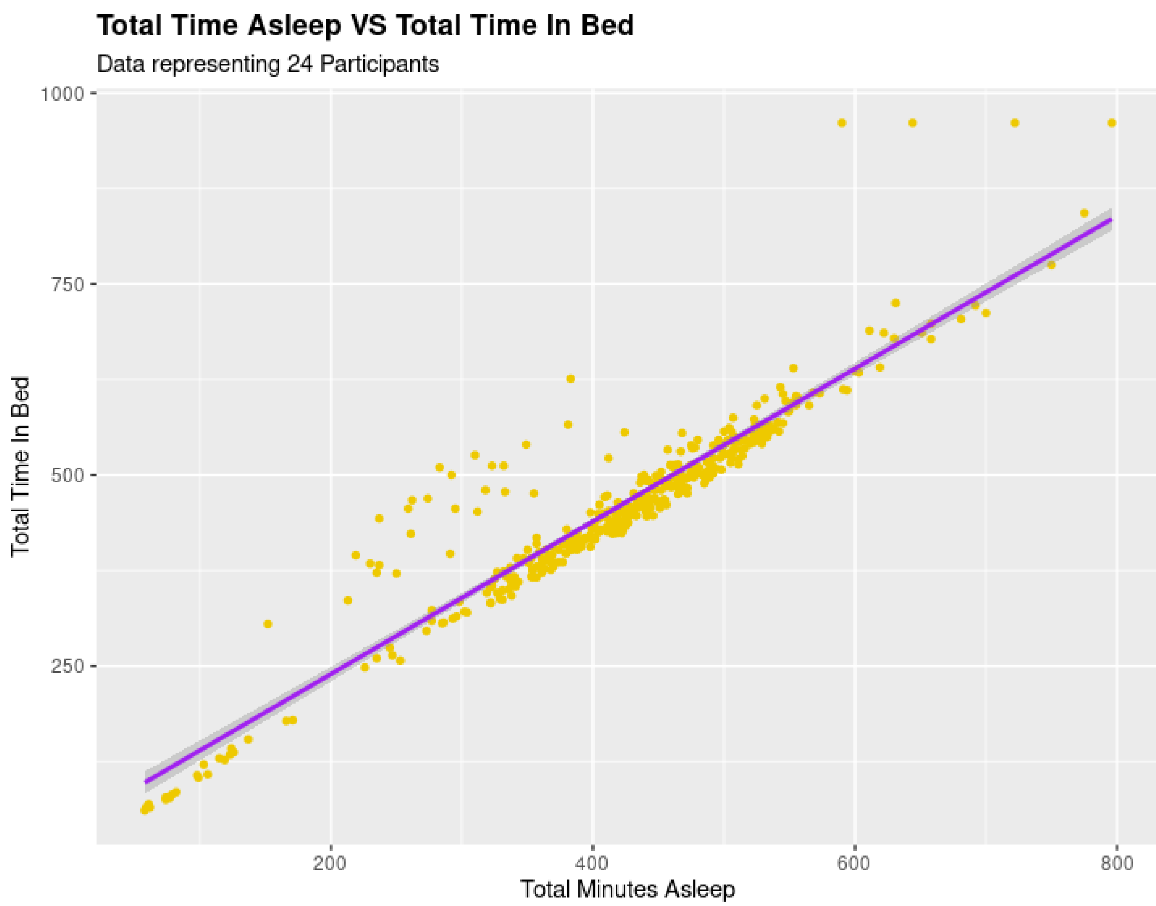
```
#### Creating a tibble to visualize the mean of each variable:

sleep_user_pref <- sleep_day %>%
  group_by(Id) %>%
  summarize(mean_TotalMinutesAsleep = mean(TotalMinutesAsleep),
            mean_TotalTimeInBed = mean(TotalTimeInBed))
head(sleep_user_pref)
```

```
#### Visualizing the relationship between total minutes asleep VS in bed:

ggplot(sleep_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed) +
  geom_point(size=1.25, color="gold2")) +
  geom_smooth(method=lm, color="purple") +
  labs(title="Total Time Asleep VS Total Time In Bed",
       subtitle="Data representing 24 Participants",
       x="Total Minutes Asleep", y="Total Time In Bed")+
  theme(legend.position="none") +
  theme(plot.title = element_text(face="bold"))
```

<u>OUTPUT:</u>



Conclusion: This graph is an example of the total time spent in bed by smart device wearers. The pattern shows that most participants spend about the same amount of total time in bed as they do actively sleeping. A couple outlying data points suggest that some users do spend a bit more time in bed while they aren't sleeping.

6. Determining how many people fall into each activity intensity category:

```
#### Creating a tibble showing the average daily intensities for each participant:

activity_user_type <- daily_intensities %>%
  group_by(Id) %>%
  summarize(mean_SedentaryMinutes = mean(SedentaryMinutes),
            mean_LightlyActiveMinutes = mean(LightlyActiveMinutes),
            mean_FairlyActiveMinutes = mean(FairlyActiveMinutes),
            mean_VeryActiveMinutes = mean(VeryActiveMinutes))
head(activity_user_type)
```

OUTPUT:

```
# A tibble: 6 × 5
          Id mean_SedentaryMinutes mean_LightlyActiveMinutes mean_FairlyActiveMinutes mean_VeryActiveMinut…
       <dbl>                 <dbl>                     <dbl>                    <dbl>                 <dbl>
1 1503960366                  848.                      220.                   19.2                  38.7
2 1624580081                 1258.                      153.                    5.81                  8.68
3 1644430081                 1162.                      178.                   21.4                   9.57
4 1844505072                 1207.                      115.                    1.29                  0.129
5 1927972279                 1317.                       38.6                   0.774                 1.32
6 2022484408                 1113.                      257.                   19.4                  36.3
```

```
#### Averaging daily intensities per category type:

user_type <- c("Sedentary","Lightly Active","Fairly Active","Very Active")
average_intensity <-
c((mean(activity_user_type$mean_SedentaryMinutes, na.rm = TRUE)),

mean(activity_user_type$mean_LightlyActiveMinutes, na.rm = TRUE),

mean(activity_user_type$mean_FairlyActiveMinutes, na.rm = TRUE),

mean(activity_user_type$mean_VeryActiveMinutes, na.rm = TRUE))
intensity_user_type <- data.frame(user_type,average_intensity)
View(intensity_user_type)
```

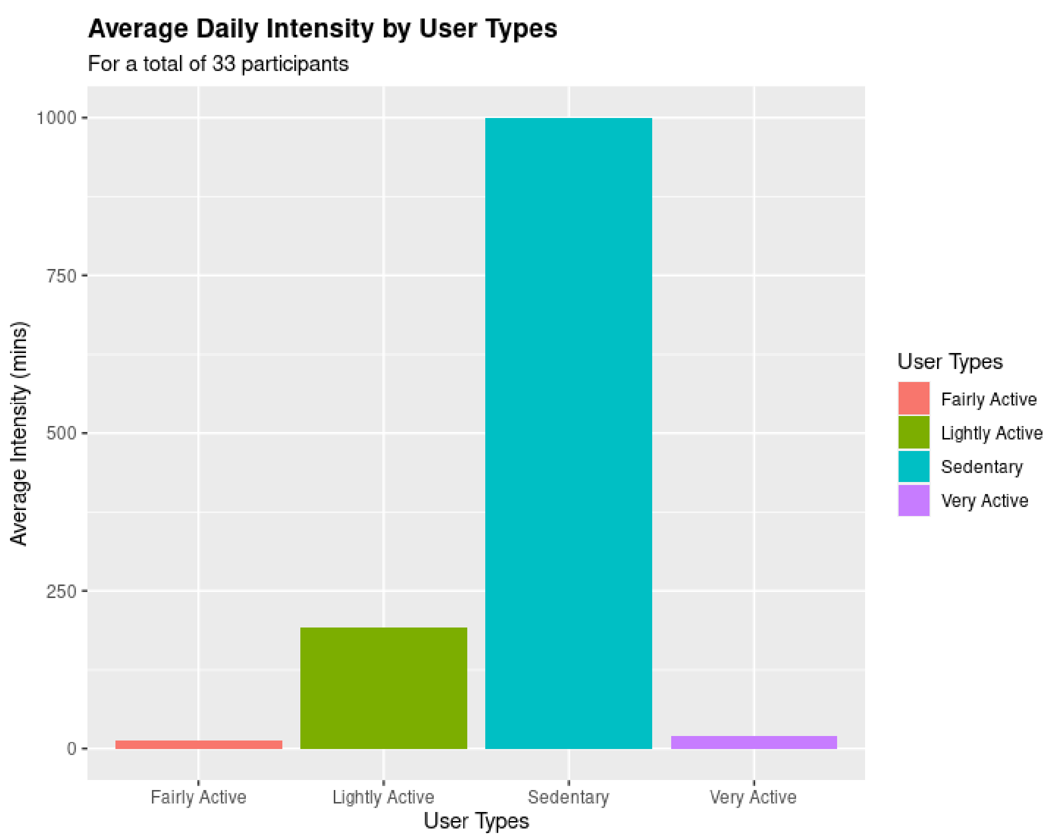
OUTPUT:

	user_type	average_intensity
1	Sedentary	999.89471
2	Lightly Active	191.75036
3	Fairly Active	13.27269
4	Very Active	20.31239

```
#### Visualizing Results:

ggplot(intensity_user_type, aes(x=user_type,y=average_intensity,
fill=user_type, group=1))+
  geom_bar(stat="identity")+
  labs(fill="User Types", title="Average Daily Intensity Types",
  subtitle="For a total of 33 participants",
  x= "User Types", y="Average Intensity (mins)") +
theme(plot.title = element_text(face="bold"))
```

OUTPUT:



Conclusion: From this graph we can determine that out of 33 users most of them are sedentary. This is interesting because smart devices are oftentimes used for physical activity, however these 'FitBit' users are not very active which means they might not only be interested in features related to fitness.

VII. Phase 6 - Act:

- ***What are your final conclusions based on your analysis?***

I will now address the trends and patterns found during the analysis based on the visualizations.

1. **People are mostly interested in their daily activities and intensities**:

While analyzing multiple variables it is clear to see that most participants had entries for daily activities rather than sleep time or weight information. This could be due to the fact that the smart device automatically registers this data, but it cannot check for weight, or sleep patterns if people don't manually enter data or don't wear their smart devices at all times, day and night.

Recommended Action: ***Push Notifications***. To provide people with the best use of their smart device I think that they should be sent a reminder to log in their weights daily if they wish to. This would be part of a push notification system that would remind them to step on the scale to log in their weight and monitor important items, such as BMI, that the device has to offer.

As far as sleep, I don't believe there are enough incentives for people to wear their devices at night. Knowing how long they've slept or spent in bed can easily be determined with a clock or timer. I think creating new informative ways to monitor sleep patterns such as sleep apnea, light sleep vs. deep sleep, snoring or sleep position would entice people to use this feature.

2. **There is a positive correlation between steps taken and calories burned:**

While this finding might seem obvious I think it's important to note that all participants have logged data about daily steps which means that the interest exists. A company who is looking for opportunities to improve their products should look at winning features like daily steps and understand that people are interested in seeing a number being put on burned calories.

Recommended Action: ***Fitness Pal App***. If people are interested in the correlation between their daily steps and calories burned then there is an opportunity to create a category within the app that would show in real time the calories people burn while performing specific exercises. If people are already liking this feature, they will love to be able to take control of their health in real time. This would be successful for people who need to visualize their progress to feel good about themselves. This feature would be a neat addition to the already existing membership.

3. **The weight log feature is not being used to its full potential:**

Besides the fact that people need to be reminded to log in their weights if they wish, I think revisiting the feature entirely would allow for it to be more successful. To understand the data better, out of 33 total participants only 8 logged in their weights. Out of the 8, only 2 logged in their weights for over 24 days while everyone else had less than 4 entries total during a 2-month survey. This tells us that the feature isn't optimized for use and is an opportunity for Bellabeats to make improvements to their own in-app weight log system.

Recommended Action: ***Better AI and Incentives****.* Since people have the option to log in their weight automatically this means that there must be an algorithm able to calculate weight fluctuations. This would be a very high-end feature that would require people to input a lot of daily data in order to be accurate. Based on our analysis so far we know that isn't always the case. Most people prefer to log in their weight manually rather than automatically. This could be due to the fact that the automatic feature isn't accurate enough and they feel that they need to manually correct it.

I think the option to input weight automatically should only be available to people who have enough data in order to be accurate. On the other hand, people who report their weights manually should receive better incentives to collect data by adding interactive features that would remind them to log in data points so that they could benefit from the automatic reporting. This would ultimately increase the use of the weight log feature. I also think that visualization is important and that there should be clear graphs showing weight fluctuation over time.

4. **People sleep the most on Wednesdays and Saturdays:**

Smart device users in this survey seem to sleep the most on Wednesdays and Saturdays. While this data isn't much help to improve features, it is an important reminder that people need rest days and should be encouraged to take care of themselves in ways that aren't related to physical activity. This could be through encouraging quality sleep, but also through drinking enough water, ways to release stress, coping with frustrations etc…

Recommended Action. ***Encouraging Rest Days & Mental Health Days****.* A smart device should not just be about fitness tracking, but also about overall wellbeing which is what Bellabeats wants for its target audience. Reaching goals and visualizing progress is important, however taking time to rest the mind is easily overlooked and yet very impactful for a better quality of life. Bellabeats could really take over the market by presenting empowering features that would put women's mental health first so they can be inspired to be the best version of themselves. Bellabeats could easily position themselves to be a leader in the global market with such progressive features that not every basic fitness tracker will have.

5. **The smart device users from this survey are not sleeping enough:**

After calculating the relationship between total minutes asleep vs. total minutes in bed I noticed that on average people are sleeping 6.98 hours a night. That is less than the recommended 7-9 hours of sleep per night. This indicates that there is an opportunity for Bellabeats to push their sleep feature by showing users what getting better sleep could do for them.

Recommended Action: ***Daily Reminders and Information***. While Bellabeats might already be sending out lots of useful information to their users, it is extremely important for people to understand why they need to care more about their health. For example, telling people that they need to sleep 8h/night is not enough. They need to know 'WHY' and 'HOW' getting good sleep can improve their overall mood, mental health, encourage them to be more active during the day, etc.. Sending out daily tips and tricks to all users would be a great addition to the program as it would encompass all aspects and features the Bellabeats smart devices have to offer.

6. **Most smart devices users are sedentary:**

This is probably the most surprising data point I have found in this analysis. There is a misconception that smart device wearers use their devices for health and fitness only and that if they invest in such a device they must only use it during physical activity to collect data points. This simplistic idea does not correlate with the main activity intensity levels found in this analysis. Based on the findings, users are sedentary for almost 1000 minutes/day (16.5 hours), lightly active for 192 minutes/day (3 hours), fairly active for 13 minutes/day and very active for 20 minutes/day.

Recommended Action: ***Setting Goals & Rewards***. While people seem to be content with using smart devices for features other than intense activity, I think it's important to remember that one of Bellabeats business' goals is to empower women with knowledge about their health habits. For this reason, I recommend pushing for features that would align with the CDC recommendation of at least 30 minutes of moderate exercise a day. This could take the shape of notifications, or light vibrations to let the user know that it's time to get up or get moving. Alternatively, I think there could be a system set up where people are rewarded by completing daily activity challenges such as: running a mile, doing aerobics for 20 minutes, going on a short walk, completing 10,000 steps, etc. This would give incentives to users to get moving!

- ***<u>Is there additional data you could use to expand on your findings?</u>***

This case study actually had an enormous amount of data and many more findings could lead to an even deeper analysis. For example, I would love to dive into the relationship between sleep patterns and daily activity intensity to better understand how poor sleep habits affect overall activity. I would also like to take a closer look at the logged distances in the daily activities and how they compare to steps taken daily and heart rate monitoring.

Overall, the biggest limitation of this case study was the lack of information about the users in the Fitbit data. Bellabeats wants to focus on women's health however the Fitbit data does not give us information about the sex of users which is important to keep in mind. We also don't have an age range so analyzing sleep patterns and daily activity are all based on the assumption that the survey was taken by adults ages 18-60. Moreover, the analysis was based on FitBit data and not directly on Bellabeats smart device usage data.

Finally, I also believe that in order to be a well-rounded company that cares about the well-being of women Bellabeats should really emphasize on mental health. Taking care of one's body health and function is very important, however the mind is equally as powerful when it comes to making progress and reaching goals, which in the end could only be beneficial to a company that is interested in growth opportunities.