

Bellabeats Case Study

Daniel James

2024-12-31

Contents

Introduction	1
Datasets	1
Cleaning our Data	2
Getting Started In R	2
Time to Analyze our Data	5
Initial Findings	6
Time to Visualize	6
Further Analysis Needed	9
Recommendations	9

Introduction

Business Task

To analyze smart device fitness data to gain insight into how consumers are using their smart devices. Then provide ways to help guide Bellabeats marketing strategy. Then we will focus on one specific Bellabeats product to provide recommendations for.

The Stakeholders

- Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
- Sando Mur: Mathematician and Bellabeat's cofounder
- Bellabeat marketing analytics team

Datasets

The Datasets that we will be using during this case study.

- For this case study we will be using FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius). Specifically, we are going to be focusing on the datasets from 4/12/2016 through 5/12/2016 as it is the newest data that we have been provided to us.
- I have confirmed that I will be specifically focusing on three data sets but looked at the basic data for five. HeartRate_seconds and weighLogInfo datasets were reviewed but not chosen for further evaluation.

Why were these datasets chosen?

- I am focusing on the three mentioned datasets for a number of reasons but the main points are:
 - The daily activity dataset is the most comprehensive and I expect it to provide the most insight
 - sleep_day though it has less than the 30 recommended participants for a case study will provide insights into how sleeping and being active are related.
 - hourly_intensity has been chosen because it will provide us an idea of when consumers are most active and how Bellabeats can continue to help those that are active stay active and those who are not with ways to try and possibly improve.

Why we did not further investigate the Heartrate_seconds and weighLogInfo datasets

- After reviewing the datasets for weight_log and heartrate_seconds, there were not enough participants.
- The weight_log dataset did not provide information regarding age or gender or other defining factors to help provide a in depth summary.

Cleaning our Data

- I noticed that in the datasets that I would be using, the date columns were not uniform in regards to column name or format for the date. Before uploading the csv files into RStudio I wanted to make sure to correct this issue.
- I created a new column in the hourly_intensities dataset labeled HourOfDay and created a Date Column. I made sure that the date for each dataset was M/DD/YYYY. For the HourOfDay column I changed the time to the military time format.

Getting Started In R

Install the Packages

```
install.packages('tidyverse')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'  
## (as 'lib' is unspecified)
```

```
install.packages('knitr')
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'  
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --  
## v dplyr      1.1.4      v readr      2.1.5  
## v forcats    1.0.0      v stringr    1.5.1  
## v ggplot2     3.5.1      v tibble     3.2.1  
## v lubridate  1.9.4      v tidyr      1.3.1  
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag()     masks stats::lag()  
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(knitr)
```

Upload the CSV Files that we will be using for analysis

(note that weight_log and heartrate_seconds datasets are mentioned. This is just to show the work I did in my analysis).

```
daily_activity <- read.csv("dailyActivity_merged.csv")
sleep_day <- read.csv("sleepDay_merged.csv")
weight_log <- read.csv("weightLogInfo_merged.csv")
heartrate_seconds <- read.csv("heartrate_seconds_merged.csv")
hourly_intensity <- read.csv("hourlyIntensities_merged.csv")
```

Confirm the datasets were loaded in correctly

```
head(daily_activity)
```

```
##           Id      Date TotalSteps TotalDistance TrackerDistance
## 1 1503960366 4/12/2016      13162           8.50           8.50
## 2 1503960366 4/13/2016       10735           6.97           6.97
## 3 1503960366 4/14/2016       10460           6.74           6.74
## 4 1503960366 4/15/2016        9762           6.28           6.28
## 5 1503960366 4/16/2016       12669           8.16           8.16
## 6 1503960366 4/17/2016        9705           6.48           6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                      0                1.88                   0.55
## 2                      0                1.57                   0.69
## 3                      0                2.44                   0.40
## 4                      0                2.14                   1.26
## 5                      0                2.71                   0.41
## 6                      0                3.19                   0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                   0                25
## 2                4.71                   0                21
## 3                3.91                   0                30
## 4                2.83                   0                29
## 5                5.04                   0                36
## 6                2.51                   0                38
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                 13                328                728    1985
## 2                 19                217                776    1797
## 3                 11                181               1218    1776
## 4                 34                209                726    1745
## 5                 10                221                773    1863
## 6                 20                164                539    1728
```

```
head(sleep_day)
```

```
##           Id      Date TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 1503960366 4/12/2016                1                327           346
## 2 1503960366 4/13/2016                2                384           407
## 3 1503960366 4/15/2016                1                412           442
## 4 1503960366 4/16/2016                2                340           367
## 5 1503960366 4/17/2016                1                700           712
## 6 1503960366 4/19/2016                1                304           320
##   SleepTimeVsTimeInBedAverage
## 1                      19
```

```
## 2 23
## 3 30
## 4 27
## 5 12
## 6 16
```

```
head(weight_log)
```

```
##      Id      Date WeightKg WeightPounds Fat   BMI IsManualReport
## 1 1503960366 5/2/2016 23:59    52.6    115.9631 22 22.65          TRUE
## 2 1503960366 5/3/2016 23:59    52.6    115.9631 NA 22.65          TRUE
## 3 1927972279 4/13/2016 1:08   133.5    294.3171 NA 47.54         FALSE
## 4 2873212765 4/21/2016 23:59    56.7    125.0021 NA 21.45          TRUE
## 5 2873212765 5/12/2016 23:59    57.3    126.3249 NA 21.69          TRUE
## 6 4319703577 4/17/2016 23:59    72.4    159.6147 25 27.45          TRUE
##      LogId
## 1 1.46223e+12
## 2 1.46232e+12
## 3 1.46051e+12
## 4 1.46128e+12
## 5 1.46310e+12
## 6 1.46094e+12
```

```
head(heartrate_seconds)
```

```
##      Id      Date Time Value
## 1 2022484408 4/12/16 7:21    97
## 2 2022484408 4/12/16 7:21   102
## 3 2022484408 4/12/16 7:21   105
## 4 2022484408 4/12/16 7:21   103
## 5 2022484408 4/12/16 7:21   101
## 6 2022484408 4/12/16 7:22    95
```

```
head(hourly_intensity)
```

```
##      Id      Date HourOfDay TotalIntensity AverageIntensity
## 1 1503960366 4/12/2016    0:00             20      0.333333
## 2 1503960366 4/12/2016    1:00              8      0.133333
## 3 1503960366 4/12/2016    2:00              7      0.116667
## 4 1503960366 4/12/2016    3:00              0      0.000000
## 5 1503960366 4/12/2016    4:00              0      0.000000
## 6 1503960366 4/12/2016    5:00              0      0.000000
```

Let us see how many participants are in these uploaded datasets

```
n_distinct(daily_activity$Id)
```

```
## [1] 33
```

```
n_distinct(sleep_day$Id)
```

```
## [1] 24
```

```
n_distinct(weight_log$Id)
```

```
## [1] 8
```

```
n_distinct(heartrate_seconds$Id)
```

```
## [1] 7
```

```
n_distinct(hourly_intensity$Id)
```

```
## [1] 33
```

Time to Analyze our Data

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

Daily Activity Summary

```
##      TotalSteps      TotalDistance      VeryActiveMinutes      FairlyActiveMinutes  
## Min.       :    0      Min.       : 0.000      Min.       :  0.00      Min.       :  0.00  
## 1st Qu.: 3790      1st Qu.: 2.620      1st Qu.:  0.00      1st Qu.:  0.00  
## Median : 7406      Median : 5.245      Median :  4.00      Median :  6.00  
## Mean   : 7638      Mean   : 5.490      Mean    : 21.16      Mean    : 13.56  
## 3rd Qu.:10727      3rd Qu.: 7.713      3rd Qu.: 32.00      3rd Qu.: 19.00  
## Max.    :36019      Max.    :28.030      Max.     :210.00      Max.     :143.00  
## LightlyActiveMinutes SedentaryMinutes  
## Min.       :  0.0      Min.       :  0.0  
## 1st Qu.:127.0      1st Qu.: 729.8  
## Median :199.0      Median :1057.5  
## Mean   :192.8      Mean    : 991.2  
## 3rd Qu.:264.0      3rd Qu.:1229.5  
## Max.    :518.0      Max.     :1440.0
```

Sleep Day Summary

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

```
##      TotalSleepRecords      TotalMinutesAsleep      TotalTimeInBed  
## Min.       :1.000      Min.       : 58.0      Min.       : 61.0  
## 1st Qu.:1.000      1st Qu.:361.0      1st Qu.:403.0  
## Median :1.000      Median :433.0      Median :463.0  
## Mean   :1.119      Mean   :419.5      Mean   :458.6  
## 3rd Qu.:1.000      3rd Qu.:490.0      3rd Qu.:526.0  
## Max.    :3.000      Max.    :796.0      Max.    :961.0  
## SleepTimeVsTimeInBedAverage  
## Min.       :  0.00
```

```
## 1st Qu.: 17.00
## Median : 25.00
## Mean   : 39.17
## 3rd Qu.: 40.00
## Max.   :371.00
```

Initial Findings

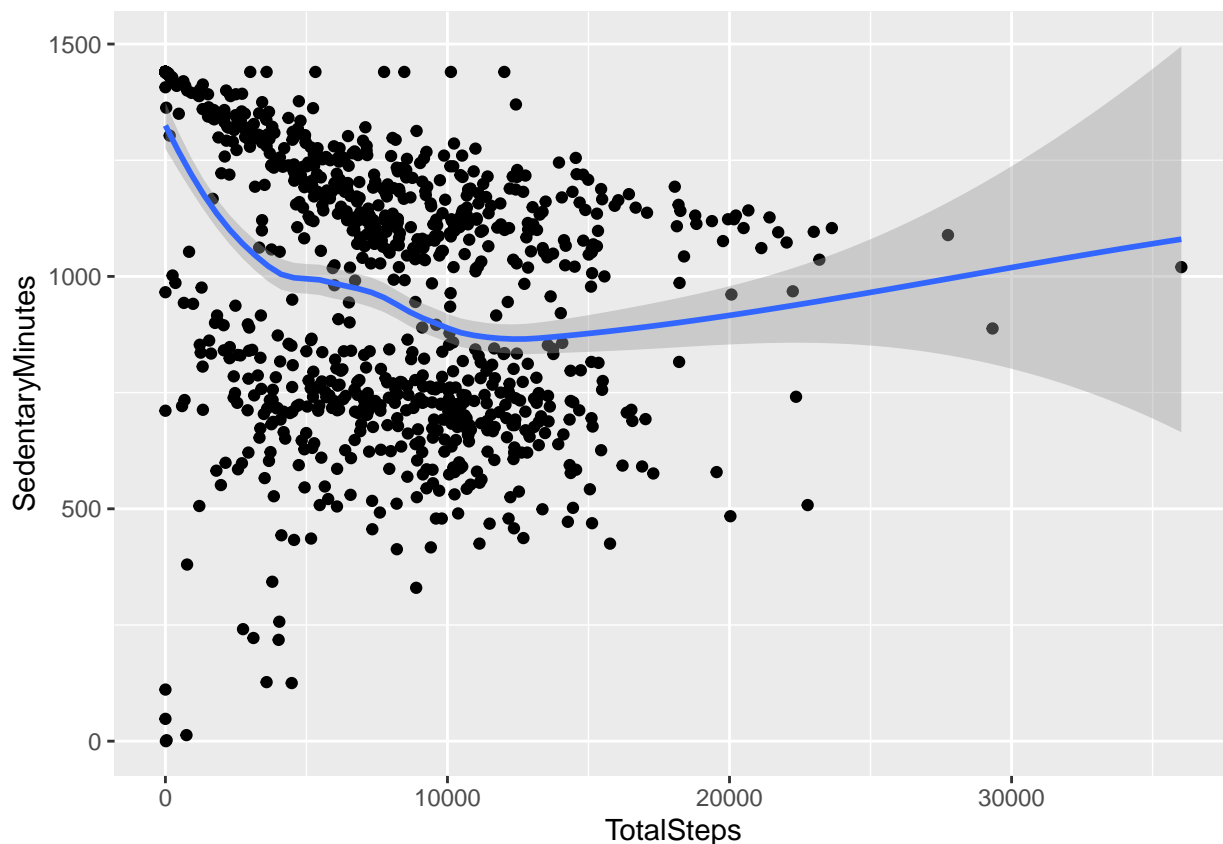
- That out of the people studied they spend the most time being sedentary (Average of 991.2 minutes per day) with lightly active a far second (Average of 192.8 Minutes).
- On average people are sleeping just under the 7 hours of recommended sleep (7 hours is recommended for adults according to www.mayoclinic.org)
- What really stood out is that on average people are in bed 39 minutes longer than the amount of time they are sleeping.

Time to Visualize

Comparing number of Steps vs. Sedentary Minutes Scatter Plot

```
ggplot(data=daily_activity, aes(x=TotalSteps, y=SedentaryMinutes)) + geom_point() + geom_smooth()

## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

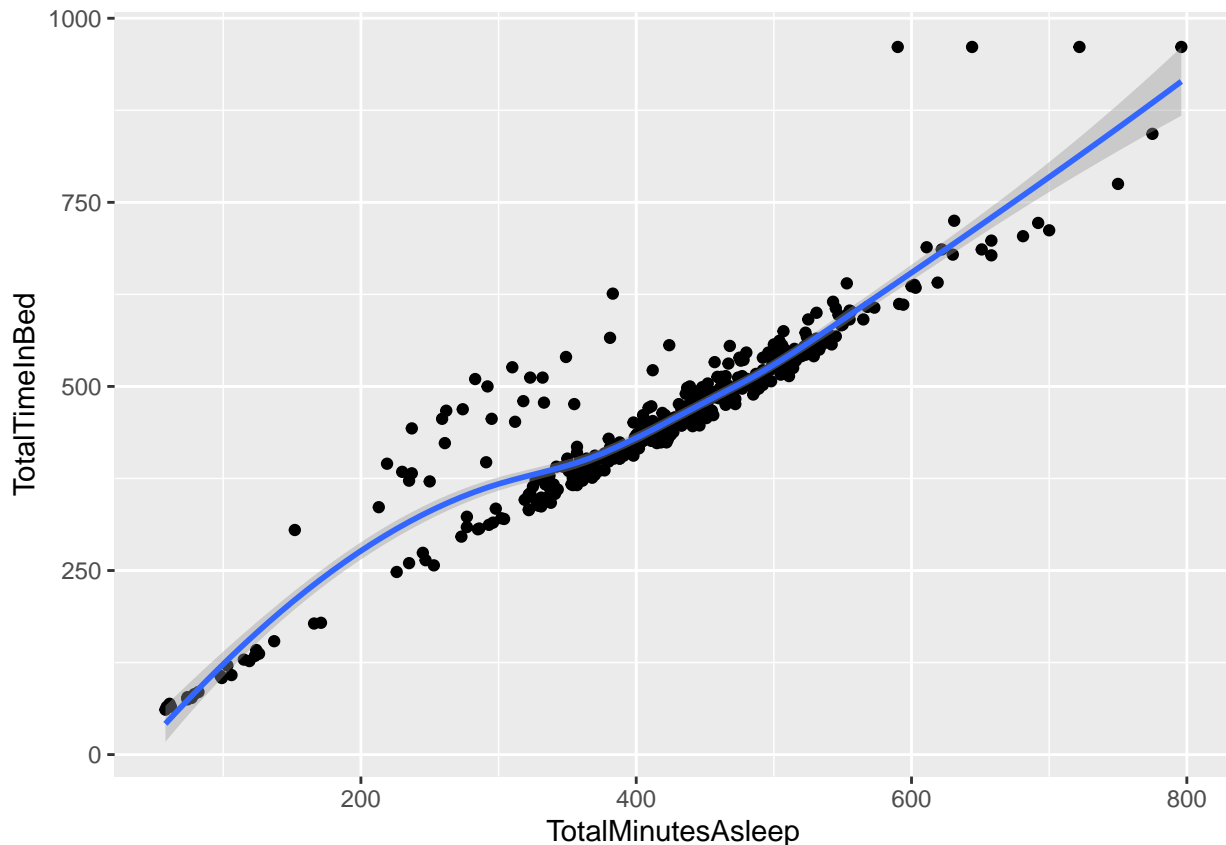


* For this data you can see that there is a trend of the reduction of sedentary minutes between the step amounts of 5,000 and 10,000 but once 20,000 steps is reached it seems a slight increase in sedentary minutes does occur. This would be something to further evaluate.

Comparing Time in Bed Vs Minutes Asleep Scatter Plot

```
ggplot(data=sleep_day, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) + geom_point() + geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



* On average a vast majority of people slept between 375 minutes and 575 minutes and the time in bed seemed to be slightly higher for this time frame.

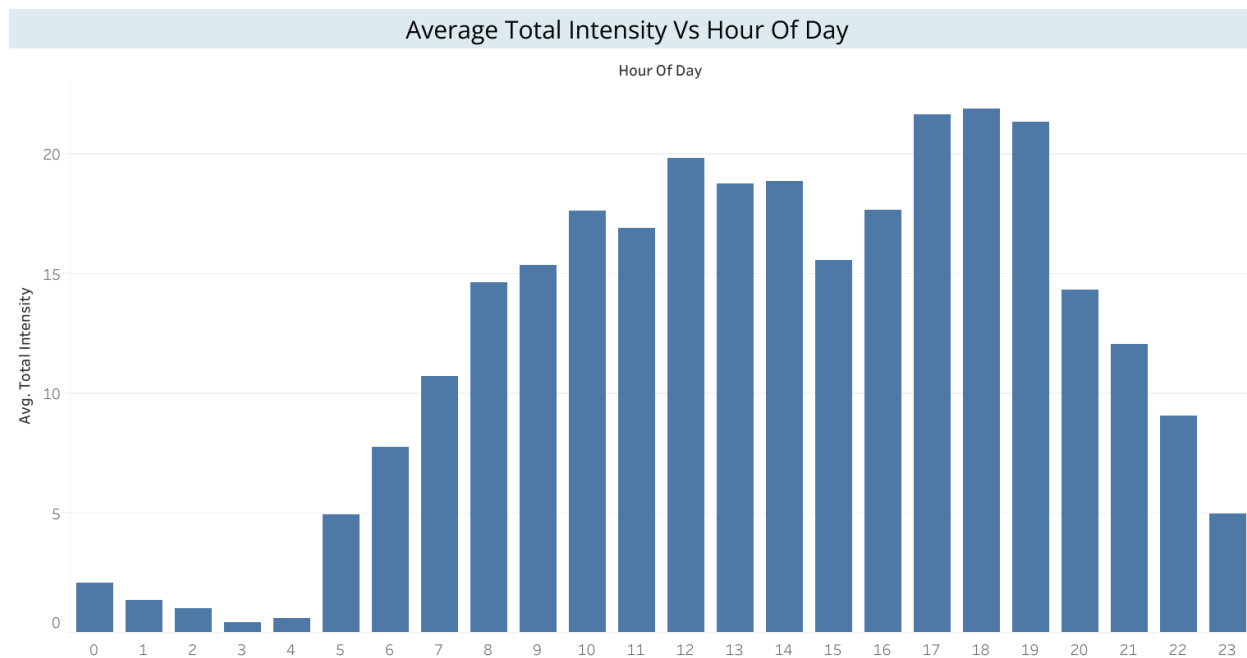
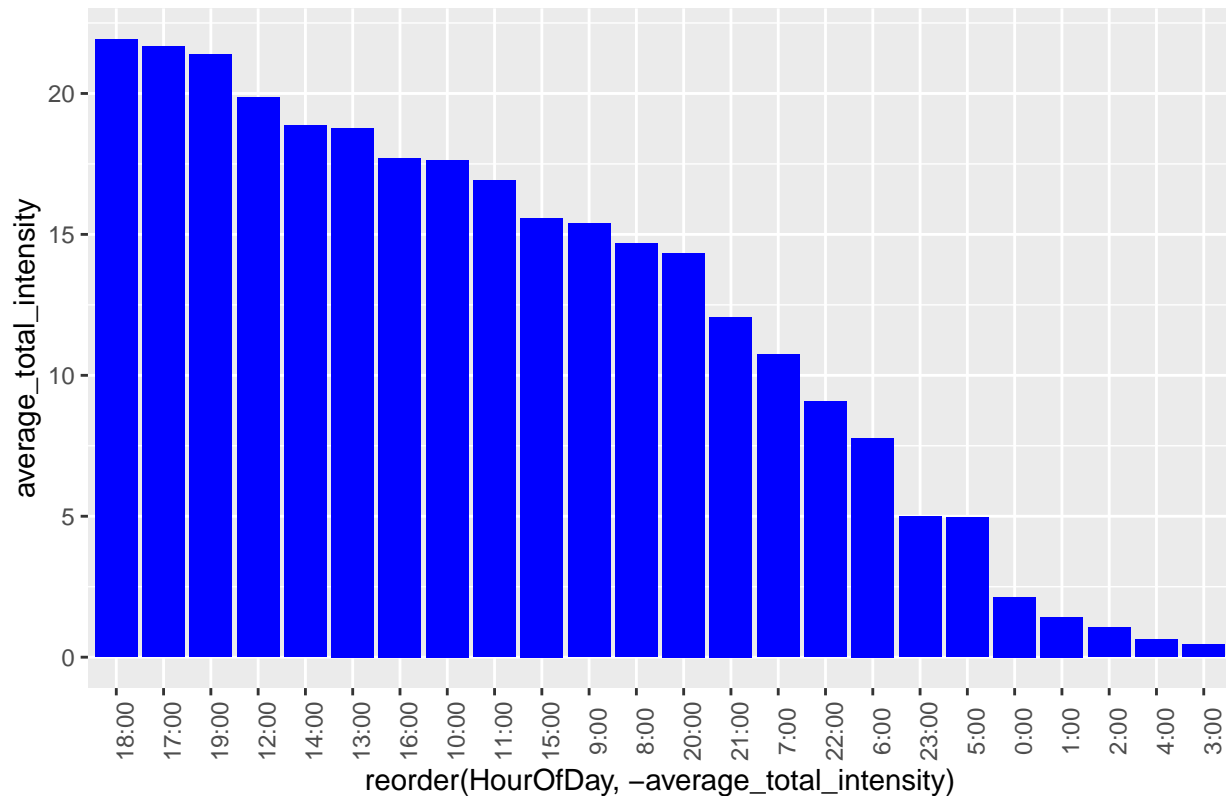
* There is a little bit of an increase in total time in bed though between the sleeping hours of 250-375 where people were staying in bed much longer than the norm.

Average Hourly Intensity vs Time of Day(In Descending Order)

```
average_intensity_per_day <- hourly_intensity %>%  
  group_by(HourOfDay) %>%  
  summarise(average_total_intensity = mean(TotalIntensity))  
view(average_intensity_per_day)
```

```
ggplot(data=average_intensity_per_day, aes(x = reorder(HourOfDay, -average_total_intensity), y =average.  
geom_bar(stat = "identity", fill = ('blue'))+ theme(axis.text.x = element_text(angle = 90)) + labs(title=
```

Total Average Hourly Intensity vs Time of Day(In Descending Order)



* People are most active between the hours of 5 PM through 7 PM. With the hours of Noon to 2 PM being the second most active grouping.

* People are the least active between the hours of Midnight to 5 AM. This does make sense given that most people are sleeping during this time.

Further Analysis Needed

After reviewing the data I wanted to point out a few things that would help benefit our marketing efforts and provide more insight into our clients.

1. Trying to collect more data regarding our clients such as age, and activities they are participating in so that we can better find our target audience for our products.
2. Looking at the sleep data there was an average of 39 minutes where participants were staying in bed longer than they were actually asleep. I find it would be helpful to further look into if this is because they are having trouble falling asleep, having trouble getting up, or are they just in bed doing other activities such as reading or watching tv?

Recommendations

Just to remind everyone the goal that we set out at the beginning of this case study was to analyze smart device fitness data to gain insight into how consumers are using their smart devices. Then provide ways to help guide Bellabeats marketing strategy. Then we will focus on one specific Bellabeats product to provide recommendations for.

As Bellabeats is a fast growing company, only being around a total of 11 years they have made massive strides in helping women focus on fitness and their daily lives by providing top notch products.

Do not stop what you are doing because it is working!

With that being said after analyzing the data I have a few recommendations that will help the company. I am going to be focusing on the BellaBeats app specifically but my recommendations could work for any of the wearable products as well such as the Leaf Tracker and the Time Watch.

1. As consumers are most active during the times of 5 PM to 7 PM, we want to make sure that we keep people active and meet their goals. Having a reminder sent out to our clients during this time to “be active let’s go for a walk or a simple”let us get moving” will keep the current clients active during this time but also promote new users to also start a trend of being active.
2. With consumers being in bed 39 minutes then they are asleep I think it is important that we find a way to help these users. Whether it is because they are having trouble falling asleep or they are staying in bed after they wake up we can help both clientele. For those sleepers who have a hard time getting to sleep, let us send out a reminder 30 minutes before they usually go to sleep saying “It is time for bed” that way our clients can start getting ready and spend less time tossing and turning. For our clients that have a hard time getting out of bed let us send a simple “It is time to get up reminder if they do not get out of bed 10 minutes after waking up. This will make sure they are not just laying around being sedentary and get clients motivated to start moving.
3. Per the Comparing number of Steps vs. Sedentary Minutes Scatter Plot participants who were between 5,000 and 10,000 steps saw a decrease in the minutes they spent being sedentary. I would recommend a “It is time to stand up” reminder to make sure that our users are remembering to get up from their desk, couch, or a chair and remember to get some steps in. Another recommendation could be if a user is not going to meet a certain number of steps by noon send a reminder of “How about we go for a walk?” to convince our users to be more active and get in those steps