Case Study 2: How Can a Wellness Technology Company (Bellabeat) Play It Smart?

Martin U

2023-02-11

Introduction

This report is a requirement of the capstone program in fulfillment of the partial requirement of the Google Data Analytics course. The report focuses on case study 2 with the title headlined above. The objective is to analyze smart device data to gain insight into how consumers use their smart devices and to use discovered insights to guide the marketing strategy for Bellabeat. Also, to present high-level recommendations for Bellabeat's marketing strategy to the executive team.

About the company

Bellabeat is a high-tech company that manufactures health-focused innovative products that inform and inspire women worldwide by using data collected from smart devices to empower women with knowledge about their health and habits. The company was founded in 2013 by Urška Sršen and Sando Mur. Although Bellabeat is a successful small company, however, has the potential to become a more prominent player in the global smart device market. They have smart device products such as: the Bellabeat app – which provides users with health data related to their activities; Leaf – a wellness tracker worn as a bracelet or necklace; Time – a smart device for tracking activity and telling time; a Spring–water bottle for tracking daily water intake; and the Bellabeat membership.

Business Task:

The business objective is to analyze smart device usage data to understand how consumers use non-Bellabeat smart devices. Then, select one Bellabeat product to apply these insights to in your presentation, with answers to the following questions: 1. What are some trends in smart device usage? 2. How could these trends apply to Bellabeat customers? 3. How could these trends help influence Bellabeat's marketing strategy?

Prepare

The data was stored in a Kaggle open-source repository, and it has 18 CSV files which are organized in a long format. The data is open-source with no copy-write, licensing, or privacy restrictions. According to the data's author, the datasets were gathered from respondents of a distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible Fitbit users consented to submit personal tracker data (although thirty-three unique user IDs were noticed), including minute-level output for physical activity, heart rate, and sleep monitoring. R's RStudio is the primary tool that will be used to process and clean the data, there in preparation for the task ahead some relevant libraries are first loaded to the session.

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v purrr   1.0.1
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
```

```
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## Loading required package: timechange
##
##
## Attaching package: 'lubridate'
##
##
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
##
##
## here() starts at C:/Users/Martin/OneDrive/Documents/CambrianClasses/coursera_folder
##
##
## Attaching package: 'janitor'
##
##
## The following objects are masked from 'package:stats':
##
##     chisq.test, fisher.test
```

Given that this is supposed to help provide insight to advise on marketing decisions, sample data of only thirty-three respondents is insufficient to provide a representative analysis. Also, the data does not provide any demographic information. Hence, it's difficult to determine if the data represents the company's intended focus - women. Four of the eighteen CSV files provided daily data about various interesting attributes.

```r
# Read all relevant files into R and assign them to easy to use names
activity <- read.csv("dailyActivity_merged.csv")
calories <- read.csv("dailyCalories_merged.csv")
intensity <- read.csv("dailyIntensities_merged.csv")
sleep <- read.csv("sleepDay_merged.csv")
steps <- read.csv("dailySteps_merged.csv")
weight <- read.csv("weightLogInfo_merged.csv")
hr <- read.csv("heartrate_seconds_merged.csv")
```

Still, the others offered minute-level data and other areas of interest, such as weight (8 respondents) and heart rate (14 respondents). Unfortunately, drawing any conclusion from these dataset categories is impossible because only a third or less of the respondents provided data for these attributes. Sleep_day (24 respondents), Steps (33 respondents), intensities (33 days)

```
## [1] 33

## [1] 33

## [1] 33

## [1] 24

## [1] 33

## [1] 8

## [1] 14
```

Process

R's RStudio is the primary tool for this report, although a preliminary check of the files in Excel will be done first to take a cursory look at the data. Five of the eighteen were deemed germane to the task at hand. The other files' actual contents were either already contained in some of the five relevant CSV files or were considered unusable due to insufficient respondents.

```
# Start the data cleaning
#skim_without_charts(activity)     #Gives a comprehensive summary of the data set

glimpse(activity)       #Gives a a quick idea summary of the data set
```

```
## Rows: 940
## Columns: 15
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDate             <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ TotalSteps               <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019~
## $ TotalDistance            <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ TrackerDistance          <dbl> 8.50, 6.97, 6.74, 6.28, 8.16, 6.48, 8.59, 9.8~
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ VeryActiveMinutes        <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ FairlyActiveMinutes      <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ LightlyActiveMinutes     <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ SedentaryMinutes         <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ Calories                 <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 203~
```

```
glimpse(calories)
```

```
## Rows: 940
## Columns: 3
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ Calories    <int> 1985, 1797, 1776, 1745, 1863, 1728, 1921, 2035, 1786, 1775~
```

```
glimpse(intensity)
```

```
## Rows: 940
## Columns: 10
## $ Id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ ActivityDay              <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/~
## $ SedentaryMinutes         <int> 728, 776, 1218, 726, 773, 539, 1149, 775, 818~
## $ LightlyActiveMinutes     <int> 328, 217, 181, 209, 221, 164, 233, 264, 205, ~
## $ FairlyActiveMinutes      <int> 13, 19, 11, 34, 10, 20, 16, 31, 12, 8, 27, 21~
## $ VeryActiveMinutes        <int> 25, 21, 30, 29, 36, 38, 42, 50, 28, 19, 66, 4~
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ LightActiveDistance      <dbl> 6.06, 4.71, 3.91, 2.83, 5.04, 2.51, 4.71, 5.0~
## $ ModeratelyActiveDistance <dbl> 0.55, 0.69, 0.40, 1.26, 0.41, 0.78, 0.64, 1.3~
## $ VeryActiveDistance       <dbl> 1.88, 1.57, 2.44, 2.14, 2.71, 3.19, 3.25, 3.5~
```

```
glimpse(sleep)
```

```
## Rows: 413
## Columns: 5
```

```
## $ Id                 <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 150~
## $ SleepDay           <chr> "4/12/2016 12:00:00 AM", "4/13/2016 12:00:00 AM", "~
## $ TotalSleepRecords  <int> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ TotalMinutesAsleep <int> 327, 384, 412, 340, 700, 304, 360, 325, 361, 430, 2~
## $ TotalTimeInBed     <int> 346, 407, 442, 367, 712, 320, 377, 364, 384, 449, 3~
```

```
glimpse(steps)
```

```
## Rows: 940
## Columns: 3
## $ Id          <dbl> 1503960366, 1503960366, 1503960366, 1503960366, 1503960366~
## $ ActivityDay <chr> "4/12/2016", "4/13/2016", "4/14/2016", "4/15/2016", "4/16/~
## $ StepTotal   <int> 13162, 10735, 10460, 9762, 12669, 9705, 13019, 15506, 1054~
```

```
glimpse(weight)
```

```
## Rows: 67
## Columns: 8
## $ Id             <dbl> 1503960366, 1503960366, 1927972279, 2873212765, 2873212~
## $ Date           <chr> "5/2/2016 11:59:59 PM", "5/3/2016 11:59:59 PM", "4/13/2~
## $ WeightKg       <dbl> 52.6, 52.6, 133.5, 56.7, 57.3, 72.4, 72.3, 69.7, 70.3, ~
## $ WeightPounds   <dbl> 115.9631, 115.9631, 294.3171, 125.0021, 126.3249, 159.6~
## $ Fat            <int> 22, NA, NA, NA, NA, 25, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ BMI            <dbl> 22.65, 22.65, 47.54, 21.45, 21.69, 27.45, 27.38, 27.25,~
## $ IsManualReport <chr> "True", "True", "False", "True", "True", "True", "True"~
## $ LogId          <dbl> 1.462234e+12, 1.462320e+12, 1.460510e+12, 1.461283e+12,~
```

```
glimpse(hr)
```

```
## Rows: 2,483,658
## Columns: 3
## $ Id    <dbl> 2022484408, 2022484408, 2022484408, 2022484408, 2022484408, 2022~
## $ Time  <chr> "4/12/2016 7:21:00 AM", "4/12/2016 7:21:05 AM", "4/12/2016 7:21:~
## $ Value <int> 97, 102, 105, 103, 101, 95, 91, 93, 94, 93, 92, 89, 83, 61, 60, ~
```

The data structure, attributes' data types conformation to their values and consistency of naming was checked, and "N/A" values were dropped if contained in the files. These were done to ensure the data was clean and ready for analysis. All cleaning processes were documented and can be shared via the R Markdown file.

```
##        Id             ActivityDate         TotalSteps    TotalDistance
##  Min.   :1.504e+09   Length:940         Min.   :    0   Min.   : 0.000
##  1st Qu.:2.320e+09   Class :character   1st Qu.: 3790   1st Qu.: 2.620
##  Median :4.445e+09   Mode  :character   Median : 7406   Median : 5.245
##  Mean   :4.855e+09                      Mean   : 7638   Mean   : 5.490
##  3rd Qu.:6.962e+09                      3rd Qu.:10727   3rd Qu.: 7.713
##  Max.   :8.878e+09                      Max.   :36019   Max.   :28.030
##  TrackerDistance  LoggedActivitiesDistance VeryActiveDistance
##  Min.   : 0.000   Min.   :0.0000           Min.   : 0.000
##  1st Qu.: 2.620   1st Qu.:0.0000           1st Qu.: 0.000
##  Median : 5.245   Median :0.0000           Median : 0.210
##  Mean   : 5.475   Mean   :0.1082           Mean   : 1.503
##  3rd Qu.: 7.710   3rd Qu.:0.0000           3rd Qu.: 2.053
##  Max.   :28.030   Max.   :4.9421           Max.   :21.920
##  ModeratelyActiveDistance LightActiveDistance SedentaryActiveDistance
##  Min.   :0.0000           Min.   : 0.000      Min.   :0.000000
##  1st Qu.:0.0000           1st Qu.: 1.945      1st Qu.:0.000000
```

```
##  Median :0.2400           Median : 3.365      Median :0.000000
##  Mean   :0.5675           Mean   : 3.341      Mean   :0.001606
##  3rd Qu.:0.8000           3rd Qu.: 4.782      3rd Qu.:0.000000
##  Max.   :6.4800           Max.   :10.710      Max.   :0.110000
##  VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  0.0        Min.   :   0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0        1st Qu.: 729.8
##  Median :  4.00    Median :  6.00      Median :199.0        Median :1057.5
##  Mean   : 21.16    Mean   : 13.56      Mean   :192.8        Mean   : 991.2
##  3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0        3rd Qu.:1229.5
##  Max.   :210.00    Max.   :143.00      Max.   :518.0        Max.   :1440.0
##     Calories         date
##  Min.   :   0   Min.   :2016-04-12
##  1st Qu.:1828   1st Qu.:2016-04-19
##  Median :2134   Median :2016-04-26
##  Mean   :2304   Mean   :2016-04-26
##  3rd Qu.:2793   3rd Qu.:2016-05-04
##  Max.   :4900   Max.   :2016-05-12
```

Analyze

The required files for this task are the activity and sleep files, which were read into R and named accordingly. The sleep and activity files were merged, the date type was formatted from a string/character to a date type, and the column names were cleaned for consistency. A quick peek at the resultant merge is shown below.

```
##           Id       date ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366 2016-04-12    4/12/2016      13162          8.50            8.50
## 2 1503960366 2016-04-13    4/13/2016      10735          6.97            6.97
## 3 1503960366 2016-04-15    4/15/2016       9762          6.28            6.28
## 4 1503960366 2016-04-16    4/16/2016      12669          8.16            8.16
## 5 1503960366 2016-04-17    4/17/2016       9705          6.48            6.48
## 6 1503960366 2016-04-19    4/19/2016      15506          9.88            9.88
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.14                     1.26
## 4                        0               2.71                     0.41
## 5                        0               3.19                     0.78
## 6                        0               3.53                     1.32
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                2.83                       0                29
## 4                5.04                       0                36
## 5                2.51                       0                38
## 6                5.03                       0                50
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  34                  209              726     1745
## 4                  10                  221              773     1863
## 5                  20                  164              539     1728
## 6                  31                  264              775     2035
##               SleepDay TotalSleepRecords TotalMinutesAsleep TotalTimeInBed
## 1 4/12/2016 12:00:00 AM                 1                327            346
```

```
## 2 4/13/2016 12:00:00 AM                 2                 384            407
## 3 4/15/2016 12:00:00 AM                 1                 412            442
## 4 4/16/2016 12:00:00 AM                 2                 340            367
## 5 4/17/2016 12:00:00 AM                 1                 700            712
## 6 4/19/2016 12:00:00 AM                 1                 304            320
```

It was surprising that the data did not include demographic information and geographic details, making it difficult to determine if it represented our desired market segment (women). Also, it was observed that not all the respondents recorded all items of interest because "the variation between output represents the use of different types of Fitbit trackers and individual tracking behaviors/preferences," as noted by the data's author.

Transforming and Focusing on Desired Fields

The earlier merged file is now trimmed to include only the attributes we'll be using for analysis. The column names was renamed to simplicity and cleaned for consistency. The sedentary and sleep values were also transformed from minutes to hour.

```
##           id       date steps distance calories sedentary sleep sedentary_hour
## 1 1503960366 2016-04-12 13162     8.50     1985       728   327      12.133333
## 2 1503960366 2016-04-13 10735     6.97     1797       776   384      12.933333
## 3 1503960366 2016-04-15  9762     6.28     1745       726   412      12.100000
## 4 1503960366 2016-04-16 12669     8.16     1863       773   340      12.883333
## 5 1503960366 2016-04-17  9705     6.48     1728       539   700       8.983333
## 6 1503960366 2016-04-19 15506     9.88     2035       775   304      12.916667
##   sleep_hour   weekday
## 1   5.450000   Tuesday
## 2   6.400000 Wednesday
## 3   6.866667    Friday
## 4   5.666667  Saturday
## 5  11.666667    Sunday
## 6   5.066667   Tuesday
```

```
##        id                  date                steps          distance
##  Min.   :1.504e+09   Min.   :2016-04-12   Min.   :   17   Min.   : 0.010
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   1st Qu.: 5206   1st Qu.: 3.600
##  Median :4.703e+09   Median :2016-04-27   Median : 8925   Median : 6.290
##  Mean   :5.001e+09   Mean   :2016-04-26   Mean   : 8541   Mean   : 6.039
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:11393   3rd Qu.: 8.030
##  Max.   :8.792e+09   Max.   :2016-05-12   Max.   :22770   Max.   :17.540
##     calories      sedentary          sleep       sedentary_hour
##  Min.   : 257   Min.   :   0.0   Min.   : 58.0   Min.   : 0.00
##  1st Qu.:1850   1st Qu.: 631.0   1st Qu.:361.0   1st Qu.:10.52
##  Median :2220   Median : 717.0   Median :433.0   Median :11.95
##  Mean   :2398   Mean   : 712.2   Mean   :419.5   Mean   :11.87
##  3rd Qu.:2926   3rd Qu.: 783.0   3rd Qu.:490.0   3rd Qu.:13.05
##  Max.   :4900   Max.   :1265.0   Max.   :796.0   Max.   :21.08
##    sleep_hour        weekday
##  Min.   : 0.9667   Length:413
##  1st Qu.: 6.0167   Class :character
##  Median : 7.2167   Mode  :character
##  Mean   : 6.9911
##  3rd Qu.: 8.1667
##  Max.   :13.2667
```

```
## 'data.frame':    413 obs. of  10 variables:
```

```
##  $ id            : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ date          : Date, format: "2016-04-12" "2016-04-13" ...
##  $ steps         : int  13162 10735 9762 12669 9705 15506 10544 9819 14371 10039 ...
##  $ distance      : num  8.5 6.97 6.28 8.16 6.48 ...
##  $ calories      : int  1985 1797 1745 1863 1728 2035 1786 1775 1949 1788 ...
##  $ sedentary     : int  728 776 726 773 539 775 818 838 732 709 ...
##  $ sleep         : int  327 384 412 340 700 304 360 325 361 430 ...
##  $ sedentary_hour: num  12.13 12.93 12.1 12.88 8.98 ...
##  $ sleep_hour    : num  5.45 6.4 6.87 5.67 11.67 ...
##  $ weekday       : chr  "Tuesday" "Wednesday" "Friday" "Saturday" ...
```

Aggregating the Data

The trimmed dataframe is aggregated so it's useful and accessible for our analysis, as shown below.

```
## `summarise()` has grouped output by 'id', 'date'. You can override using the
## `.groups` argument.
```

```
## # A tibble: 6 x 8
## # Groups:   id, date [6]
##           id date       weekday  avg_steps avg_dist avg_cal avg_sedent~1 avg_s~2
##        <dbl> <date>     <chr>        <dbl>    <dbl>   <dbl>        <dbl>   <dbl>
## 1 4388161847 2016-05-07 Saturday     22770     17.5    4022         8.47    7.87
## 2 8053475328 2016-04-23 Saturday     22359     17.2    3554        12.4     5.52
## 3 6962181067 2016-04-23 Saturday     20031     13.2    2571         8.07    8.8
## 4 8053475328 2016-05-07 Saturday     19769     15.7    3331        17.9     1.23
## 5 6117666160 2016-04-21 Thursday     19542     15.0    4900         9.65    8.47
## 6 1644430081 2016-04-30 Saturday     18213     13.2    3846        13.6     2.07
## # ... with abbreviated variable names 1: avg_sedentary, 2: avg_sleep
```

```
##        id                 date              weekday            avg_steps
##  Min.   :1.504e+09   Min.   :2016-04-12   Length:410         Min.   :   17
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   Class :character   1st Qu.: 5189
##  Median :4.703e+09   Median :2016-04-27   Mode  :character   Median : 8913
##  Mean   :4.995e+09   Mean   :2016-04-26                      Mean   : 8515
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04                      3rd Qu.:11370
##  Max.   :8.792e+09   Max.   :2016-05-12                      Max.   :22770
##     avg_dist          avg_cal      avg_sedentary     avg_sleep
##  Min.   : 0.010   Min.   : 257   Min.   : 0.00   Min.   : 0.9667
##  1st Qu.: 3.592   1st Qu.:1841   1st Qu.:10.52   1st Qu.: 6.0167
##  Median : 6.270   Median :2207   Median :11.95   Median : 7.2083
##  Mean   : 6.012   Mean   :2389   Mean   :11.87   Mean   : 6.9862
##  3rd Qu.: 8.005   3rd Qu.:2920   3rd Qu.:13.05   3rd Qu.: 8.1667
##  Max.   :17.540   Max.   :4900   Max.   :21.08   Max.   :13.2667
```

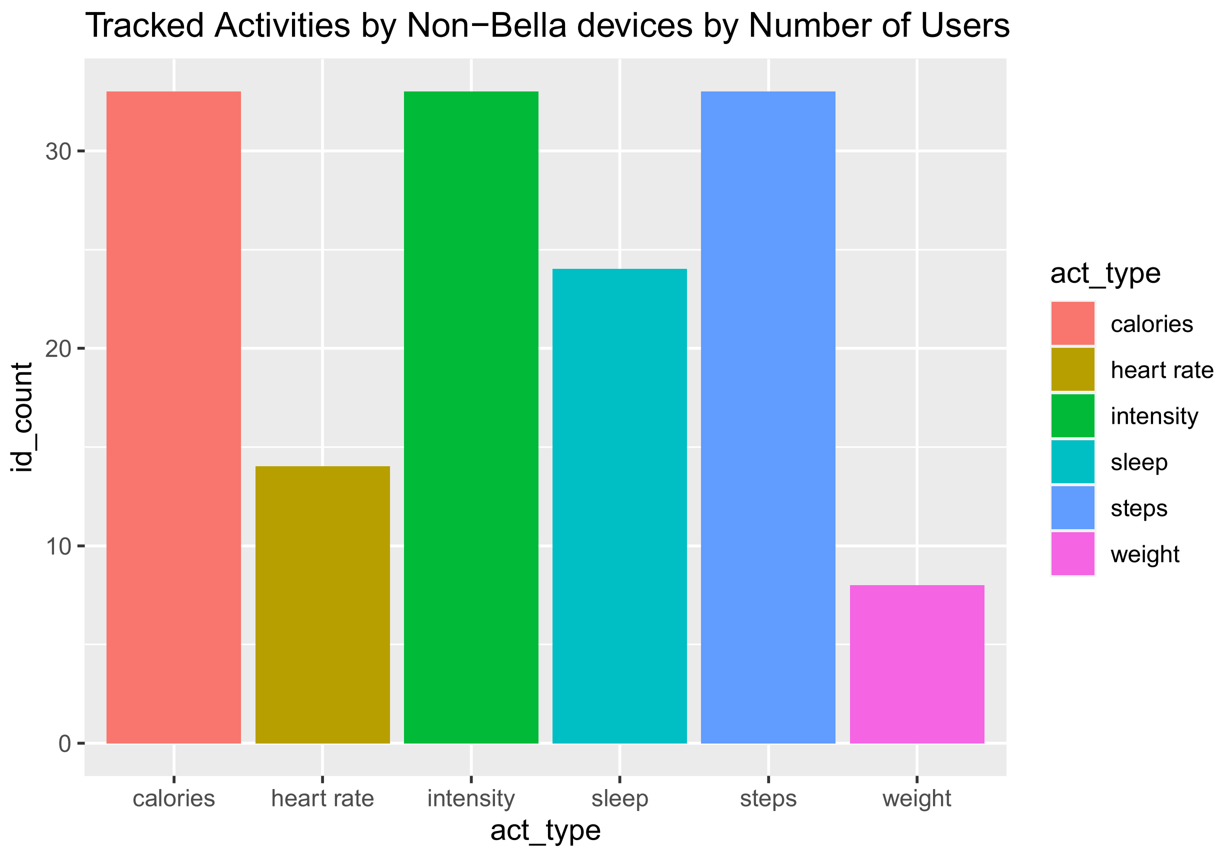
It was observed that all the participants recorded steps, intensity, and calories, while over 70% of them recorded data about their sleep. Most respondents either don't have devices that can record or chose not to record heart rate and weight data, as only 42% had data on heart rate and less than 24% documented weight data.

A dataframe is created to tabulate how users currently use their smart devices by looking at activities that users prefer or were able to document.

```
##     act_type id_count
## 1      steps       33
## 2   calories       33
## 3      sleep       24
```

```
## 4     weight       8
## 5 heart rate      14
## 6  intensity      33
```

Figure 1. Chart of Popular Tracked Activities



Share

Resultant Visualizations from Analysis.

Relationship between Total Steps, Distance and Calories Burn The data also show a strong positive relationship between the number of steps taken, distance traveled, and the calories expended daily.

Figure 2. Relationship Between Distance walked and Calories

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

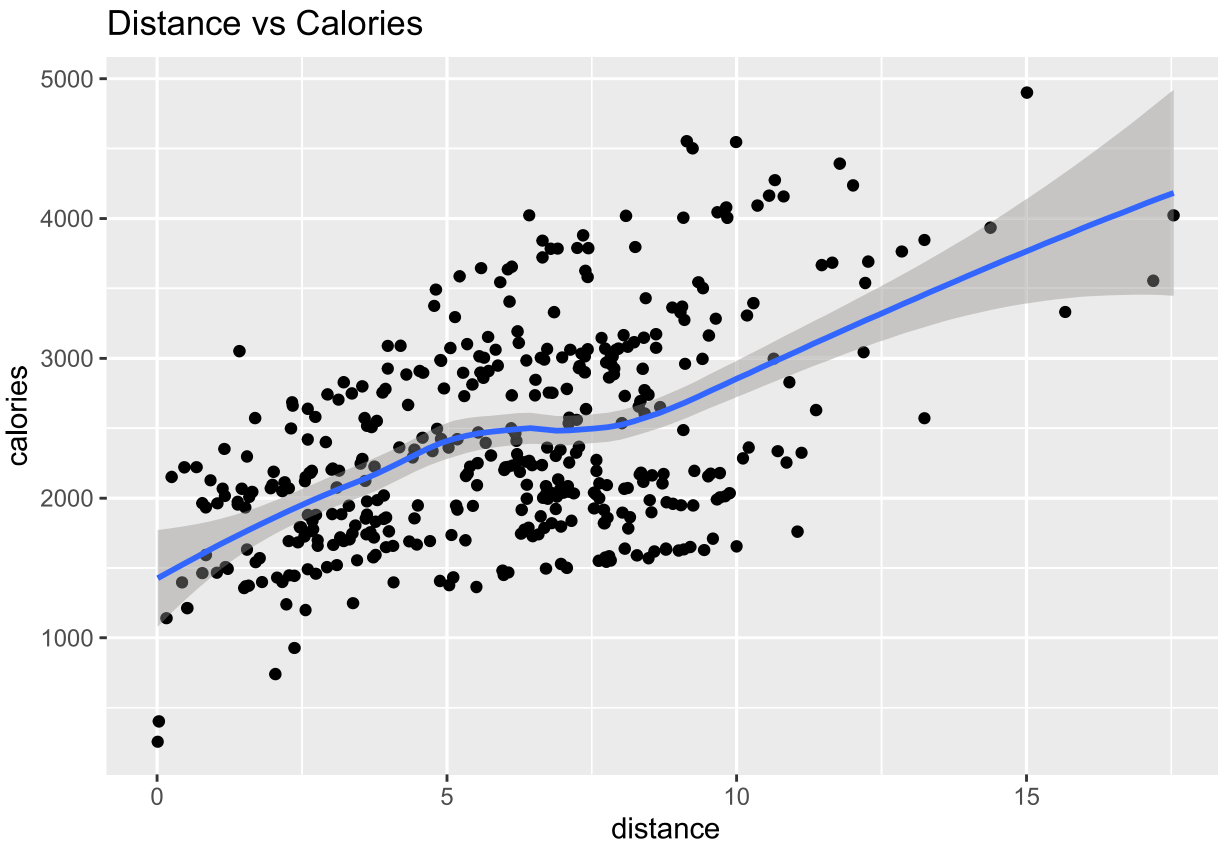
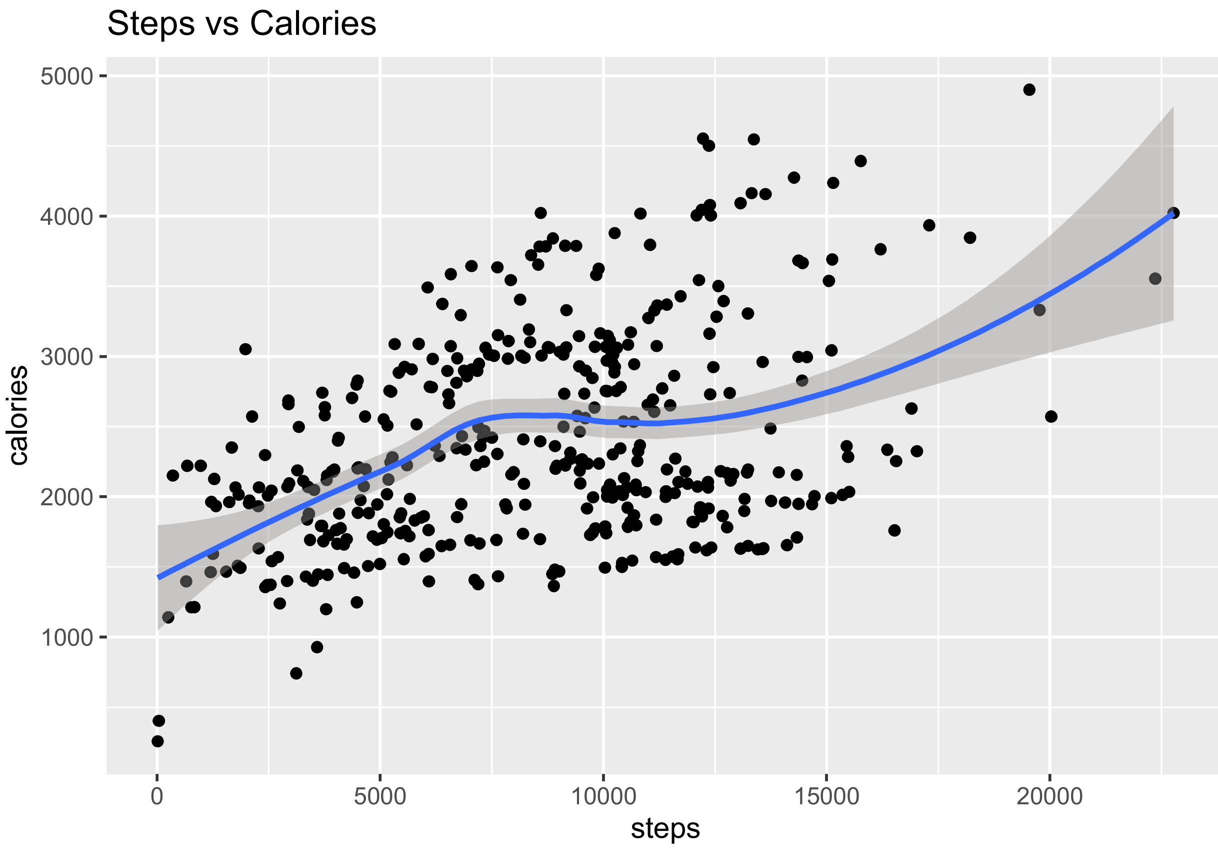


Figure 3. Relationship Between Steps and Calories

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



Relationship between Total Steps, Distance and Sleep The chart showed a negative relationship between the number of steps taken, distance traveled, and the hours of sleep daily – this was surprising as the reverse will have been perceived to be the case.

Figure 4. Relationship Between Distance walked and Hours of Sleep

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

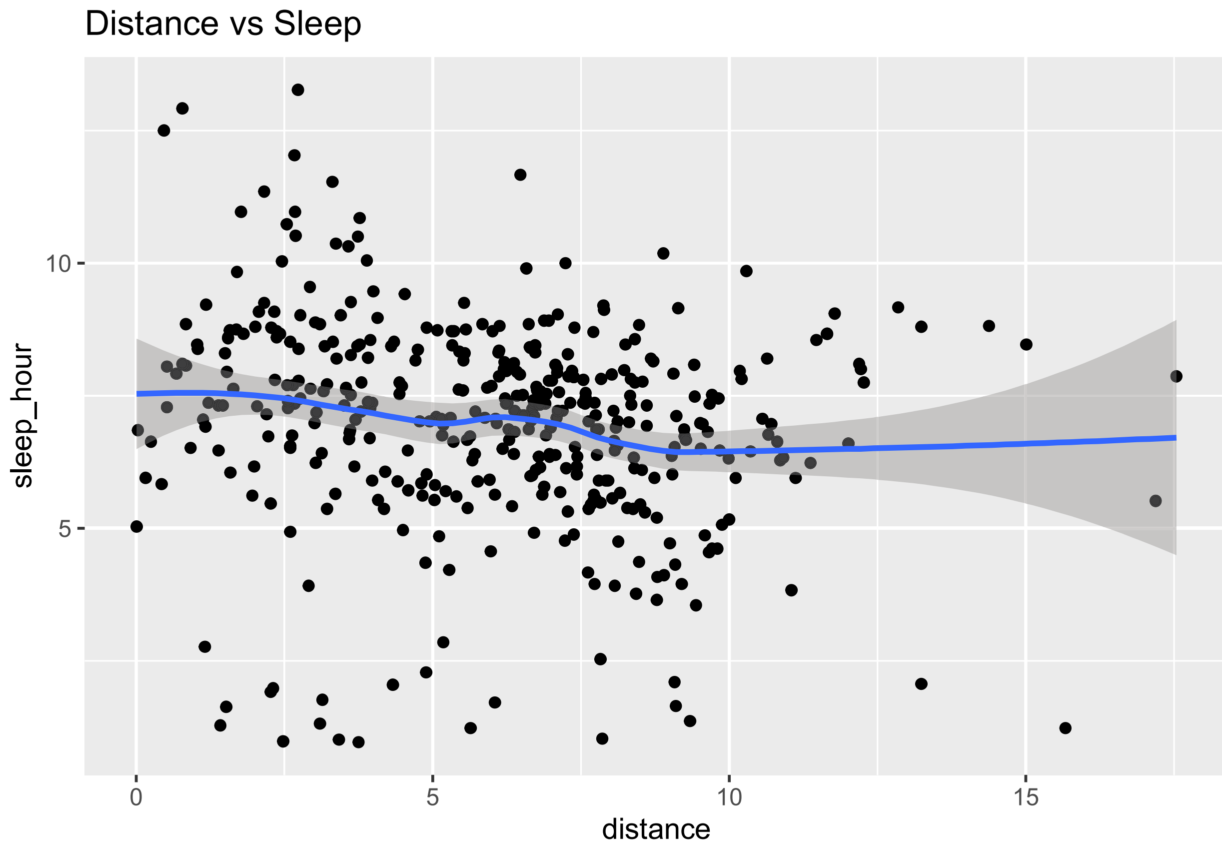
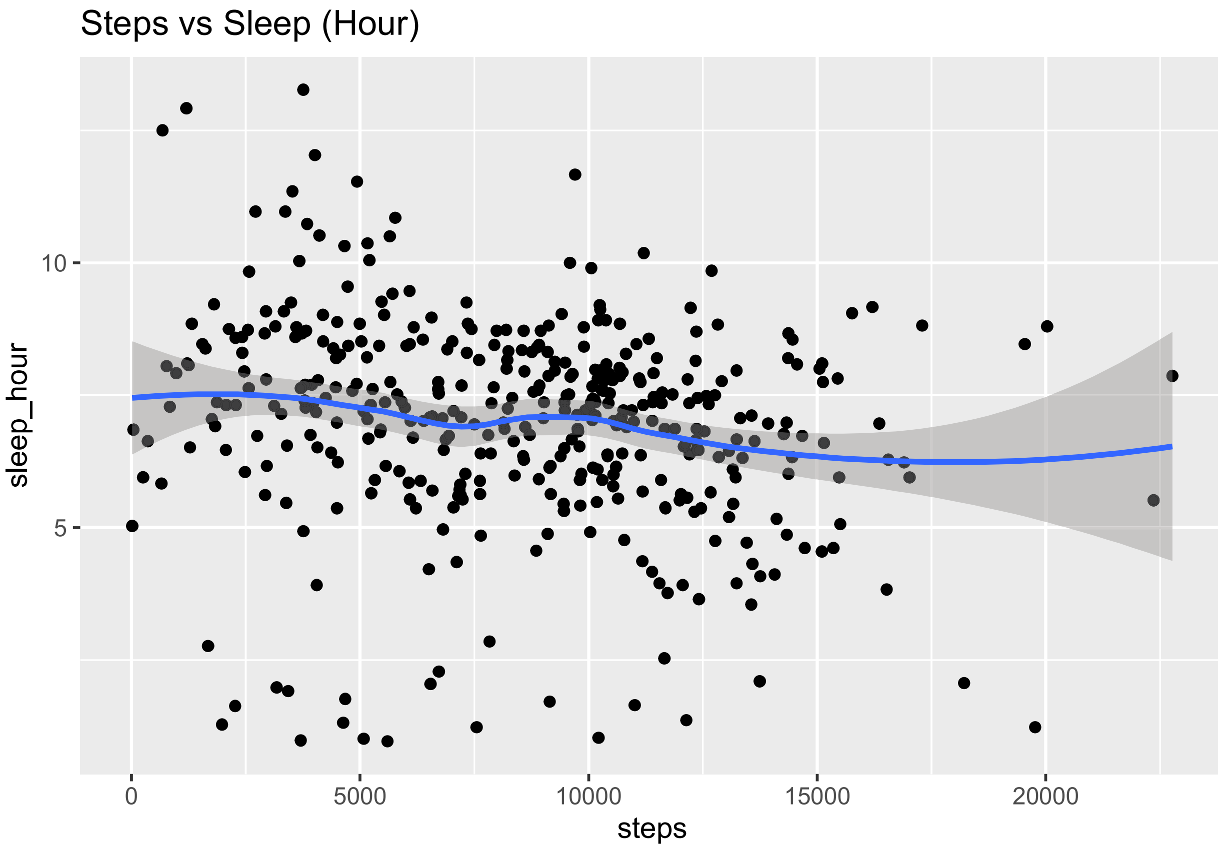


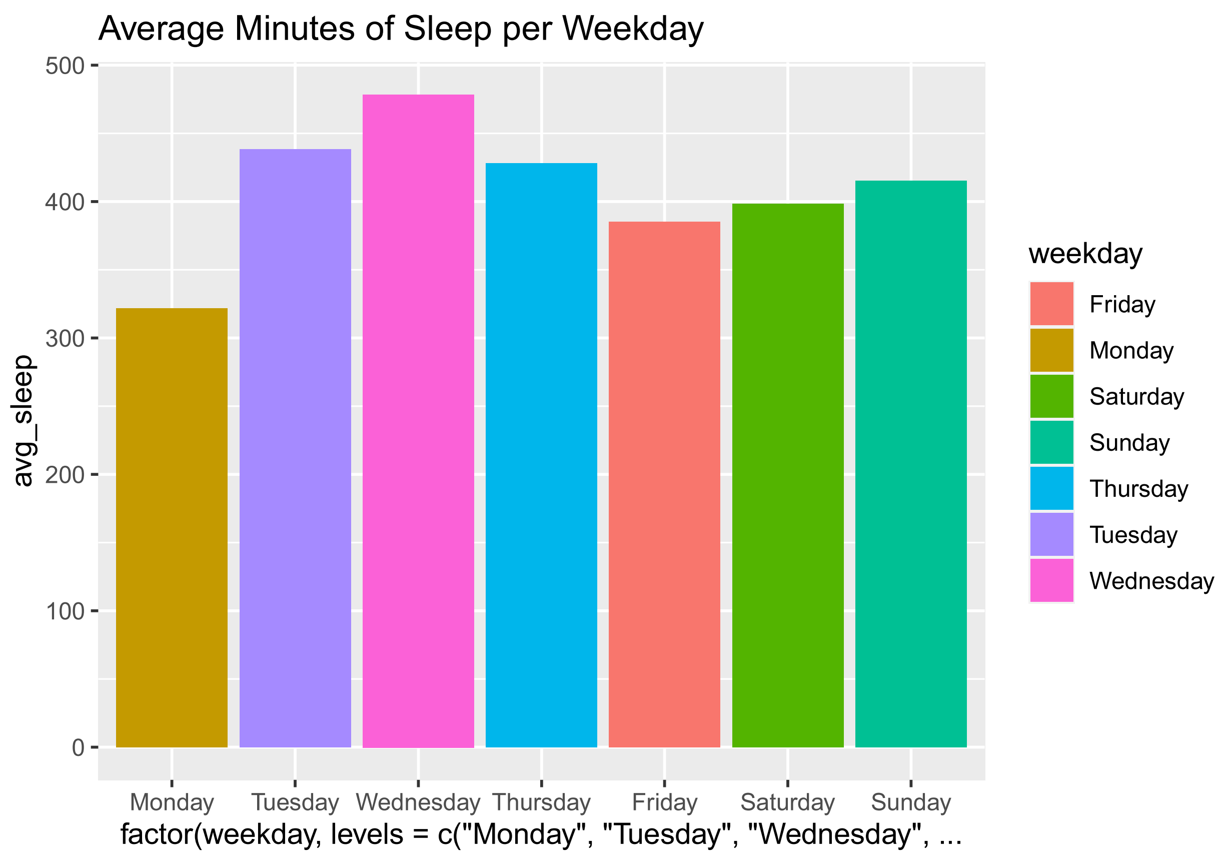
Figure 5. Relationship Between Total Steps and Hours of Sleep

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



Furthermore, the data showed that participants, on average, slept in the longest on Wednesdays. The average most extended sleep hours were on weekdays, Wednesdays, Tuesdays, and Thursdays.

Figure 6. Average Sleep in Minutes by day of the Week



Relationship between Sedentary and Calories Burn Another interesting chart is the Sedentary vs. Calories chart, which shows an initial strong positive relationship that later stemmed and flattened.

It is interesting to see that although sedentary is seen as inactivity, people still tend to expend a significant amount of calories while working at their desks in the office - I am assuming most of the respondents work in the office. This assumption that a considerable number of respondents have some form of an office job is buttressed by the sedentary vs. calories chart, which showed the highest average sedentary minutes during the week are Tuesday through Friday.

Figure 7. Relationship Between Inactivity and Calories Burned

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

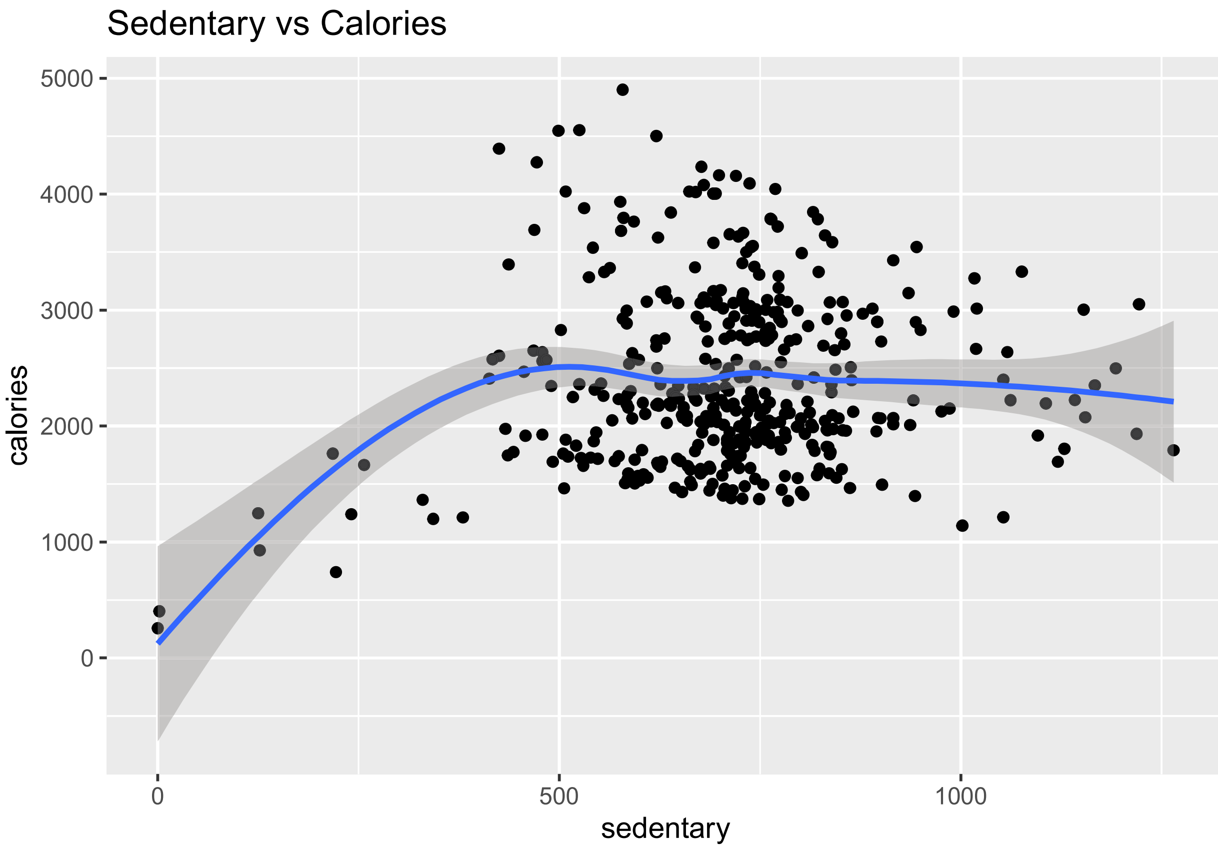


Figure 8. Average Inactivity in Minutes by day of the Week

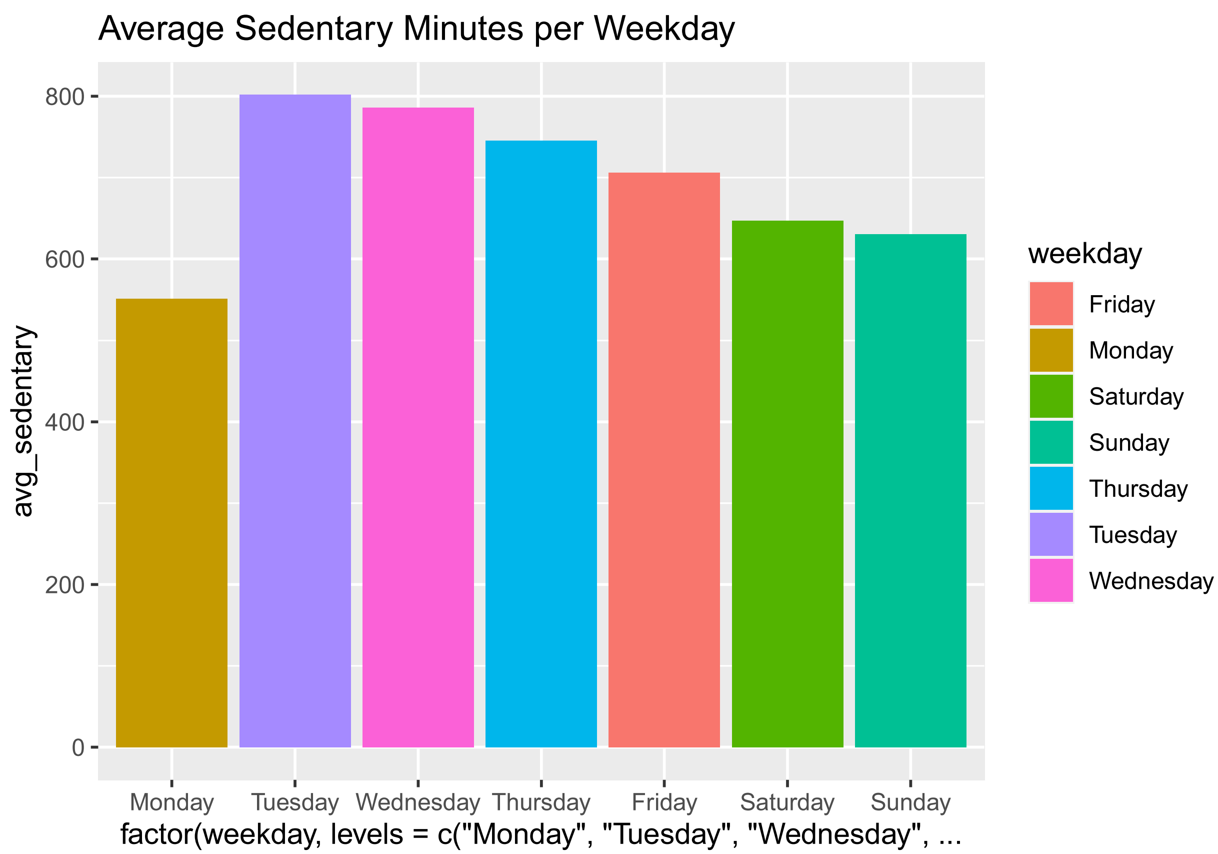
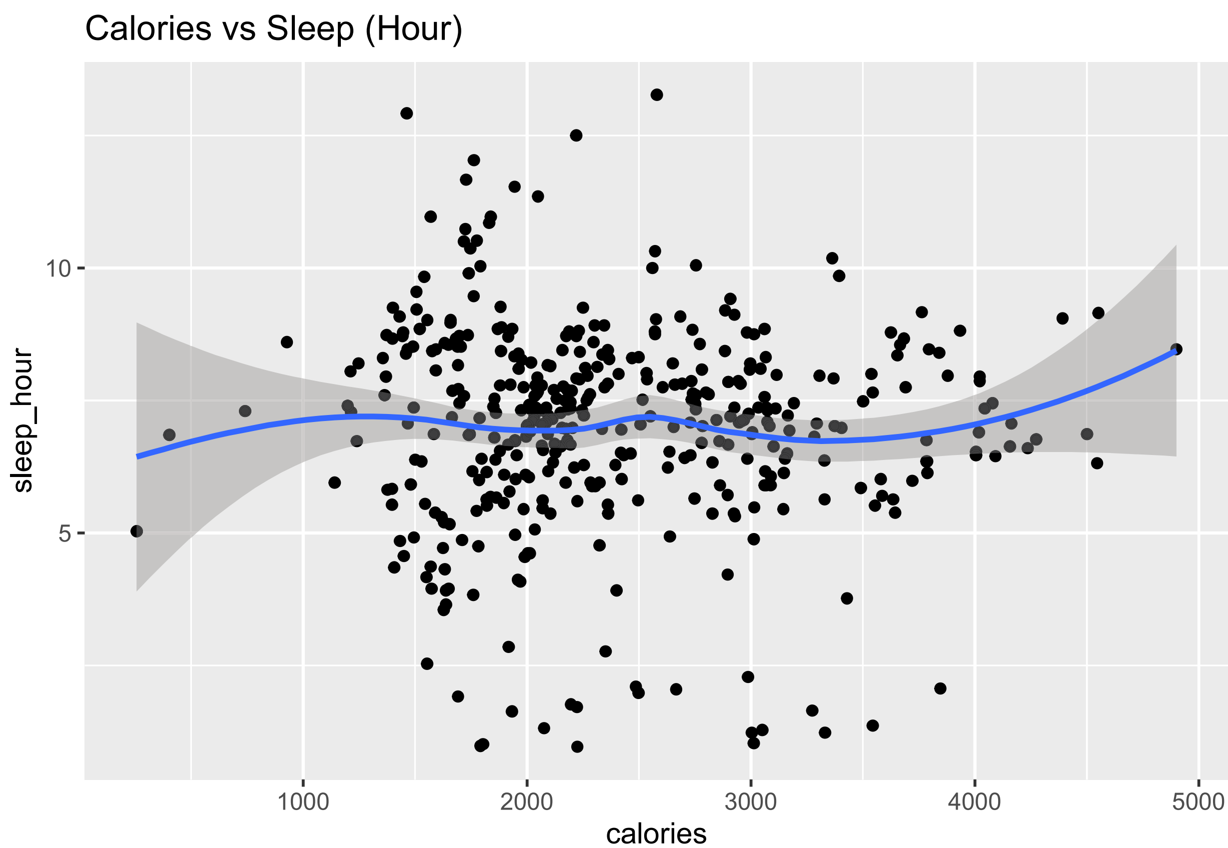


Figure 9. Relationship Between Calories Burned and Hours of Sleep

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



Summary

After analyzing the Fitbit fitness tracker survey data, I found some interesting insights that will help shape Bellabeat's marketing strategy.

Before proceeding with the case study, I had to make some assumptions. Bellabeat's ambition is to be a global leader in her industry, ideal this would require a sample size of over 2,400 participants; however, our dataset had only thirty-three participants. Therefore, my first assumption is that this sample size is acceptable. Another assumption is that the data represents the desired segment of the population, that is, women. The data did not have any demographic data; however, I assumed that all the survey participants were women of various age groups. The author noted, "Variation between output represents the use of different types of Fitbit trackers and individual tracking behaviors/preferences." This use pattern showed in the number of distinct users who tracked various activities (see figure 1); the most popularly tracked activities included steps, intensity, and calories. The least tracked activity was weight or BMI, and heart rate. The analysis revealed how various daily activities correlate to hours of sleep, calories burned, and the day of the week people tend to be most active. This information can help us choose which area of the spectrum to focus on depending on the individual market segment we'll be interested in. Figure 8 suggests that most of the participants are office workers or have some form of desk jobs, and this insight can help shape our marketing campaign. However, more information will be required to ascertain for sure.

Recommendation

1. Bellabeat customers already have access to a range of products; however, the Bellabeat app needs to provide a medium for data such as weight, heart rate, and sleep alongside other famous tracking metrics like steps and calories. Suppose Bellabeat has access to a wider variety of datapoint than the competition. In that case, Bellabeat might be able to use this information to make better tailor-made recommendations for customers or even help them with goal setting through data analytics, thereby

keeping their customers connected to the brand.

2. Should Bellabeat be interested in understanding her market much better, they should try having their survey through primary or secondary methods to ensure sample size, representation, and unbiased is possible. A sample size of at least 2,400 women randomly to have a representative sample since the company's goal is to be a significant player in the global market.