Bellabeat

Marisa Kuyava

How can a wellness company play it smart?

1. Introduction

Company

Bellabeat

- Bellabeat develops beautifully designed, high-tech, health-focused smart products for women.

Stakeholders

- Urška Sršen: Chief Creative Officer and Bellabeat Cofounder
- Sando Mu: Key member of executive team and Bellabeat Cofounder
- The Bellabeat Marketing Analytics team: Responsible for collecting, analyzing, and reporting data

2. Ask

Scenario

As a data analyst on the Bellabeat marketing analyst team I have been tasked with analyzing public smart device data to gain insight into how smart devices are being used by consumers.

Goal

Provide insights on trends in smart device usage that will help unlock new growth opportunities for the company and help guide marketing strategies.

3. Prepare

Libraries Utilized:

- tidyverse
- dplyr
- here
- janitor
- lubridate
- skimr
- ggplot2

Storage and Accessibility

The data set being analyzed is FitBit Fitness Tracker Data provided on Kaggle. The data set was generated by thirty eligible FitBit user respondents, to a distributed survey via Amazon Mechanical Turn, between the dates of April 12, 2016 and May 12, 2016. These users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring.

Organization and ROCCC

- The data is provided in 18 CSV files, after reviewing the data it was determined that the data in the files imported below contain the most complete and relevant data. The data was cleaned, including checking for duplicate files, and formats were standardized to ensure completeness and accuracy.
- The data set is obsolete for analyzing current trends as it was gathered in 2016.
- Data set sample size is limited to 30 individuals; this is the smallest sample size possible to be statistically in representing the population. A larger sample size would provide a higher confidence level, a decrease margin of error and a greater statistical significance.
- Data is possibly not relevant for the customers Bellabeat targeted audience as the data provided does not provide user respondents gender.
- Some data provided was manually entered and could contain error of be incorrect information.

Variables Measured

- Steps
- Sleep
- Calories
- Intensity

Data Import

- 6 of 18 files will be used as data provided in the other files are a duplication.
- Data was imported and each file was confirmed as successfuly uploaded.

Daily Activity

```
## 'data.frame':    940 obs. of  15 variables:
##  $ Id                      : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09
...
##  $ ActivityDate            : chr  "4/12/2016" "4/13/2016" "4/14/2016"
"4/15/2016" ...
##  $ TotalSteps              : int  13162 10735 10460 9762 12669 9705 13019
15506 10544 9819 ...
##  $ TotalDistance           : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance         : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance      : num  1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance     : num  6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes       : int  25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes     : int  13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes    : int  328 217 181 209 221 164 233 264 205 211
...
##  $ SedentaryMinutes        : int  728 776 1218 726 773 539 1149 775 818
838 ...
##  $ Calories                : int  1985 1797 1776 1745 1863 1728 1921 2035
1786 1775 ...
```

Daily Sleep

```
## 'data.frame':    413 obs. of  5 variables:
##  $ Id                : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr  "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00
AM" "4/15/2016 12:00:00 AM" "4/16/2016 12:00:00 AM" ...
##  $ TotalSleepRecords : int  1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: int  327 384 412 340 700 304 360 325 361 430 ...
##  $ TotalTimeInBed    : int  346 407 442 367 712 320 377 364 384 449 ...
```

Hourly Calories

```
## 'data.frame':    22099 obs. of  3 variables:
##  $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr  "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM"
"4/12/2016 2:00:00 AM" "4/12/2016 3:00:00 AM" ...
##  $ Calories    : int  81 61 59 47 48 48 48 47 68 141 ...
```

Hourly Intensity

```
## 'data.frame':    22099 obs. of  4 variables:
##  $ Id              : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour    : chr  "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM"
"4/12/2016 2:00:00 AM" "4/12/2016 3:00:00 AM" ...
##  $ TotalIntensity  : int  20 8 7 0 0 0 0 0 13 30 ...
##  $ AverageIntensity: num  0.333 0.133 0.117 0 0 ...
```

Hourly Steps

```
## 'data.frame':    22099 obs. of  3 variables:
##  $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr  "4/12/2016 12:00:00 AM" "4/12/2016 1:00:00 AM"
"4/12/2016 2:00:00 AM" "4/12/2016 3:00:00 AM" ...
##  $ StepTotal   : int  373 160 151 0 0 0 0 0 250 1864 ...
```

Weight

```
## 'data.frame':    67 obs. of  8 variables:
##  $ Id            : num  1.50e+09 1.50e+09 1.93e+09 2.87e+09 2.87e+09 ...
##  $ Date          : chr  "5/2/2016 11:59:59 PM" "5/3/2016 11:59:59 PM"
"4/13/2016 1:08:52 AM" "4/21/2016 11:59:59 PM" ...
##  $ WeightKg      : num  52.6 52.6 133.5 56.7 57.3 ...
##  $ WeightPounds  : num  116 116 294 125 126 ...
##  $ Fat           : int  22 NA NA NA NA 25 NA NA NA NA ...
##  $ BMI           : num  22.6 22.6 47.5 21.5 21.7 ...
##  $ IsManualReport: chr  "True" "True" "False" "True" ...
##  $ LogId         : num  1.46e+12 1.46e+12 1.46e+12 1.46e+12 1.46e+12 ...
```

4. Process

Data Cleaning
- Data set was checked for duplicates and missing values.
- Duplicate data was removed from the dataset
- Dates and date column name within the data set were all normalized
- Individual files were merged for easier analysis

5. Analyze

Data Analysis

Data Summary - Distances

```
##    Steps_Total      Distance_Total    High_Activity      Mid_Activity
## Min.    :    17    Min.    : 0.010   Min.    : 0.000   Min.    :0.0000
## 1st Qu.: 5189      1st Qu.: 3.592    1st Qu.: 0.000    1st Qu.:0.0000
## Median : 8913      Median : 6.270    Median : 0.570    Median :0.4200
## Mean   : 8515      Mean    : 6.012   Mean    : 1.446   Mean    :0.7439
## 3rd Qu.:11370      3rd Qu.: 8.005    3rd Qu.: 2.360    3rd Qu.:1.0375
## Max.   :22770      Max.    :17.540   Max.    :12.540   Max.    :6.4800
##    Low_Activity
## Min.    :0.010
## 1st Qu.:2.540
## Median :3.665
## Mean    :3.791
## 3rd Qu.:4.918
## Max.    :9.480
```

Standard Deviations - Distances

```
## [1] "Steps_Total:            4157.3756"

## [1] "Distance_Total:         3.0439"

## [1] "High_Activity_Distance: 1.9926"

## [1] "Mid_Activity_Distance:  0.9998"

## [1] "Low_Activity_Distance:  1.7208"
```

Data Summary - Time

```
##  High_Active_Min   Mid_Activive_Min  Low_Activive_Min  Sedentary_Min
## Min.    :  0.00   Min.    :  0.00   Min.    :  2.0    Min.    :   0.0
## 1st Qu.:  0.00    1st Qu.:  0.00    1st Qu.:158.0     1st Qu.: 631.2
## Median :  9.00    Median : 11.00    Median :208.0     Median : 717.0
## Mean    : 25.05   Mean    : 17.92   Mean    :216.5    Mean    : 712.1
## 3rd Qu.: 38.00    3rd Qu.: 26.75    3rd Qu.:263.0     3rd Qu.: 782.8
## Max.    :210.00   Max.    :143.00   Max.    :518.0    Max.    :1265.0
##    Asleep_Min        InBed_Min         Calorie
## Min.    : 58.0    Min.    : 61.0    Min.    : 257
## 1st Qu.:361.0     1st Qu.:403.8     1st Qu.:1841
## Median :432.5     Median :463.0     Median :2207
## Mean    :419.2    Mean    :458.5    Mean    :2389
## 3rd Qu.:490.0     3rd Qu.:526.0     3rd Qu.:2920
## Max.    :796.0    Max.    :961.0    Max.    :4900
```

Standard Deviations - Time

```
## [1] "High_Activity_Minutes:    36.2182"

## [1] "Mid_Activity_Minutes:    22.4186"

## [1] "Low_Activity_Minutes:    86.7053"

## [1] "Sedentary_Minutes:      166.1794"

## [1] "Asleep_Minutes:         118.6359"

## [1] "InBed_Minutes:          127.4551"

## [1] "Calories Burned:        758.435"
```

Observations from statistical Summary

- Users average a distance of 6.27 miles and 8515 steps daily
- There is a significant variation between users activity as the standard deviation for distance is 3 and for steps is 4157
- Users spend the most time doing Low activity, this correlates with low activity distance being the greatest
- Users burn an average of 2389 calories per day
- Users average sleep time is 6.98 hours with time in bed at 7.65 hours

Activity Summary

```
##            Count Percentage                                Steps
## Low Steps    96   23.41463                     Less than 5000
## Mid Steps   149   36.34146 More than 5000 and Less than 1000
## High Steps  165   40.24390     Equal or Greater than 10000
```

Observations from Activity Summary

- The majority of users have mid to high activity at 76%
- 40% reaching the recommended 10K steps per day
- 23% of users are well below the recommended 10K steps per day

Sleep Correlations

```
##                    Correlation of Minutes Asleep
## TotalSteps                         -0.19034392
## TotalDistance                      -0.17693625
## VeryActiveMinutes                  -0.08812658
## FairlyActiveMinutes                -0.24920793
## LightlyActiveMinutes                0.02758336
## SedentaryMinutes                   -0.60107314
## Calories                           -0.03169899
```
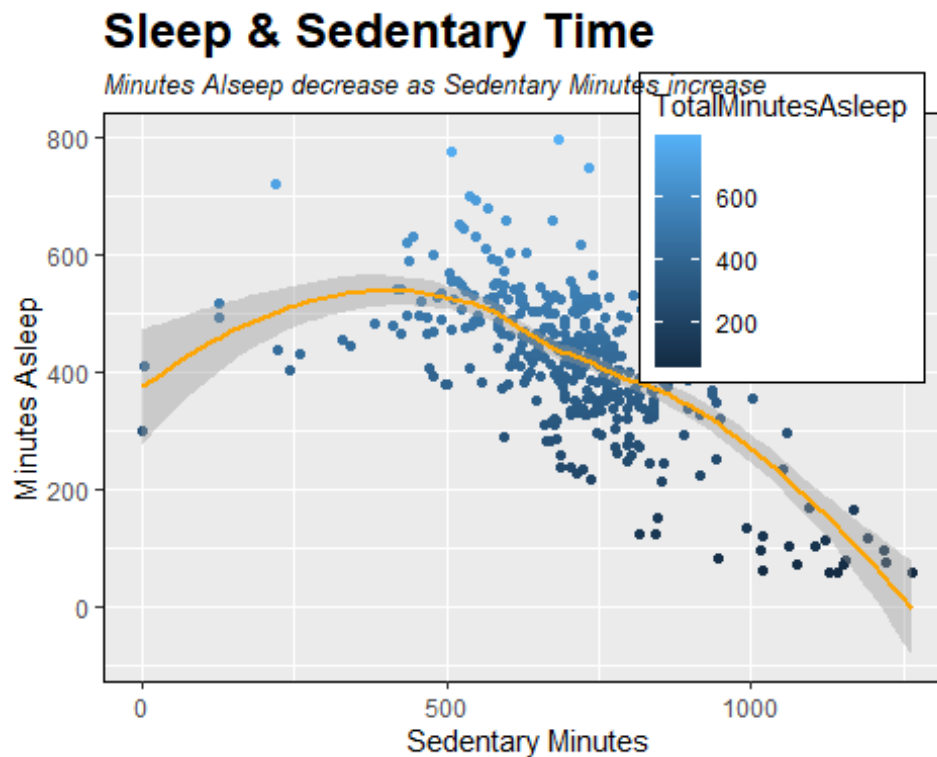
```
## TotalMinutesAsleep                          1.00000000
## TotalTimeInBed                               0.93042239
```

Observations from Sleep Correlations

- There is a moderate negative correlation between sleep and Sedentary Minutes, this
  suggests that users sleep better when their Sedentary Minutes are lower.

Calorie Correlations

```
##                      Correlation of Calories
## TotalSteps                       0.40630068
## TotalDistance                    0.52121615
## VeryActiveMinutes                0.61119829
## FairlyActiveMinutes              0.17598776
## LightlyActiveMinutes             0.11376614
## SedentaryMinutes                 0.09865571
## Calories                         1.00000000
## TotalMinutesAsleep              -0.03169899
## TotalTimeInBed                  -0.13480458
```

Observations from Calories Correlations

- There is a strong positive correlation between Total Steps, Distance, Active Minutes
  and Calories burned; as expected.
- Total Distance and Total Steps both have a positive correlation with Calories
  burned; distance having a slightly stronger correlation
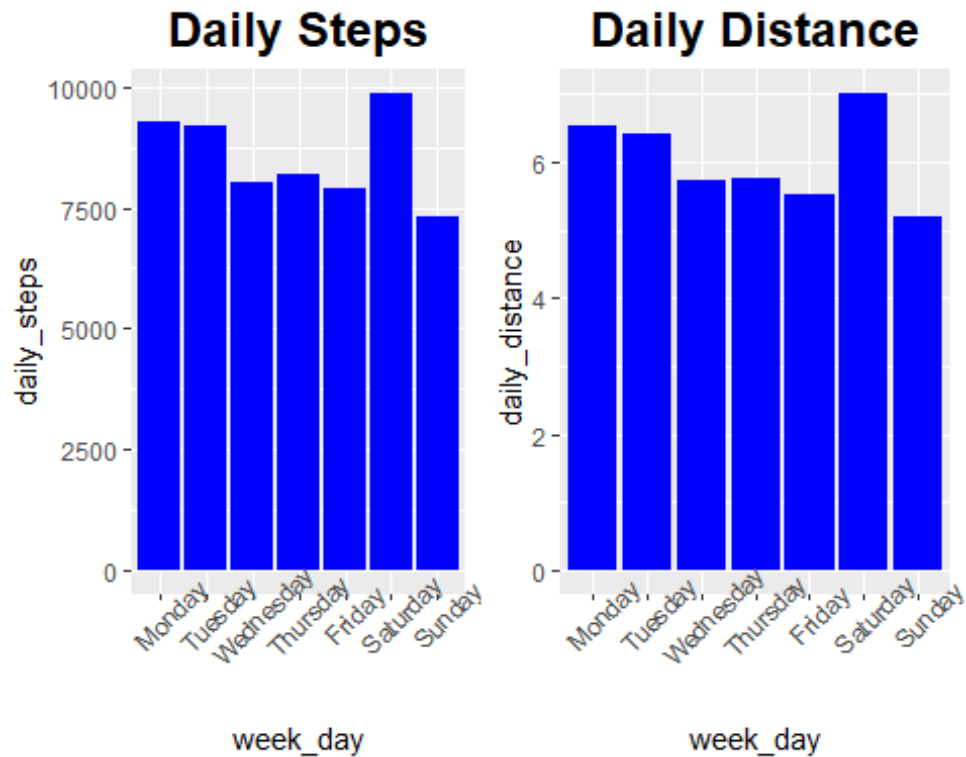
Most Active Times

User activity has a correlation to the time of day and day of the week
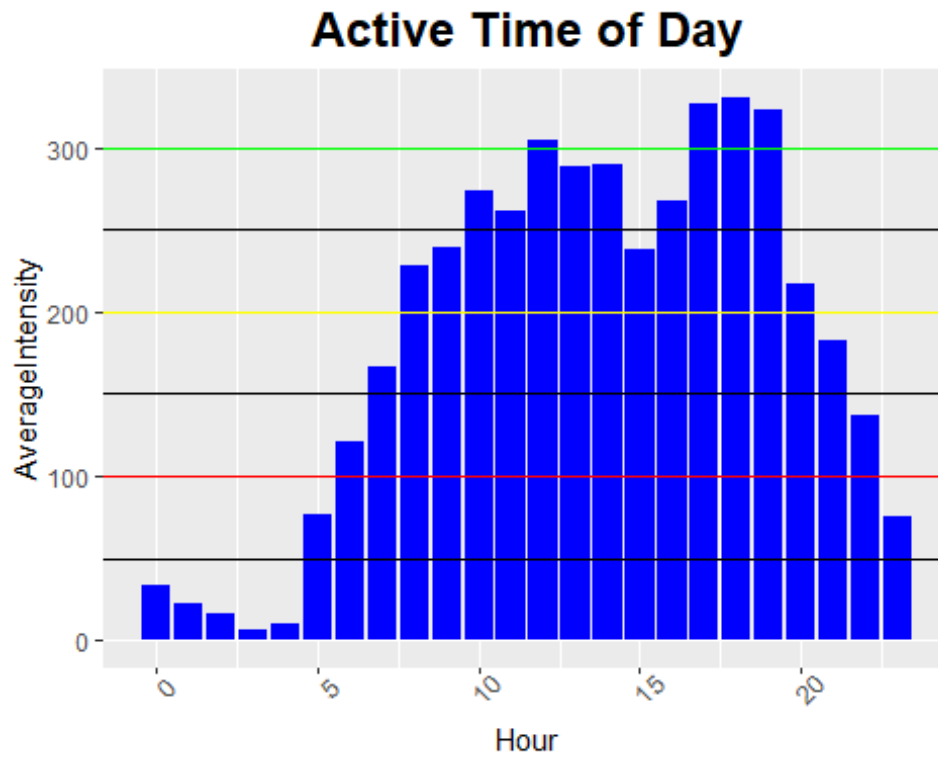
Active Hours

- Very Active 08:00 -20:00
- Mid Active 06:00 - 08:00 / 20:00 -22:00
- Low Active 23:00 - 6:00

Active Days

- Users are most active on Saturday, followed by Monday and Tuesday respectively
- Mid-week (Wednesday, Thursday, Friday) has slightly lower activity
- Sunday is the lowest activity day

6. Act

Findings and Recommendations

Observations

- Users average a daily distance of 6.27 miles and 8515, however as the standard deviation is 3 miles and 4157 steps there is a significant variation between the activities of the users
- Users spend the most time in low impact activity and only 40% are meeting or exceeding the recommended 10K daily steps
- Users with higher sedentary time have lower time spent asleep
- Users are most active on Saturdays, Mondays and Tuesdays with mid day and evening hours having the highest activity

Recommendations

Functionality:

- Gamification: Incorporating both individual and competitive goals, rewards, streaks, and achievements can help increase user engagement and activity.

- Recommendations for active lifestyle and quality sleep strategies can add additional value for the user

- Social Networking: Providing in-app chat ability and user interaction can promote exercise habits and enhance user engagement. Have a community provides accountability and motivation for the users to be active.

- Alerts:

    - Give users the ability to set alerts and reminders
        - Daily exercise and sleep reminders
        - Step count notifications through out the day
        - Extended period sedentary time warning for a break reminder and to encourage activity

Data

- The data analysis was done on a very limited and dated data set - Additional analysis with Bellabeat's tracking data recommended