Case Study: Wellness Product Marketing

**Ojaswini Patankar**

**2023-05-11**

**Case Study**

## How can a wellness Technology Company Improve Marketing Strategy?

***CONTEXT***:

Bellabeat, a high-tech manufacturing of health focused products for women. It was founded in 2013.Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Cofounder and chief creative of Bellabeat, believes that analyzing smart fitness data could help unlock new growth opportunities for the company. Bellabeat has the following products.

1. **Bellabeat app**: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits.
2. **Leaf**: Bellabeat’s classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
3. **Time**: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress.
4. **Spring**: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day.
5. **Bellabeat membership**: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

## Ask: This question will guide to clearly specify the Business task.

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 marketing strategy?

***Business Task*:**

How to improve the Marketing strategy so that it will increase Bellabeat products usage.

1. Analyze smart device usage data in order to gain insight into how consumers use non Bellabeat smart devices.
2. Then select one Bellabeat product and apply these insights to that product.
3. Finally help to guide Bellabeat marketing strategy.

***Stakeholders***:

○ Urška Sršen: Bellabeat’s cofounder and Chief Creative Officer

○ Sando Mur: Mathematician and Bellabeat’s cofounder; key member of the Bellabeat executive team.

## Prepare:

**Data Provenance-CCO: Public Domain**

FitBit Fitness Tracker Data<https://www.kaggle.com/datasets/arashnic/fitbit> (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users. This dataset was generated by respondents to distributed survey via Amazon Mechanical Turk between 03.12.2016-05.12.2016. Thirty eligible fitbit users consented to the submission of personal tracker data, including min-level o/p for physical activity, heart rate and sleep monitoring.

**Expected Update Frequency: Annually**

* Variation between o/p represents use of different types of fitbit trackers and individual tracking behaviors/preferences.
* Data is interoperable, means the ability of data system and services to openly connect and share data.
* The dataset has 18 files for the different information as daily activity records, sleeping time, heart rate, calories, and weight. It is comprehensive.

Limitation of data: Dataset has only 30 people’s information. It shows sampling bias, data is not representative of the population as a whole. There is no information on Gender, so not assure that data information belongs to women only.

## Process:

I checked the dataset in Excel spreadsheet and observed the problems in timestamp data, so I need to format timestamp data before analysis. I am choosing R as my cleaning, analysis, and visualization tool.

install.packages("tidyverse")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

install.packages("dplyr")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

install.packages("skimr")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

install.packages("janitor")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

install.packages("lubridate")

## Installing package into '/cloud/lib/x86\_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)

# Loading Packages

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)  
library(skimr)  
library(janitor)

##   
## Attaching package: 'janitor'  
##   
## The following objects are masked from 'package:stats':  
##   
## chisq.test, fisher.test

library(lubridate)  
library(ggplot2)

##Impoting Dataset

I am importing 4 files here to work with, as they include all the information needed. Exploring data to get high level insight.

daily\_activity<-read.csv("dailyActivity\_merged.csv")  
head(daily\_activity)

## Id ActivityDate 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

weight\_info<-read.csv("weightLogInfo\_merged.csv")  
head(weight\_info)

## Id Date WeightKg WeightPounds Fat BMI  
## 1 1503960366 5/2/2016 11:59:59 PM 52.6 115.9631 22 22.65  
## 2 1503960366 5/3/2016 11:59:59 PM 52.6 115.9631 NA 22.65  
## 3 1927972279 4/13/2016 1:08:52 AM 133.5 294.3171 NA 47.54  
## 4 2873212765 4/21/2016 11:59:59 PM 56.7 125.0021 NA 21.45  
## 5 2873212765 5/12/2016 11:59:59 PM 57.3 126.3249 NA 21.69  
## 6 4319703577 4/17/2016 11:59:59 PM 72.4 159.6147 25 27.45  
## IsManualReport LogId  
## 1 True 1.462234e+12  
## 2 True 1.462320e+12  
## 3 False 1.460510e+12  
## 4 True 1.461283e+12  
## 5 True 1.463098e+12  
## 6 True 1.460938e+12

sleep\_day<-read.csv("sleepDay\_merged.csv")  
glimpse(sleep\_day)

## 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…

hourly\_intensity<-read.csv("hourlyIntensities\_merged.csv")  
head(hourly\_intensity)

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

Formatting the timestamp data.Cleaning and arranging data before analysis.

sleep\_day$SleepDay<-parse\_date\_time(sleep\_day$SleepDay,"%m/%d/%y %I:%M:%S %p")  
  
weight\_info$Date<-parse\_date\_time(weight\_info$Date, "%m/%d/%y %I:%M:%S %p")

colnames(daily\_activity)

## [1] "Id" "ActivityDate"   
## [3] "TotalSteps" "TotalDistance"   
## [5] "TrackerDistance" "LoggedActivitiesDistance"  
## [7] "VeryActiveDistance" "ModeratelyActiveDistance"  
## [9] "LightActiveDistance" "SedentaryActiveDistance"   
## [11] "VeryActiveMinutes" "FairlyActiveMinutes"   
## [13] "LightlyActiveMinutes" "SedentaryMinutes"   
## [15] "Calories"

daily\_Activity<-daily\_activity%>%filter(TotalSteps!=0)  
head(daily\_Activity)

## Id ActivityDate 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

str(sleep\_day)

## '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 : POSIXct, format: "2016-04-12" "2016-04-13" ...  
## $ 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 ...

daily\_activity\_new<-daily\_activity%>%arrange(-VeryActiveMinutes)  
  
head(daily\_activity\_new)

## Id ActivityDate TotalSteps TotalDistance TrackerDistance  
## 1 5577150313 4/24/2016 15764 11.78 11.78  
## 2 5577150313 4/30/2016 12363 9.24 9.24  
## 3 5577150313 4/17/2016 12231 9.14 9.14  
## 4 5577150313 5/1/2016 13368 9.99 9.99  
## 5 1624580081 5/1/2016 36019 28.03 28.03  
## 6 5577150313 4/16/2016 14269 10.66 10.66  
## LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance  
## 1 0 7.65 2.15  
## 2 0 5.83 0.79  
## 3 0 5.98 0.83  
## 4 0 5.31 1.44  
## 5 0 21.92 4.19  
## 6 0 6.64 1.28  
## LightActiveDistance SedentaryActiveDistance VeryActiveMinutes  
## 1 1.98 0.00 210  
## 2 2.61 0.00 207  
## 3 2.32 0.00 200  
## 4 3.24 0.00 194  
## 5 1.91 0.02 186  
## 6 2.73 0.00 184  
## FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories  
## 1 65 141 425 4392  
## 2 45 163 621 4501  
## 3 37 159 525 4552  
## 4 72 178 499 4546  
## 5 63 171 1020 2690  
## 6 56 158 472 4274

colnames(daily\_activity\_new)[colnames(daily\_activity\_new)=="ActivityDate"]<-"date"  
daily\_activity\_new$date<-as.Date(daily\_activity\_new$date,format="%m/%d/%y")  
  
sleep\_day$date<-as.Date(sleep\_day$SleepDay, format="%m/%d/%y")

hourly\_intensity$date<-as.Date(hourly\_intensity$ActivityHour,format ="%m/%d/%y")  
  
weight\_info$Date1<-as.Date(weight\_info$Date,format="%m/%d/%y")

skim\_without\_charts(daily\_activity\_new)

**Data summary**

|  |  |
| --- | --- |
| Name | daily\_activity\_new |
| Number of rows | 940 |
| Number of columns | 15 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Column type frequency: |  |
| Date | 1 |
| numeric | 14 |
| \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |
| Group variables | None |

**Variable type: Date**

| skim\_variable | n\_missing | complete\_rate | min | max | median | n\_unique |
| --- | --- | --- | --- | --- | --- | --- |
| date | 0 | 1 | 2020-04-12 | 2020-05-12 | 2020-04-26 | 31 |

**Variable type: numeric**

| skim\_variable | n\_missing | complete\_rate | mean | sd | p0 | p25 | p50 | p75 | p100 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Id | 0 | 1 | 4.855407e+09 | 2.424805e+09 | 1503960366 | 2.320127e+09 | 4.445115e+09 | 6.962181e+09 | 8.877689e+09 |
| TotalSteps | 0 | 1 | 7.637910e+03 | 5.087150e+03 | 0 | 3.789750e+03 | 7.405500e+03 | 1.072700e+04 | 3.601900e+04 |
| TotalDistance | 0 | 1 | 5.490000e+00 | 3.920000e+00 | 0 | 2.620000e+00 | 5.240000e+00 | 7.710000e+00 | 2.803000e+01 |
| TrackerDistance | 0 | 1 | 5.480000e+00 | 3.910000e+00 | 0 | 2.620000e+00 | 5.240000e+00 | 7.710000e+00 | 2.803000e+01 |
| LoggedActivitiesDistance | 0 | 1 | 1.100000e-01 | 6.200000e-01 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 4.940000e+00 |
| VeryActiveDistance | 0 | 1 | 1.500000e+00 | 2.660000e+00 | 0 | 0.000000e+00 | 2.100000e-01 | 2.050000e+00 | 2.192000e+01 |
| ModeratelyActiveDistance | 0 | 1 | 5.700000e-01 | 8.800000e-01 | 0 | 0.000000e+00 | 2.400000e-01 | 8.000000e-01 | 6.480000e+00 |
| LightActiveDistance | 0 | 1 | 3.340000e+00 | 2.040000e+00 | 0 | 1.950000e+00 | 3.360000e+00 | 4.780000e+00 | 1.071000e+01 |
| SedentaryActiveDistance | 0 | 1 | 0.000000e+00 | 1.000000e-02 | 0 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.100000e-01 |
| VeryActiveMinutes | 0 | 1 | 2.116000e+01 | 3.284000e+01 | 0 | 0.000000e+00 | 4.000000e+00 | 3.200000e+01 | 2.100000e+02 |
| FairlyActiveMinutes | 0 | 1 | 1.356000e+01 | 1.999000e+01 | 0 | 0.000000e+00 | 6.000000e+00 | 1.900000e+01 | 1.430000e+02 |
| LightlyActiveMinutes | 0 | 1 | 1.928100e+02 | 1.091700e+02 | 0 | 1.270000e+02 | 1.990000e+02 | 2.640000e+02 | 5.180000e+02 |
| SedentaryMinutes | 0 | 1 | 9.912100e+02 | 3.012700e+02 | 0 | 7.297500e+02 | 1.057500e+03 | 1.229500e+03 | 1.440000e+03 |
| Calories | 0 | 1 | 2.303610e+03 | 7.181700e+02 | 0 | 1.828500e+03 | 2.134000e+03 | 2.793250e+03 | 4.900000e+03 |

**Checking for duplicate, by using distinct functions. It will give only unique ids.**

n\_distinct(daily\_activity\_new$Id)

## [1] 33

n\_distinct(hourly\_intensity$Id)

## [1] 33

n\_distinct(weight\_info$Id)

## [1] 8

n\_distinct(sleep\_day$Id)

## [1] 24

*Summary statistics of dataset*

daily\_activity\_new%>%  
 select(TotalSteps,VeryActiveMinutes,Calories,SedentaryMinutes)%>%  
 summary()

## TotalSteps VeryActiveMinutes Calories SedentaryMinutes  
## Min. : 0 Min. : 0.00 Min. : 0 Min. : 0.0   
## 1st Qu.: 3790 1st Qu.: 0.00 1st Qu.:1828 1st Qu.: 729.8   
## Median : 7406 Median : 4.00 Median :2134 Median :1057.5   
## Mean : 7638 Mean : 21.16 Mean :2304 Mean : 991.2   
## 3rd Qu.:10727 3rd Qu.: 32.00 3rd Qu.:2793 3rd Qu.:1229.5   
## Max. :36019 Max. :210.00 Max. :4900 Max. :1440.0

daily\_activity\_new%>%  
 select(VeryActiveMinutes,FairlyActiveMinutes,LightlyActiveMinutes)%>%  
 summary()

## VeryActiveMinutes FairlyActiveMinutes LightlyActiveMinutes  
## Min. : 0.00 Min. : 0.00 Min. : 0.0   
## 1st Qu.: 0.00 1st Qu.: 0.00 1st Qu.:127.0   
## Median : 4.00 Median : 6.00 Median :199.0   
## Mean : 21.16 Mean : 13.56 Mean :192.8   
## 3rd Qu.: 32.00 3rd Qu.: 19.00 3rd Qu.:264.0   
## Max. :210.00 Max. :143.00 Max. :518.0

sleep\_day%>%  
 select(TotalSleepRecords,TotalMinutesAsleep,TotalTimeInBed)%>%  
 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

weight\_info%>%  
 select(WeightKg,BMI)%>%  
 summary()

## WeightKg BMI   
## Min. : 52.60 Min. :21.45   
## 1st Qu.: 61.40 1st Qu.:23.96   
## Median : 62.50 Median :24.39   
## Mean : 72.04 Mean :25.19   
## 3rd Qu.: 85.05 3rd Qu.:25.56   
## Max. :133.50 Max. :47.54

###Intresting Facts discoveries from this summary:It provide the high level information about the data.

1.) The very active min on an avg is 21.16min, which is very less than the standard minutes.People should do 75to 150 min of vigorous intensity physical activity. 2.) The majority of the participants are lighly active. 3.) BMI has a mean 25.19 ,which is greater than the standard (18-24),so as weight.Most of the people are overweight. 4.) The average sleeping hour is approximately 7 hour.

merge2<-merge(daily\_activity\_new,sleep\_day,by="Id")  
   
intensity<-hourly\_intensity%>%group\_by(ActivityHour)%>%drop\_na()%>%summarize(mean\_intensity=mean(TotalIntensity))

##visualization:

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

A picture containing text, screenshot, diagram, line

Description automatically generated

The scatter plot shows the proportional relation between two variables. If you increase the daily steps you will burn more calories.

## `geom\_smooth()` using method = 'loess' and formula = 'y ~ x'

A picture containing map, text, screenshot, diagram

Description automatically generated

The plot shows the negative relation between sedentary minutes and calories.

A picture containing text, screenshot, black and white

Description automatically generated

Physical activity improves the sleeping pattern.

A picture containing screenshot, diagram, text, plot

Description automatically generated

The plot shows the frequency of physical activity.

A picture containing text, screenshot, diagram, plot

Description automatically generated

The line graph shows that very few people are vigorously active and most of the people are lightly active.

A picture containing screenshot, text, diagram, plot

Description automatically generated

The plot shows that lots of people are overweight and have a obesity issues.

**Act:**

### Findings:

1. There are very few people who are physically active. Physical activity has significant health benefits for hearts, bodies, and minds.
2. Sedentary minutes are high. Sedentary behavior increases adiposity and reduce the sleep duration.
3. People have obesity issues. Regular physical activity can help maintain a healthy body weight, reduce the risk of hypertensions, coronary heart disease, stroke, diabetes, various types of cancer and depression.

### Recommendations:

1. Bellabeat company should use gamification, which keeps people invested and interested in moving forward, reaching goals and winning awards.
2. Bellabeat app can give daily notification for setting small goals that are specific, measurable, action oriented, realistic and time sensitive. It will increase self-efficacy in people watching that goal will activate the reward system in the brain and eventually motivate the people.