Bella Beat Case Study

Yipu Lerina

16 November 2021

About the company

Urška Sršen and **Sando Mur** founded **Bellabeat**,a high-tech company that manufactures health-focused smart products.They offer different smart devices that collect data on activity, sleep, stress, and reproductive health to empower women with knowledge about their own health and habits. Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

Business Task

Identify trends in smart device usage and provide recommendations to improve Bellabeat marketing strategy.

1. Loading packages

```
library(tidyverse)
library(lubridate)
library(janitor)
library(here)
```

2. Importing datasets

I'll be using three datasets for making my analysis

```
activity <- read_csv(here("Data","dailyActivity_merged.csv"))
sleep <- read_csv(here("Data", "sleepDay_merged.csv"))
steps <- read_csv(here("Data", "hourlySteps_merged.csv"))
```

3. Clean Dataset

Making necessary changes for easier data exploration.

```
activity <- activity%>%
  clean_names()%>%
  drop_na()%>%
  distinct()
```

```
sleep <- sleep%>%
  clean_names()%>%
  distinct()%>%
  drop_na()

steps<-steps%>%
  distinct()%>%
  clean_names()%>%
  drop_na()
```

4. Formatting dates

```
activity <- activity %>%
  rename(date = activity_date) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y"))


sleep<-sleep%>%
  separate(col = sleep_day, into  = c('date', 'time'), sep = ' ')

sleep$date = mdy(sleep$date)
sleep$time = hms(sleep$time)


steps <- steps%>%
  rename(date_time = activity_hour)%>%
  mutate(date_time = as.POSIXct(date_time, format = "%m/%d/%Y %I:%M:%S %p", tz = Sys.timezone()))


head(activity)
```

```
## # A tibble: 6 x 15
##       id date       total_steps total_distance tracker_distance logged_activiti~
##    <dbl> <date>           <dbl>          <dbl>            <dbl>             <dbl>
## 1 1.50e9 2016-04-12       13162           8.5              8.5                 0
## 2 1.50e9 2016-04-13       10735           6.97             6.97                0
## 3 1.50e9 2016-04-14       10460           6.74             6.74                0
## 4 1.50e9 2016-04-15        9762           6.28             6.28                0
## 5 1.50e9 2016-04-16       12669           8.16             8.16                0
## 6 1.50e9 2016-04-17        9705           6.48             6.48                0
## # ... with 9 more variables: very_active_distance <dbl>,
## #   moderately_active_distance <dbl>, light_active_distance <dbl>,
## #   sedentary_active_distance <dbl>, very_active_minutes <dbl>,
## #   fairly_active_minutes <dbl>, lightly_active_minutes <dbl>,
## #   sedentary_minutes <dbl>, calories <dbl>
```

```
head(steps)
```

```
## # A tibble: 6 x 3
##           id date_time           step_total
```

```
##        <dbl> <dttm>                   <dbl>
## 1 1503960366 2016-04-12 00:00:00       373
## 2 1503960366 2016-04-12 01:00:00       160
## 3 1503960366 2016-04-12 02:00:00       151
## 4 1503960366 2016-04-12 03:00:00         0
## 5 1503960366 2016-04-12 04:00:00         0
## 6 1503960366 2016-04-12 05:00:00         0
```

```
head(sleep)
```

```
## # A tibble: 6 x 6
##       id date       time      total_sleep_rec~ total_minutes_a~ total_time_in_b~
##    <dbl> <date>     <Period>             <dbl>            <dbl>            <dbl>
## 1 1.50e9 2016-04-12 12H 0M 0S                1              327              346
## 2 1.50e9 2016-04-13 12H 0M 0S                2              384              407
## 3 1.50e9 2016-04-15 12H 0M 0S                1              412              442
## 4 1.50e9 2016-04-16 12H 0M 0S                2              340              367
## 5 1.50e9 2016-04-17 12H 0M 0S                1              700              712
## 6 1.50e9 2016-04-19 12H 0M 0S                1              304              320
```

5. Merge dataset

```
activity_sleep <- merge(activity, sleep, by = c("id", "date"))
glimpse(activity_sleep)
```

```
## Rows: 410
## Columns: 19
## $ id                         <dbl> 1503960366, 1503960366, 1503960366, 1503960~
## $ date                       <date> 2016-04-12, 2016-04-13, 2016-04-15, 2016-0~
## $ total_steps                <dbl> 13162, 10735, 9762, 12669, 9705, 15506, 105~
## $ total_distance             <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6~
## $ tracker_distance           <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6~
## $ logged_activities_distance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ very_active_distance       <dbl> 1.88, 1.57, 2.14, 2.71, 3.19, 3.53, 1.96, 1~
## $ moderately_active_distance <dbl> 0.55, 0.69, 1.26, 0.41, 0.78, 1.32, 0.48, 0~
## $ light_active_distance      <dbl> 6.06, 4.71, 2.83, 5.04, 2.51, 5.03, 4.24, 4~
## $ sedentary_active_distance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0~
## $ very_active_minutes        <dbl> 25, 21, 29, 36, 38, 50, 28, 19, 41, 39, 73,~
## $ fairly_active_minutes      <dbl> 13, 19, 34, 10, 20, 31, 12, 8, 21, 5, 14, 2~
## $ lightly_active_minutes     <dbl> 328, 217, 209, 221, 164, 264, 205, 211, 262~
## $ sedentary_minutes          <dbl> 728, 776, 726, 773, 539, 775, 818, 838, 732~
## $ calories                   <dbl> 1985, 1797, 1745, 1863, 1728, 2035, 1786, 1~
## $ time                       <Period> 12H 0M 0S, 12H 0M 0S, 12H 0M 0S, 12H 0M ~
## $ total_sleep_records        <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1~
## $ total_minutes_asleep       <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361~
## $ total_time_in_bed          <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384~
```

Dropping columns that mostly contains null values and are not relevant for our analysis.

```
activity_sleep <- activity_sleep%>%
  select(-c("logged_activities_distance", "tracker_distance", "sedentary_active_distance"))
```

Let's create a new dataset for further analysis

```
new_average <- activity_sleep%>%
  group_by(id)%>%
  summarise(mean_hours_sleep = mean(total_minutes_asleep)/60,
            mean_daily_steps = mean(total_steps),
            mean_daily_calories = mean(calories))
head(new_average)
```

```
## # A tibble: 6 x 4
##           id mean_hours_sleep mean_daily_steps mean_daily_calories
##        <dbl>            <dbl>            <dbl>               <dbl>
## 1 1503960366             6.00           12406.               1872.
## 2 1644430081             4.9             7968.               2978.
## 3 1844505072            10.9             3477                1676.
## 4 1927972279             6.95            1490                2316.
## 5 2026352035             8.44            5619.               1541.
## 6 2320127002             1.02            5079                1804
```

6. Visualizations

The fun part:)

6.1 Quality of sleep

Now, we'll try to find the quality of users sleep. This is important to understand how bellabeat products can be of use for helping users to improve their sleep.

```
sleep_quality <- new_average%>%
  mutate(sleep_quality = case_when(
    mean_hours_sleep <5 ~ "poor",
    mean_hours_sleep >=5 & mean_hours_sleep <7 ~ "moderate",
    mean_hours_sleep >=7 ~ "healthy"
  ))

sleep_quality_percent <- sleep_quality%>%
  group_by(sleep_quality)%>%
  summarise(total = n())%>%
  mutate(totals = sum(total))%>%
  group_by(sleep_quality)%>%
  summarise(total_percent = scales::percent(total/totals))

head(sleep_quality_percent)
```

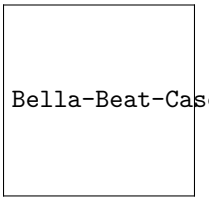
```
## # A tibble: 3 x 2
##   sleep_quality total_percent
##   <chr>         <chr>
```

```
## 1 healthy       46%
## 2 moderate      29%
## 3 poor          25%
```

```
# change 'chr' format to 'factor' format
sleep_quality_percent$sleep_quality<- factor(sleep_quality_percent$sleep_quality,
                                             levels = c("healthy", "moderate", "poor"))
```

Now, the visuals:)

```
sleep_quality_percent%>%
  ggplot(aes(x = "", y=total_percent, fill= sleep_quality))+
  geom_bar( stat = "identity", width = 1)+
  coord_polar("y", start = 0)+
  theme_minimal()+
  theme(axis.title = element_blank(),
        axis.text.x = element_blank(),
        plot.title = element_text(hjust =0.5, size=14, face= "bold"))+
  scale_fill_manual(values = c("#85e085", "#ffd480","#ff8080"),
                    labels = c("Healthy", "Moderate", "Poor"))+
  geom_text(aes(label = total_percent),
            position = position_stack(vjust=0.5))+
  labs(title = "Sleep Quality of users")
```



Great! majority(46%) of our users get a healthy amount of sleep as per the data we have.

Still 25% of users have poor quality of sleep. Bellabeat has a wide range of wonderful products that could definitely help users improve their sleep.

6.2 User Type

Now, lets categorize the users based on their activity level.

```
user_type<- new_average%>%
  mutate(user_type = case_when(
    mean_daily_steps <5000 ~"sedentary",
    mean_daily_steps >=5000 & mean_daily_steps <7500 ~"lightly active",
    mean_daily_steps >=7500 & mean_daily_steps <10000 ~"fairly active",
    mean_daily_steps >=10000 ~"very active"
  ))

user_type_percent <- user_type%>%
  group_by(user_type)%>%
  summarise(total = n())%>%
  mutate(totals = sum(total))%>%
  group_by(user_type)%>%
```

```
  summarise(total_percent = scales::percent(total/totals))
head(user_type_percent)
```

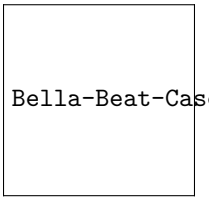
```
## # A tibble: 4 x 2
##   user_type      total_percent
##   <chr>          <chr>
## 1 fairly active  38%
## 2 lightly active 21%
## 3 sedentary      21%
## 4 very active    21%
```

```
#Change 'chr' format to 'factor' format.

user_type_percent$user_type <- factor(user_type_percent$user_type,
                                      levels = c("fairly active", "lightly active", "sedentary", "very a
```

Visuals:)

```
user_type_percent%>%
  ggplot(aes(x="",y=total_percent, fill = user_type))+
  geom_bar(stat = "identity", width = 1)+
  coord_polar("y", start= 0)+
  theme_minimal()+
  theme(axis.title = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks = element_blank(),
        plot.title = element_text(hjust = 0.5, size = 14, face = "bold"))+
  scale_fill_manual(values = c("#85e085","#0075ab","#ff6583","#ffa600"))+
  geom_text(aes(label = total_percent), position = position_stack(vjust = 0.5))+
  labs(title = "Types Of User")
```



Majority of the users combined are not really active, this result is not good. Bellabeat can definitely improve this.

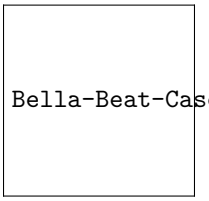
6.3 Steps per hour

Let's determine the average amount of steps that users walk on an hourly basis. This will help us understand when users are most active, which is really important for improving Bellabeat's marketing strategy.

```
steps <- steps%>%
  separate(col = "date_time", into = c('date','time'), sep =' ')


steps%>%
```

```
  group_by(time)%>%
  summarise(mean_steps = mean(step_total))%>%
  ggplot()+
  geom_col(aes(x = time, y= mean_steps, fill= mean_steps))+
  labs(title = "Steps VS Hour", x="", y="")+
  scale_fill_gradient(low = "yellow", high = "red")+
  theme(axis.text.x = element_text(angle =90))
```



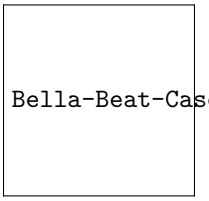
As we can see above, users are most active between:

a) 12 - 2pm &

b) 5- 7pm

6.4 Total Steps VS Calories

```
ggplot(data=activity_sleep, aes(x=total_steps, y=calories)) +
  geom_point(color= '#66b2b2') +
  geom_smooth(color = 'purple', method = 'loess', formula = 'y~x') +
  labs(title="Total Steps vs. Calories", x="Total Steps", y= "Calories")
```



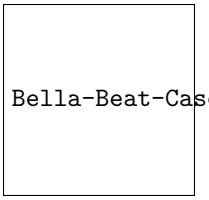
Great !! There is a positive correlation betweeen total steps and calories burned.

The more we excercise the healthier we get:)

6.5 Sleep VS Sedentary

Let's find the correlation between the users amount of sleep and sedentary minutes.

```
ggplot(activity_sleep, aes(x = total_minutes_asleep ,y= sedentary_minutes))+
  geom_jitter(color ='#66b2b2')+
  geom_smooth(color = "blue", method = 'loess', formula = 'y~x')+
  labs(title = "Sedentary Minutes VS Minutes Asleep", x = "Minutes Asleep", y= "Sedentary Minutes")
```



There is a negative correlation between the two.

Less amount of sedentary minutes should be encouraged for better sleep.

7. Recommendations

a) The data shows that a total 42% users [lightly active(21%) + sedentary(21%)] are not meeting the required quota of daily steps i.e.,8000 steps per day for a healthy lifestyle. Bellabeat could use this data to encourage users that are lagging behind to increase their activity level and improve their health.

b) The data clearly shows a positive relation between the total steps and the calories burned. If user intends to burn calories and lose weight, Bellabeat could help them achieve their goals with the use of Bellabeat products. The data collected through bellabeat products can be of great use.

c) The data also shows how higher sedentary minutes leads to lower sleep quality. Bellabeat can emphasize on how it's products can improve users sleep quality for it's marketing strategy.

d) Users are most active between 5-7 pm, Bellabeat could focus on educating and encouraging users to engage in physical activities like going for a walk or play outdoor games during this time period to improve their health.