Bellabeat Case Study

Emmanuel Bobmanuel

2024-07-09

Setting up my enviroment

Notes: setting up my R environment by loading 'tidyverse' package

```r
library(tidyverse)
library("ggpubr")
```

Executive Summary

Bellabeat is a technology-driven manufacturer specializing in health-focused products for women. This case study examines the diverse range of products offered by Bellabeat and explores insights derived from analyzing data extracted from product usage. By identifying trends in customer fitness data, the company aims to unlock new growth opportunities.

Introduction

Bellabeat, a high-tech company that manufactures health-focused smart products. This case study focuses on analyzing smart device data to gain insights into how consumers use Bella smart devices. The founders recognize growth opportunities and aim to extract valuable insights from the data. These insights will inform marketing strategies and guide the promotion of new products or ideas to enhance company growth. Specifically, we'll explore trends in smart device usage and consider how they apply to Bellabeat customers, ultimately shaping the marketing strategy.

Business Objective

Analyze smart device data for one of Bellabeat's products, to gain insights into how consumers use these devices

Stakeholders

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

ABOUT DATA

This data used IS from is for Bellabeat Time Device. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness. The data is organized in a Long format. It is data collected from Thirty eligible Fitbit users who consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring

DATA STORAGE AND CREDIBILITY

The data is publicly available. It is stored in Kaggle and the dataset is made available through Mobius. The data has a Kaggle calculated score of 100% for Completeness and credibility. However there are only thirty participants and 2 months of data hence we will be focusing on the big picture to guide in our decision making and recommendations.

LICENSING, PRIVACY, SECURITY AND ACCESSIBILITY

The document is not subject to copyright. The individual who linked this work to this deed has relinquished all rights to the work worldwide under copyright law, including related and neighboring rights, to the extent permitted by law. You are free to copy, modify, distribute, and use the work, even for commercial purposes, without seeking permission.

Data Analysis

From the datasets we have, Our analysis will encompass daily_Activities, hourly_Tracker, and sleep datasets. From this analysis, we will derive insights and formulate recommendations for management to inform our marketing approach. The data has been meticulously organized into rows and columns, covering various activities over a 2-month period. It includes information on average daily steps taken by customers, calories burned, sleep patterns, and other relevant metrics that will guide the marketing team's decision-making process.

```
summary(Merge_sleep_Activity <-
read_csv("C:/Users/USER/Desktop/merge/Merge_sleep_Activity.csv"))

##       Id              ActivityDate        TotalSleepRecords
TotalMinutesAsleep
##  Min.   :1.504e+09   Length:410          Min.   :1.00        Min.   : 58.0
##  1st Qu.:3.977e+09   Class :character    1st Qu.:1.00        1st Qu.:361.0
##  Median :4.703e+09   Mode  :character    Median :1.00        Median :432.5
##  Mean   :4.995e+09                       Mean   :1.12        Mean   :419.2
##  3rd Qu.:6.962e+09                       3rd Qu.:1.00        3rd Qu.:490.0
##  Max.   :8.792e+09                       Max.   :3.00        Max.   :796.0
##  TotalTimeInBed   TotalSteps     TotalDistance     TrackerDistance
##  Min.   : 61.0   Min.   :   17   Min.   : 0.010   Min.   : 0.010
##  1st Qu.:403.8   1st Qu.: 5189   1st Qu.: 3.592   1st Qu.: 3.592
##  Median :463.0   Median : 8913   Median : 6.270   Median : 6.270
##  Mean   :458.5   Mean   : 8515   Mean   : 6.012   Mean   : 6.007
##  3rd Qu.:526.0   3rd Qu.:11370   3rd Qu.: 8.005   3rd Qu.: 7.950
##  Max.   :961.0   Max.   :22770   Max.   :17.540   Max.   :17.540
##  LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
##  Min.   :0.0000           Min.   : 0.000     Min.   :0.0000
##  1st Qu.:0.0000           1st Qu.: 0.000     1st Qu.:0.0000
##  Median :0.0000           Median : 0.570     Median :0.4200
##  Mean   :0.1024           Mean   : 1.446     Mean   :0.7439
##  3rd Qu.:0.0000           3rd Qu.: 2.360     3rd Qu.:1.0375
##  Max.   :4.0000           Max.   :12.540     Max.   :6.4800
##  LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
```

```
##  Min.    :0.010        Min.    :0.0000000        Min.    :  0.00
##  1st Qu.:2.540        1st Qu.:0.0000000        1st Qu.:  0.00
##  Median :3.665        Median :0.0000000        Median :  9.00
##  Mean   :3.791        Mean   :0.0009268        Mean   : 25.05
##  3rd Qu.:4.918        3rd Qu.:0.0000000        3rd Qu.: 38.00
##  Max.   :9.480        Max.   :0.1100000        Max.   :210.00
##  FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes    Calories
##  Min.   :  0.00      Min.   :  2.0        Min.   :   0.0    Min.   : 257
##  1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.2    1st Qu.:1841
##  Median : 11.00      Median :208.0        Median : 717.0    Median :2207
##  Mean   : 17.92      Mean   :216.5        Mean   : 712.1    Mean   :2389
##  3rd Qu.: 26.75      3rd Qu.:263.0        3rd Qu.: 782.8    3rd Qu.:2920
##  Max.   :143.00      Max.   :518.0        Max.   :1265.0    Max.   :4900
```

```r
spec(hourlytracker <-
read_csv("C:/Users/USER/Desktop/merge/hourlytracker.csv"))
```

```
## cols(
##   Id = col_double(),
##   TotalIntensity = col_double(),
##   AverageIntensity = col_double(),
##   Calories = col_double(),
##   hourlyStepsTotal = col_double(),
##   time = col_time(format = ""),
##   Activitydate = col_character()
## )
```

DATA CLEANING

- The data was cleaned and merged in Excel

- The dataset arrived with 18 files with data about the participants behavior. It included activities divided into daily, hourly and by minute activities. For the analysis, we will focus on the daily and hourly activity.

- The "sleepDay_merged" and dailyActivity_merged were merge and created as a new file called "Merge_sleep_Activity". There were 3 duplicates which were removed.

- The dataset included hourly activity in 3 different files with data on Steps taken, intensities, and calories burned per hour. I combined them into 1 file named "hourlytracker". There were no duplicates.

- The activity date and time were in 1 column. I created a separate column for them called "Activitydate" and "time".

```r
#We will first take at how going for a walk and the amount of steps taken
daily affect the mount of calories burned vs. not being very active and
having a high amount of Sedentary Minutes daily

ggarrange(
ggplot(data=Merge_sleep_Activity,
```

```
aes(x=TotalSteps,y=Calories))+geom_point()+geom_smooth() +
  geom_jitter() +
  labs(title = "TotalSteps vs Calories"),
ggplot(data=Merge_sleep_Activity,
aes(x=SedentaryMinutes,y=Calories))+geom_point()+geom_smooth() +
  geom_jitter() +
  labs(title = "SedentaryMinutes vs Calories")
)
```
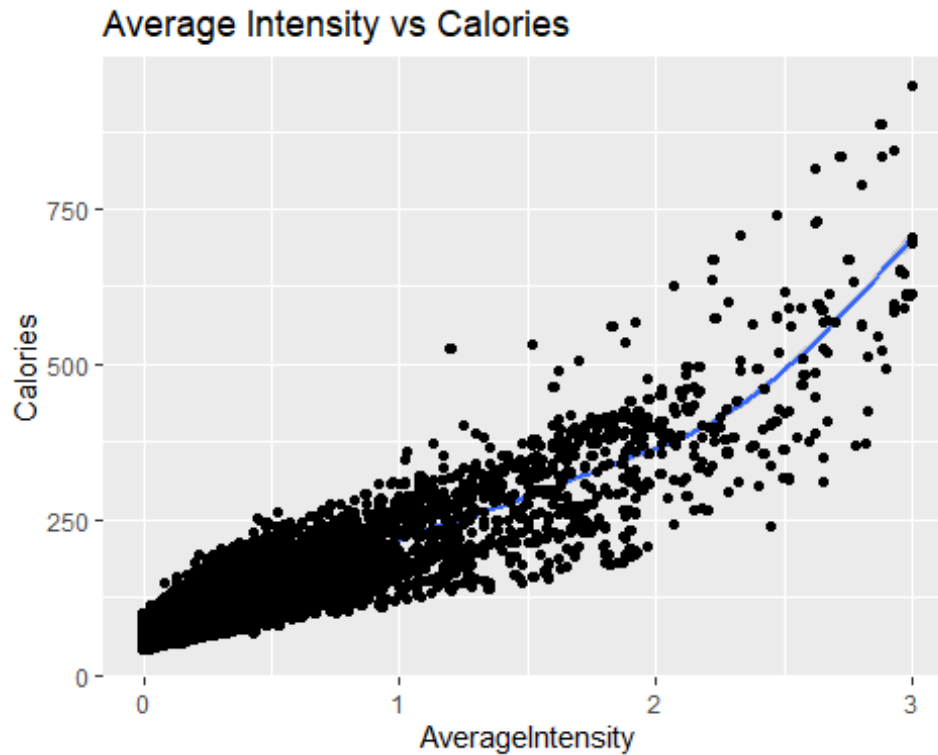

Analysis

- The data reveals a significant correlation between daily step count and calories burned. As the number of steps increases, so does the calorie expenditure. Regular walking contributes to improved physical fitness.
- Conversely, sedentary minutes are inversely related to calorie burn—more sedentary time leads to fewer calories expended. This underscores the importance of an active lifestyle for overall health benefits. Consistently taking more steps daily results in higher calorie burn, while prolonged inactivity negatively impacts lifestyle

```
#Here we will see how the intensity, presumably walking speed or running
affect calories burned

ggplot(data=hourlytracker, aes(x= AverageIntensity,
Calories))+geom_point()+geom_smooth() +
  geom_jitter() +
  labs(title = "Average Intensity vs Calories")
```

## Average Intensity vs Calories

Analysis

- Upon analyzing the intensity dataset, it becomes evident that as the workout or activity intensity increases, so does the number of calories burned. Consequently, to achieve greater calorie expenditure, we should elevate the intensity of our workouts. This information is crucial for determining the desired level of intensity during walks or runs

```
#Here will look and how having an inactive lifestyle or not exercising and
having a high amount of Sedentary Minutes could affect your sleep

ggplot(data=Merge_sleep_Activity, aes(x= SedentaryMinutes, y=
TotalMinutesAsleep))+geom_point()+geom_smooth() +
  geom_jitter() +
  labs(title = "Inactivity vs Amount of Sleep")
```

Inactivity vs Amount of Sleep
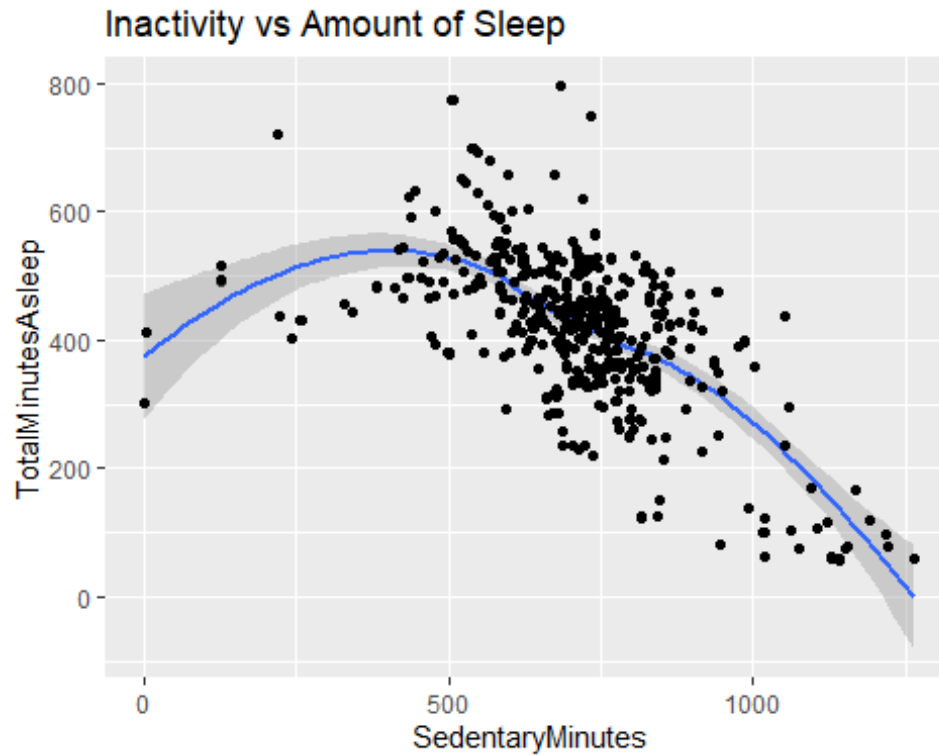
Analysis

- Based on our data analysis, it is clear that a sedentary lifestyle has adverse effects on sleep duration. Specifically, individuals with lower activity levels tend to experience poorer sleep quality. This insufficient sleep can also impact productivity and overall health. Therefore, it is essential to minimize sedentary minutes throughout the day to promote better sleep

```r
#Here we will look at the data to see what time participant are most active
during the day.
hr_tracker_new <- hourlytracker%>%
  group_by(time) %>%
  drop_na() %>%
  summarize(average_steps = mean(hourlyStepsTotal))

ggplot(data= hr_tracker_new, aes(x=time, y=average_steps)) +
geom_histogram(stat = "identity", fill='blue') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title="Average Steps/Hr. vs. Time")
```

Average Steps/Hr. vs. Time
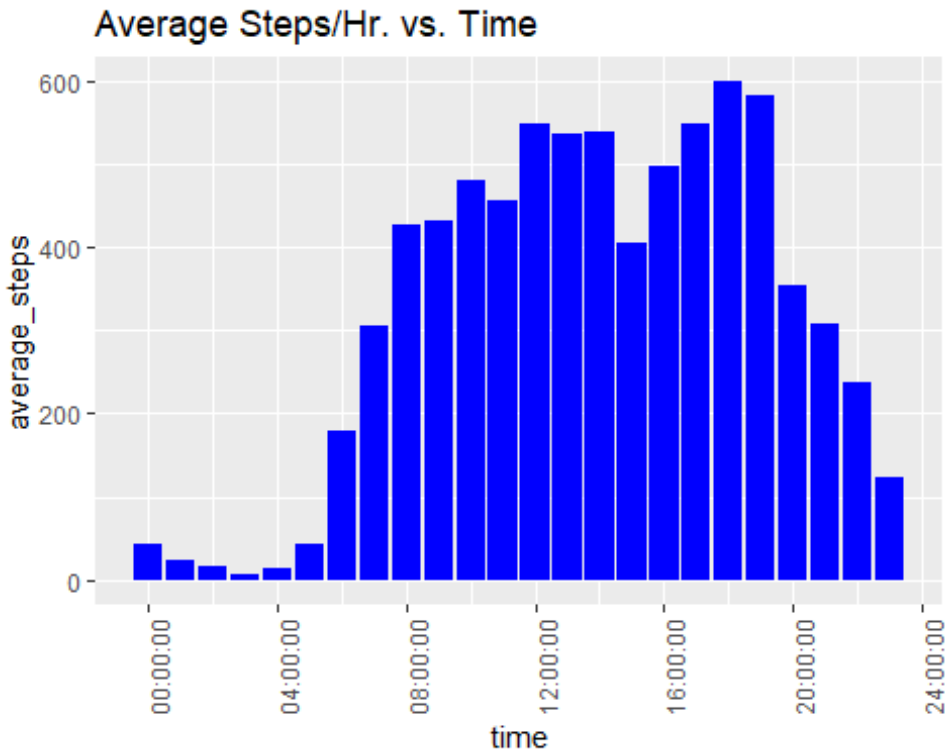
Analysis

- The data analysis reveals that the highest step activity occurs between 8 a.m. and 7 p.m., with peak walking times concentrated from 5 to 7 p.m. This pattern likely corresponds to individuals engaging in walks or exercise during their leisure hours

Conclusions and Recommendations.

After thorough data analysis, we have extracted valuable insights that can drive our company's growth. The data reveals opportunities to enhance our current client experience and outlines strategies for expanding our client base in the future. Allow me to share my recommendations for fostering growth by improving our products and services to benefit our clients

- Develop a comprehensive program centered around combating a sedentary lifestyle. This initiative will provide valuable information on the adverse effects of inactivity, including obesity, diabetes, and cardiovascular diseases. Additionally, consider integrating the following features into your smartwatch:

  1. Sedentary Reminder: Implement a feature that alerts users when they've been sitting for an extended period. The watch can vibrate or send notifications, encouraging them to move
  2. Sleep Tracking: Prioritize sleep quality by incorporating sleep-tracking capabilities. Smartwatches can analyze sleep cycles, aiding users in managing stress and overall health.

- Implement promotional campaigns or offer discounts to attract buyers. Additionally, consider strategic collaborations with third parties to enhance brand exposure. Highlight the watch's features and their positive impact on lifestyle and overall health.

- We should craft a compelling brand narrative that conveys the company's origin, rooted in our unwavering passion for human health. By doing so, we aim to create an association where individuals think of our company when considering a healthy lifestyle

Thank you.