## **Bellabeat Case-study**

## **INTRODUCTION**

In this case study, I will perform data analysis for Bellabeat, a high-tech manufacturer of health-focused products for women. I will analyze the smart device data to gain insight into how consumers are using their smart devices. My analysis will help guide future marketing strategies for their team. Along the way, I will perform numerous real-world tasks of a junior data analyst by following the steps of the data analysis process: ***Ask, Prepare, Process, Analyze, Share, and Act*.** By the end of the process, I will share my analysis conclusion and recommendation that will provide Bellabeat with insights that will help guide marketing strategy for the company.

## **SCENARIO**

Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market. Urška Sršen, co-founder and Chief Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunities for the company. The company asked to focus on one of Bellabeat’s products and analyze smart device data to gain insight into how consumers are using their smart devices. Data has been provided on their products such as the Bellabeat app, Leaf, Time, Spring and Bellabeat memberships.

***Case Study:*** **How Can a Wellness Technology Company Play It Smart?**

**ASK**  
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 Can a Wellness Technology Company Play It Smart?***

**PREPARE**  
 Sršen encourages you to use public data that explores smart device users’ daily habits. She points you to a specific data set:**FitBit Fitness Tracker Data** (CC0: Public Domain, dataset made available through Mobius): This Kaggle data set contains personal fitness tracker from thirty fitbit users. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. It includes information about daily activity, steps, and heart rate that can be used to explore users’ habits.

The data is located in-<https://www.kaggle.com/datasets/arashnic/fitbit>

**PROCESS**  
I used Excel sheets to clean the data ,R Programming and used Tableau to create visualizations.

# ***Install and load packages-***

install.packages("ggplot2")

library(ggplot2)

install.packages("tidyverse")

library(tidyverse)

install.packages("janitor")

library(janitor)

install.packages("skimr")

library(skimr)

install.packages("here")

library(here)

install.packages("dplyr")

library(dplyr)

# ***Organize and Re-name csv datasets-***

Calories <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/dailyCalories\_merged.csv")

Intensities <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/dailyIntensities\_merged.csv")

Steps <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/dailySteps\_merged.csv")

HeartRate <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/heartrate\_seconds\_merged.csv")

SleepPerDay <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/sleepDay\_merged.csv")

Weight <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/weightLogInfo\_merged.csv")

Activity <- read\_csv("/cloud/project/bellabeat\_case\_study\_3\_27\_2023/dailyActivity\_merged.csv")

# ***sGet a glimpse of the data set -***

glimpse(Calories)

glimpse(Intensities)

glimpse(Steps)

glimpse(HeartRate)

glimpse(SleepPerDay)

glimpse(Weight)

glimpse(Activity)

# ***Cleaning -Check for duplicates in data set-***

# sum(duplicated(Activity))

sum(duplicated(Weight))

sum(duplicated(SleepPerDay))

sum(duplicated(Intensities))

sum(duplicated(Steps))

sum(duplicated(Calories))

sum(duplicated(HeartRate))

# ***SleepPerDay datasets conatins duplicates. Remove the duplicates-***

SleepPerDay <- SleepPerDay%>%

distinct()

sum(duplicated(SleepPerDay))

**ANALYZE**

Activity%>%

select(TotalSteps,

TotalDistance,

SedentaryMinutes,

Calories)%>%

summary()

Intensities%>%

select(LightlyActiveMinutes,

FairlyActiveMinutes,

VeryActiveMinutes)%>%

summary()

Weight%>%

select(WeightPounds,

IsManualReport,

BMI)%>%

summary()

SleepPerDay%>%

select(TotalMinutesAsleep,

TotalTimeInBed)%>%

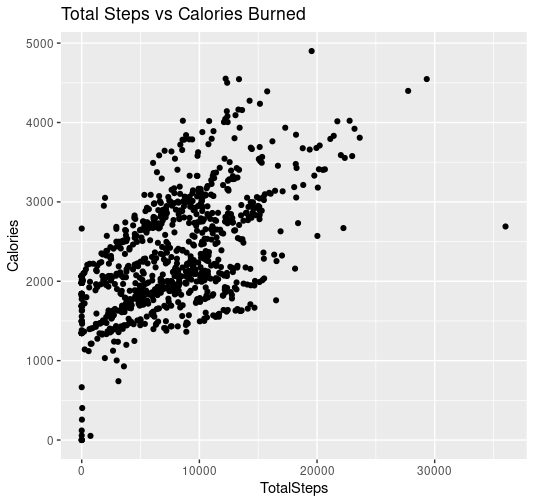
summary()

**SHARE**

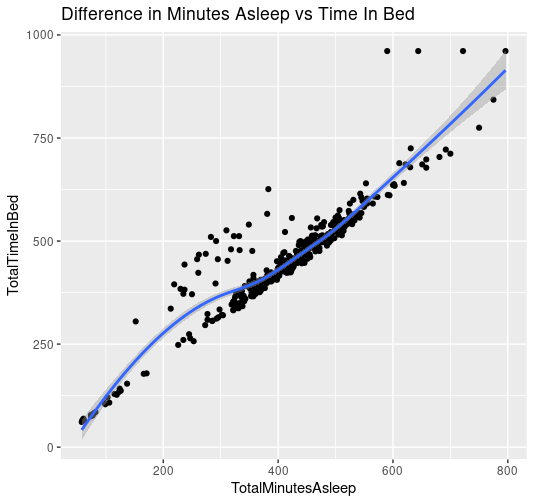
# ***Visualizations-***

1. ggplot(data=SleepPerDay)+geom\_point(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+ geom\_smooth(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+ labs(title="Difference in Minutes Asleep vs Time In Bed")
2. ggplot(data=SleepPerDay)+ geom\_point(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+ geom\_smooth(mapping = aes(x=TotalMinutesAsleep, y=TotalTimeInBed))+labs(title="Difference in Minutes Asleep vs Time In Bed")
3. ggplot(data=Activity)+geom\_point(mapping = aes(x=TotalSteps, y=LightlyActiveMinutes)) + geom\_smooth(mapping = aes(x=TotalSteps, y=LightlyActiveMinutes)) + labs(title="Total Steps vs Lightly Active Minutes")
4. ggplot(data=Activity)+ geom\_point(mapping = aes(x=TotalSteps, y=FairlyActiveMinutes)) + geom\_smooth(mapping = aes(x=TotalSteps, y=FairlyActiveMinutes)) + labs(title="Total Steps vs Fairly Active Minutes")
5. ggplot(data=Activity)+geom\_point(mapping = aes(x=TotalSteps, y=VeryActiveMinutes)) +geom\_smooth(mapping = aes(x=TotalSteps, y=VeryActiveMinutes)) +labs(title="Total Steps vs Very Active Minutes")

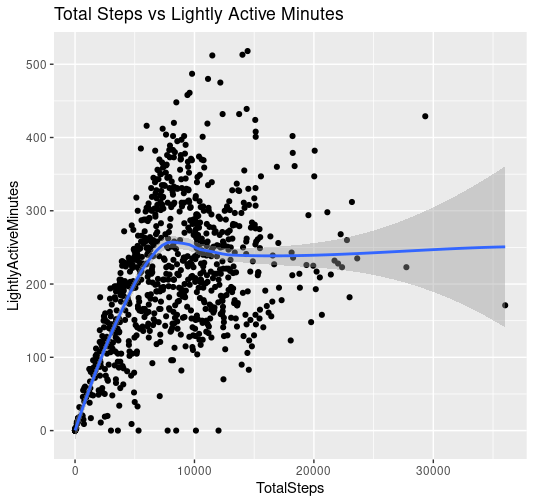
**OUTPUTS 1.**



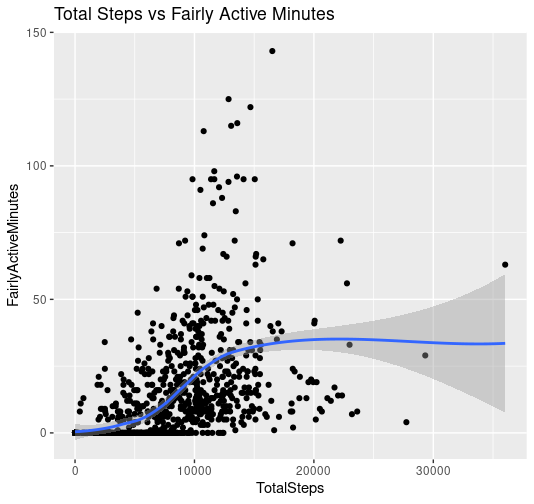
**2.**

****

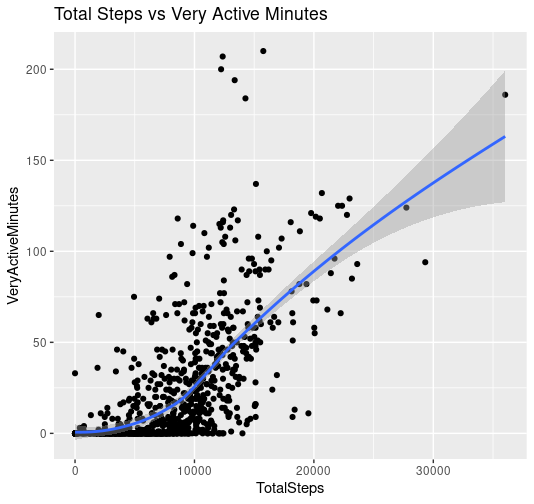
**3.**

****

**4.**

****

**5.**

****

### **Conclusions**

From a dataset of 30 users of different wearable technology products, the median user:

1. spends the majority of tracked active minutes performing light activity, presumably walking.
2. participates in 20 minutes of combined time in fairly active and very active zones per tracked time.
3. negative correlation between sleep and steps
4. no correlation between time spent in vigorous activity and sedentary time
5. slight negative correlation between time spent in light activity and sedentary time

This dataset is applicable to Bellabeat products, as the dataset represents a mix of users and products. This provides a profile of the general industry rather than a specific subset of industry.

### **Recommendations**

Marketing strategy should be targeted to the customer who is already active and spends most of their active time performing light activity.

\*In addition, targeting user groups other than this with different strategies may expand the customer base.

### **Next Steps**[**¶**](https://www.kaggle.com/code/bryankoewler/bellabeat-with-r#Next-Steps)

* Develop primary marketing strategy towards the population segment that matches the median user.
* Develop secondary strategies towards those who are on not median user: those who are sedentary and those who track high amounts of activity.
* Obtain more user data for wearable technology, and break down the data by product and price to obtain more specific detail on the type of user of each product.