Antasha D. Lamb

Bellabeat: How can a wellness company play it smart?

12/04/2024

Click here for Presentation

Introduction

Welcome to my Capstone/Case Study project. My name is Antasha D. Lamb, I have come to the conclusion of my data analytics certificate program through Merit America. As a junior data analyst I have been tasked with performing a real-life task of analyzing datasets in order to identify a trend in smart device usage. For a company named Bellabeat; a high-tech manufacturer of health-focused products for women. https://bellabeat.com/

I will conduct an analysis on the FitBit Fitness Tracker Data (CC0: Public Domain, dataset made available through Mobius): https://www.kaggle.com/datasets/arashnic/fitbit
This Kaggle data set contains personal fitness trackers 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. Sršen, the co-founder of Bellabeat states that this data set might have some limitations, and encourages me to consider adding another dataset to help address those limitations.

With this information I retrieved an additional dataset named Apple Watch and Fitbit data, this dataset was collected from 46 participants using an apple watch and fitbit for 65 minutes.

I intend to analyze both datasets to determine a trend in the use of these devices.

Ask

Sršen asks you to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat sma devices. She then wants you to select one Bellabeat product to apply these insights to in your presentation. These questions will guide your analysis:

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?

Prepare

Key Stakeholders: ***Urška Sršen***: Bellabeat's co founder and Chief Creative Officer

Sando Mur: Mathematician and Bellabeat cofounder; key member of the Bellabeat executive team

Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and repoing data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat''s mission and business goals — as well as how you, as a junior data analyst, can help Bellabeat achieve them.

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) : files: dailyActivity_Merged, because this table contains the same information as the other tables, however the data is combined. Then I chose the sleep_day table, this information isn't in any of the other tables and could hold great insight. Finally,clean_weightLogInfo_merged table because I'd like to analyze the weight log to see if the steps are paying off for the participants of the study. As well as take a look at the sleep table to determine correlation.

Sršen tells you that this data set might have some limitations, and encourages you to consider adding another data to help address those limitations as you begin to work more with this data.

With this information I retrieved an additional dataset named Apple Watch and Fitbit data, this dataset was collected from 46 participants using an apple watch and fitbit for 65 minutes. File: aw_fb_data, I chose this table because it contains all of the information merged into one.
https://www.kaggle.com/datasets/aleespinosa/apple-watch-and-fitbit-data

Process

The selected tool for this analysis is R, as I am working with two substantial datasets. This software will facilitate an efficient and error-free outcome, reducing processing time. Initially, I employed Excel to examine the tables for misspellings, duplicates, and empty fields in the files: dailyActivity_Merged and clean_weightLogInfo_merged. The files were subsequently cleaned within R Studio, resulting in no errors such as missing data, blank spaces, or duplicates. Additionally, the files: data_for_weka_aw, aw_fb_data, and data_for_weka_fb were cleaned using Google Sheets, where minor issues related to inconsistent formatting and spelling were identified. No duplicates were detected, and unnecessary spaces were removed.

Data Cleaning of Daily_Activity

```{r}
library(here)
library(skimr)
library(janitor)
library(dplyr)
skim_without_charts(daily_activity)
colnames(daily_activity)
view(daily_activity)
clean_names(daily_activity)
head(daily_activity)
sample(daily_activity)
str(daily_activity)
summary(daily_activity)
```

Cleaning of Sleep_Score

```{r}
library(tidyverse)
install.packages("here")
library("here")
install.packages("skimr")
library("skimr")
install.packages("janitor")
library("janitor")
install.packages("dplyr")
library("dplyr")
View(sleep_score)
clean_names(sleep_score)
skim_without_charts(sleep_score)
```

Data Cleaning of Weight Log

{r}
```
library(tidyverse)
install.packages("here")
library("here")
install.packages("skimr")
library("skimr")
install.packages("janitor")
library("janitor")
install.packages("dplyr")
library("dplyr")
attach(weight)
clean_names(weight)
head(weight)
```

```
skim_without_charts(weight)
glimpse(weight)
```

Cleaning of Combined Data

```
{r}
library(tidyverse)
library(dplyr)
head(combined_data)
n_distinct(combined_data$Id)
nrow(combined_data)
summary(combined_data)
```

Cleaning of Apple Watch Fitbit Data

```
{r}
library(tidyverse)
library("here")
library("skimr")
library("janitor")
clean_names(aw_fb_data)
head(aw_fb_data)
colnames(aw_fb_data)
```

Credibility of the data:

CRAAP Test: This method provides you with a set of criteria that make a source more or less credible. The criteria are:

Currency: 2020-Apple Watch and Fitbit Data
Relevance: Research
Authority: Harvard Dataverse
Accuracy: CCO Public Domain
Purpose: Medicine: Health and Life Sciences; Computer and Information Science.

Currency: 2016-Fitbit Fitness Tracker Data
Relevance: "Human temporal routine behavioral analysis and pattern recognition.This dataset generated by respondents to a 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 minute-level output for physical activity, heart rate, and sleep monitoring."
Authority: Mobius
Accuracy: CCO Public Domain
Purpose: Fitbit trackers and individual tracking behaviors / preferences.

Analyze

Data was analyzed using RStudio. Dataframes were created to save, store, and retrieve cleaned files.

Analysis of Daily_Activity

```
{r}
library(tidyverse)
library(dplyr)
library(ggplot2)
attach(daily_activity)
summary(daily_activity)

 LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
LightActiveDistance
 Min.   :0.0000           Min.   : 0.000     Min.   :0.0000           Min.   :
0.000
 1st Qu.:0.0000           1st Qu.: 0.000     1st Qu.:0.0000           1st Qu.:
1.945
 Median :0.0000           Median : 0.210     Median :0.2400           Median :
3.365
 Mean   :0.1082           Mean   : 1.503     Mean   :0.5675           Mean   :
3.341
 3rd Qu.:0.0000           3rd Qu.: 2.053     3rd Qu.:0.8000           3rd Qu.:
4.782
 Max.   :4.9421           Max.   :21.920     Max.   :6.4800           Max.
:10.710
 SedentaryActiveDistance VeryActiveMinutes FairlyActiveMinutes
LightlyActiveMinutes
 Min.   :0.000000        Min.   :  0.00    Min.   :  0.00      Min.   :  0.0
 1st Qu.:0.000000        1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:127.0
 Median :0.000000        Median :  4.00    Median :  6.00      Median :199.0
 Mean   :0.001606        Mean   : 21.16    Mean   : 13.56      Mean   :192.8
 3rd Qu.:0.000000        3rd Qu.: 32.00    3rd Qu.: 19.00      3rd Qu.:264.0
 Max.   :0.110000        Max.   :210.00    Max.   :143.00      Max.   :518.0
 SedentaryMinutes    Calories
 Min.   :   0.0   Min.   :   0
 1st Qu.: 729.8   1st Qu.:1828
 Median :1057.5   Median :2134
 Mean   : 991.2   Mean   :2304
```

```
 3rd Qu.:1229.5   3rd Qu.:2793
 Max.   :1440.0   Max.   :4900
```

The Daily Activity Table has 15 columns, and I decided to look at how active minutes stack up against different distances and the actual calories burned. The dataset includes 33 participants. A cumulative total of 4,900 calories burned has been reported. The total distance covered is 39.21 units.
The overall number of steps taken amounts to 36,019.
The maximum distances for various activity levels are as follows: Very Active has a maximum distance of 21.92, Moderately Active has a maximum distance of 6.48, Light Active has a maximum distance of 10.71, and Sedentary has a maximum distance of 0.11. This data is collected over a timeframe of 31 days, specifically from 4th December 2016 to 5th December 2016.

Analysis of Sleep Log

```
{r}
sleep_score %>%
  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
```

The sleep score dataset comprises five columns and 413 rows of data.
The maximum recorded time spent in bed was 961 minutes, while the maximum total sleep duration was 796 minutes. The study involved 24 participants and covered the period from 4th December 2016 to 9th May 2016. The average time spent in bed was 458.6 minutes, and the average time asleep was 419.5 minutes.

Analysis of Combined Data

```
{r}
combined_data %>%
  select(TotalSteps,
         TotalTimeInBed,
```

```
          TotalDistance,
          Calories)%>%
  summary()
```

```
TotalSteps    TotalTimeInBed  TotalDistance       Calories
 Min.   :    0   Min.   : 61.0   Min.   : 0.000   Min.   :   0
 1st Qu.: 4660   1st Qu.:402.0   1st Qu.: 3.180   1st Qu.:1783
 Median : 8596   Median :463.0   Median : 6.120   Median :2162
 Mean   : 8117   Mean   :458.4   Mean   : 5.735   Mean   :2329
 3rd Qu.:11317   3rd Qu.:526.0   3rd Qu.: 7.920   3rd Qu.:2865
 Max.   :22988   Max.   :961.0   Max.   :17.950   Max.   :4900
```

This table was created by merging the daily activity and sleep score tables using RStudio. It includes data from 24 participants and comprises 940 rows of information.

Analysis of Weight Log Data

```
```{r}
weight %>%
  select(Fat,
         BMI,
         WeightPounds) %>%
  summary()
```
```

```
Fat              BMI          WeightPounds
 Min.   :22.00   Min.   :21.45   Min.   :116.0
 1st Qu.:22.75   1st Qu.:23.96   1st Qu.:135.4
 Median :23.50   Median :24.39   Median :137.8
 Mean   :23.50   Mean   :25.19   Mean   :158.8
 3rd Qu.:24.25   3rd Qu.:25.56   3rd Qu.:187.5
 Max.   :25.00   Max.   :47.54   Max.   :294.3
 NA's   :65
```

There are 8 participants included in this table. The time frame spans from 12th April 2016 to 9th May 2016. The minimum weight recorded is 116 pounds, while the maximum weight is 294.3 pounds, with a standard deviation of 30.69 pounds.

The minimum Body Mass Index (BMI) is 21.45, and the maximum BMI is 47.54, with a standard deviation of 3.06.

Analysis of Apple watch Fitbit Data

```
{r}
aw_fb_data %>%
  select(height, weight)
```

A tibble: 6,264 × 2

height <dbl>	weight <dbl>
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4
168	65.4

Previous
1-10 of 6,264 rows

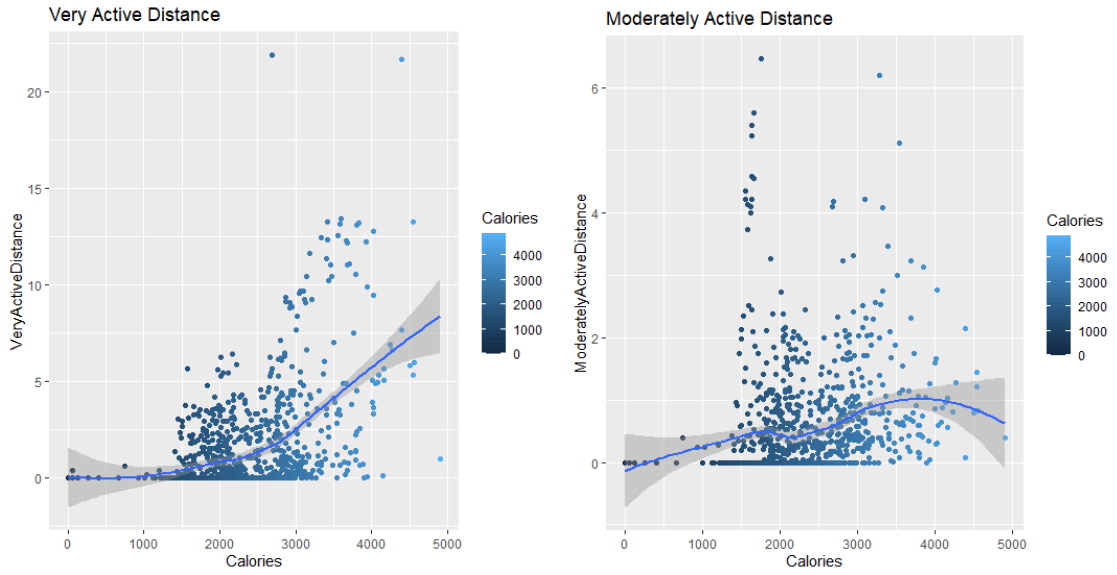
The table comprises 20 columns and 6,264 rows of data. The age range of participants is from 18 to 56 years. Participants include both females and males. The maximum recorded distance is 335 units, with a maximum calorie expenditure of 97.5. The highest number of steps taken is 1,714, and the maximum weight is 115 kg, while the maximum height is 191 cm. The product of maximum steps and distance is 51,520.00.

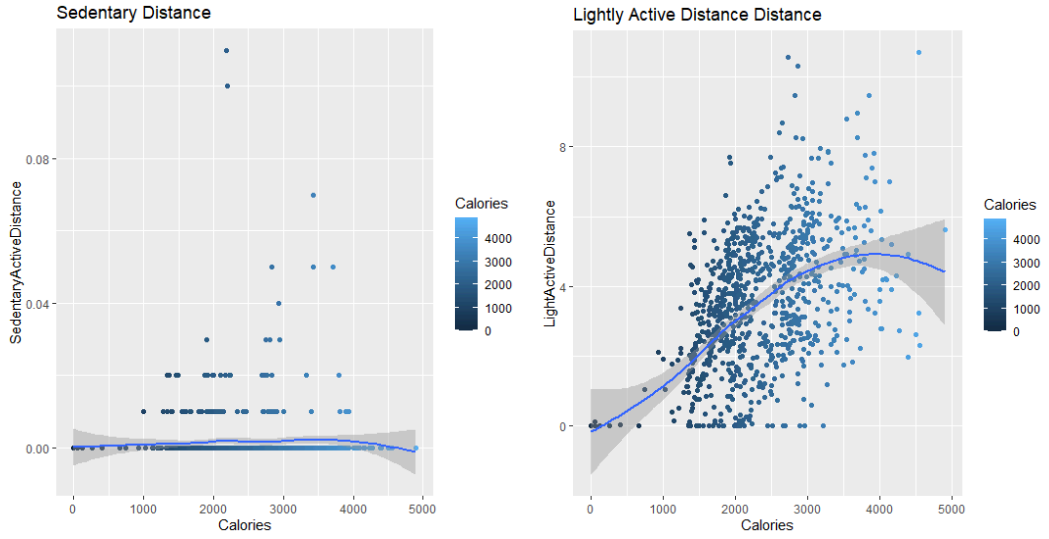
In terms of averages, the average distance is 13.83 units, with an average of 109.56 steps taken. The average weight is 69.61 kg, and the average height is 169.7 cm. The average product of steps and distance is 590.04, with an average calorie expenditure of 19.47.

The standard deviation for steps is 222.79, for distance is 45.94, for calories is 27.3, and for the product of steps and distance is 4,063.83.

Share

Representation of distances and calories associated with varying levels of activity, from highly active to sedentary.



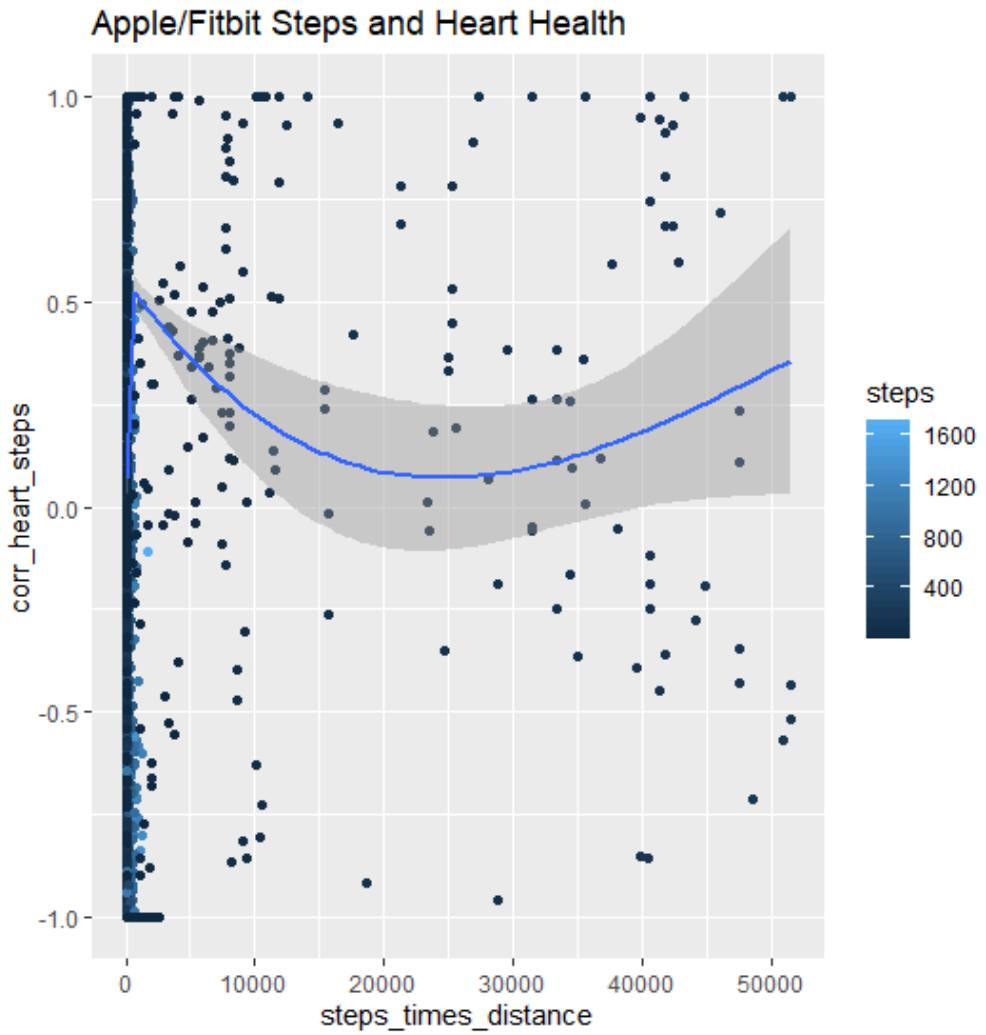


By means of visualization, I have analyzed the distances associated with Very, Moderately, Lightly, and Sedentary active categories to ascertain whether a correlation exists between the number of steps taken and the calories expended across these four groups.

I have concluded that the category of very active distance exhibits a significant correlation between distance covered and calories expended. Furthermore, lightly active distance demonstrates a stronger

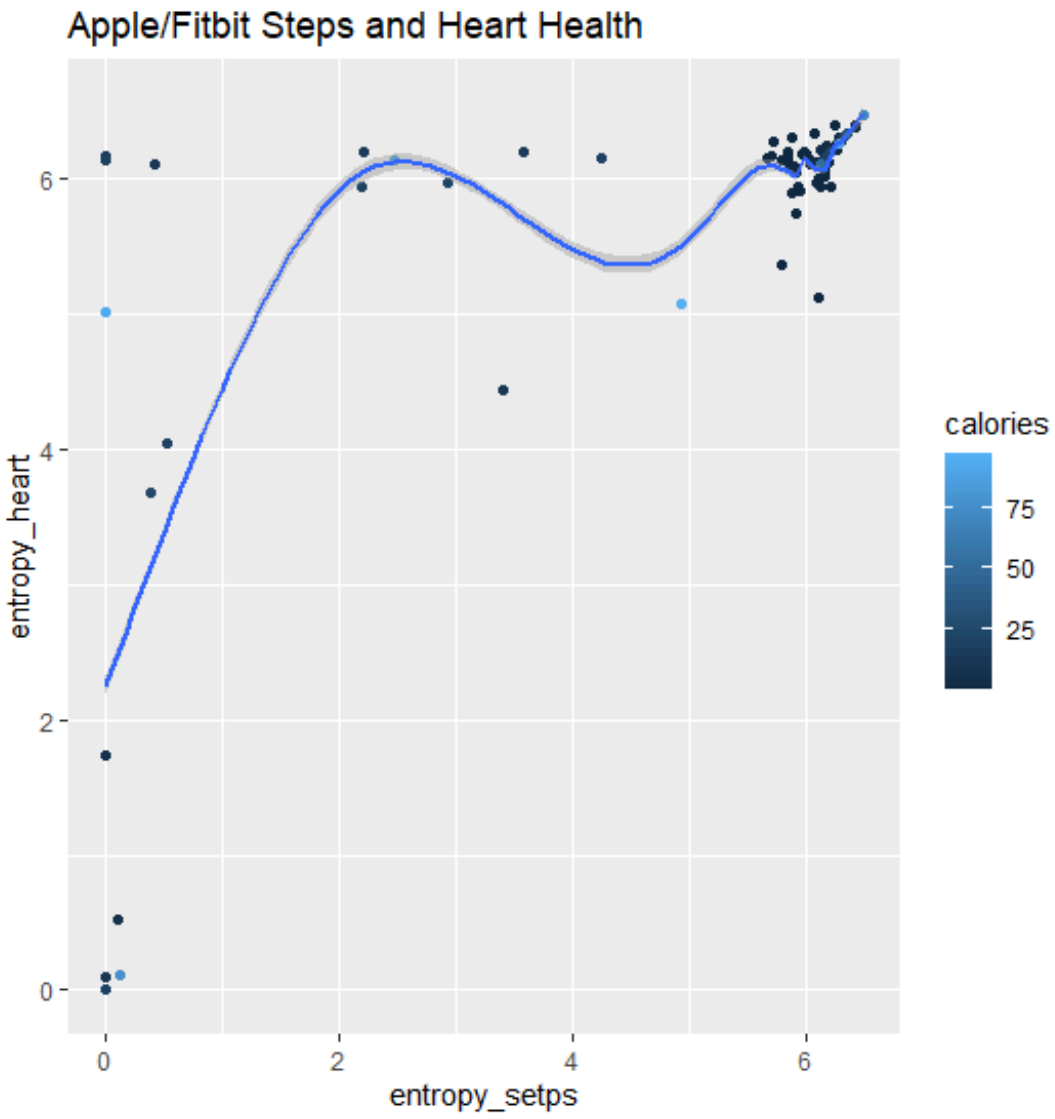
correlation compared to moderately active distance in relation to calories burned. In contrast, sedentary behavior shows no correlation whatsoever.

Visualization of steps times distances and correlated heart steps



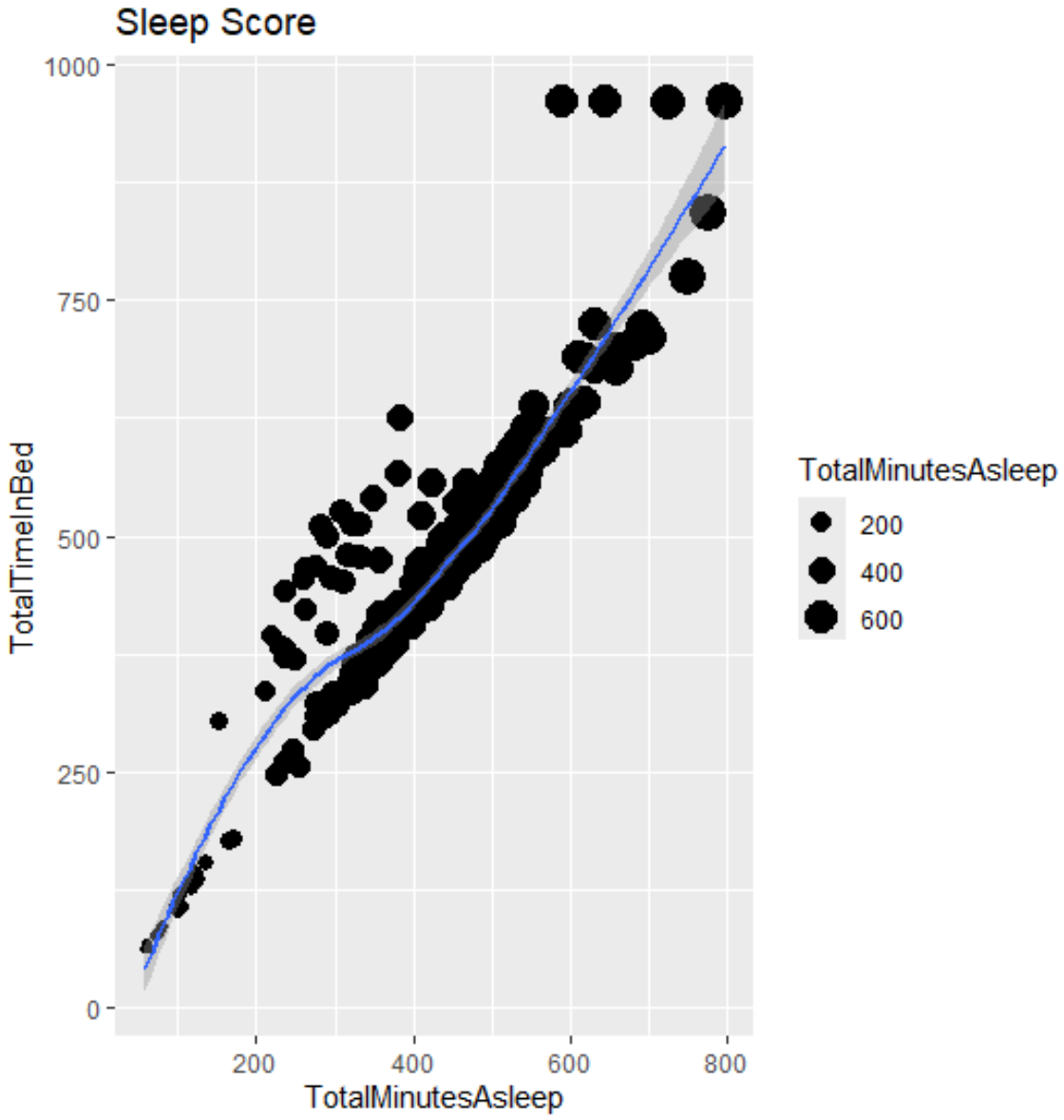
This chart illustrates a polynomial trendline that depicts the variations presented in this table. It also indicates that even minimal physical activity results in an increase in heart rate, which is beneficial for cardiovascular health. Furthermore, it highlights that greater distances correspond to a higher number of calories expended.

Visualization of Entropy heart rate and steps



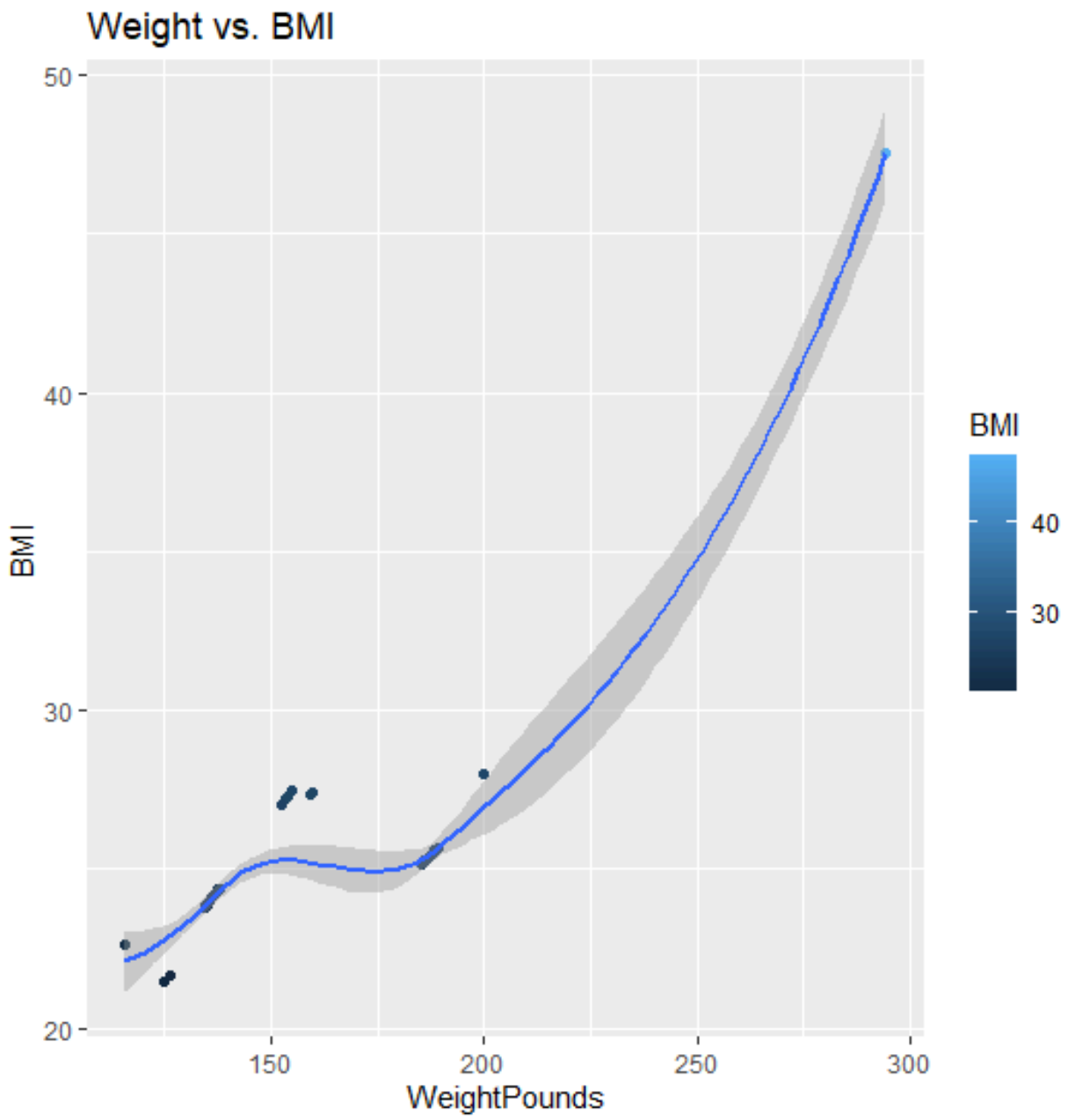
The presented chart illustrates a cubic trend line, characterized by two curves, one corresponding to 6 entropy heart- rate and 3 entropy steps. Subsequently, it ascends once more at 6 entropy heart-rate and 6 entropy steps. This pattern indicates a positive correlation between vigorous physical activity and an enhanced heart-rate.

Visualization of Sleep and efficiency



This chart illustrates a positive linear correlation within the presented table. The total duration of sleep and the total time spent in bed exhibit a positive relationship, indicating that a significant portion of this sample demonstrates greater efficiency when both sleep and rest are at optimal levels.

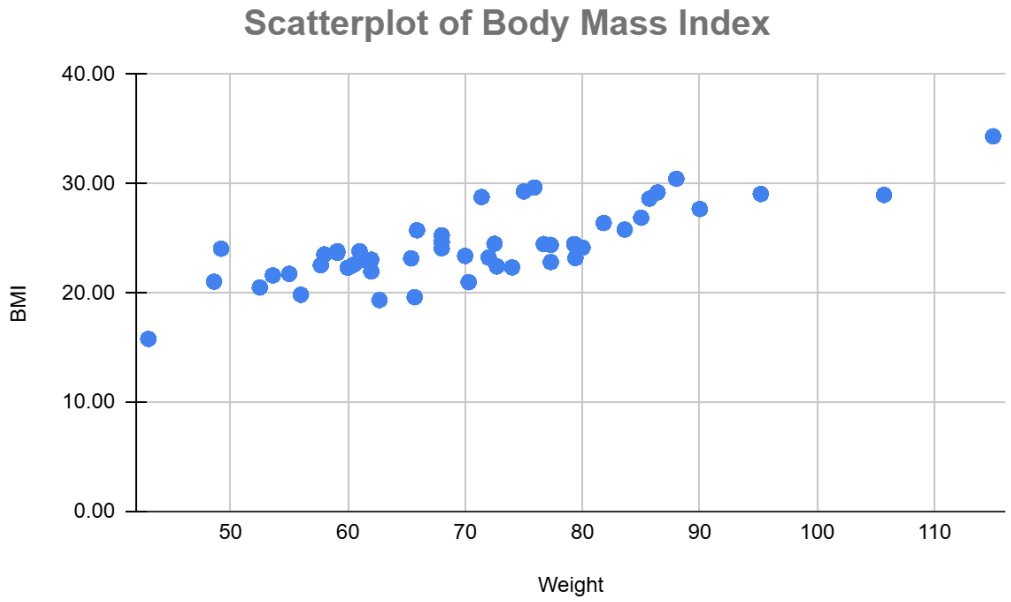
Visualization of Weigh-In

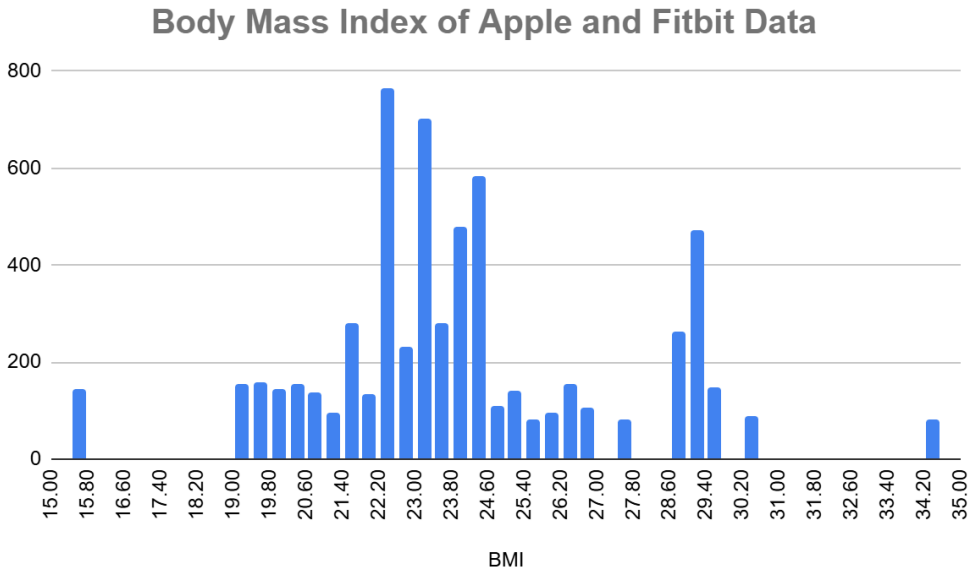


The chart presented was derived from the Fitbit weigh-in data. Body Mass Index (BMI) was employed to identify trends within this sample in relation to body weight measured in pounds, thereby assessing the fitness level of the group. There were several outliers, indicating that some individuals were classified as underweight while others were categorized as overweight. Nevertheless, the majority of participants adhered to the recommended BMI range, which is associated with numerous health advantages.

Visualization of Body Mass Index

The charts presented were generated using Excel by adding an extra column labelled BMI, which was absent from the original dataset. I computed the Body Mass Index (BMI) for each participant in the dataset using the formula weight divided by height squared (weight/height²).





The weights in these tables are expressed in kilograms (KG). I opted to examine the same variables as those in the preceding table, which illustrates the correlation between weight and body mass index (BMI). The scatterplot indicates a positive correlation, while the histogram reveals that most participants are sustaining a healthy body weight as a result of varying levels of physical activity.

Act

Final Conclusion

As stated by the Centers for Disease Control and Prevention in 2020, “Regular physical activity can improve overall health and prevent negative health outcomes in people of all ages.” Furthermore, engaging in exercise enhances cardiovascular health, boosts lung function, aids in weight management, reduces blood pressure, and improves cholesterol levels. In conclusion, my analysis indicates that individuals who are lightly to very active can effectively burn calories and sustain a healthy body mass index.

What does this mean for Bellabeat? As a recommendation, the target market should be the very active group. This group has shown consistent activity, with great results according to the data. Bellabeat could implement a goal system for each individual, to move and cover distance in order to shed calories and maintain a healthy weight.

What is body mass index (BMI)?

According to the American Cancer Society, 2024.

Body mass index, or BMI, is one way to look at whether a person is at a healthy weight. BMI is a number based on your weight and height. In general, the higher the number, the more body fat a person has. BMI is often used as a screening tool to decide if your weight might be putting you at risk for health problems such as heart disease, diabetes, and cancer.

BMI is used to broadly define different weight groups in adults 20 years old or older.

- Underweight: BMI is less than 18.5
- Normal weight: BMI is 18.5 to 24.9
- Overweight: BMI is 25 to 29.9
- Obese: BMI is 30 or more

Bellabeat can integrate this type of monitoring by allowing users to enter their weight and height into the device or application; allowing continuous monitoring, while encouraging a healthy habits, such as diet, and exercise routines

I also examined the relationship between sleep and efficiency. It was found that the majority of participants exhibited healthy sleep patterns. This aspect is monitored by Bellabeat products and can be leveraged for marketing purposes. While maintaining good sleep is optional for a healthy lifestyle, consumers can be incentivised or reminded to prioritise it. Furthermore, adequate sleep is crucial for both physical health and emotional well-being; thus, achieving sufficient sleep duration and quality is vital for restorative sleep. The required amount of sleep varies with age.

It may be beneficial to reimagine the Ivy and develop an Ivy 2.0 that incorporates improved functionalities for tracking steps, exercise, body mass index (BMI), sleep patterns, and heart rate notifications. It is essential to retain all existing features while fostering a sense of healthy competition among users. Additionally, the introduction of a display that periodically conveys total steps and distance, along with heart rate, menstrual cycle tracking, mood, BMI, and healthy sleep practices, could be advantageous. Incentivizing healthy behaviors and establishing alerts when activity levels drop below the normal range would also be valuable enhancements.

In conclusion, the present price stands at $380.00 when compared to the current market landscape. Fitbit products are priced between $95 and $180, while Apple Watches range from $99 to $799. Notably, Apple possesses a larger market share compared to many of its rivals. It is recommended that Bellabeat develop the Ivy 2.0 featuring a robust and comfortable band, priced at 25% less than its existing price. Marketing efforts should commence three months prior to Christmas, intensifying throughout December.

References

https://www.kaggle.com/datasets/arashnic/fitbit

https://www.kaggle.com/datasets/aleespinosa/apple-watch-and-fitbit-data

https://www.cdc.gov/nchs/products/databriefs/db443.htm

https://www.cancer.org/cancer/risk-prevention/diet-physical-activity/body-weight-and-cancer-risk/body-mass-index-bmi-calculator.html