

BellaFit Análise

Patrick Cuervo Rinzo

2022-07-06

Sobre BellaBeat

Bellabeat é uma empresa que possui um ecossistema de produtos focados na saúde da mulher. Por meio da captura de dados com dispositivos móveis e monitores biométricos tem a proposta de ajudar as mulheres a entender melhor o próprio corpo e fazer escolhas melhores. Urška Sršen co-fundadora e chefe criativa da marca, acredita no poder da análise de dados para oportunidades de crescimento da empresa. Site: <https://us.bellabeat.com/about-us/>

Objetivos

1. Quais são algumas tendências no uso de dispositivos inteligentes?
2. Como essas tendências podem se aplicar aos clientes Bellabeat?
3. Como essas tendências podem ajudar a influenciar a estratégia de marketing da Bellabeat?

Sršen, incentiva fortemente o uso de datasets disponíveis de maneira pública pela empresa, respeitando certos quesitos de confiabilidade.

Para esse estudo o conjunto de dados foi obtido no Kaggle: <https://www.kaggle.com/datasets/arashnic/fitbit>
Análises baseadas nos estudos de caso de Anastasiia Chebotina e Julen Aranguren.

Bibliotecas Necessárias

Após o download dos arquivos csv e da instalação das principais bibliotecas para análise é preciso inicia-las:

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.3      v purrr   0.3.4
## v tibble  3.1.1      v dplyr   1.0.8
## v tidyr   1.2.0      v stringr 1.4.0
## v readr   2.1.2      v forcats 0.5.1
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(lubridate)
```

```
##  
## Attaching package: 'lubridate'  
  
## The following objects are masked from 'package:base':  
##  
##     date, intersect, setdiff, union
```

```
library(dplyr)  
library(ggplot2)  
library(tidyr)
```

Atribuindo datasets a variáveis

Para esse estudo, destaquei os seguinte conjunto de dados:

```
dActivity <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/  
dCalories <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/  
dIntensities <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/  
hCalories <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/  
hIntensities <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/  
hSteps <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/h  
sleep<- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/sleep  
weight <- read.csv("C:/Users/patri/OneDrive/Área de Trabalho/GoogleAnalytcsCourse/CaseStudy2Bellafit/we
```

Visualizando

```
View(dActivity)  
View(dCalories)  
View(dIntensities)  
View(hCalories)  
View(hIntensities)  
View(hSteps)  
View(sleep)  
View(weight)
```

Fase Exploratória dos Dados:

A seguir é possível verificar alguns resumos estatísticos dos dados:

```
n_distinct(dActivity$Id)
```

```
## [1] 33
```

```
n_distinct(dCalories$Id)
```

```
## [1] 33
```

```
n_distinct(dIntensities$Id)
```

```
## [1] 33
```

```
n_distinct(hCalories$Id)
```

```
## [1] 33
```

```
n_distinct(hIntensities$Id)
```

```
## [1] 33
```

```
n_distinct(hSteps$Id)
```

```
## [1] 33
```

```
n_distinct(sleep$Id)
```

```
## [1] 24
```

```
n_distinct(weight$Id)
```

```
## [1] 8
```

Por meio da função `n_distinct` pode-se observar que a quantidade de registros de peso por Id são insuficientes, comparadas aos demais, sendo assim insuficiente para relacionar de forma adequada com os outros dados.

Atividades

```
dActivity %>%  
  select(TotalSteps,  
         TotalDistance,  
         SedentaryMinutes, Calories) %>%  
  summary()
```

```
##      TotalSteps      TotalDistance      SedentaryMinutes      Calories  
##  Min.   :    0      Min.   : 0.000      Min.   :    0.0      Min.   :    0  
## 1st Qu.: 3790      1st Qu.: 2.620      1st Qu.: 729.8      1st Qu.:1828  
## Median : 7406      Median : 5.245      Median :1057.5      Median :2134  
## Mean   : 7638      Mean   : 5.490      Mean   : 991.2      Mean   :2304  
## 3rd Qu.:10727      3rd Qu.: 7.713      3rd Qu.:1229.5      3rd Qu.:2793  
## Max.   :36019      Max.   :28.030      Max.   :1440.0      Max.   :4900
```

```
dActivity %>%
  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
```

Esse resumo estatístico das atividades nos apresenta usuários com minutos de sedentarismo muito alto e realizando atividades de baixa intensidade.

Calorias

```
#calories
dCalories %>%
  select(Calories) %>%
  summary()
```

```
## Calories
## Min. : 0
## 1st Qu.:1828
## Median :2134
## Mean :2304
## 3rd Qu.:2793
## Max. :4900
```

Horas de Sono

```
sleep %>%
  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
```

Peso

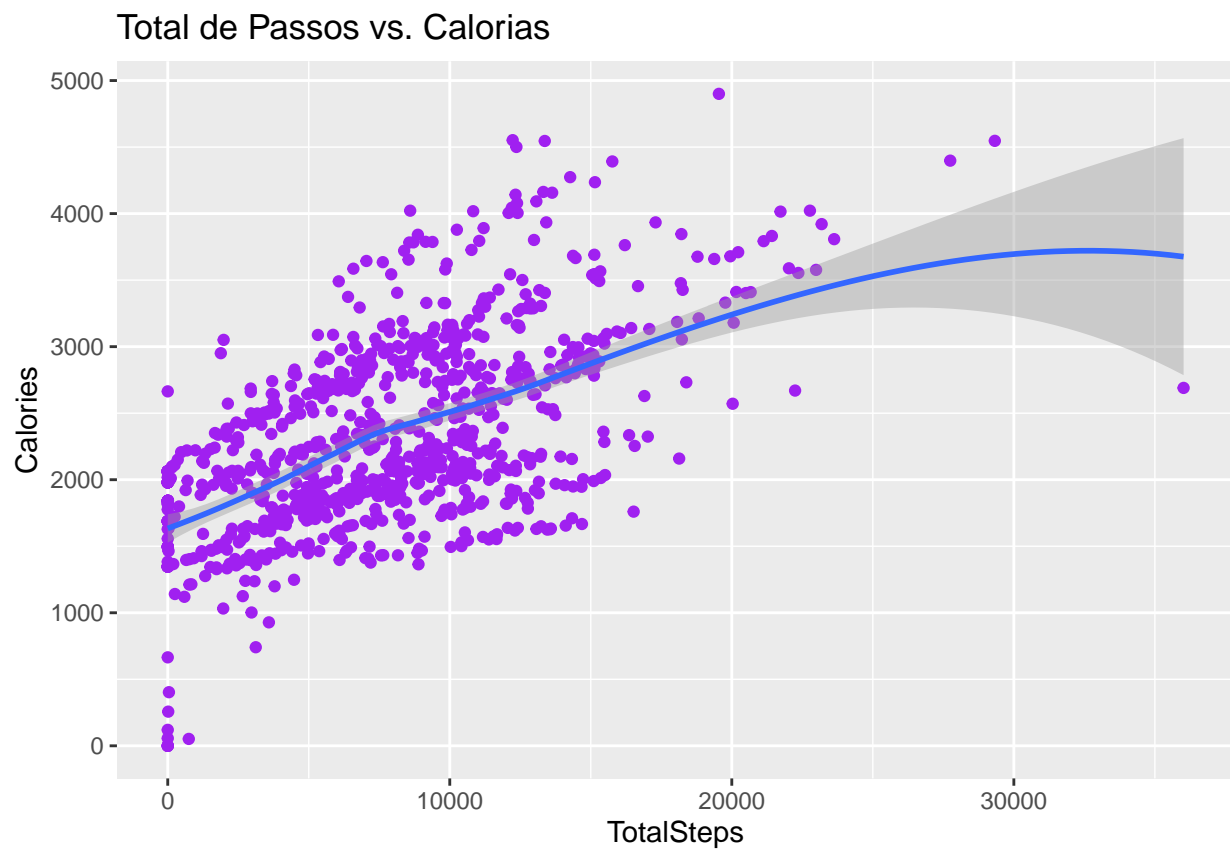
```
weight %>%
  select(WeightKg, BMI) %>%
  summary()
```

```
##      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
```

Criando Visualizações

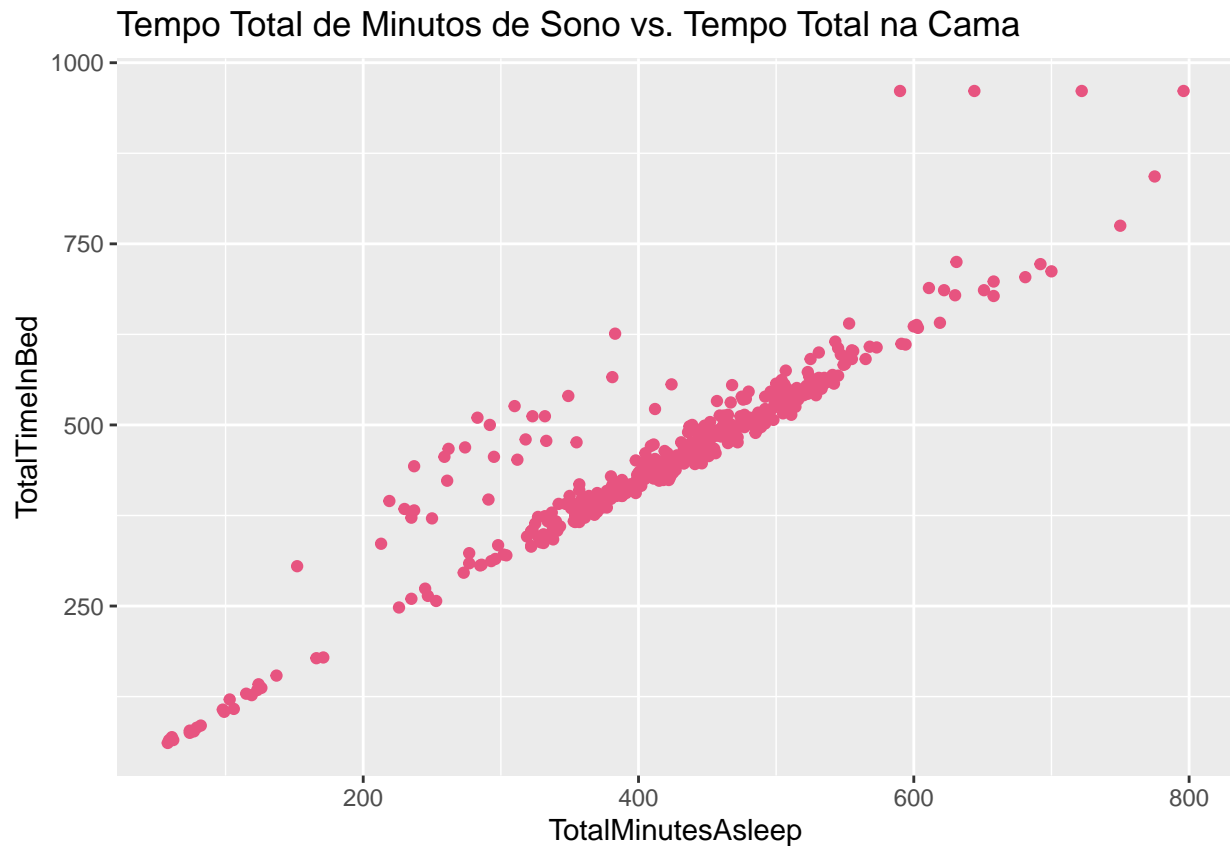
```
ggplot(data=dActivity, aes(x=TotalSteps, y=Calories)) +
  geom_point(colour="purple") + geom_smooth() + labs(title="Total de Passos vs. Calorias")
```

```
## 'geom_smooth()' using method = 'loess' and formula 'y ~ x'
```



Correlação positiva entre passos e calorias.

```
ggplot(data=sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +
  geom_point(colour="#E75480")+ labs(title="Tempo Total de Minutos de Sono vs. Tempo Total na Cama")
```



É possível notar uma linearidade entre o tempo de sono e o tempo que o usuário permanece na cama, apresentando uma média de 7 a 8 horas dormidas.

Para o gráfico abaixo foi necessário antes modificar o formato de hora do dataset de Intesidade:

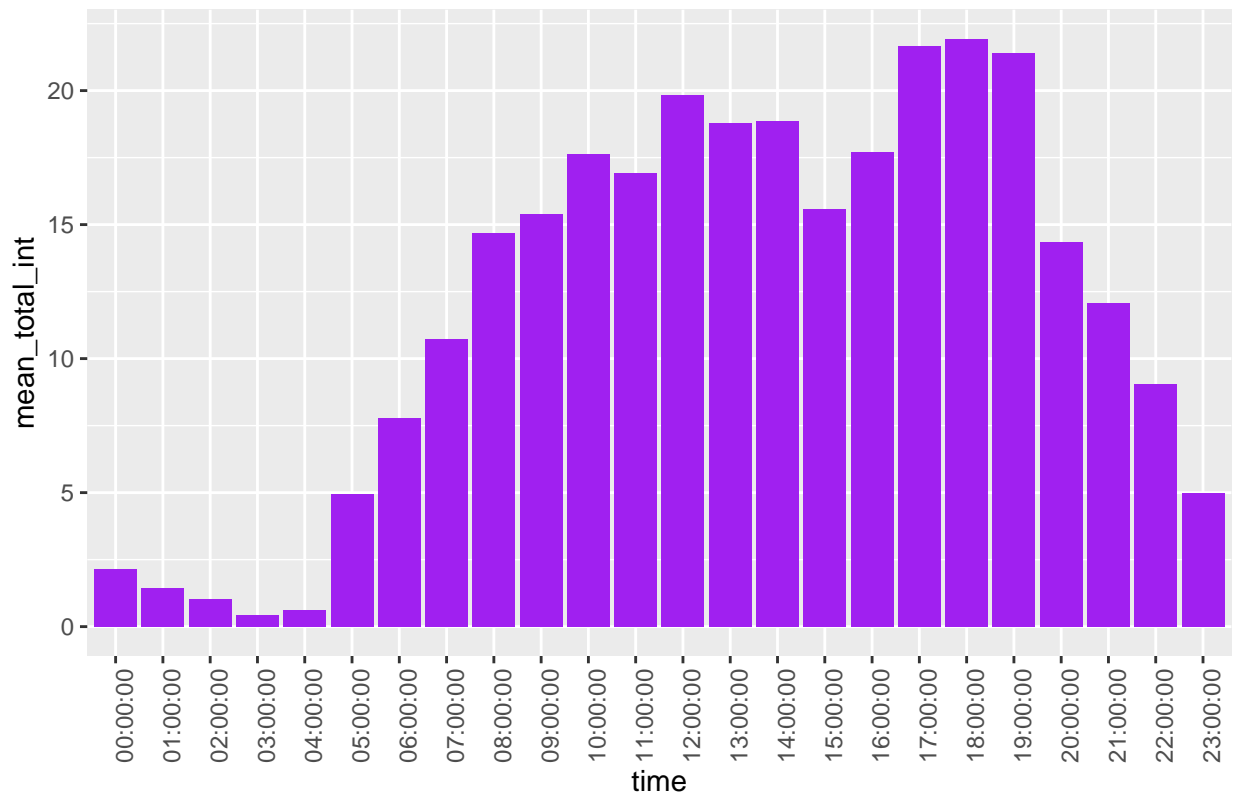
```
#Convertendo as Horas
# intensities
hIntensities$ActivityHour=as.POSIXct(hIntensities$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.tz)
hIntensities$time <- format(hIntensities$ActivityHour, format = "%H:%M:%S")
hIntensities$date <- format(hIntensities$ActivityHour, format = "%m/%d/%y")

#Agrupando as Horas
int_new <- hIntensities %>%
  group_by(time) %>%
  drop_na() %>%
  summarise(mean_total_int = mean(TotalIntensity))

ggplot(data=int_new, aes(x=time, y=mean_total_int)) + geom_histogram(stat = "identity", fill='purple') +
  theme(axis.text.x = element_text(angle = 90)) +
  labs(title="Média Total de Período Ativo vs. Tempo")
```

```
## Warning: Ignoring unknown parameters: binwidth, bins, pad
```

Média Total de Período Ativo vs. Tempo



Nota-se que grande parte dos usuários se exercitam por volta das 17 as 19 horas, horário que geralmente as pessoas saem do trabalho.

Conclusão

Não existem dados suficientes sobre o peso, mas seria interessante ter alguma proposta que encoraje o usuário a introduzir estes valores, uma vez que a compreensão da perda calórica está relacionada com o peso e o metabolismo. É possível verificar que o tempo relacionado ao sedentarismo é muito longo, o que pode justificar os dados, a tendência é de um grupo com excesso de peso, BMI maior que 25. O número de passos é também relativamente baixo em referência ao padrão sugerido pela comunidade de saúde para uma pessoa ativa, que é de 8000 passos. Hoje em dia temos sistemas de recompensa com distintivos, seria bom para o usuário ter esta informação. Os dados indicam também que os usuários fazem exercício entre as 17h e as 19h, provavelmente depois do trabalho, o que é talvez um momento ideal para enviar mensagens relacionadas com o marketing. Seria interessante criar um sistema de pontos relacionados com calorias queimadas em que se pudesse trocar por descontos em produtos Bellabeat ou spas, salões de unhas, produtos saudáveis, etc. Acho que esta seria uma forma de encorajar os usuários a diminuir o sedentarismo.