

Bellabeat Case Study

Kristoffer

2023-01-30

Företaget

Bellabeat är ett företag inriktat på träningsprodukter för kvinnor. De är ett litet bolag men Bellabeat har potential att bli en större aktör på marknaden för smarta produkter. Urška Sršen som är Chief Creative Officer för Bellabeat tror att analys av data från smarta klockor kan hjälpa till att skapa nya tillväxtmöjligheter för företaget.

Frågor inför analysen

- Vilka trender finns inom användningen av smarta klockor?
- Vad innebär dessa trender för Bellabeats kunder?
- Hur kan dessa trender påverka Bellabeats marknadsföringsstrategi?

Uppgiften

Identifiera tillväxtmöjligheter samt komma med rekommendationer för förbättringar av Bellabeats marknadsföringsstrategi baserat på trender inom användningen av smarta klockor.

Paket

```
library(tidyverse)

## — Attaching packages ————— tidyverse
1.3.2 —
## ✓ ggplot2 3.4.1      ✓ purrr  1.0.1
## ✓ tibble  3.1.8      ✓ dplyr  1.1.0
## ✓ tidyr   1.3.0      ✓ stringr 1.5.0
## ✓ readr   2.1.4      ✓ forcats 1.0.0
## — Conflicts —————
tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ 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)
```

Importera data

Data jag använder kommer från FitBit Fitness Tracker Data.

```
activity <- read.csv("daily_Activity_merged.csv")
calories <- read.csv("hourly_Calories_merged.csv")
intensities <- read.csv("hourly_Intensities_merged.csv")
sleep <- read.csv("sleep_Day_merged.csv")
weight <- read.csv("weight_Log_Info_merged.csv")
```

Jag har kontrollerat data i Excel och lagt märke till att tidsangivelserna varierar mellan filerna så jag behöver konvertera till enhetligt datum och tid format samt dela upp dessa. För att se att allt importerades som det ska använder jag View() funktionen.

```
View(activity)
```

Formatering

```
# intensiteter
intensities$ActivityHour=as.POSIXct(intensities$ActivityHour,
format="%m/%d/%Y %I:%M:%S %P" , tz=Sys.timezone())
intensities$time <- format(intensities$ActivityHour, format = "%H:%M:%S")
intensities$date <- format(intensities$ActivityHour, format = "%m/%d/%y")
# kalorier
calories$ActivityHour=as.POSIXct(calories$ActivityHour, format="%m/%d/%Y
%I:%M:%S %p", tz=Sys.timezone())
calories$time <- format(calories$ActivityHour, format = "%m/%d/%y")
calories$date <- format(calories$ActivityHour, format = "%m/%d/%y")
# aktivitet
activity$ActivityDate=as.POSIXct(activity$ActivityDate, format="%m/%d/%Y",
tz=Sys.timezone())
activity$date <- format(activity$ActivityDate, format = "%m/%d/%y")
# sömn
sleep$SleepDay=as.POSIXct(sleep$SleepDay, format="%m/%d/%Y %I:%M:%S %p",
tz=Sys.timezone())
sleep$date <- format(sleep$SleepDay, format = "%m/%d/%y")
```

Sammanfattning

```
n_distinct(activity$Id)
## [1] 33
n_distinct(calories$Id)
## [1] 33
n_distinct(intensities$Id)
## [1] 33
```

```
n_distinct(sleep$Id)
## [1] 24

n_distinct(weight$Id)
## [1] 8
```

Antalet deltagare i varje datamängd varierar stort. Det finns 33 deltagare i aktivitets, kalori och intensitetsdataseten, 24 i sömn men bara 8 i viktdatasetet. 8 deltagare är för lågt antal för att kunna presentera några datadrivna rekommendationer samt att dra slutsatser utifrån.

aktivitet

```
activity %>% 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
```

antalet aktiva minuter per kategori

```
activity %>% 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
```

kalorier

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

```
##      Calories
## Min.   : 42.00
## 1st Qu.: 63.00
## Median : 83.00
## Mean   : 97.39
## 3rd Qu.:108.00
## Max.   :948.00
```

sömn

```
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
```

vikt

```
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
```

Det intressanta med sammanfattningen

- Det genomsnittliga stillasittandet är 991 minuter eller 16 timmar/dag!
- Majoriteten av deltagarna är "lightly active".
- I genomsnitt sover deltagarna 1 gång/dygn i 7 timmar.
- 7638 steg är det dagliga genomsnittet, vilket är lite mindre än vad som rekommenderas av Världshälsoorganisationen.

Sammanfoga data

Innan jag börjar visualisera data måste jag slå samman två datamängder. Jag kommer att slå samman "aktivitet" och "sömn" genom kolumnerna id och datum (den som jag konverterade till datum och tidsformat).

```
merged_data <- merge(sleep, activity, by=c('Id', 'date'))
head(merged_data)
```

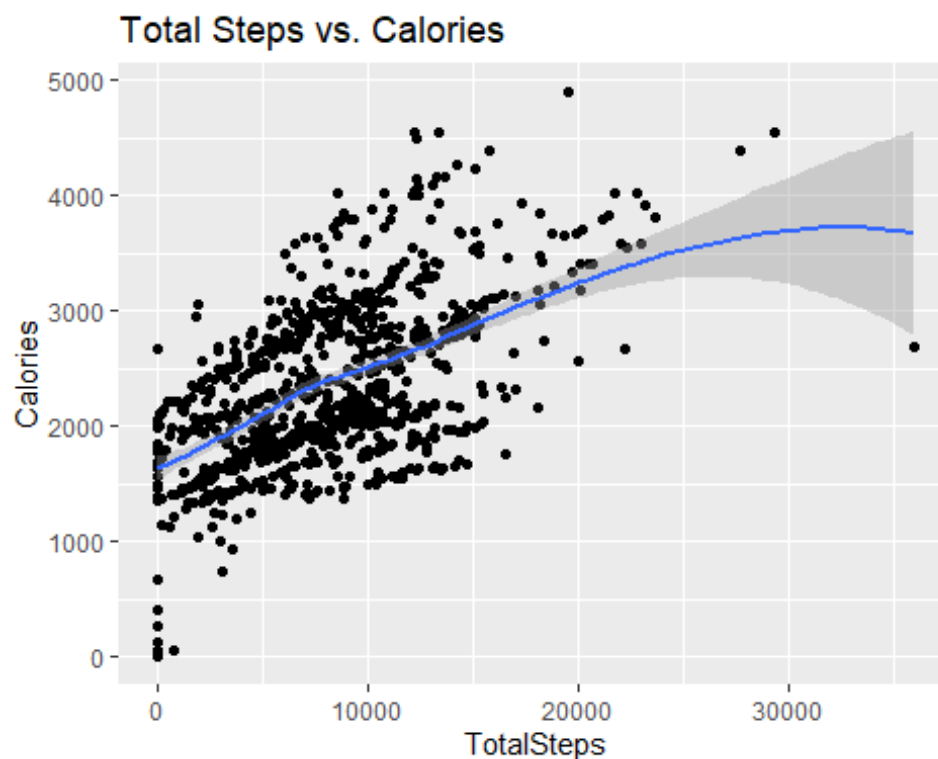
```
##      Id      date SleepDay TotalSleepRecords TotalMinutesAsleep
## 1 1503960366 04/12/16 2016-04-12              1              327
## 2 1503960366 04/13/16 2016-04-13              2              384
## 3 1503960366 04/15/16 2016-04-15              1              412
## 4 1503960366 04/16/16 2016-04-16              2              340
## 5 1503960366 04/17/16 2016-04-17              1              700
## 6 1503960366 04/19/16 2016-04-19              1              304
##      TotalTimeInBed ActivityDate TotalSteps TotalDistance TrackerDistance
## 1              346   2016-04-12      13162           8.50           8.50
## 2              407   2016-04-13      10735           6.97           6.97
## 3              442   2016-04-15       9762           6.28           6.28
## 4              367   2016-04-16      12669           8.16           8.16
## 5              712   2016-04-17       9705           6.48           6.48
## 6              320   2016-04-19      15506           9.88           9.88
##      LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
```

```
## 1      0      1.88      0.55
## 2      0      1.57      0.69
## 3      0      2.14      1.26
## 4      0      2.71      0.41
## 5      0      3.19      0.78
## 6      0      3.53      1.32
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1          6.06              0              25
## 2          4.71              0              21
## 3          2.83              0              29
## 4          5.04              0              36
## 5          2.51              0              38
## 6          5.03              0              50
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1          13          328          728      1985
## 2          19          217          776      1797
## 3          34          209          726      1745
## 4          10          221          773      1863
## 5          20          164          539      1728
## 6          31          264          775      2035
```

Visualisering

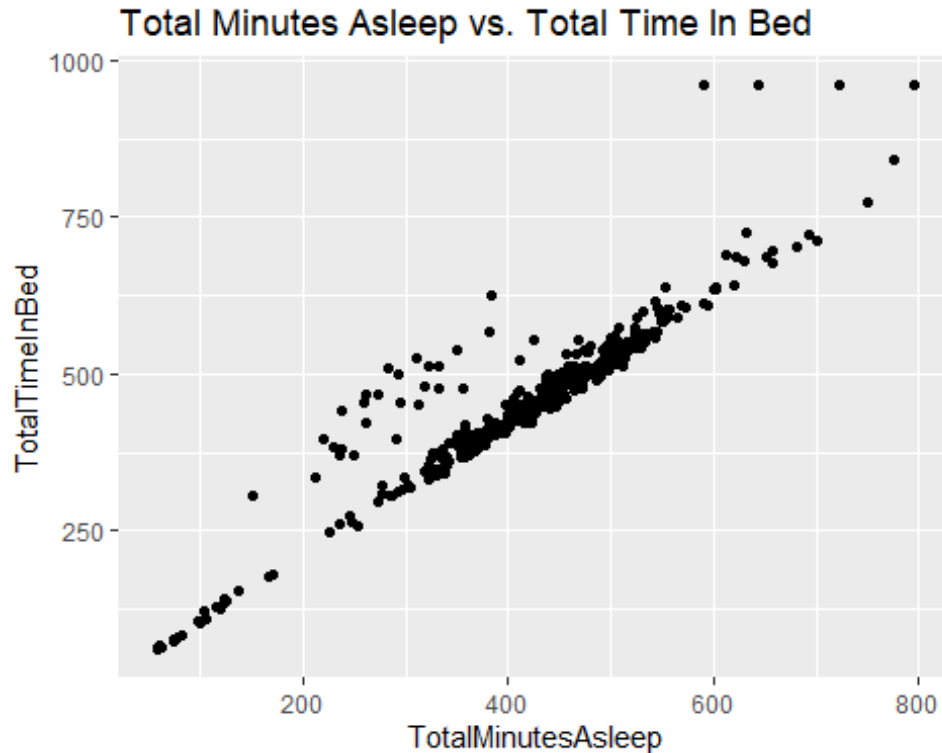
```
ggplot(data=activity, aes(x=TotalSteps, y=Calories)) + geom_point() +
geom_smooth() + labs(title="Total Steps vs. Calories")
```

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



Det finns en positiv korrelation här mellan totalt antal steg och kalorier, uppenbart såklart eftersom ju mer aktiva vi är, desto fler kalorier bränner vi.

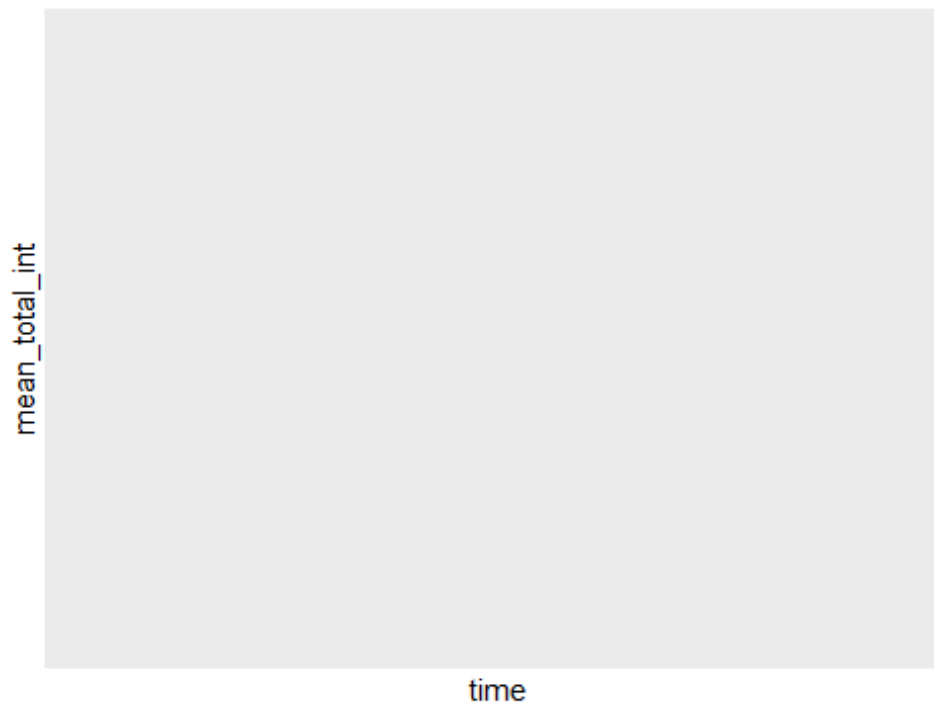
```
ggplot(data=sleep, aes(x=TotalMinutesAsleep, y=TotalTimeInBed)) +  
geom_point() + labs(title="Total Minutes Asleep vs. Total Time In Bed")
```



Förhållandet mellan totala minuter av sömn och totala minuter i sängen ser linjärt ut. Så om Bellabeats användare vill förbättra sin sömn bör man överväga att använda någon sorts påminnelse om att gå och lägga sig.

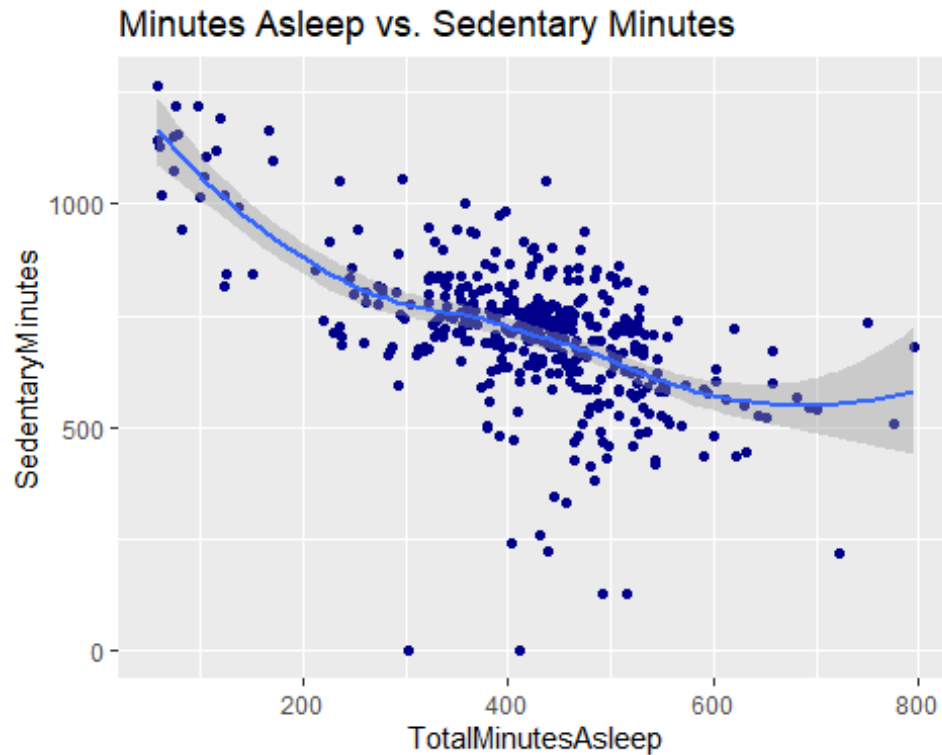
```
int_new <- intensities %>%  
  group_by(time) %>%  
  drop_na() %>%  
  summarise(mean_total_int = mean(TotalIntensity))  
  
ggplot(data=int_new, aes(x=time, y=mean_total_int)) + geom_histogram(stat =  
"identity", fill='darkblue') + theme(axis.text.x = element_text(angle = 90))  
+ labs(title = "Average Total Intensity vs. Time")  
  
## Warning in geom_histogram(stat = "identity", fill = "darkblue"): Ignoring  
## unknown parameters: `binwidth`, `bins`, and `pad`
```

Average Total Intensity vs. Time



I visualiseringen ser man tydligt att de flesta är aktiva mellan 17.00 och 19.00, jag antar att folk t.ex. tar en promenad eller går till gymmet efter jobbet. Under den här tiden kan man använda Bellabeat-appen för att påminna och försöka motivera användarna till att träna eller promenera.

```
ggplot(data=merged_data, aes(x=TotalMinutesAsleep, y=SedentaryMinutes)) +  
  geom_point(color='darkblue') + geom_smooth() + labs(title = "Minutes Asleep  
vs. Sedentary Minutes")  
  
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



I förhållandet mellan totala minuter sömn och stillasittande minuter kan man tydligt se ett negativt förhållande mellan dessa två. Om Bellabeats användare vill försöka förbättra sin sömn kan appen rekommendera olika sätt att minska stillasittande tid. En sådan slutsats behöver dock stödjas med mer data eftersom korrelationen här inte behöver betyda att det finns ett orsakssamband.

Efter att ha analyserat FitBit Fitness Tracker Data är detta mina rekommendationer för Bellabeats marknadsföringsstrategi.

Bellabeats målgrupp är heltidsarbetande kvinnor (enligt intensitetsdata/timme) och spenderar mycket tid stillasittande (även det enligt data). Dessa kvinnor utför "lätt aktivitet" för att hålla sig friska (enligt aktivitetsanalysen). För att nå en mer hälsosam livsstil behöver de öka sin vardagliga aktivitet. Här kan appen fungera som en vän som förmedlar kunskap om hälsosamma vanor samt att kunna bidra med motivation för att bibehålla de man redan utför.

Konkreta förslag

- Det genomsnittliga antalet steg per dag är 7638 vilket är lite mindre än 8000 som är Världshälsoorganisationens rekommendation. Bellabeat-appen kan uppmuntra användarna att försöka ta minst 8 000 genom att förklara hälsofördelarna.
- Om användarna vill förbättra sin sömn bör Bellabeat överväga att använda påminnelser för att gå och lägga sig. Appen kan också på olika sätt rekommendera att minska stillasittandet.
- Den mesta träningen sker mellan klockan 17.00 och 19.00. Bellabeat-appen kan under denna tid påminna och motivera sina användare att träna eller springa.

- Om användaren vill gå ner i vikt nyttjar man funktionen att kontrollera sin dagliga kaloriförbrukning. Bellabeat-appen kan då innehålla förslag på luncher och middagar.