BellaBeat

Pauline M

2022-05-13

**Introduction**

This is an analysis requested by BellaBeat executives. BellaBeat is a high-tech manufacturer of health-focused products for women. Bellabeat’s products include; BellaBeat App, Leaf, Time, Spring and BellaBeat membership.

**##Business Task**

To reveal opportunities for growth by gaining insight into consumer smart device usage and applying insights to develop a marketing strategy for a BellaBeat product.

**#Questions to answer**

\* What are some trends in smart device usage?

\* How could these trends apply to BellaBeat customers?

\* How could these trends help influence Bellabeat Marketing Strategy

**##About Data**

**#Fitbit Fitness Tracker Data by Mobius**1

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. Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.

***\*\*There is no personal identifying information within the datasets used\*\****

**#Data Limitations**

\* Study date shows 03/12/2016-05/12/2016 but results actually run from 04/12/2016-05/12/2016

\* Insufficient data with only 33 participants - note that the study had shown that there were 30 participants

\* No demographic data was available which would have been useful since Bellabeat targets the female market \* Data is not current - 2016

*\** Insufficient weight data for analysis as only 8 participants provided this data*.*

*First I start by loading the dataframes*

After loading dataframes, I prepared the data. I had checked the data for duplicates, blanks, formatting errors in excel prior to upload, further cleaning was done within R. Heartrate dataset had several duplicates. Removed these by creating another dataset using ‘distinct’.

Activity\_summary <- dailyActivity\_merged %>%  
 group\_by(Id) %>%   
 drop\_na() %>%  
 summarize(max\_steps=max(TotalSteps),max\_distance=max(TotalDistance),max\_calories=max(Calories),Avg\_steps=mean(TotalSteps),Avg\_distance=mean(TotalDistance),Avg\_calories=mean(Calories))

heartrate <- heartrate\_seconds\_merged %>% distinct  
  
heartrate\_summary <- heartrate%>%  
 group\_by(Id,Date)%>%  
 summarize(daily\_avg\_rate=mean(Value, na.rm = FALSE))

## `summarise()` has grouped output by 'Id'. You can override using the `.groups`  
## argument.

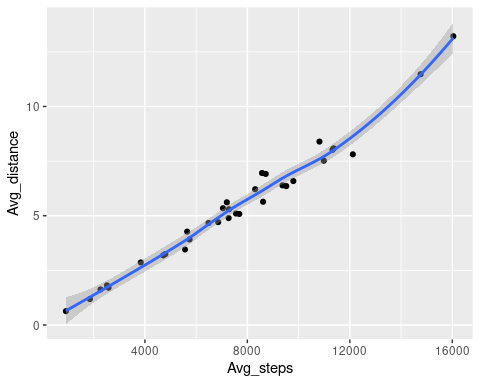
# Analyze the activity levels of the participants

ggplot(data=Activity\_summary)+  
 geom\_point(mapping=aes(x=Avg\_steps, y=Avg\_distance))+  
 geom\_smooth(mapping=aes(x=Avg\_steps, y=Avg\_distance))+

labs(title ="Steps Vs Distance")

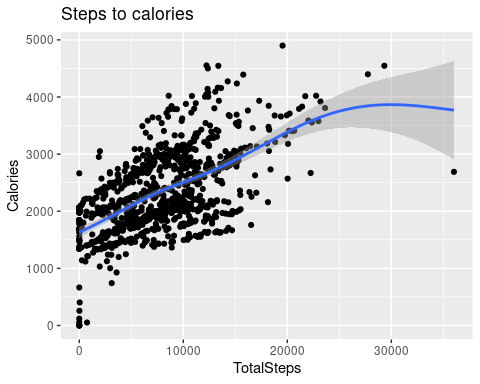
geom\_smooth()` using method = 'loess' and formula 'y ~ x'

## $title  
## [1] "Steps Vs Distance"  
##   
## attr(,"class")  
## [1] "labels"



Result shows that there is a positive correlation between distance traveled and steps. The higher the average distance traveled the more the average steps taken.

ggplot(data=dailyActivity\_merged)+  
geom\_point(mapping=aes(x=TotalSteps, y=Calories))+  
geom\_smooth(mapping=aes(x=TotalSteps, y=Calories), method ="gam", formula = y~s(x))+  
labs(title="Steps to calories")



The results indicate that the more steps taken, the higher the calories burned.

**#Number of participants that meet and exceed commonly used 10K daily steps limit**

Activity\_summary%>%  
 count(Avg\_steps>=10000)

## # A tibble: 2 × 2  
## `Avg\_steps >= 10000` n  
## <lgl> <int>  
## 1 FALSE 26  
## 2 TRUE 7

Activity\_summary%>%  
 count(Avg\_steps>=8000)

## # A tibble: 2 × 2  
## `Avg\_steps >= 8000` n  
## <lgl> <int>  
## 1 FALSE 19  
## 2 TRUE 14

7 out of the 33 participants have met the 10000 steps per day average, 21%. In addition, only 14 out of 33 participants are taking more than 8000 average steps, 42%, lower than half of them. Majority of participants have lower than the 10K steps. This is an area of opportunity….

<https://www.nih.gov/news-events/nih-research-matters/number-steps-day-more-important-step-intensity>a According to the above article, ‘Compared with people who took 4,000 steps a day, those who took 8,000 steps a day at the start of the study had a 50% lower risk of dying from any cause during follow-up. People who took 12,000 steps a day had a 65% lower risk of dying than those who took only 4,000. Higher step counts were also associated with lower rates of death from heart disease and cancer. These benefits were consistent across age, sex, and race groups.’ *Conclusion: the more steps taken, the lower the mortality rate.*

**#Participants with average calories burned of >400 per day**

Speaking to Newsweek, Cemal Ozemek**b**, a clinical exercise physiologist certified by the American College of Sports Medicine (ACSM) recommended burning 400-500 calories per day, and an average of 2000 calories a week.

Activity\_summary%>%  
 count(Avg\_calories>=400)

## # A tibble: 1 × 2  
## `Avg\_calories >= 400` n  
## <lgl> <int>  
## 1 TRUE 33

Avg\_calories >= 400 n lgl> 1 TRUE 33

Despite the fact that most participants don’t meet the daily steps recommended, all 33 participants met the recommended calories burned of 400-500 calories per day. This indicates that they are undertaking other activities that are not related to the number of steps they take daily that still allows them to burn higher calories.

**#Review the level of activity by intensity and time**

dailyActivity\_merged%>%  
 summarize(avg\_fairlyactive=mean(FairlyActiveMinutes),  
 avg\_lightlyactive=mean(LightlyActiveMinutes),  
 avg\_veryactive=mean(VeryActiveMinutes))

## # A tibble: 1 × 3  
## avg\_fairlyactive avg\_lightlyactive avg\_veryactive  
## <dbl> <dbl> <dbl>  
## 1 13.6 193. 21.2

Most participants spent a large part of their time being lightly active(light exercise 1-3 days a week), a moderate amount of time being very active and the least time being fairly active based on daily averages. Based on the above number it means that a weekly average of moderate activity(fairly active) would amount to an average of 95.2 minutes and for vigorous (very active) minutes, 148.4.

The Physical Activity Guidelines for Americans recommend that adults get at least 150 minutes of moderate-intensity aerobic physical activity or 75 minutes of vigorous-intensity physical activity, or an equivalent combination each week.

Participants are exceeding the recommended level of moderate to vigorous physical activity.

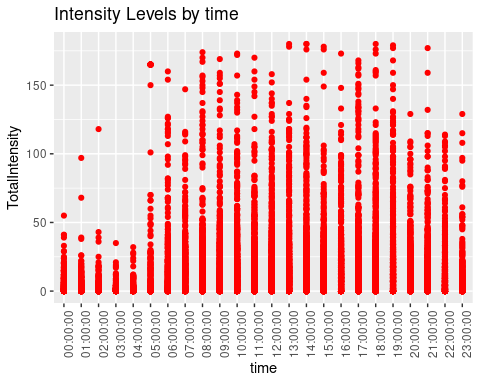
**#Intensity Levels**

***First we needed to separate date and time:***

intensities <- hourlyIntensities\_merged  
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")

**#Intensities plot**

ggplot(data=intensities, aes(x=time, y=TotalIntensity)) +  
 geom\_point(color="red")+  
 theme(axis.text.x=element\_text(angle=90))+  
 labs(title="Intensity Levels by time")



Intensity levels show to be at their highest between 1:00pm-3:00pm and 6:00pm-7:00pm but levels are sporadic between 5:00am and 9:00pm. Higher intensity levels at the two intervals may be due to lunch breaks and having left work in the evening when people get time to work out.

**#Analyze calories**

***#Separate date and time for hourly calories***

calories <- hourlyCalories\_merged  
calories$ActivityHour=as.POSIXct(calories$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())  
calories$time <- format(calories$ActivityHour, format = "%H:%M:%S")  
calories$date <- format(calories$ActivityHour, format = "%m/%d/%y")

***#Separate date and time for hourly steps***

steps <- hourlySteps\_merged  
steps$ActivityHour=as.POSIXct(steps$ActivityHour, format="%m/%d/%Y %I:%M:%S %p", tz=Sys.timezone())  
steps$time <- format(steps$ActivityHour, format = "%H:%M:%S")  
steps$date <- format(steps$ActivityHour, format = "%m/%d/%y")

***#Create new dataframe for hourly data to analyse hourly trends***

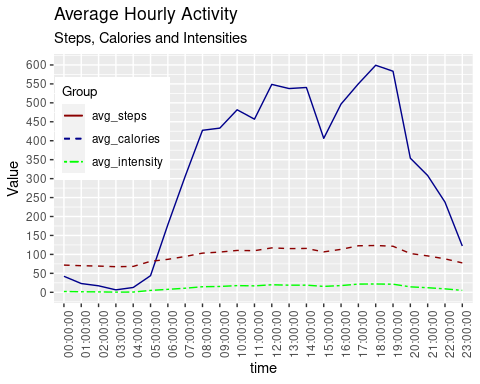
hourly\_data <-merge(steps,calories,intensities, by.x=c('Id','date','time'), by.y=c('Id','date','time'))   
hourly\_combined\_data <-merge(hourly\_data,intensities, by.x=c('Id','date','time'), by.y=c('Id','date','time'))

***#Prep - plotting the line chart***

avg\_hourly<-hourly\_combined\_data %>%  
group\_by(time) %>%  
summarize(avg\_steps=mean(StepTotal),avg\_calories=mean(Calories),avg\_intensity=mean(TotalIntensity))  
  
variable <- c("avg\_steps" = "solid", "avg\_calories" = "dashed", "avg\_intensity" = "twodash")  
color\_variable <- c("darkred", "darkblue", "green")

***#Calories,Steps and Intensities Plot***

ggplot(data = avg\_hourly,aes(x=time, group=1))+  
 geom\_line(mapping=aes(y = avg\_steps, color = "darkred", linetype = "avg\_steps")) +   
 geom\_line(mapping=aes(y = avg\_calories, color ="darkblue", linetype = "avg\_calories"))+  
 geom\_line(mapping=aes(y = avg\_intensity, color = "green", linetype = "avg\_intensity"))+  
 scale\_y\_continuous(breaks=seq(0,600,50))+  
 labs(x="time", y="Value") +  
 theme(axis.text.x=element\_text(angle=90))+  
 labs(title="Average Hourly Activity", subtitle="Steps, Calories and Intensities")+  
 scale\_linetype\_manual(values=variable, name = "Group", labels = c("avg\_steps", "avg\_calories", "avg\_intensity"))+  
 scale\_colour\_manual(values=color\_variable, name = "Group", labels = c("avg\_steps", "avg\_calories", "avg\_intensity"))+  
 theme(legend.position=c(0.14,0.7), legend.title=element\_text(size = 10))



Participants are shown to be most active in steps, calories burned and intensities between 12:00 noon -2:00pm and 5:00-7:00pm. This could be due to people working out or taking walks during lunch and after work. There is an opportunity here to send users reminders and exercise suggestions during these times.

***#Analyze stress levels by steps and heartrate.***

avg\_daily\_steps <- dailyActivity\_merged %>%  
group\_by(Id, ActivityDate)%>%  
drop\_na()%>%  
summarize(daily\_avg\_steps=mean(TotalSteps))

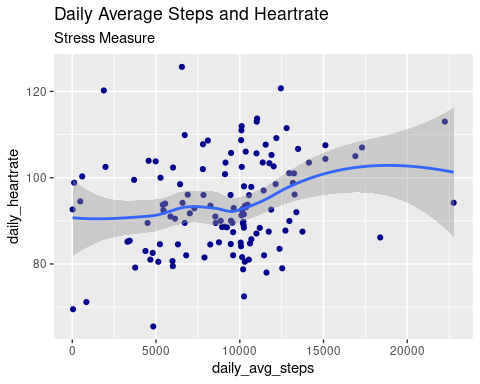
## `summarise()` has grouped output by 'Id'. You can override using the `.groups`  
## argument.

Avg\_heartrate\_summary <- heartrate\_summary %>%  
rename(daily\_heartrate = daily\_avg\_rate)  
merged\_stepsbyheartrate <- merge(avg\_daily\_steps, Avg\_heartrate\_summary, by.x=c('Id', 'ActivityDate'),by.y=c('Id','Date'))

***#Heartrate Plot***

ggplot(data=merged\_stepsbyheartrate) +  
geom\_point(mapping=aes(x=daily\_avg\_steps, y=daily\_heartrate), color="darkblue") +  
geom\_smooth(mapping=aes(x=daily\_avg\_steps, y=daily\_heartrate)) +   
labs(title="Daily Average Steps and Heartrate", subtitle="Stress Measure")

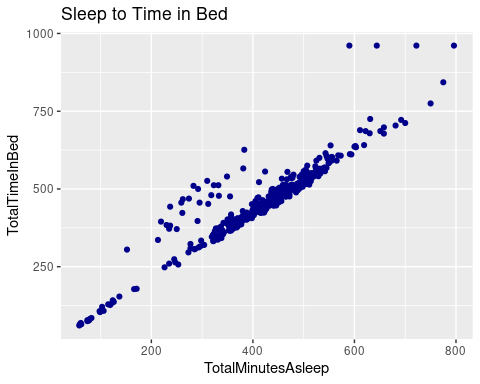
## `geom\_smooth()` using method = 'loess' and formula 'y ~ x'



Heart rate steadily increases with increased steps taken. We are limited on analyzing heart rate as we don’t have the age of the participants that is a key determinant for ideal maximum heart rate.

***#Sleep Analysis***

ggplot(data=sleepDay\_merged) +  
geom\_point(mapping=aes(x=TotalMinutesAsleep, y=TotalTimeInBed), color="darkblue")+  
labs(title="Sleep to Time in Bed")



The more time spent in bed the longer one sleeps. Users can be sent reminders on bed times and when to come out of bed to increase sleep. Sleep pattern reports can also be sent to them to track their sleep patterns with recommendations on when to go to bed and when to get up.

***Since the Fitbit Fitness Tracker Data has several limitations, I found a similar dataset to analyze.***

**## PMData**

PMData is a dataset that combines traditional lifelogging data with sports-activity data. PMData combines input from Fitbit Versa 2 smartwatch wristbands, the PMSys sports logging smartphone application, and Google forms. PMData contains data collected from 16 persons: twelve men and three women, in the age range 25–60 years, with an average ageof 34 years. The reporting period is from the start of November 2019 to the end of March 2020. The participants range from a broad background with regards to training and exercises. Some are active athletes, some previous athletes, and some rarely exercised at all. An overview of the participants’ demographic information is provided in the participant-overview.xlsx file where information like age, height, gender, measured max heart rate, test run results, and walk and run stride lengths are included.

*Cleaned CSV datasets provided by Salah; PMData: Fitbit Versa Activity Data2*

*Original PMData : https://osf.io/vx4bk/3*

# Limitations of PMData dataset

* There are only 3 females out of a population of 16
* There are only 16 participants

There were no duplicates in the data, however there were a few days in the pmdata\_daily\_activity\_merged where zero activity was recorded for some of the participants. This were not sufficient to alter the results significantly.

I also created several pivottables to analyse: \* Activity by gender \* Activity by Day \* Activity vs Sleep and Age \* Steps and Calories \* Time in bed, Sleep and Sedentary Minutes \* Sleep vs Activity

The link below provides access to my excel analysis of the PMData set.

[..\Desktop\DA Course Exercises\pmdata\_analyzed.xlsx](file:///C:\Users\Akula\Desktop\DA%20Course%20Exercises\pmdata_analyzed.xlsx)

Below are some of the results:

|  |
| --- |
| Women are more active than men in general but men put in more very-active and moderately-active minutes than women. |
| Participants are more lightly active than moderately or very active. Activities done by the women may be more related to chores or errands. |
| Reminders for set work out times to increase very active minutes would be useful especially for the ladies. |

Most active days are Thursday, Friday and Saturday. Reminders can be sent for days Monday-Wednesday to encourage participants to work out more often. Least active day is Sunday

|  |  |
| --- | --- |
| Participants between the ages of 31-42 seem to have the highest level of lightly active minutes and the least minutes asleep. | |
| This may be due to likely having more family responsibilities. They would benefit from bed time reminders to increase amount of sleep. | |
| 2 of the participants with the highest very active minutes are over the age of 54. They may have more time to work out. | |
| Younger participants need to be encouraged to spend more time working out.  8 out of 16 participants have an average of 10K steps(50%) , 10 out of the 16 took over 8K steps (62.5%). All the participants averaged over 2K calories burned. 3 Participants aged 26, 27 and 60 respectively had the highest steps taken. No correlation between age, average calories burnt and average steps taken. There is, however, a positive correlation between steps taken and calories burnt. | |
| There is a positive correlation between time spent in bed and minutes asleep. The longer participants stayed in bed, the longer they slept. |
| There is a negative correlation between minutes asleep and sedentary minutes. The more sedentary time participants had the less sleep they got. |

The higher the number of very active to moderate activity minutes put in, the more sleep minutes the participant had.

More activity results in more sleep and a more rested person.

Increase very active to moderate minutes to increase sleep and reduce stress

**KEY INSIGHTS**

\* Both Fitbit and PMData datasets support the conclusion that the more steps you take the more calories burned. The Fitbit group has 21% of their members taking over 10K steps and 42% taking over 8K steps. The PMData group had over 50% of their group taking over 10K steps and 63% with steps over 8K.

\* Both groups of participants were mostly lightly active. In the PMData group, the women were overall more active than men but men had higher moderately to very active minutes. PMData group also appeared to be more active on Thursday-Saturday and least active on Sunday. In addition, the participants aged 31-42 had the highest level of lightly active minutes and the least minutes asleep.

\*Both datasets supported the conclusion that those that spent more time in bed slept more. For the PMData group there was a negative correlation between sleep and sedentary minutes. Participants in this group that had higher moderate to vigorous active minutes slept more.

\*It would be beneficial to gather more data that is recent, includes demographics and has more information on heart rate and weight. This would shed more light on stress levels and effect of activity on weight management.

**RECOMMENDATIONS**

Based on the findings, I chose the Time product the best product to focus our marketing strategy on. The Time is a watch that also tracks activity, sleep and stress. It connects to the app to track daily wellness.

\* Users of Time can be sent reminders on slow activity days to encourage them to work out on those days as well, thus increase overall activity levels. These reminders should also be accompanied by physical activity recommendations to help guide users to become more moderately or vigorously active. I have included a link in which the CDC categorizes some moderate to vigorous activities that can be used as a guide for recommended workoutsc.

\* Time should also allow tracking of users heart rate based on their age information that should be provided by the user. Ideal heart rate should be calculated based on level of activity and users should be alerted if their heart rate is over the maximum. At rest heart rate should also be tracked as lower heart rates have been found to reduce mortalityd. Heart rate is a key measure of stress, therefore the ability to track user’s heart rate customized by age could be very useful and perhaps even lifesaving.

\*With Time, users can be encouraged to increase sleep by increasing time in bed, reducing sedentary time, and increase moderate to active minutes. Reminders or alarms can be set for these events to help the user track sleep patterns.

**Resources**

[1https://www.kaggle.com/datasets/arashnic/fitbit](#_Limitations_of_PMData)

[2https://www.kaggle.com/datasets/arashnic/fitbit/discussion/313589](#_Limitations_of_PMData) *(permission to use private data within note)*

[2https://www.kaggle.com/code/salihobaid/pmdata-fitbit-versa-activity-data/notebook](#_Limitations_of_PMData)

[*3https://osf.io/vx4bk/*](#_Limitations_of_PMData)

a<https://www.nih.gov/news-events/nih-research-matters/number-steps-day-more-important-step-intensity>

[bhttps://www.newsweek.com/how-many-calories-burn-workout-fitness-advice-1673009](#_Limitations_of_PMData)

[chttps://www.cdc.gov/nccdphp/dnpa/physical/pdf/pa\_intensity\_table\_2\_1.pdf](#_Limitations_of_PMData)

[dhttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4754196/](#_Limitations_of_PMData)