Fitbit Wellness Report

Derrick

2023-03-19

How Can a Wellness Technology Company Play It Smart?

As a Bella beat Data Analyst working closely with the Bella Beat Marketing Team, I have been tasked with finding answers to three business problems that will inform the marketing team on what to do next.

Business Task

1. What are some trends in smart device usage?
2. How could these trends apply to Bella beat customers?
3. How could these trends help influence Bella beat marketing strategy?

Data Sets

To answer these questions, a public kaggle data set of fit bit users will be used. The data set comprises of multiple csv files but after a close inspection of the business task as well as the time constraints of delivering in a week, three of the data sets will be used.

1. Daily Activity Data set
2. Hourly Activity Data set
3. Weight log Data set

1. Daily Activity Data set

The Daily Activity Data set is a collection of distance, time calories and steps data for each participant.

column names:

- Id,ActivityDate, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDistance, VeryActiveDistance, ModeratelyActiveDistance LightActiveDistance, SedentaryActiveDistance, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes, Calories

2. Hourly Activity Data Set

For the Hourly datset it contains data about the intensities during the day,steps and calories that the participant takes throughtout a day.

column names:

- Id, ActivityHour, Calories, TotalIntensity, AverageIntensity, StepTotal

3. Weight log Data Set

The weight Data set is a collection of different weight metrics from kg to lbs aswell as the BMI

column names:

- Id, Date, WeightKg, WeightPounds, Fat, BMI, IsManualReport, LogId

Data Processing/Cleaning

Loading Libraries

```
library(dplyr)
library(skimr)
library(tidyr)
library(ggplot2)
library(gridExtra)
```

Loading Datasets

```
##------------------------------ DATA SETS ------------------------------

## daily activity data
daily_activity <- read.csv("dailyActivity_merged.csv")

## hourly data

# uniting hourly calories, intensities and steps data sets
hourly_activity <-
inner_join(inner_join(read.csv("hourlyCalories_merged.csv"),

read.csv("hourlyIntensities_merged.csv")
                                          ),
                              read.csv("hourlySteps_merged.csv")
                              )
## Weight data

weightlog_info <- read.csv("weightLogInfo_merged.csv")
```

Data set Cleaning

```
##------------------------------ Data set Cleaning -------------------------
----



## Daily_Activity Data set

# viewing data
#str(daily_activity)
#colnames(daily_activity)
#skim_without_charts(daily_activity)
```

```r
# splitting date and changing columns to integer from character

# splitting columns
daily_activity <-
separate(daily_activity,ActivityDate,c("Month","Day","Year"),sep = "/")

# columns from character to integer
columns_to_integers <- c("Day","Year")
daily_activity[columns_to_integers] <-
lapply(daily_activity[columns_to_integers], as.integer)

# changing numeric month numbers to character
daily_activity$Month[daily_activity$Month == "4"] <- "April"
daily_activity$Month[daily_activity$Month == "5"] <- "May"

## Hourly activity Data set

# viewing data
#str(hourly_activity)
#colnames(hourly_activity)
#skim_without_charts(hourly_activity)

# splitting date and changing columns to integer from character

# splitting columns
hourly_activity <-
separate(hourly_activity,ActivityHour,c("MDY","Time","AM_PM"),sep = " ")
hourly_activity <- separate(hourly_activity,MDY,c("Month","Day","Year"),sep =
"/")
hourly_activity <-
separate(hourly_activity,Time,c("Hour","Minute","Second"),sep = ":")

# columns from character to integer
columns_to_integers3 <- c("Day","Year","Hour","Minute","Second")
hourly_activity[columns_to_integers3] <-
lapply(hourly_activity[columns_to_integers3], as.integer)

# changing numeric month numbers to character
hourly_activity$Month[hourly_activity$Month == "4"] <- "April"
hourly_activity$Month[hourly_activity$Month == "5"] <- "May"

# deleting unnecessary columns
hourly_activity <- hourly_activity %>%
  select(-Minute,-Second)


## weight log Data set

# viewing data
```

```r
#str(weightlog_info)
#colnames(weightlog_info)
#skim_without_charts(weightlog_info)

# splitting date and changing columns to integer from character

# splitting columns
weightlog_info <- separate(weightlog_info,Date,c("MDY","Time","AM_PM"),sep =
" ")
weightlog_info <- separate(weightlog_info,MDY,c("Month","Day","Year"),sep =
"/")
weightlog_info <-
separate(weightlog_info,Time,c("Hour","Minute","Second"),sep = ":")

# columns from character to integer
columns_to_integers4 <- c("Day","Year","Hour","Minute","Second")
weightlog_info[columns_to_integers4] <-
lapply(weightlog_info[columns_to_integers4], as.integer)

# changing numeric month numbers to character
weightlog_info$Month[weightlog_info$Month == "4"] <- "April"
weightlog_info$Month[weightlog_info$Month == "5"] <- "May"

# deleting unnecessary columns
weightlog_info$Fat[is.na(weightlog_info$Fat)] <- 0
```

Analysis/Findings

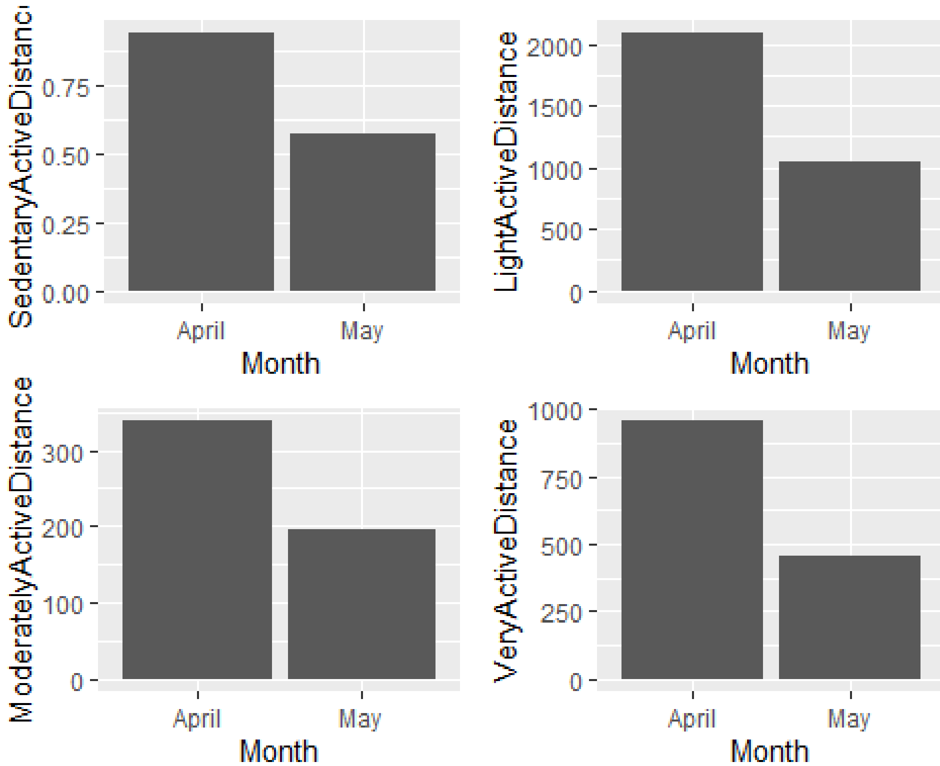
Distance Findings

From the distance columns it is clear that most of the exercise that people are doing is light active by a wide margin.

The data in May is Little bit less showing that there was not much exercise during May which is interesting. Also the distance covered by the majority is 3 to 5 KM of light exercise

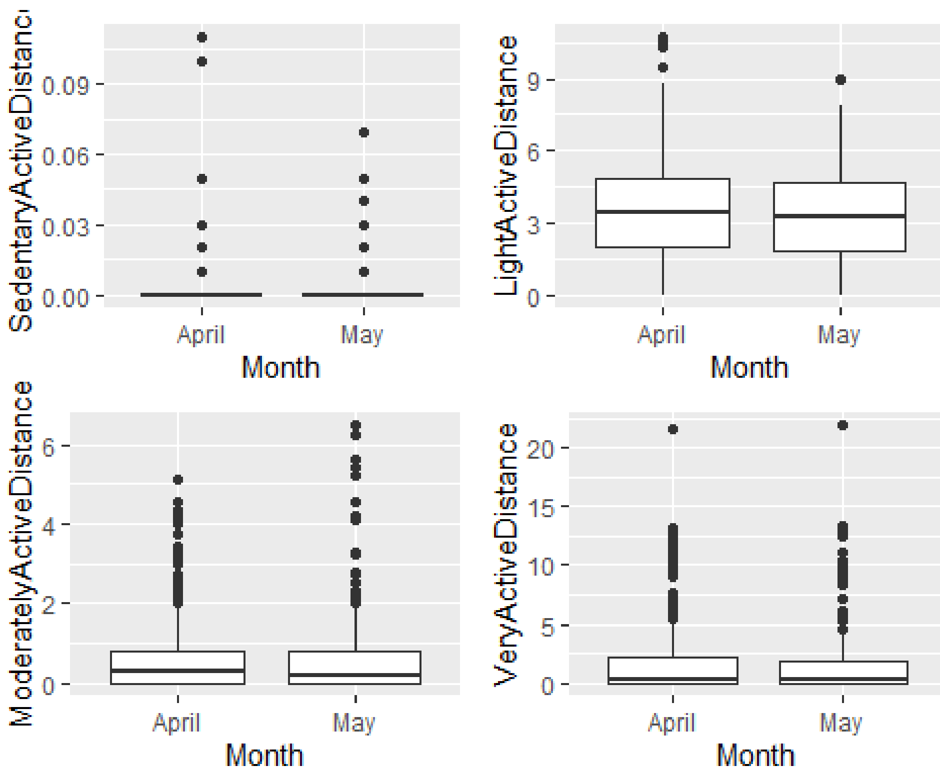
```r
# Distance Analysis
Sedentary_Active_Distance2 <- ggplot(daily_activity,
                                     aes(Month,
                                         SedentaryActiveDistance
                                         )
                                     )+geom_col()
Light_Active_Distance2 <- ggplot(daily_activity,
                                     aes(Month,
                                         LightActiveDistance
                                         )
                                     )+geom_col()
Moderately_Active_Distance2 <- ggplot(daily_activity,
                                     aes(Month,
                                         ModeratelyActiveDistance
```

```
                                         )
                                     )+geom_col()
Very_Active_Distance2 <- ggplot(daily_activity,
                                     aes(Month,
                                         VeryActiveDistance
                                         )
                                     )+geom_col()
grid.arrange(Sedentary_Active_Distance2,Light_Active_Distance2,
             Moderately_Active_Distance2,Very_Active_Distance2,
             ncol = 2)
```




```
# Distance Analysis
Sedentary_Active_Distance <- ggplot(daily_activity,
                                    aes(Month,
                                        SedentaryActiveDistance
                                        )
                                    )+geom_boxplot()

light_Active_Distance <- ggplot(daily_activity,
                                aes(Month,
                                    LightActiveDistance
                                    )
                                )+geom_boxplot()

Moderately_Active_Distance <- ggplot(daily_activity,
                                aes(Month,
                                    ModeratelyActiveDistance
```

```
                                    )
                                )+geom_boxplot()

Very_Active_Distance <- ggplot(daily_activity,
                                     aes(Month,
                                         VeryActiveDistance
                                         )
                                     )+geom_boxplot()
grid.arrange(Sedentary_Active_Distance,light_Active_Distance,
             Moderately_Active_Distance,Very_Active_Distance,
             ncol = 2)
```



Time

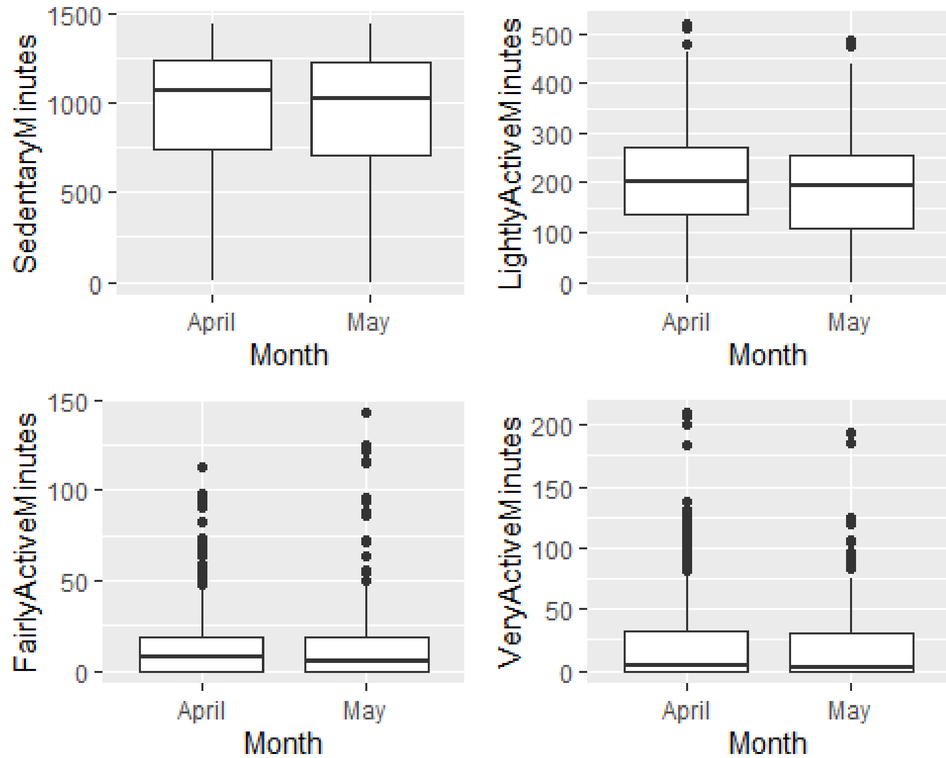
Analysis Findings

Looking at the Time graphs it is clear that there is alot of sedentary time and not alot of Very active time which could mean that the exercise is less intense for most people or doesn't consume alot of time for example a short gym session or a jog in the park.

This is beneficial for the marketing team to know the target audience

```
# Time analysis
Sedentary_Minutes <- ggplot(daily_activity,
                            aes(Month,SedentaryMinutes
                                )
                            )+geom_boxplot()
Lightly_Active_Minutes <- ggplot(daily_activity,
                            aes(Month,LightlyActiveMinutes
                                )
```

```
                            )+geom_boxplot()
Fairly_Active_Minutes <- ggplot(daily_activity,
                            aes(Month,FairlyActiveMinutes
                                )
                            )+geom_boxplot()
Very_Active_Minutes <- ggplot(daily_activity,
                            aes(Month,VeryActiveMinutes
                                )
                            )+geom_boxplot()

grid.arrange(Sedentary_Minutes,Lightly_Active_Minutes,
             Fairly_Active_Minutes,Very_Active_Minutes,
             ncol = 2)
```



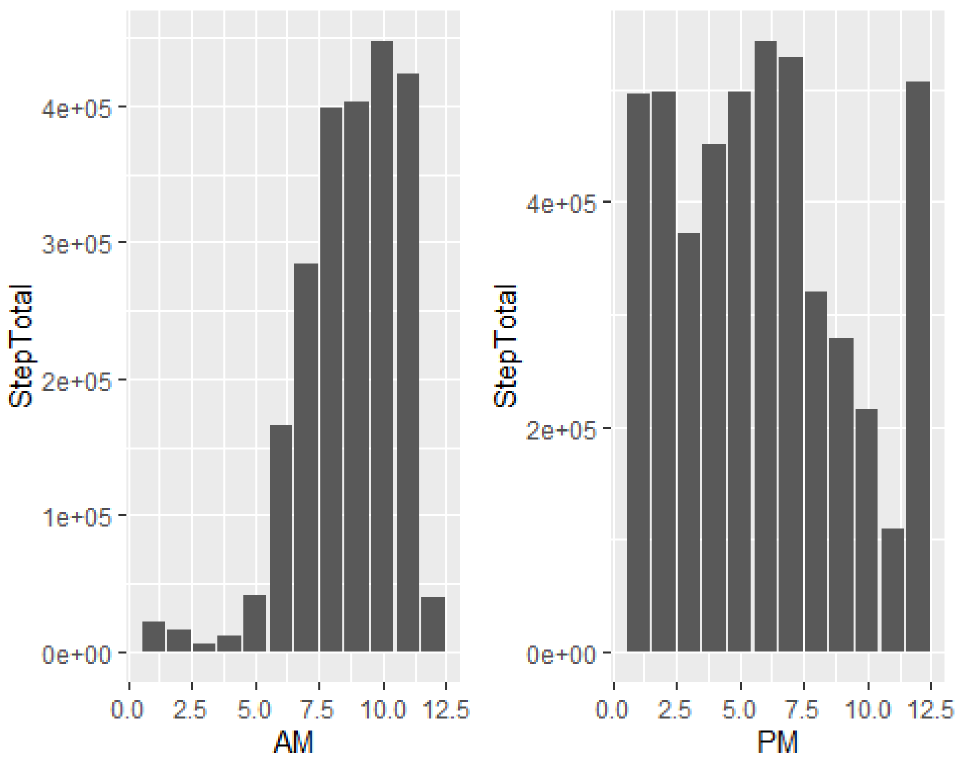
Hours in the Day

It is clear that there are favorite times where most of the steps of the day are taken. In the morning from 7 to 10 roughly and in the evening from 5 to 7 there are spikes in the steps which coincides with the normal 9-5 work day.

Also theirs a spike at close to midnight

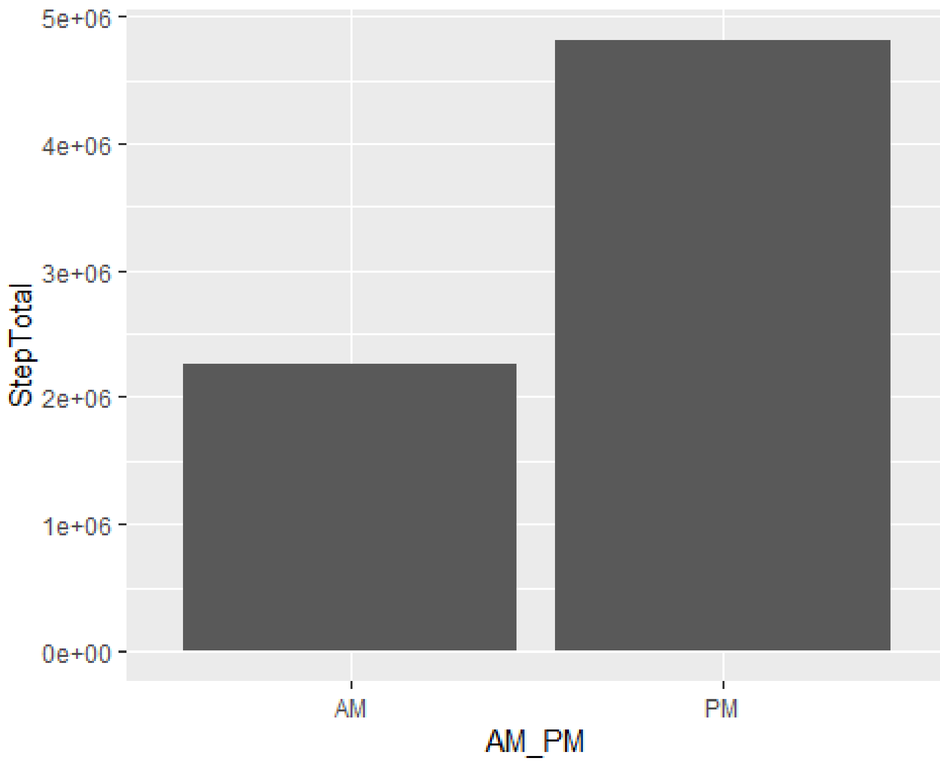
```
##---------------- Hourly dataset ---------------
# Hourly analysis
subset_of_hourly_AM <- subset(hourly_activity,AM_PM == "AM")
subset_of_hourly_PM <- subset(hourly_activity,AM_PM == "PM")
```

```
TimeofDay_AM <- ggplot(subset_of_hourly_AM,
                    aes(Hour,StepTotal
                        )
                    )+geom_col()+
  labs(x="AM")

TimeofDay_PM <- ggplot(subset_of_hourly_PM,
                       aes(Hour,StepTotal
                           )
                       )+geom_col()+
  labs(x="PM")


grid.arrange(TimeofDay_AM,
             TimeofDay_PM,
             ncol = 2)
```



It is clear that most steps are taken in the evening PM time Average amount of steps is 7600~

```
AM_PM <- ggplot(hourly_activity,
                aes(AM_PM,StepTotal
                    )
                )+geom_col()
AM_PM
```




```
mean(daily_activity$TotalSteps)

## [1] 7637.911
```

Recommendations

After analyzing the data and gaining some insight on trends, i can go back and answer the business task questions confidently and with actionable insights

1. What are some trends in smart device usage?

The Trends that stook out to me are the fact that the evening time is when most of the Total steps are accumulated as well as that for the majority of the time the participants are sedentary and most of the time are doing light exercises which could signal that maybe the device is more of a fashion statement for most than a exercise aid.

2. How could these trends apply to Bella beat customers?

These Trends could apply to Bella beat customers by providing a newsletter or weekly email that would show the stats and give information on why they need to be more active for health benefits and reduce sedentary time

3. How could these trends help influence Bella beat marketing strategy?

For the Marketing Strategy the findings could help narrow down the target audience for the campaign to a few factors

- The Target audience does light exercise and is sedentary most of the time
- The exercise timing is mostly in the evenings after work time or before work
- The Average distance is around 3 to 5 KM with steps Averaging 7600~

With that information the marketing team knows the best times for the campaign to air as well as the target audience for the campaign.