Bellabeat

Yusrah

2022-09-20

title: "Bellabeat Case Study" author: "Yusrah" date: "2022-09-17" output: pdf_document

Introduction

I am a junior data analyst working on the marketing analyst team at Bellabeat, a high-tech manufacturer of health-focused products for women. Bellabeat is a successful small company, but they have the potential to become a larger player in the global smart device market.At the end of this analysis I will present my insights and recommendation to the executive team to help them guide marketing strategy for the company.

Bellabeat Products

○ Bellabeat app: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.

○ Leaf: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.

○ Time: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.

○ Spring: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.

○ Bellabeat membership: Bellabeat also offers a subscription-based membership program for users.Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

Ask

I've been asked to analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices.

Prepare

I made use of the FitBit Fitness Tracker Data for this case study and the data set was stored in kaggle. It contained 18 CSV documents.Every user has a unique ID and different rows

since data is tracked by day and time. Each document represents different quantitative data tracked by Fitbit. The data is considered long since each row is one time point per subject, so each subject will have data in multiple rows.

Data integrity

The dataset set contains information about 33 users, this sample size is very small and we can see sampling bias here.

Process

installing packages

install.packages ("ggplot2")

install.packages ("tidyverse")

install.packages("scales")

install.packages("lubridate")

install.packages("janitor")

install.packages("dplyr")

install.packages("here")

install.packages("skimr")

Loading packages

```
library (ggplot2)
library (tidyverse)

## — Attaching packages ——————————————————————————————— tidyverse 1.
3.2 —
## ✓ tibble  3.1.8      ✓ dplyr   1.0.10
## ✓ tidyr   1.2.0      ✓ stringr 1.4.1
## ✓ readr   2.1.2      ✓ forcats 0.5.2
## ✓ purrr   0.3.4
## — Conflicts ————————————————————————————————— tidyverse_conflict
s() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()

library (scales)

##
## Attaching package: 'scales'
##
## The following object is masked from 'package:purrr':
##
##     discard
```

```
## 
## The following object is masked from 'package:readr':
## 
##     col_factor

library (lubridate)

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

library (janitor)

## 
## Attaching package: 'janitor'
## 
## The following objects are masked from 'package:stats':
## 
##     chisq.test, fisher.test

library (dplyr)
library (here)

## here() starts at C:/Users/user/Documents

library (skimr)
```

Importing the datasets
```
daily_activity <- read.csv("dailyActivity_merged.csv")
hourly_steps <- read.csv("hourlySteps_merged.csv")
```

Preview dataset
```
head(daily_activity)

##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    4/12/2016      13162          8.50            8.50
## 2 1503960366    4/13/2016      10735          6.97            6.97
## 3 1503960366    4/14/2016      10460          6.74            6.74
## 4 1503960366    4/15/2016       9762          6.28            6.28
## 5 1503960366    4/16/2016      12669          8.16            8.16
## 6 1503960366    4/17/2016       9705          6.48            6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
```

```
## 1               6.06                      0              25
## 2               4.71                      0              21
## 3               3.91                      0              30
## 4               2.83                      0              29
## 5               5.04                      0              36
## 6               2.51                      0              38
##    FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

head(hourly_steps)

```
##           Id    ActivityHour StepTotal
## 1 1503960366 4/12/2016 0:00       373
## 2 1503960366 4/12/2016 1:00       160
## 3 1503960366 4/12/2016 2:00       151
## 4 1503960366 4/12/2016 3:00         0
## 5 1503960366 4/12/2016 4:00         0
## 6 1503960366 4/12/2016 5:00         0
```

str(daily_activity)

```
## 'data.frame':    940 obs. of  15 variables:
##  $ Id                     : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09
...
##  $ ActivityDate           : chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/1
5/2016" ...
##  $ TotalSteps             : int  13162 10735 10460 9762 12669 9705 13019
15506 10544 9819 ...
##  $ TotalDistance          : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance        : num  8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance     : num  1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance    : num  6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes      : int  25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes    : int  13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes   : int  328 217 181 209 221 164 233 264 205 211
...
##  $ SedentaryMinutes       : int  728 776 1218 726 773 539 1149 775 818 83
8 ...
##  $ Calories               : int  1985 1797 1776 1745 1863 1728 1921 2035
1786 1775 ...
```

str(hourly_steps)

```
## 'data.frame':    22099 obs. of  3 variables:
##  $ Id          : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityHour: chr  "4/12/2016 0:00" "4/12/2016 1:00" "4/12/2016 2:00" "
4/12/2016 3:00" ...
##  $ StepTotal   : int  373 160 151 0 0 0 0 0 250 1864 ...
```

Cleaning

Checking for unique value

```
length(unique(daily_activity$Id))
```

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

```
length(unique(hourly_steps$Id))
```

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

Checking for duplicates

```
sum(duplicated(daily_activity))
```

```
## [1] 0
```

```
sum(duplicated(hourly_steps))
```

```
## [1] 0
```

Making column name consistent

```
clean_names(head(daily_activity))
```

```
##            id activity_date total_steps total_distance tracker_distance
## 1 1503960366     4/12/2016        13162           8.50             8.50
## 2 1503960366     4/13/2016        10735           6.97             6.97
## 3 1503960366     4/14/2016        10460           6.74             6.74
## 4 1503960366     4/15/2016         9762           6.28             6.28
## 5 1503960366     4/16/2016        12669           8.16             8.16
## 6 1503960366     4/17/2016         9705           6.48             6.48
##   logged_activities_distance very_active_distance moderately_active_distan
ce
## 1                          0                 1.88                       0.
55
## 2                          0                 1.57                       0.
69
## 3                          0                 2.44                       0.
40
## 4                          0                 2.14                       1.
26
## 5                          0                 2.71                       0.
41
## 6                          0                 3.19                       0.
78
```

```
##    light_active_distance sedentary_active_distance very_active_minutes
## 1                   6.06                         0                  25
## 2                   4.71                         0                  21
## 3                   3.91                         0                  30
## 4                   2.83                         0                  29
## 5                   5.04                         0                  36
## 6                   2.51                         0                  38
##    fairly_active_minutes lightly_active_minutes sedentary_minutes calories
## 1                     13                    328               728     1985
## 2                     19                    217               776     1797
## 3                     11                    181              1218     1776
## 4                     34                    209               726     1745
## 5                     10                    221               773     1863
## 6                     20                    164               539     1728
```

```
clean_names(head(hourly_steps))
```

```
##            id   activity_hour step_total
## 1 1503960366 4/12/2016 0:00         373
## 2 1503960366 4/12/2016 1:00         160
## 3 1503960366 4/12/2016 2:00         151
## 4 1503960366 4/12/2016 3:00           0
## 5 1503960366 4/12/2016 4:00           0
## 6 1503960366 4/12/2016 5:00           0
```

Formatting to Date data type.

```
daily_activity <- daily_activity %>%
    mutate(ActivityDate = as.Date(ActivityDate, format = "%m/%d/%Y"))
```

```
head(daily_activity)
```

```
##           Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       9705          6.48            6.48
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                4.71                       0                21
## 3                3.91                       0                30
## 4                2.83                       0                29
```

```
## 5                        5.04                              0                       36
## 6                        2.51                              0                       38
##    FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

head(hourly_steps)

```
##            Id    ActivityHour StepTotal
## 1 1503960366 4/12/2016 0:00       373
## 2 1503960366 4/12/2016 1:00       160
## 3 1503960366 4/12/2016 2:00       151
## 4 1503960366 4/12/2016 3:00         0
## 5 1503960366 4/12/2016 4:00         0
## 6 1503960366 4/12/2016 5:00         0
```

head(daily_activity)

```
##            Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366   2016-04-12      13162          8.50            8.50
## 2 1503960366   2016-04-13      10735          6.97            6.97
## 3 1503960366   2016-04-14      10460          6.74            6.74
## 4 1503960366   2016-04-15       9762          6.28            6.28
## 5 1503960366   2016-04-16      12669          8.16            8.16
## 6 1503960366   2016-04-17       9705          6.48            6.48
##    LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.57                     0.69
## 3                        0               2.44                     0.40
## 4                        0               2.14                     1.26
## 5                        0               2.71                     0.41
## 6                        0               3.19                     0.78
##    LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                 6.06                       0                25
## 2                 4.71                       0                21
## 3                 3.91                       0                30
## 4                 2.83                       0                29
## 5                 5.04                       0                36
## 6                 2.51                       0                38
##    FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories
## 1                  13                  328              728     1985
## 2                  19                  217              776     1797
## 3                  11                  181             1218     1776
## 4                  34                  209              726     1745
## 5                  10                  221              773     1863
## 6                  20                  164              539     1728
```

```
head(hourly_steps)

##           Id    ActivityHour StepTotal
## 1 1503960366 4/12/2016 0:00       373
## 2 1503960366 4/12/2016 1:00       160
## 3 1503960366 4/12/2016 2:00       151
## 4 1503960366 4/12/2016 3:00         0
## 5 1503960366 4/12/2016 4:00         0
## 6 1503960366 4/12/2016 5:00         0
```

Merging both datasets on the Id column

```
activities_data <- merge(daily_activity,hourly_steps, by=c("Id"))
n_distinct(activities_data$Id)

## [1] 33

glimpse(activities_data)

## Rows: 654,126
## Columns: 17
## $ Id                      <dbl> 1503960366, 1503960366, 1503960366, 15039
6036…
## $ ActivityDate            <date> 2016-04-12, 2016-04-12, 2016-04-12, 2016
-04-…
## $ TotalSteps              <int> 13162, 13162, 13162, 13162, 13162, 13162,
131…
## $ TotalDistance           <dbl> 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8
.5, …
## $ TrackerDistance         <dbl> 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8.5, 8
.5, …
## $ LoggedActivitiesDistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, …
## $ VeryActiveDistance      <dbl> 1.88, 1.88, 1.88, 1.88, 1.88, 1.88, 1.88,
1.8…
## $ ModeratelyActiveDistance <dbl> 0.55, 0.55, 0.55, 0.55, 0.55, 0.55, 0.55,
0.5…
## $ LightActiveDistance     <dbl> 6.06, 6.06, 6.06, 6.06, 6.06, 6.06, 6.06,
6.0…
## $ SedentaryActiveDistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, …
## $ VeryActiveMinutes       <int> 25, 25, 25, 25, 25, 25, 25, 25, 25, 25, 2
5, 2…
## $ FairlyActiveMinutes     <int> 13, 13, 13, 13, 13, 13, 13, 13, 13, 13, 1
3, 1…
## $ LightlyActiveMinutes    <int> 328, 328, 328, 328, 328, 328, 328, 328, 3
28, …
## $ SedentaryMinutes        <int> 728, 728, 728, 728, 728, 728, 728, 728, 7
28, …
## $ Calories                <int> 1985, 1985, 1985, 1985, 1985, 1985, 1985,
198…
```

```
## $ ActivityHour             <chr> "4/12/2016 0:00", "4/12/2016 1:00", "4/12
/201…
## $ StepTotal                <int> 373, 160, 151, 0, 0, 0, 0, 0, 250, 1864,
676,…
```

Analyze and Share

```
activities_data %>%
  select(TotalSteps ,TotalDistance, Calories , StepTotal ) %>%
  summary()
```

```
##     TotalSteps      TotalDistance        Calories        StepTotal
##  Min.   :    0   Min.   : 0.000   Min.   :   0   Min.   :    0.0
##  1st Qu.: 3761   1st Qu.: 2.600   1st Qu.:1827   1st Qu.:    0.0
##  Median : 7443   Median : 5.280   Median :2156   Median :   41.0
##  Mean   : 7674   Mean   : 5.521   Mean   :2314   Mean   :  321.2
##  3rd Qu.:10771   3rd Qu.: 7.750   3rd Qu.:2800   3rd Qu.:  359.0
##  Max.   :36019   Max.   :28.030   Max.   :4900   Max.   :10554.0
```

1.) What is the relationship between activities and the amount of calories burnt

```
ggplot(data=activities_data) +
  geom_point(mapping =aes(TotalSteps, Calories) ) +labs(title="Relationship b
etween total steps and calories", subtitle="There is an increase in the numbe
r of calories burnt as steps increase.")+theme(plot.title = element_text(size
 = 15),plot.subtitle = element_text(size = 10,color = "green"))
```
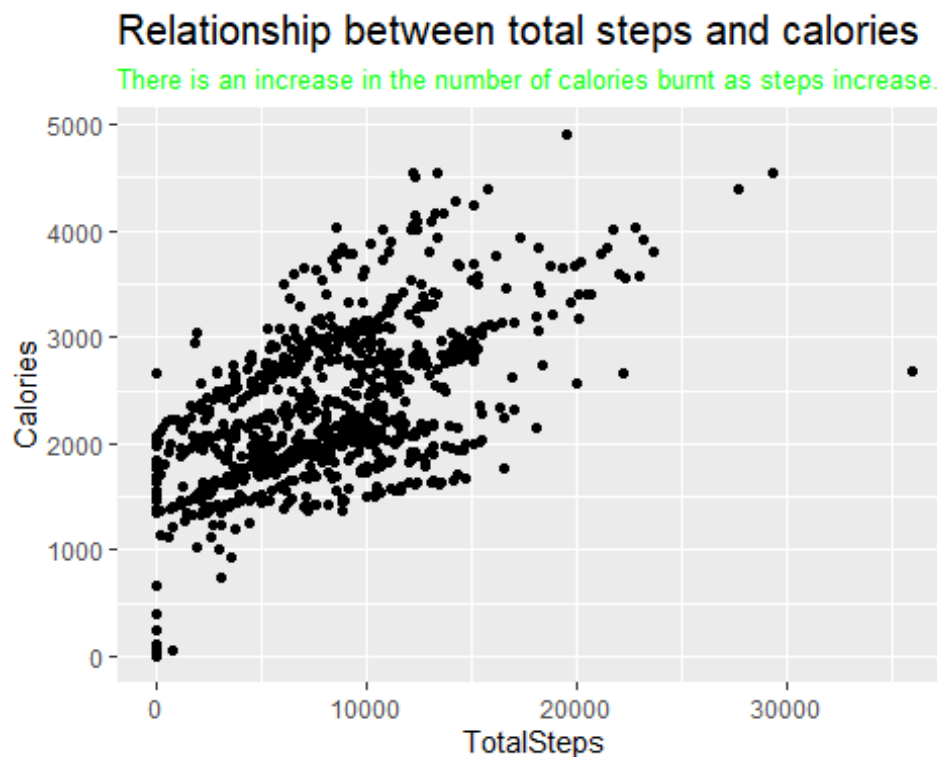

2.) What is the average daily steps for each user

```
daily_average <- daily_activity %>%
  group_by(Id) %>%
  summarise (mean_daily_steps = mean(TotalSteps))

head(daily_average)

## # A tibble: 6 × 2
##            Id mean_daily_steps
##         <dbl>            <dbl>
## 1 1503960366           12117.
## 2 1624580081            5744.
## 3 1644430081            7283.
## 4 1844505072            2580.
## 5 1927972279             916.
## 6 2022484408           11371.
```

3.)How active are the different types of users? I'll be classifying the users into different types: Sedentary - < 5000 steps a day,Lightly active - Between 5000 and 7499 steps a day, Moderately active - Between 7500 and 9999 steps a day, Very active - > 10000 steps a day.

```
user_type <- daily_average %>%
  mutate(user_type = case_when(
    mean_daily_steps < 5000 ~ "sedentary",
    mean_daily_steps >= 5000 & mean_daily_steps < 7499 ~ "lightly active",
    mean_daily_steps >= 7500 & mean_daily_steps < 9999 ~ "moderately active",
    mean_daily_steps >= 10000 ~ "very active"
  ))

head(user_type)

## # A tibble: 6 × 3
##            Id mean_daily_steps user_type
##         <dbl>            <dbl> <chr>
## 1 1503960366           12117. very active
## 2 1624580081            5744. lightly active
## 3 1644430081            7283. lightly active
## 4 1844505072            2580. sedentary
## 5 1927972279             916. sedentary
## 6 2022484408           11371. very active

ggplot(data=user_type) +
  geom_bar(mapping =aes(user_type)) +
  labs(title="User distribution by average daily steps" )
```
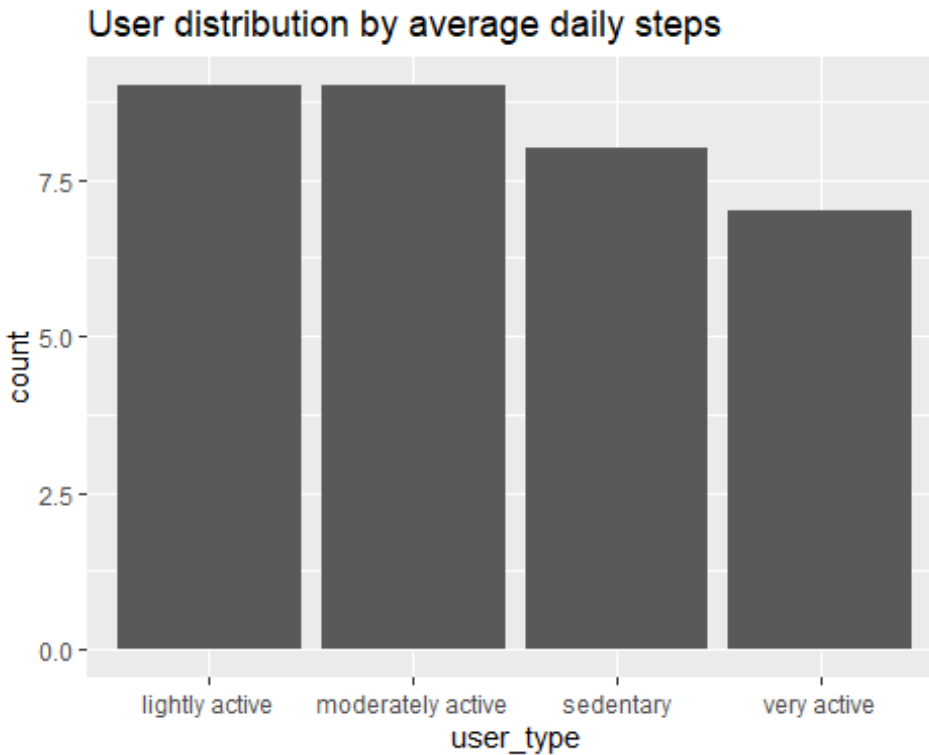
User distribution by average daily steps

4.) What day of the week are the users most active?

```
weekday_steps <- activities_data %>%
  mutate(weekday = weekdays(ActivityDate ))

weekday_steps$weekday <-ordered(weekday_steps$weekday, levels=c("Monday", "Tu
esday", "Wednesday", "Thursday","Friday", "Saturday", "Sunday"))

 weekday_steps <-weekday_steps%>%
  group_by(weekday) %>%
  summarize (mean_daily_steps = mean(TotalSteps))

head(weekday_steps)

## # A tibble: 6 × 2
##   weekday    mean_daily_steps
##   <ord>                 <dbl>
## 1 Monday                7800.
## 2 Tuesday               8198.
## 3 Wednesday             7560.
## 4 Thursday              7459.
## 5 Friday                7536.
## 6 Saturday              8163.

ggplot(weekday_steps) +
      geom_col(aes(x=weekday, y=mean_daily_steps)) +
      geom_hline(yintercept = 7500) +
```
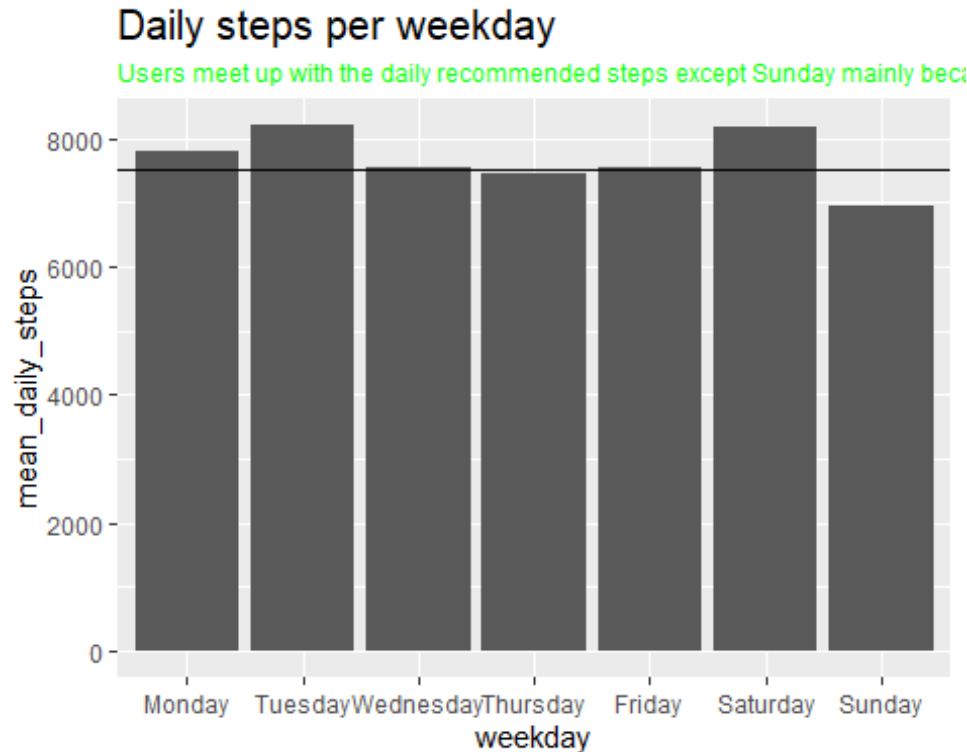
```
        labs(title = "Daily steps per weekday",subtitle = "Users meet up with t
he daily recommended steps except Sunday mainly because it's restday for most
people")+theme(plot.title = element_text(size = 15),plot.subtitle = element_t
ext(size = 9,color = "green"))
```

Daily steps per weekday
Users meet up with the daily recommended steps except Sunday mainly beca

5.) What is the average minutes spent in each active category?

```
average_minutes <- activities_data %>%
  summarise(mean(SedentaryMinutes),mean(FairlyActiveMinutes),mean(LightlyActi
veMinutes),mean(VeryActiveMinutes))
head(average_minutes)

##   mean(SedentaryMinutes) mean(FairlyActiveMinutes) mean(LightlyActiveMinut
es)
## 1               988.3805                   13.5653                    192.5
141
##   mean(VeryActiveMinutes)
## 1                21.53375
```

We can see that the amount of time users spend being sedentary is quite significant, the users spends about 16.5 hrs a day being sedentary, 3.2hrs being lightly active,13.5 minutes being fairly active and 21.5 minutes being very active.

6.) What hour of the day are the users most active?

```
hourly_steps <- activities_data %>%
  separate(ActivityHour, into = c("date", "time"), sep= " ") %>%
  mutate(date = ymd(date))
```

```
## Warning: All formats failed to parse. No formats found.

head(hourly_steps)

##             Id ActivityDate TotalSteps TotalDistance TrackerDistance
## 1 1503960366    2016-04-12      13162           8.5             8.5
## 2 1503960366    2016-04-12      13162           8.5             8.5
## 3 1503960366    2016-04-12      13162           8.5             8.5
## 4 1503960366    2016-04-12      13162           8.5             8.5
## 5 1503960366    2016-04-12      13162           8.5             8.5
## 6 1503960366    2016-04-12      13162           8.5             8.5
##   LoggedActivitiesDistance VeryActiveDistance ModeratelyActiveDistance
## 1                        0               1.88                     0.55
## 2                        0               1.88                     0.55
## 3                        0               1.88                     0.55
## 4                        0               1.88                     0.55
## 5                        0               1.88                     0.55
## 6                        0               1.88                     0.55
##   LightActiveDistance SedentaryActiveDistance VeryActiveMinutes
## 1                6.06                       0                25
## 2                6.06                       0                25
## 3                6.06                       0                25
## 4                6.06                       0                25
## 5                6.06                       0                25
## 6                6.06                       0                25
##   FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes Calories date
time
## 1                  13                  328              728     1985 <NA>
0:00
## 2                  13                  328              728     1985 <NA>
1:00
## 3                  13                  328              728     1985 <NA>
2:00
## 4                  13                  328              728     1985 <NA>
3:00
## 5                  13                  328              728     1985 <NA>
4:00
## 6                  13                  328              728     1985 <NA>
5:00
##   StepTotal
## 1       373
## 2       160
## 3       151
## 4         0
## 5         0
## 6         0

order <- c("0:00","1:00","2:00","3:00","4:00","5:00","6:00","7:00","8:00","9:
00","10:00","11:00","12:00","13:00","14:00","15:00","16:00","17:00","18:00","
19:00","20:00","21:00","22:00","23:00")
```

```
hourly_steps %>%
  group_by(time) %>%
  summarize(average_steps = mean(StepTotal))%>%
  ggplot() +
  geom_col(mapping = aes(factor(time,level=order), average_steps, fill = aver
age_steps)) +labs(title = "Distribution of average hourly steps", x="", y="",
subtitle = "Users most active hours is between 8am - 9pm.")+theme(plot.title
= element_text(size = 15),plot.subtitle = element_text(size = 10,color = "gre
en"))  +
  scale_fill_gradient(low = "orange", high = "green")+
  theme(axis.text.x = element_text(angle = 45))
```
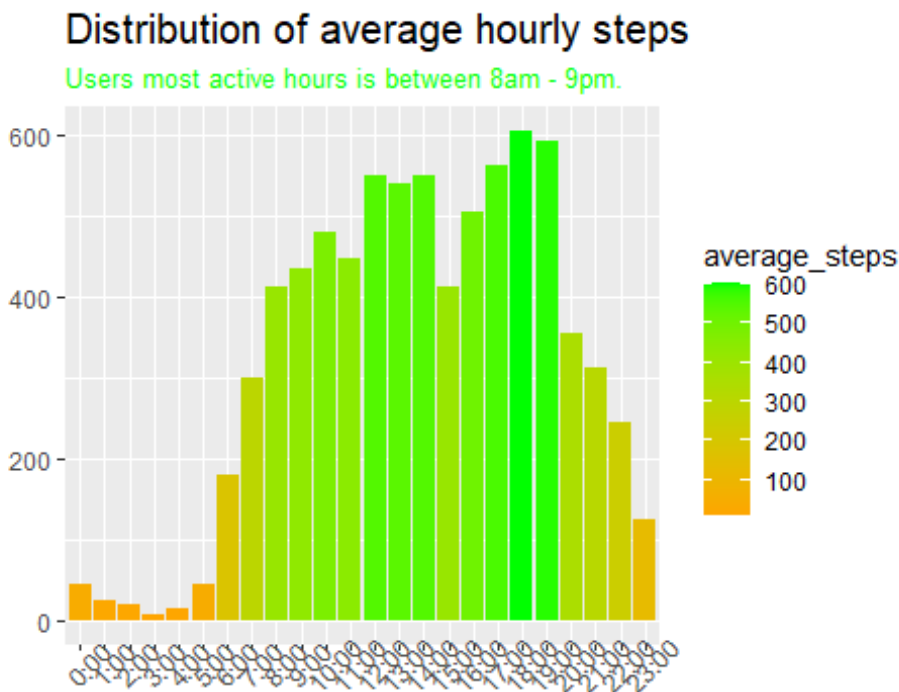

Act

Key Insights

- There is a positive relationship between the number of steps taken and number of calories burnt.

- Users are consistent with their steps during the week,The most active being Saturday and Least active being Sunday.

- Although the average user is very active for over 21 minutes, they still spend 85% of their time being sedentary.

- Users actively start their day by 6am and their most active hours is between 8am - 7pm.

Recommendations

The following recommendations were carefully created to improve bellabeat's marketing strategy:

- Personified notifications : As a result, the CDC recommend that most adults should aim for 10,000 steps per day. Bellabeat to incorporate personified notifications on its app to motivate users to keep moving throughout the day, they should also include number of steps taken and how many more steps that is needed to be taken to get to their daily goals.

- Weekly and Monthly reports : This will keep the users more motivated, bellabeat should provide weekly and monthly report so that the users can know how far they have been going. The app could also send congratulatory messages to those who keep up with their good habits as well as motivational tips to improve incase the user is lagging.

- Discounts : Another way to keep the customers going is to offer special discounts on the different bellabeat's products and their premium membership, especially users that hit daily target at the end of the year.