Bellabeat Case Study

Summary

A hightech company **Bellabeat** that manufactures health focus smart products founded by *Urska Srsen* and *Sando Mur* collecting data on activity, sleep, stress and reproductive health has allowed bellabeat to empowe women with knowledge about their own health and habits since 2013.

Stakeholders

1.Urska Srsen - Cofounder and Chief Creative

2.Sando Mur - Cofounder and Mathematician

3.Bellabeat's Marketing AnalyticsTeam

Ask Phase

Bussiness Task

1.Analyze smart device usage data in order to gain insight into how can women use non-bellabeat smart devices.

2.Select One Product from bellabeat app, leaf, time, spring and apply insights.

Key Questions

1.What are some trends in smart device usage?

2.How could these trends apply to Bellabeat customers?

3.How could these trends help influence Bellabeat marketing strategy?

Prepare Phase

##Data Source - Fitbit Fitness Tracker Data ## About Data

the data set contains personal fitness tracker from thirty user. 30 eligible fitbit user consented to the submission of personal tracker data. including files about minute level output for physical activity, heart rate, sleep monitoring etc.

Survey via - Amazon Mechanical Turk b/w 03/12/2016 to 05/12/2016

Data Limitations

1.Information such as location, lifestyle, weather, tempreture, humidity etc is not provided.

2.Key demographic data such as gender, age were not identified.

3.Small sample size is provided, thirty users is not an ideal sample size.

4.Data Collected in 2016 so it is outdated and it can not represent present trends correctly.

Process Phase

Installing Packages and Opening Libraries

For our analysis we have to install the following packages

1.tidyverse

2.here

3.skimr

4.janitor

5.lubridate

6.ggplot2

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6      v purrr   0.3.5
## v tibble  3.1.8      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.4.1
## v readr   2.1.3      v forcats 0.5.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(here)
```

```
## here() starts at /cloud/project
```

```
library(skimr)
library(janitor)
```

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

```
library(lubridate)
```

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

```
library(ggplot2)
```

Upload Data

now we will upload the following data files for our analysis.

1.dailyActivity_merged.csv

2.sleepDay_merged.csv

- now we will name the variables

```
Daily_activity <- read_csv("/cloud/project/dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
Daily_sleep <- read_csv("/cloud/project/sleepDay_merged.csv")
```

```
## Rows: 413 Columns: 5
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Reviewing our dataset

now we will apply **head** and **str** functions to get few starting rows and structure of our data.

```
head(Daily_activity)
```

```
## # A tibble: 6 x 15
##       Id Activ~1 Total~2 Total~3 Track~4 Logge~5 VeryA~6 Moder~7 Light~8 Seden~9
##    <dbl> <chr>      <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 1.50e9 4/12/2~   13162    8.5     8.5         0    1.88   0.550    6.06       0
## 2 1.50e9 4/13/2~   10735    6.97    6.97        0    1.57   0.690    4.71       0
## 3 1.50e9 4/14/2~   10460    6.74    6.74        0    2.44   0.400    3.91       0
## 4 1.50e9 4/15/2~    9762    6.28    6.28        0    2.14   1.26     2.83       0
## 5 1.50e9 4/16/2~   12669    8.16    8.16        0    2.71   0.410    5.04       0
## 6 1.50e9 4/17/2~    9705    6.48    6.48        0    3.19   0.780    2.51       0
## # ... with 5 more variables: VeryActiveMinutes <dbl>,
## #   FairlyActiveMinutes <dbl>, LightlyActiveMinutes <dbl>,
## #   SedentaryMinutes <dbl>, Calories <dbl>, and abbreviated variable names
## #   1: ActivityDate, 2: TotalSteps, 3: TotalDistance, 4: TrackerDistance,
## #   5: LoggedActivitiesDistance, 6: VeryActiveDistance,
## #   7: ModeratelyActiveDistance, 8: LightActiveDistance,
## #   9: SedentaryActiveDistance
```

```
str(Daily_activity)
```

```
## spc_tbl_ [940 x 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
```

```
##  $ Id                       : num [1:940] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ ActivityDate             : chr [1:940] "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
##  $ TotalSteps               : num [1:940] 13162 10735 10460 9762 12669 ...
##  $ TotalDistance            : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ TrackerDistance          : num [1:940] 8.5 6.97 6.74 6.28 8.16 ...
##  $ LoggedActivitiesDistance: num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveDistance       : num [1:940] 1.88 1.57 2.44 2.14 2.71 ...
##  $ ModeratelyActiveDistance: num [1:940] 0.55 0.69 0.4 1.26 0.41 ...
##  $ LightActiveDistance      : num [1:940] 6.06 4.71 3.91 2.83 5.04 ...
##  $ SedentaryActiveDistance : num [1:940] 0 0 0 0 0 0 0 0 0 0 ...
##  $ VeryActiveMinutes        : num [1:940] 25 21 30 29 36 38 42 50 28 19 ...
##  $ FairlyActiveMinutes      : num [1:940] 13 19 11 34 10 20 16 31 12 8 ...
##  $ LightlyActiveMinutes     : num [1:940] 328 217 181 209 221 164 233 264 205 211 ...
##  $ SedentaryMinutes         : num [1:940] 728 776 1218 726 773 ...
##  $ Calories                 : num [1:940] 1985 1797 1776 1745 1863 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   ActivityDate = col_character(),
##   ..   TotalSteps = col_double(),
##   ..   TotalDistance = col_double(),
##   ..   TrackerDistance = col_double(),
##   ..   LoggedActivitiesDistance = col_double(),
##   ..   VeryActiveDistance = col_double(),
##   ..   ModeratelyActiveDistance = col_double(),
##   ..   LightActiveDistance = col_double(),
##   ..   SedentaryActiveDistance = col_double(),
##   ..   VeryActiveMinutes = col_double(),
##   ..   FairlyActiveMinutes = col_double(),
##   ..   LightlyActiveMinutes = col_double(),
##   ..   SedentaryMinutes = col_double(),
##   ..   Calories = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

```
head(Daily_sleep)
```

```
## # A tibble: 6 x 5
##           Id SleepDay              TotalSleepRecords TotalMinutesAsleep TotalT~1
##        <dbl> <chr>                             <dbl>              <dbl>    <dbl>
## 1 1503960366 4/12/2016 12:00:00 AM                 1                327      346
## 2 1503960366 4/13/2016 12:00:00 AM                 2                384      407
## 3 1503960366 4/15/2016 12:00:00 AM                 1                412      442
## 4 1503960366 4/16/2016 12:00:00 AM                 2                340      367
## 5 1503960366 4/17/2016 12:00:00 AM                 1                700      712
## 6 1503960366 4/19/2016 12:00:00 AM                 1                304      320
## # ... with abbreviated variable name 1: TotalTimeInBed
```

```
str(Daily_sleep)
```

```
## spc_tbl_ [413 x 5] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ Id                : num [1:413] 1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
##  $ SleepDay          : chr [1:413] "4/12/2016 12:00:00 AM" "4/13/2016 12:00:00 AM" "4/15/2016 12:00:0
##  $ TotalSleepRecords : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
##  $ TotalMinutesAsleep: num [1:413] 327 384 412 340 700 304 360 325 361 430 ...
```

```
##  $ TotalTimeInBed    : num [1:413] 346 407 442 367 712 320 377 364 384 449 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   Id = col_double(),
##   ..   SleepDay = col_character(),
##   ..   TotalSleepRecords = col_double(),
##   ..   TotalMinutesAsleep = col_double(),
##   ..   TotalTimeInBed = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Cleaning and Formating

- First of all we will find out number of unique users in our data frame

```
n_unique(Daily_activity$Id)
```

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

```
n_unique(Daily_sleep$Id)
```

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

*now we will look for any duplicate

```
sum(duplicated(Daily_activity))
```

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

```
sum(duplicated(Daily_sleep))
```

```
## [1] 3
```

as we find out that one of our data frame contains 3 duplicate rows now we will drop those duplicate rows

```
Daily_activity <- Daily_activity %>%
  distinct() %>%
  drop_na()

Daily_sleep <- Daily_sleep %>%
  distinct() %>%
  drop_na()
```

now we will check that duplicates have been removed or not

```
sum(duplicated(Daily_activity))
```

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

now we want to merge our data frames so everything should be in same format so we will change the format of column names to lower case

```
clean_names(Daily_activity)
```

```
## # A tibble: 940 x 15
##            id activity~1 total~2 total~3 track~4 logge~5 very_~6 moder~7 light~8
##         <dbl> <chr>        <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
##  1 1503960366 4/12/2016    13162    8.5     8.5        0    1.88   0.550    6.06
##  2 1503960366 4/13/2016    10735    6.97    6.97       0    1.57   0.690    4.71
##  3 1503960366 4/14/2016    10460    6.74    6.74       0    2.44   0.400    3.91
```

```
##  4 1503960366 4/15/2016     9762    6.28    6.28       0    2.14   1.26     2.83
##  5 1503960366 4/16/2016    12669    8.16    8.16       0    2.71   0.410    5.04
##  6 1503960366 4/17/2016     9705    6.48    6.48       0    3.19   0.780    2.51
##  7 1503960366 4/18/2016    13019    8.59    8.59       0    3.25   0.640    4.71
##  8 1503960366 4/19/2016    15506    9.88    9.88       0    3.53   1.32     5.03
##  9 1503960366 4/20/2016    10544    6.68    6.68       0    1.96   0.480    4.24
## 10 1503960366 4/21/2016     9819    6.34    6.34       0    1.34   0.350    4.65
## # ... with 930 more rows, 6 more variables: sedentary_active_distance <dbl>,
## #   very_active_minutes <dbl>, fairly_active_minutes <dbl>,
## #   lightly_active_minutes <dbl>, sedentary_minutes <dbl>, calories <dbl>, and
## #   abbreviated variable names 1: activity_date, 2: total_steps,
## #   3: total_distance, 4: tracker_distance, 5: logged_activities_distance,
## #   6: very_active_distance, 7: moderately_active_distance,
## #   8: light_active_distance
```

```
Daily_activity <- rename_with(Daily_activity, tolower)

clean_names(Daily_sleep)
```

```
## # A tibble: 410 x 5
##            id sleep_day             total_sleep_records total_minutes_~1 total~2
##         <dbl> <chr>                               <dbl>            <dbl>   <dbl>
##  1 1503960366 4/12/2016 12:00:00 AM                   1              327     346
##  2 1503960366 4/13/2016 12:00:00 AM                   2              384     407
##  3 1503960366 4/15/2016 12:00:00 AM                   1              412     442
##  4 1503960366 4/16/2016 12:00:00 AM                   2              340     367
##  5 1503960366 4/17/2016 12:00:00 AM                   1              700     712
##  6 1503960366 4/19/2016 12:00:00 AM                   1              304     320
##  7 1503960366 4/20/2016 12:00:00 AM                   1              360     377
##  8 1503960366 4/21/2016 12:00:00 AM                   1              325     364
##  9 1503960366 4/23/2016 12:00:00 AM                   1              361     384
## 10 1503960366 4/24/2016 12:00:00 AM                   1              430     449
## # ... with 400 more rows, and abbreviated variable names
## #   1: total_minutes_asleep, 2: total_time_in_bed
```

```
Daily_sleep <- rename_with(Daily_sleep, tolower)
```

as we know that we want to merge data frames so we will clean the date formats in both data frame.

```
Daily_activity <- Daily_activity %>%
  rename(date = activitydate) %>%
   mutate(date = as_date(date, format = "%m/%d/%Y"))

Daily_sleep <- Daily_sleep %>%
  rename(date = sleepday) %>%
  mutate(date = as_date(date, format = "%m/%d/%Y %I:%M:%S %p" , tz=Sys.timezone()))
```

```
## Warning: `tz` argument is ignored by `as_date()`
```

now we will check our clean data frames

```
head(Daily_activity)
```

```
## # A tibble: 6 x 15
##           id date       totals~1 total~2 track~3 logge~4 verya~5 moder~6 light~7
##        <dbl> <date>        <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>   <dbl>
## 1 1503960366 2016-04-12    13162    8.5     8.5        0    1.88   0.550    6.06
```

```
## 2 1503960366 2016-04-13    10735    6.97    6.97       0    1.57   0.690    4.71
## 3 1503960366 2016-04-14    10460    6.74    6.74       0    2.44   0.400    3.91
## 4 1503960366 2016-04-15     9762    6.28    6.28       0    2.14   1.26     2.83
## 5 1503960366 2016-04-16    12669    8.16    8.16       0    2.71   0.410    5.04
## 6 1503960366 2016-04-17     9705    6.48    6.48       0    3.19   0.780    2.51
## # ... with 6 more variables: sedentaryactivedistance <dbl>,
## #   veryactiveminutes <dbl>, fairlyactiveminutes <dbl>,
## #   lightlyactiveminutes <dbl>, sedentaryminutes <dbl>, calories <dbl>, and
## #   abbreviated variable names 1: totalsteps, 2: totaldistance,
## #   3: trackerdistance, 4: loggedactivitiesdistance, 5: veryactivedistance,
## #   6: moderatelyactivedistance, 7: lightactivedistance
```

```
head(Daily_sleep)
```

```
## # A tibble: 6 x 5
##           id date       totalsleeprecords totalminutesasleep totaltimeinbed
##        <dbl> <date>                 <dbl>              <dbl>           <dbl>
## 1 1503960366 2016-04-12                 1                327             346
## 2 1503960366 2016-04-13                 2                384             407
## 3 1503960366 2016-04-15                 1                412             442
## 4 1503960366 2016-04-16                 2                340             367
## 5 1503960366 2016-04-17                 1                700             712
## 6 1503960366 2016-04-19                 1                304             320
```

Merging Datasets

now we will merge both data sets

```
Daily_activity_sleep <- merge(Daily_activity, Daily_sleep, by=c("id", "date"))
glimpse(Daily_activity_sleep)
```

```
## Rows: 410
## Columns: 18
## $ id                       <dbl> 1503960366, 1503960366, 1503960366, 150396036~
## $ date                     <date> 2016-04-12, 2016-04-13, 2016-04-15, 2016-04-~
## $ totalsteps               <dbl> 13162, 10735, 9762, 12669, 9705, 15506, 10544~
## $ totaldistance            <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6.3~
## $ trackerdistance          <dbl> 8.50, 6.97, 6.28, 8.16, 6.48, 9.88, 6.68, 6.3~
## $ loggedactivitiesdistance <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ veryactivedistance       <dbl> 1.88, 1.57, 2.14, 2.71, 3.19, 3.53, 1.96, 1.3~
## $ moderatelyactivedistance <dbl> 0.55, 0.69, 1.26, 0.41, 0.78, 1.32, 0.48, 0.3~
## $ lightactivedistance      <dbl> 6.06, 4.71, 2.83, 5.04, 2.51, 5.03, 4.24, 4.6~
## $ sedentaryactivedistance  <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
## $ veryactiveminutes        <dbl> 25, 21, 29, 36, 38, 50, 28, 19, 41, 39, 73, 3~
## $ fairlyactiveminutes      <dbl> 13, 19, 34, 10, 20, 31, 12, 8, 21, 5, 14, 23,~
## $ lightlyactiveminutes     <dbl> 328, 217, 209, 221, 164, 264, 205, 211, 262, ~
## $ sedentaryminutes         <dbl> 728, 776, 726, 773, 539, 775, 818, 838, 732, ~
## $ calories                 <dbl> 1985, 1797, 1745, 1863, 1728, 2035, 1786, 177~
## $ totalsleeprecords        <dbl> 1, 2, 1, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, ~
## $ totalminutesasleep       <dbl> 327, 384, 412, 340, 700, 304, 360, 325, 361, ~
## $ totaltimeinbed           <dbl> 346, 407, 442, 367, 712, 320, 377, 364, 384, ~
```

Analyze phase

in this phase we will analyze our data and will find trends and patterns in our results which can help to improve Bellabeat's Marketing strategies. first of all we will find the summary of our merged data.

```
summary(Daily_activity_sleep)
```

```
##        id                 date              totalsteps    totaldistance
##  Min.   :1.504e+09   Min.   :2016-04-12   Min.   :   17   Min.   : 0.010
##  1st Qu.:3.977e+09   1st Qu.:2016-04-19   1st Qu.: 5189   1st Qu.: 3.592
##  Median :4.703e+09   Median :2016-04-27   Median : 8913   Median : 6.270
##  Mean   :4.995e+09   Mean   :2016-04-26   Mean   : 8515   Mean   : 6.012
##  3rd Qu.:6.962e+09   3rd Qu.:2016-05-04   3rd Qu.:11370   3rd Qu.: 8.005
##  Max.   :8.792e+09   Max.   :2016-05-12   Max.   :22770   Max.   :17.540
##  trackerdistance  loggedactivitiesdistance veryactivedistance
##  Min.   : 0.010   Min.   :0.0000          Min.   : 0.000
##  1st Qu.: 3.592   1st Qu.:0.0000          1st Qu.: 0.000
##  Median : 6.270   Median :0.0000          Median : 0.570
##  Mean   : 6.007   Mean   :0.1089          Mean   : 1.446
##  3rd Qu.: 7.950   3rd Qu.:0.0000          3rd Qu.: 2.360
##  Max.   :17.540   Max.   :4.0817          Max.   :12.540
##  moderatelyactivedistance lightactivedistance sedentaryactivedistance
##  Min.   :0.0000           Min.   :0.010       Min.   :0.0000000
##  1st Qu.:0.0000           1st Qu.:2.540       1st Qu.:0.0000000
##  Median :0.4200           Median :3.665       Median :0.0000000
##  Mean   :0.7439           Mean   :3.791       Mean   :0.0009268
##  3rd Qu.:1.0375           3rd Qu.:4.918       3rd Qu.:0.0000000
##  Max.   :6.4800           Max.   :9.480       Max.   :0.1100000
##  veryactiveminutes fairlyactiveminutes lightlyactiveminutes sedentaryminutes
##  Min.   :  0.00    Min.   :  0.00      Min.   :  2.0        Min.   :   0.0
##  1st Qu.:  0.00    1st Qu.:  0.00      1st Qu.:158.0        1st Qu.: 631.2
##  Median :  9.00    Median : 11.00      Median :208.0        Median : 717.0
##  Mean   : 25.05    Mean   : 17.92      Mean   :216.5        Mean   : 712.1
##  3rd Qu.: 38.00    3rd Qu.: 26.75      3rd Qu.:263.0        3rd Qu.: 782.8
##  Max.   :210.00    Max.   :143.00      Max.   :518.0        Max.   :1265.0
##     calories    totalsleeprecords totalminutesasleep totaltimeinbed
##  Min.   : 257   Min.   :1.00      Min.   : 58.0      Min.   : 61.0
##  1st Qu.:1841   1st Qu.:1.00      1st Qu.:361.0      1st Qu.:403.8
##  Median :2207   Median :1.00      Median :432.5      Median :463.0
##  Mean   :2389   Mean   :1.12      Mean   :419.2      Mean   :458.5
##  3rd Qu.:2920   3rd Qu.:1.00      3rd Qu.:490.0      3rd Qu.:526.0
##  Max.   :4900   Max.   :3.00      Max.   :796.0      Max.   :961.0
```

Some intresting discoveries from this summary

1.on an average people are sleeping 7 hours.

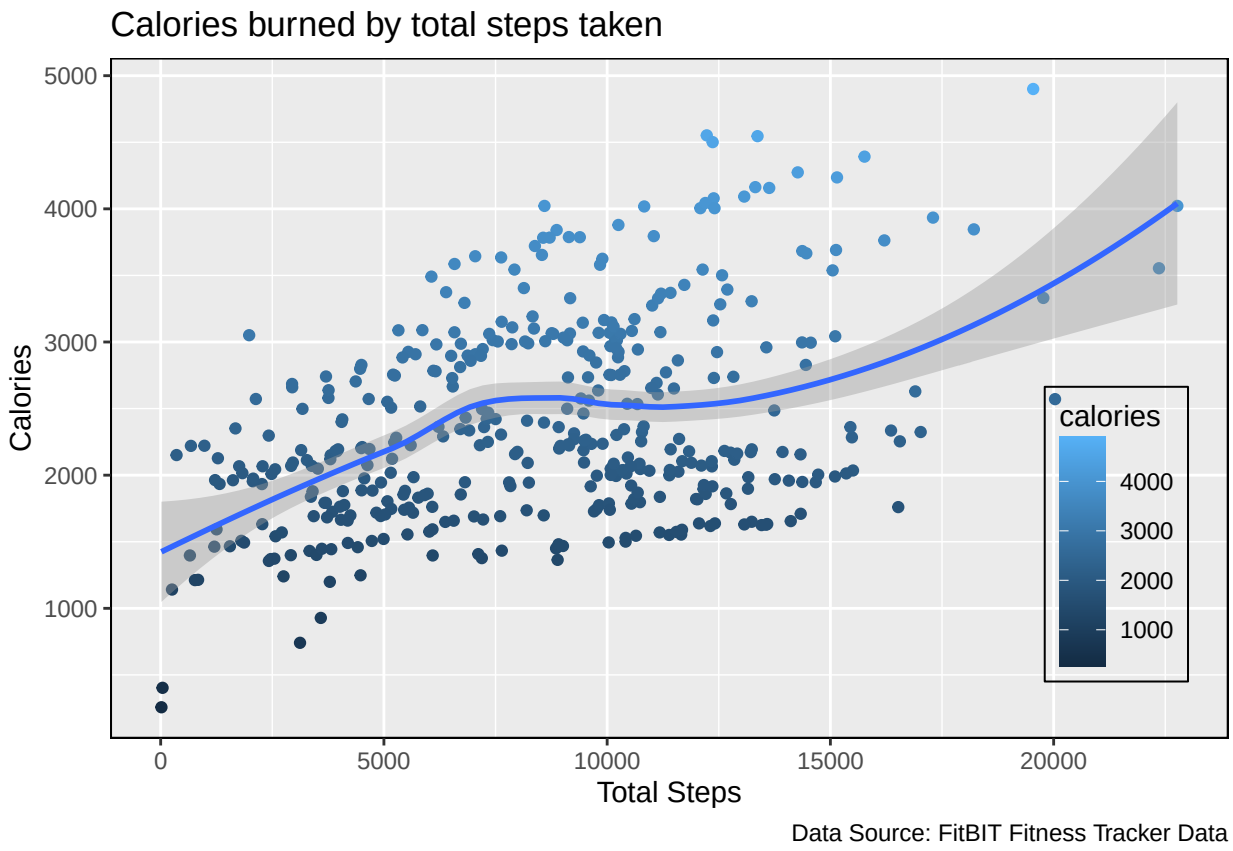
2. most of the people are lightly active.
3. The average steps per day are around 8500.

Calories burned by steps

in this analysis we will find the corelation between steps and calories burned by this we will find out how someone steps can affect their calories burn.

```
Daily_activity_sleep %>%
  group_by(totalsteps, calories) %>%
  ggplot(aes(x = totalsteps, y =calories, color = calories)) +
  geom_point() +
  geom_smooth() +
  theme(legend.position = c(.9, .3),
        legend.spacing.y = unit(1, "mm"),
        panel.border = element_rect(colour = "black",fill=NA),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black")) +
  labs(title = 'Calories burned by total steps taken',
       y = 'Calories',
       x = 'Total Steps',
       caption = 'Data Source: FitBIT Fitness Tracker Data')
```

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



by graph we can clearly see that their is a positive corelation between *Calories* and *Totalsteps* which means that the more steps taken in a day the more calories burned.

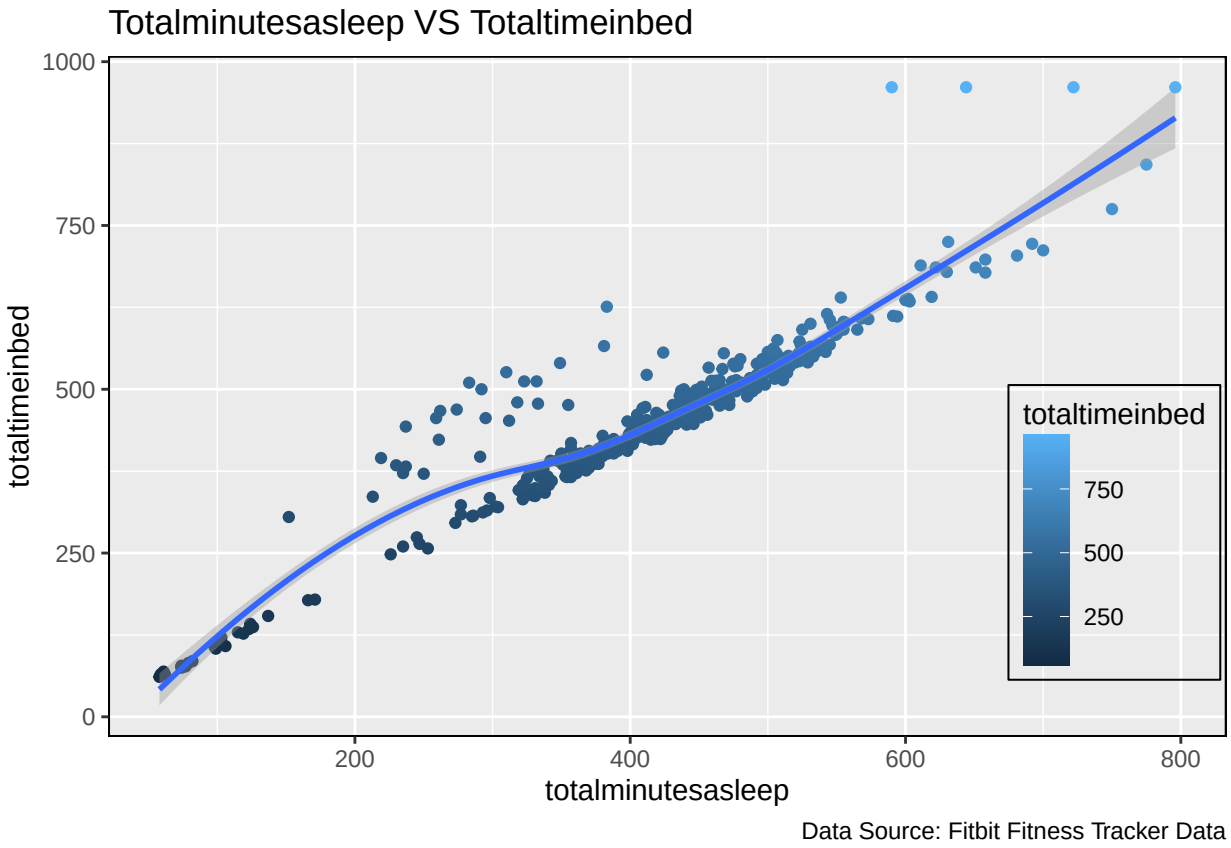
Totalminutesasleep vs Totaltimeinbed

in this analysis we will find out that how much time people are asleep with respect to how much time they are spending on bed.

```
Daily_activity_sleep %>%
  group_by(totalminutesasleep, totaltimeinbed) %>%
```

```
  ggplot(aes(x = totalminutesasleep, y = totaltimeinbed, color = totaltimeinbed)) +
  geom_point() +
  geom_smooth() +
  theme(legend.position = c(.9, .3),
        legend.spacing.y = unit(1, "mm"),
        panel.border = element_rect(colour = "black", fill='NA'),
        legend.background = element_blank(),
        legend.box.background = element_rect(colour = "black")) +
  labs(title = 'Totalminutesasleep VS Totaltimeinbed',
       y = 'totaltimeinbed',
       x = 'totalminutesasleep',
       caption = 'Data Source: Fitbit Fitness Tracker Data')
```

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



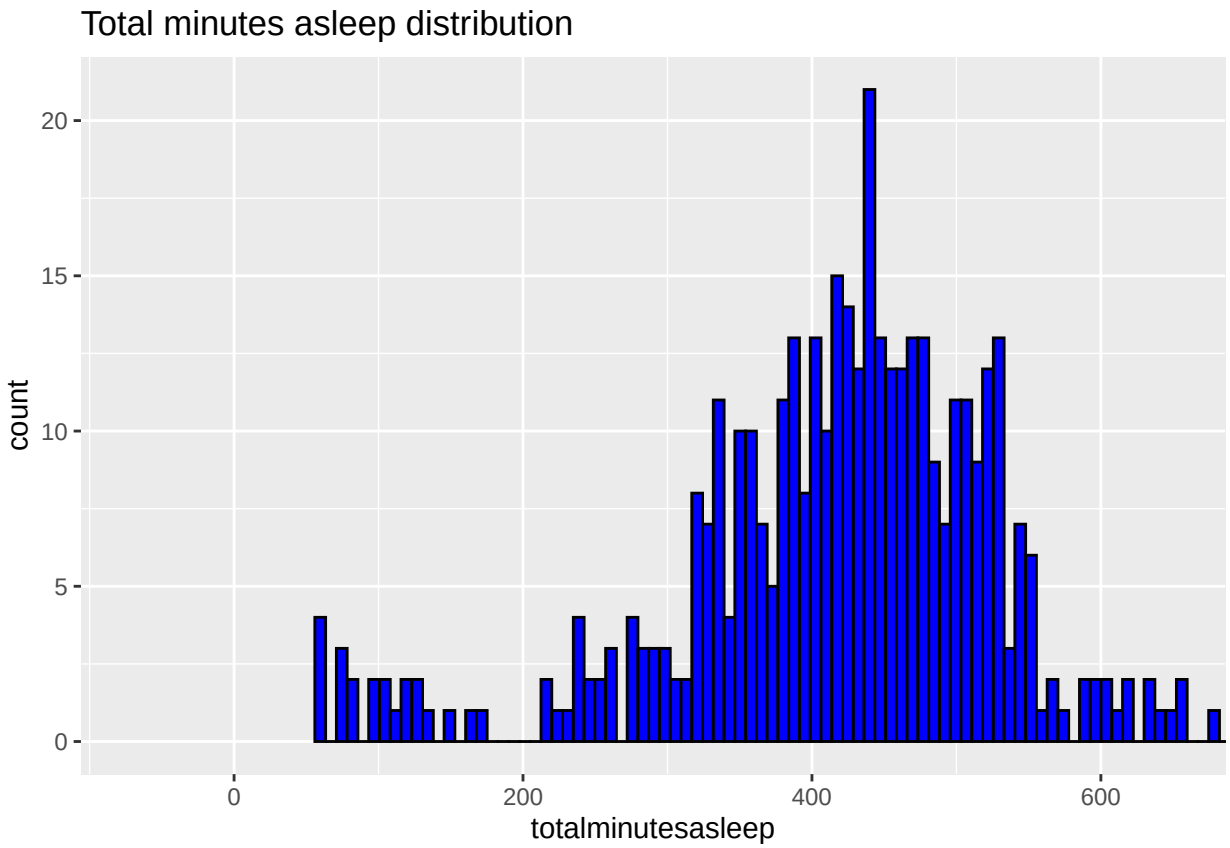
as we can see the graph shows the positive corelation between **totalminutesasleep** and **totaltimeinbed** and most of the points lies on the line graph which means that most of the people are not spending alot of their time in bed without sleeping but their are few people who are spending more time on bed than sleeping.

Totalactiveminutes VS Calories

in this analysis we will find how our active time affects our calories burn first of all we will introduce a new column called **totalactivityminutes** by using **mutate** function . to creat this new column we will add all the activeminutes columns.

```
ggplot(data = Daily_activity_sleep, mapping = aes(x = totalminutesasleep)) +
  geom_histogram(bins = 100, color = "black ", fill = "blue")  +
```

```
coord_cartesian(xlim = c(-70, 650)) +ggtitle("Total minutes asleep distribution")
```



in this histogram we can see that most of the dataset values lies between 300 to 600 minutes so most of the people are sleeping around this time on an average.

Conclusion

After analyzing FitBit fitness tracker data, I found some insights that would help influence **Bellabeat** marketing strategy.

Target Audience

Women who are doing full time job and spending alot of time working and sitting in one place and doing very light activity to maintain fitness, we can help them to improve their health habits by using **Bellabeat App**. after our analysis we have found different trends that may help our online campaign and improve **Bellabeat App**.

Discussion and Recommendations

1. By looking at the graph **Calories burned by steps** we can clearly conclude that the more steps taken in a day the more calories burned and by the research done by University of Massachusetts(article published by Hindustantimes,link of the article is given below) https://www.hindustantimes.com/lifestyle/health/walking-more-steps-a-day-can-improve-people-s-health-longevity-study-101646747500624.html in that research it is concluded that "More steps per day are better for your health and the benefit in terms of mortality risk levels of around 6000 to 8000 for older adults and 8000 to 10000 for younger adults"

Now to improve **Bellabeat's Customers** experience with **Bellabeat App** I would recommend that **Bellabeat** should include age feature in **Bellabeat App** as well and according to customers age the "daily target of steps" (which is 6000 to 8000 for older adults and 8000 to 100000 for younger adults) reminder should be send to customers and during whole day mini reminders of how much target they have achieved should be send, this will motivate them to complete their target.

2.By our second graph which is **Totalminutessleep vs Totaltimeinbed** we can conclude that most of the people are not spending their time on bed other than sleeping, which is good but there are people who are spending more time in bed other than sleeping and they are sleeping less than 8 hours or more than 8 hours. and by our third graph which is a **Histogram** of totalsleepminutes we can conclude that most of the people are sleeping on an average of 6 hours.

To improve **Bellabeat's Customers** sleeping habits I would recommend that **Bellabeat App** should include **Sleep time alarm** feature so users could set up a desired time to go to sleep and receive a notification minutes before to prepare to sleep. and after waking up they should receive a notification of for how much time they slept and on weekly basis user should receive a notification about how much time on an average they are spending on bed without sleeping.

3.Other than notifications **Bellabeat App** should recommend good relaxing music, podcasts about the importance of fitness.

4.App should sent a notification about the improvement in their fitness habits so they can evaluate their progress.