BellaBeat_Case Study

Shaima Saleh

1. Introduction:

1.1. Company Background: Bellabeat is a high-tech wellness company founded in 2013 by Urška Sršen and Sando Mur. The company focuses on creating smart, health-focused products designed specifically for women. Drawing on Sršen's background as an artist, Bellabeat combines elegant design with technology to help women monitor their activity, sleep, stress, and reproductive health.

Since its founding, Bellabeat has expanded rapidly, establishing global offices and launching several products. These products are sold through various online retailers as well as their official website. The company emphasizes digital marketing, including Google Search, social media platforms like Facebook, Instagram, and Twitter, and video and display ads on YouTube and the Google Display Network.

To support further growth, Bellabeat is analyzing usage data from its smart devices to better understand consumer behavior and enhance its marketing strategies.

1.2. Company products: Bellabeat had expanded internationally and introduced several new products. These products became available through a variety of online platforms, along with direct sales via the company's own website.These products are:

1. Bellabeat App: This central app collects and displays data related to users' physical activity, sleep patterns, stress levels, menstrual cycles, and mindfulness routines. It helps users gain deeper insights into their health habits and supports informed decision-making. The app syncs with Bellabeat's smart wellness devices.
2. Leaf: A signature wellness tracker designed to be worn as a bracelet, necklace, or clip. It monitors activity, sleep, and stress, and connects with the Bellabeat app for detailed health tracking.
3. Time: A stylish smartwatch that blends traditional design with smart features. It tracks activity, sleep, and stress, and integrates with the Bellabeat app to deliver personalized wellness data.
4. Spring: A smart water bottle that monitors daily hydration. It uses sensor technology to track water intake and syncs with the Bellabeat app to help users stay properly hydrated.
5. Bellabeat Membership: A subscription service that offers users continuous, personalized wellness support. Members receive tailored advice on nutrition, physical activity, sleep, beauty, and mindfulness based on their individual habits and health goals.

2. The purpose:

The primary purpose of company through this case study is to better understand consumer behavior and enhance its marketing strategies for supporting further growth and to become a larger player in the global smart device market.

3. Business task:

To analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices.

4. Data analysis phases:

1. Ask phase: According to the business task provided - To analyze smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices - there are some questions that can be asked to understand and address the problem we are trying to solve. These questions are:

1. What are smart devices used for the most?
2. How could the company serve its costumers by understanding how they use their smart devices?

3.How effectively could the company utilize the insights on smart devices usage to develop its market strategies?

The stackholders and executive team encompass of:

1. Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
2. Sando Mur: Mathematician and Bellabeat's cofounder; key member of the Bellabeat executive team
3. Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporing data that helps guide Bellabeat's marketing strategy.

Staclholders and executive team expect to be provided with insights on how they could achieve their goal of enhancing the company's growth and expand the business globally by examining smart devices usages and how they could improve marketing strategies that fulfill that goal. Now, we can move to the second phase (prepare).

2. Prepare phase: In the Prepare phase, I used Fitbit Fitness Tracker dataset, This dataset have some characteristics as following:

1. Publicly available through Mobius in Kaggle and licensed under the CCO Public Domain.on Kaggle.
2. Were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016–05.12.2016 and Contains data from 30 Fitbit users collected via Fitbit devices.
3. Includes daily activity, heart rate, sleep monitoring, and other health metrics.
4. Consists of 18 CSV files covering steps, sleep, calories, heart rate, and weight metrics.however, not all 18 csv files will be used in the analysis. The description of the key data files that will be used for the analysis is shown in the table below:

File Name
dailyActivity_merged.csv
sleepDay_merged.csv
dailySteps_merged.csv
hourlyCalories_merged.csv

File Name
hourlyIntensities_merged.csv
heart_rate_second.csv
minuteStepsNarrow_merged.csv
minuteMETsNarrow_merged.csv
minuteSleep_merged.csv
minuteIntensitiesNarrow_merged.csv

Once the dataset were downloaded , I saved them in a file named case-study_04.12.16-05.12.16_v02. I inspected the data and found that the data is in a long format as every IDs has different records in multiple rows. The data has been detected for bias and credibility by examining the features of good data that is, ROCCC which stand for Reliability,Originality, Comprehensive,current, Cited . I found that the data is reliable as it is obtained from a reliable and approval source (Kaggle), regarding the originality of the data, as the data is collected by third party and was published in a formal website, the data has the originality feature too. The dataset has all the data that we need to answer the business's problem, thus the data is comprehensive. As it was mentioned earlier about the source of the data and how it is obtain, data was cited.However, the data is outdated, it lost the current feature.For all above, the data can be consider as a good data and it is ROCCC.

The problems with the data can be mentioned as following:

1. The data is not complete as there were many missing values that make the data after filtering to be smaller than 30 user. This can affect the sample size and its credibility.
2. The data was collecting in 2016 and this makes it outdated data which also might affect its credibility.
3. The is no metrics that we can use to determine if the sample is representative or not.

Despite these limitations, the available data after preparing could be used to answer the business task and ready for process phase.

3. Process: For processing data, I have used spreadsheet to check the integrity of daily and hourly data and clean it as the amount of data is small. However, Minute_and second data is large, so i used R for cleaning process. I have chose the data that will be analyized as mentioned in the table in Prepare Phase: daily steps, daily sleep, daily calories, daily intensities, hourly steps,hourly intensities,hourly calories, minute_calories, minute_sleep, minute_steps,minute_intensities and minute_MENs.

The process steps using spreadsheet:

1. I have taken a look and check the csv files then found that some files have data that were already included in the daily_activity file. So, i used Vlookup function to recall the data from daily_calories, daily_intensities, and daily_steps files to the daily_activity file and make sure that the Vlookup function outcomes and the data which already exists in the daily_activity file are the same to ensure the consistency and compatibility of the data which exists in different files.
2. I populated the sleep_day data in the daily_activity file by using Vlookup function.By doing step 1 and 2 I got one file that contains comprehensively a daily data for all the activities, steps,sleep,intensities,and calories, this will make the analysis more easier.
3. I have check duplicate data for all the data that will be used in the analysis. I found a duplicate value in SleepDay data and i remedied them by removing 3 duplicate values from 414 values, remaining with 411 unique values.

4. I have reformatted all the data by choosing unified date/time format.
5. After processing heart_rate data per second using R and converted the data from data per second to daily data, I have incorporate the heart_rate daily data to the previous file of daily activity data,so that I have a complete daily data for steps, sleep, intensities, calories and heart_rate.

All the above processing steps were done for the daily and hourly data using Google_Sheet tool.

The process steps using R :

1. I have imported the csv files of *minute_calories*, *minute_intensities*, *minute_sleep*, *minute_steps*, *minute_METs* and *heart_rate* data to R by using the read.csv function

```
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/minuteCaloriesNarrow_merged.csv"
)
calories_data_per_minute<-read.csv(
   long_path,
   header = TRUE, stringsAsFactors = FALSE
)
```

```
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/minuteIntensitiesNarrow_merged.csv"
)
intensities_data_per_minute<-read.csv(
   long_path,
  header = TRUE, stringsAsFactors = FALSE
)
```

```
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/minuteMETsNarrow_merged.csv"
)
met_data_per_minute<-read.csv(
  long_path,
  header = TRUE, stringsAsFactors = FALSE
)
```

```
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/minuteSleep_merged.csv"
)
sleep_data_per_minute<-read.csv(
 long_path,
 header = TRUE, stringsAsFactors = FALSE
)
```

```r
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/minuteStepsNarrow_merged.csv"
)
steps_data_per_minute<-read.csv(
 long_path,
 header = TRUE,
 stringsAsFactors = FALSE
)
```

```r
long_path <- paste0(
  "/Users/shaimaalradai/Downloads/",
  "case_study2_03.12.16-04.12.16_v01/",
  "minutes_data/heartrate_seconds_merged.csv"
)

heart_rate_data_per_second <- read.csv(
  long_path,
  header = TRUE,
  stringsAsFactors = FALSE
)
```

2. I have checked for duplicated values by using **sum(duplicated())** function

```r
sum(duplicated(calories_data_per_minute))
```

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

```r
sum(duplicated(intensities_data_per_minute))
```

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

```r
sum(duplicated(met_data_per_minute))
```

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

```r
sum(duplicated(sleep_data_per_minute))
```

```
## [1] 543
```

```r
sum(duplicated(steps_data_per_minute))
```

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

```r
sum(duplicated(heart_rate_data_per_second))
```

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

From the result above, it was found that there are 543 duplicated values in the *SleepMinute* data, however the other data files do not have any duplicate values as the results of the **sum(duplicated())** function were zero.

3. I have removed the duplicated values from *SleepMinute* file using distinct function. And i change the name of the data from sleep_data_per_minute to sleep_data_per_minute_2 to be able to distinguish between two set of data. After applying the distinict function, I have check the data again using sum(duplicated value()) to make sure that data is free of any duplicated values and it was clean as the result was zero

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

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

```r
sleep_data_per_minute_2 <- sleep_data_per_minute %>% distinct(.keep_all = TRUE)
sum(duplicated(sleep_data_per_minute_2))
```

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

4. To ensure the consistency of the data, i checked the data format and made sure that date and time data across the different files have the same format. And i found that all the data has the same date and time format, however the date and time data were concatenated, so i have decided to split them into two separated columns as one for date with a yyyy-mm-dd format and the other for time with hh:mm:ss format and 12 hour system. For this purpose, i have installed "Lubridate" packages.

```r
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

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

```r
calories_data_per_minute_clean<- calories_data_per_minute %>%
    mutate(
   datetime_parsed = mdy_hms(ActivityMinute),      # parse to POSIXct
   date = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

I have repeated this procedure for all other data files to ensure that date and time have the same format.

```r
intensities_data_per_minute_clean<- intensities_data_per_minute %>%
    mutate(
   datetime_parsed = mdy_hms(ActivityMinute),      # parse to POSIXct
   date = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

```r
met_data_per_minute_clean<- met_data_per_minute %>%
    mutate(
   datetime_parsed = mdy_hms(ActivityMinute),      # parse to POSIXct
   date = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

```r
sleep_data_per_minute_2_clean<- sleep_data_per_minute_2 %>%
    mutate(
   datetime_parsed = mdy_hms(date),      # parse to POSIXct
   date_1 = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

```r
steps_data_per_minute_clean<- steps_data_per_minute %>%
    mutate(
   datetime_parsed = mdy_hms(ActivityMinute),      # parse to POSIXct
   date = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

```r
heart_rate_data_per_second_clean<- heart_rate_data_per_second %>%
    mutate(
   datetime_parsed = mdy_hms(Time),      # parse to POSIXct
   date = date(datetime_parsed),             # extract Date
   time = format(datetime_parsed, "%H:%M:%S")# extract time string
  ) %>%
    select(-datetime_parsed)
```

5. I have checked for missing values to ensure data integrity by using is.na function.

```r
sum(is.na(calories_data_per_minute_clean))
```

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

```r
sum(is.na(intensities_data_per_minute_clean))
```

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

```r
sum(is.na(met_data_per_minute_clean))
```

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

```r
sum(is.na(sleep_data_per_minute_2_clean))
```

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

```r
sum(is.na(steps_data_per_minute_clean))
```

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

```r
sum(is.na(heart_rate_data_per_second_clean))
```

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

6. The results above shown that the dataset is free of missing value. After processing all the data, I changed the name of the file to simplify the analysis process by assigning short names for each dataset.

```r
old_calories_data<- calories_data_per_minute_clean
new_calories_data<-"calories_minute_cleaned"
assign(new_calories_data,old_calories_data)
```

```r
old_intensities_data<- intensities_data_per_minute_clean
new_intensities_data<-"intensities_minute_cleaned"
assign(new_intensities_data,old_intensities_data)
```

```r
old_met_data<- met_data_per_minute_clean
new_met_data<-"met_minute_cleaned"
assign(new_met_data,old_met_data)
```

```r
old_sleep_data<- sleep_data_per_minute_2_clean
new_sleep_data<-"sleep_minute_cleaned"
assign(new_sleep_data,old_sleep_data)
```

```
old_steps_data<- steps_data_per_minute_clean
new_steps_data<-"steps_minute_cleaned"
assign(new_steps_data,old_steps_data)
```

```
old_heart_rate_data<- heart_rate_data_per_second_clean
new_heart_rate_data<-"heart_rate_cleaned"
assign(new_heart_rate_data,old_heart_rate_data)
```

7. I change the date from character format into Date/Time format for all the dataset using this code:

```
calories_minute_cleaned$ActivityMinute <-
 ↪ as.POSIXct(calories_minute_cleaned$ActivityMinute,
format = "%m-%d-%y %H:%M:%S")
```

```
intensities_minute_cleaned$ActivityMinute <-
 ↪ as.POSIXct(intensities_minute_cleaned$ActivityMinute, format = "%m-%d-%y %H:%M:%S")
```

```
met_minute_cleaned$ActivityMinute <- as.POSIXct(met_minute_cleaned$ActivityMinute,
format = "%m-%d-%y %H:%M:%S")
```

```
steps_minute_cleaned$ActivityMinute <- as.POSIXct(steps_minute_cleaned$ActivityMinute,
format = "%m-%d-%y %H:%M:%S")
```

```
sleep_minute_cleaned$date <- as.POSIXct(sleep_minute_cleaned$date,
format = "%m-%d-%y %H:%M:%S")
```

8. To gain insights about users' hear rate records per day, I extract daily_heart_rate and hourely_heart_rate data from heart_rate per second data using these code:

```
library(dplyr)
daily_heart_rate <- heart_rate_cleaned %>%
  group_by(Id, date) %>%
  summarise(avg_hr = mean(Value, na.rm = TRUE))
```

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

9. The heart_rate data has been processed using R, once the data was gotten ready, I exported the daily_heart_rate file as csv format to my device using *write.csv* function so that i could imported it into the googlesheet and merged it with other daily activities data and imported it to be visualized using Tableau.

```
library(dplyr)
heart_rate_cleaned$Time_parsed <- mdy_hms(heart_rate_cleaned$Time)
heart_rate_hourly <- heart_rate_cleaned %>%
  mutate(
    date = as.Date(Time_parsed),          # extract date
    hour = hour(Time_parsed)              # extract hour (0-23)
)
```

```r
heart_rate_cleaned_hourly <- heart_rate_hourly %>%
  group_by(Id, date, hour) %>%
  summarise(
    avg_hr = mean(Value, na.rm = TRUE),
    .groups = "drop"
  )
```

```r
write.csv(daily_heart_rate, "/Users/shaimaalradai/Downloads/daily_heart_rate.csv",
row.names = FALSE)
```

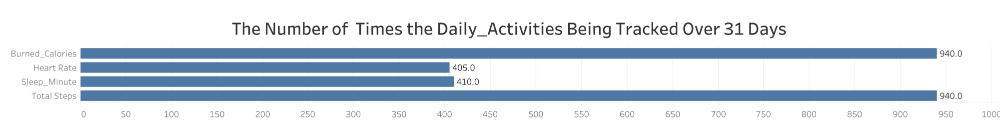
```r
write.csv(heart_rate_cleaned_hourly,
 ↪  "/Users/shaimaalradai/Downloads/heart_rate_cleaned_hourly.csv", row.names = FALSE)
```

By doing the above steps, data integrity was verified, data is clean and ready for the analysis phase.

4. Analyze Phase: For answering the business tasks, I have followed these steps:

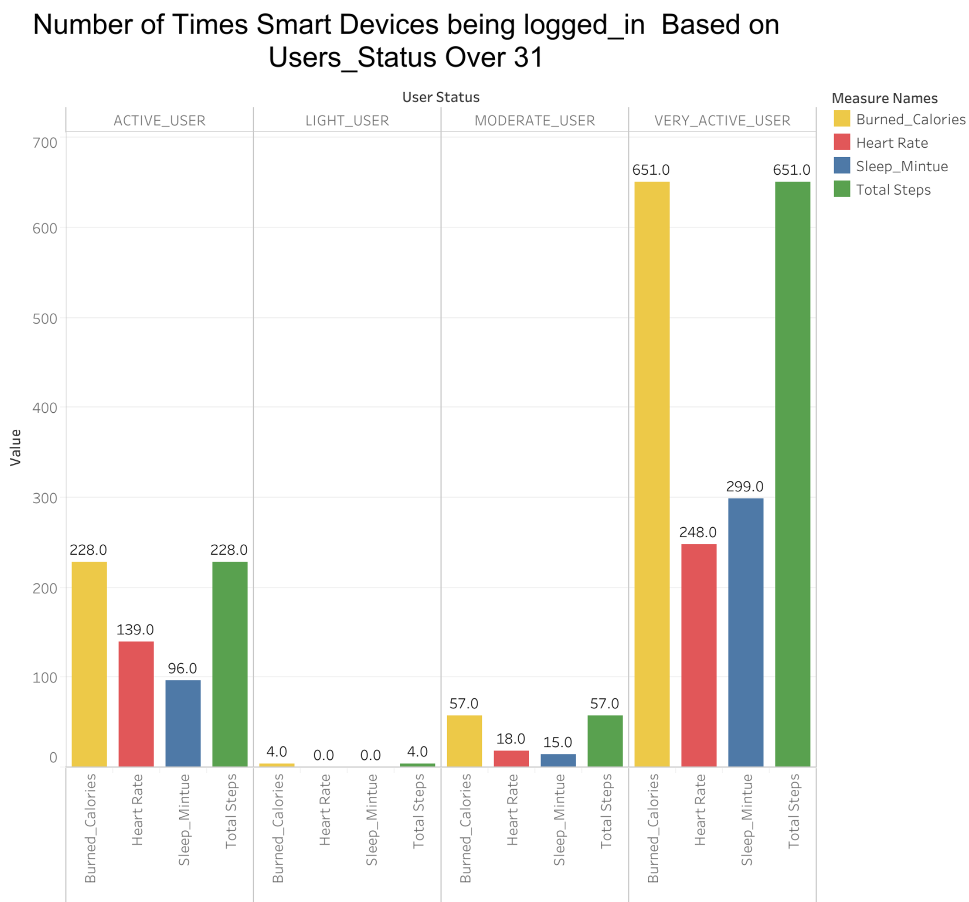
1. I have started analyzing the daily and hourly activities data that I was prepared and processed by using Googlesheet. First, I have classified the users into four categories based on how many time they were active using the smart device during the 31 days period. I have done this by using googlesheet tool. The function that have been used was **=QUERY(A1:A941, "SELECT A, COUNT(A) GROUP BY A",1)** as A column was the Id. The result was ranged between 4 and 31. that is, the highest number of logged_in was 31 and the lowest was 4. Then I classified the users based on their number of loggin Using *IFS* function. Those who logged_in 31 times, they were classified into Very_Active_User. Those who had between 30 to 26 classified into Active_User. Between 25 and 10, Moderate_User. Finally, less than 10 and more than 4, were classified as Light_User. These classifications reflect the User_Status.
2. After done with classification, I used **Vlookup** function to assign the approperiate classification for each user. The syntax was like this **VLOOKUP(A2,$U3 :V$35,2,FALSE)** as A column is for Id and the range U3:V35 was the columns that contain the results of **QUERY** function mentioned in step 1. Now every user has been assigned with his User_Status:Very_Active_User, Active_User,Moderate_User,and Light_User.
3. Now my daily and hourly activities data is ready for analyzing. I have used Tableau tool for this purpose. I have imported the daily_activities.csv and hourly_activity.csv files, that i downloaded from googlesheet to my device, to Tableau.
4. The purpose of analyzing the data is to know what is the most tracking activity the users used their smart device for.To do so, I have chosen the *Total_steps, Sleep_Minute, Burned_calories and Heart_Rate* and used the **count** to measure how many times these activities have been tracked by user using their smart devices during 31 days . First, I examine the tracked daily_activities for the whole users and then I did the same for each group of users to see which group is more active.

- I have chosen *Bar Chart* visualization to show the tracked daily_activities for the whole sample.



The chart shows us how users are used their smart devices to track their daily_health activities. The physical fitness activities (steps and calories) were the most tracked activities (940 times), the smart devices were used for by users during 31 days. However,Sleep and Heart_Rate tracking activities were the lowest particularly the Heart_Rate (410,405 times) respectively.

- I have chosen *side_by_side bar* to show the tracked daily_activity in more details for each group of users.

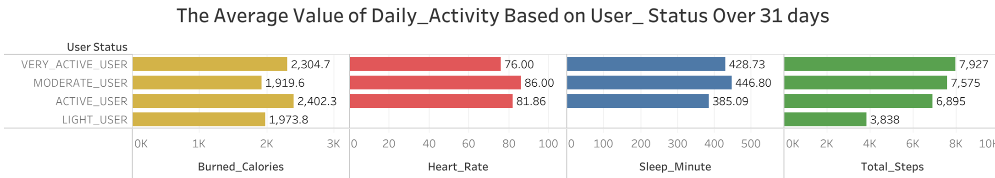


The visualization shows that the users classified by activity level (VERY_ACTIVE_USER, ACTIVE_USER, MODERATE_USER, LIGHT_USER) across four key daily activity metrics:

- Burned Calories
- Sleep Minutes
- Total Steps
- Heart_Rate

The chart shows us how many times users used their smart devices to track their daily_health activities. The physical fitness activities (steps and calories) were the most tracked activities the smart devices used for by users especially those who are VERY_ACTIVE_USERS.Sleep and Heart_Rate tracking activities are the lowest by all users compared to steps and calories activities, particularly among Active and Moderate.

5. The above findings do not mean that those higher number of logging_in higher records for daily_activities. To show more insights about this, I broke down each tracked activity based on the *User_status*. I have plotted the Average_Total_Steps, Average_Burned_Calories, Average_Sleep_Minute and Average_Heart_Rate for each and every group of users separately.

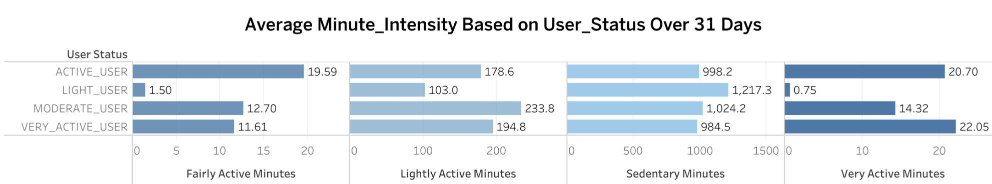


The Viz indicates that: * VERT_ACTIVE_USER group, on average, has the highest Total_steps among other activities, showing that on average the total steps is around 7927. This group recorded,on average, the lowest heart_rate (76) which indicates that they are in better health than other groups

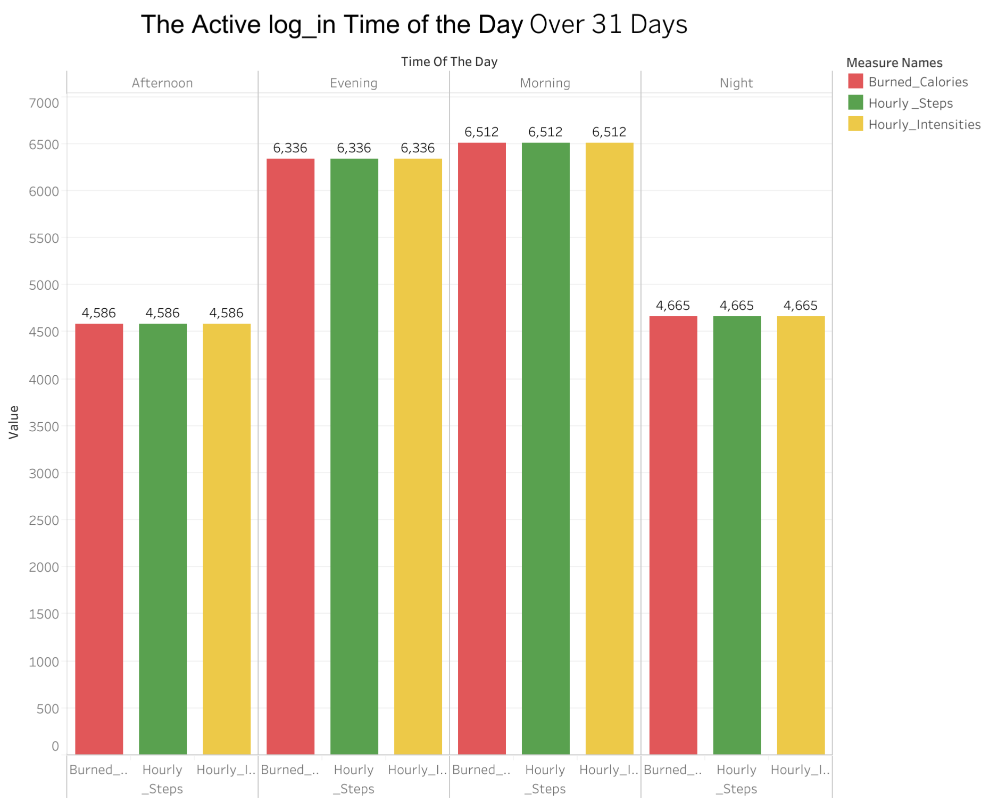
- ACTIVE_USER group record the highest average of Burned_Calories showing that users burn approximately 2402 calories daily. This group recorded the fewest sleep_minutes per day, as users only sleep 385.09 minutes per day.
- MODERATE_USER group sleep, on average, 447 minutes daily likely closer to 6.5–7.5 hours. Which is the highest Sleep_Minute among the forth group of users.Also they recorded the highest value for Heart_Rate which might indicates that they need to do more sport and exercises.
- LIGHT_USER group had, on average, the lowest records for Total_Steps, Burned_Calories, however did not track Sleep_Minute or Heart_Rate at all.

To sum up, in general, smart devices were used the most for tracking Total_Steps especially by VERY_ACTIVE_USER and the lowest for tracking Heart_Rate.

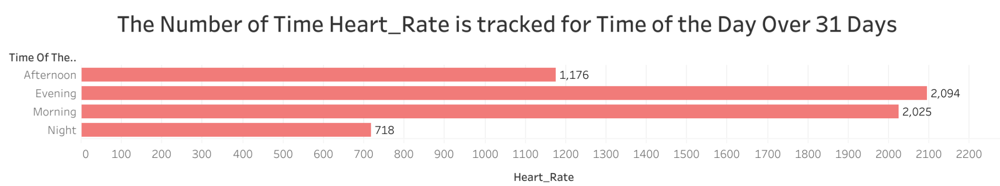
6. To get more insights about the user status, I have related the intensity_minutes records with the user status (based on their number of log_in to their smart devices). I have noticed that, on average, those how were the most active in using their smart devices, had the highest *Active_Minutes*. In contrary, those who were the lowest active had the highest *Sedentary_Minutes*.



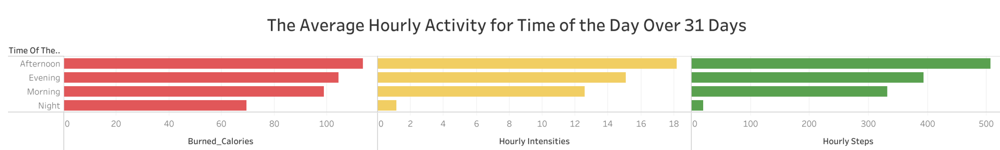
7. All the analyses that have been done so far are about daily_activity. However, We need to get more details on Time_of_The_Day for these activities. To do so, I have used Hourly_Activities records for Total_steps, Burned_Calories, and Sleep_Minutes. I have created a column for Time_of_The_Day,using *IFS* function in googlesheet, which contains fourth classifications for the time of the days: Morning(starts from 05:00 AM to 11:59 AM), Afternoon(starts from 12:00 AM To 15:59 AM), Evening(starts from 16:00 To 23:59) and Night(starts from 00:00 to 04:59).



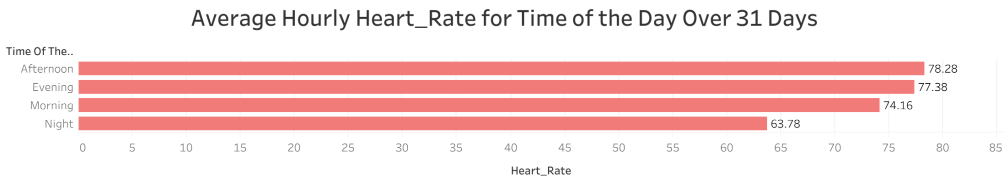
The chart shows that Morning is the most time when smart devices were used to track steps,calories and intensified minutes . However, at Night and in the Afternoon, Users were not active as much as Morning and Evening times. For Heart_Rate the case was different, as evening was the most time users tracked their heart rate.



8. To show the level of each activity based on time of the day,I have examined the average of hourly activities and the results indicate that in the afternoon, users recorded,on average, the highest value for Total_Steps, Burned_Calories and Intensity_Minute compared to morning and evening time. In contrary, night was the least active time for tracking activities.



9. I have plotted the data for heart rate per hour for Time of the Day in a separate bar chart and got the same result, Afternoon was the time when users recorded, on average, the highest value of heart rate.



10. Classification of various activities is possible by using METs so, I have used METs values per minute to classified the minutes to Active, Moderate, Light and Sedentary Minute. For the referencing click here.

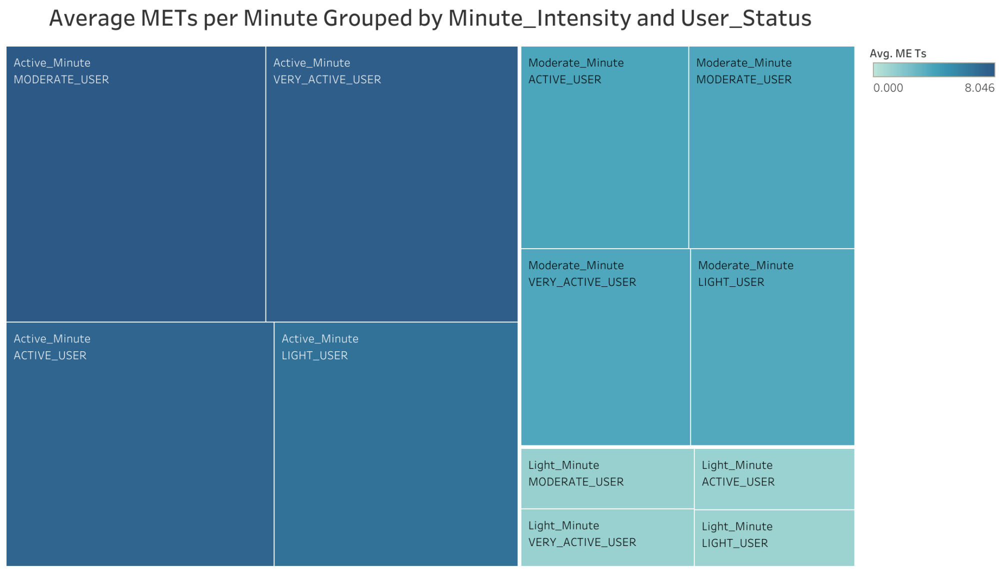
Range	Category
0	Sedentary_Minute
>0 − <3	Light_Minute
3 − <6	Moderate_Minute
6	Active_Minute

For doing this classification I used R codes. First, I divided the METs values by 10 as mentioned in FitBit Data Dictionaty click here. Second, I classified the METs values.

```
met_minute_cleaned<-met_minute_cleaned %>% mutate(METs = METs / 10)
```

```
library(dplyr)
met_minute_cleaned <- met_minute_cleaned %>%
  mutate(Minute_Intensity = case_when(
    METs == 0 ~ "Sedentary_Minute",
    METs > 0 & METs < 3 ~ "Light_Minute",
    METs >= 3 & METs < 6 ~ "Moderate_Minute",
    METs >= 6 ~ "Active_Minute"
  ))
```

Third, I plotted the average METs against each group of Users and the results show that the MODERATE_USER group has the highest average of Active_Minute based on METs, ACTIVE_USER group has the highest average of Moderate_Minute and Light_Minute, however, LIGHT_USER group has the lowest average of Active, Moderate, Light and Sedentary Minute.



5. Share and Act Phase: I have created this ppt as a presentation for sharing what I have done and what insights I have gained from the analysis. Moreover, the ppt provides the stakeholders and marketing team with recommendations that they can act on to achieve their goals.

Table of content:

Wellness Technology:

Business Task

To analyze smart device usage data in order to gain more insights about how consumers use non-Bellabeat smart devices.

What are we going to do?

To achieve the objective of this study, we are trying to answer these questions:

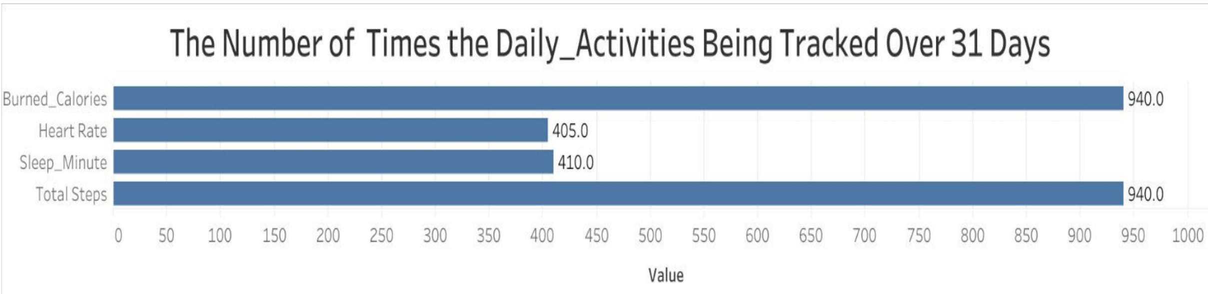
1. What are smart devices used for the most?

2. How could the company serve its customers by understanding how they use their smart devices?

3.How effectively could the company utilize the insights on smart devices usage to develop its market strategies?

Visualize the data

The trend of smart device usage (General_overview)



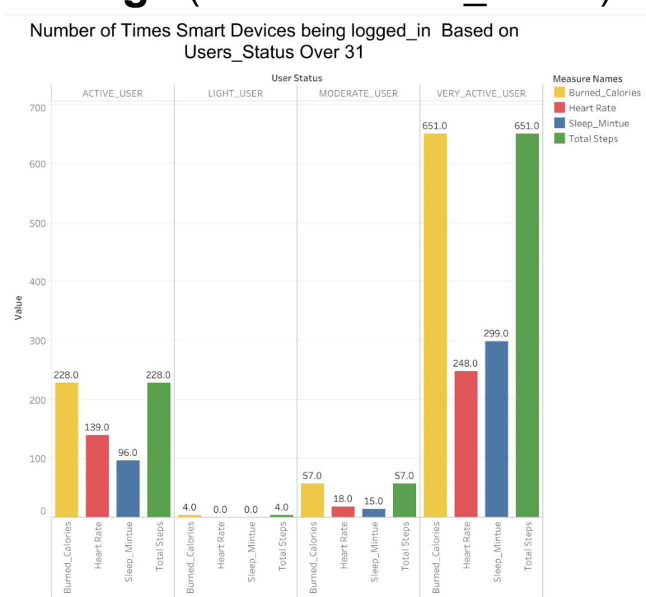
- Total Steps and Burned_calories are the most daily_activities being tracked by users over 31 days.
- Sleep and Heart Rate is the lowest tracked activity by the users over 31 days.

The trend of smart device usage (Based on User_Status)

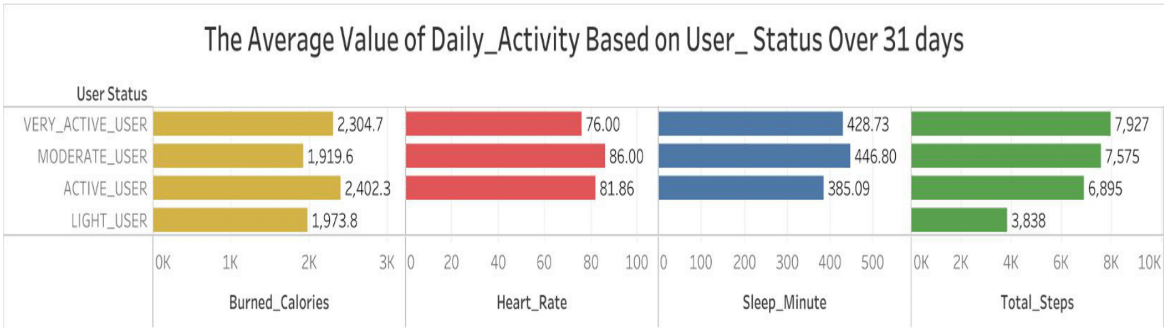
All users were using their smart device the most for tracking their **Total_Steps** and **Burned_Calories**.

VERY_ACTIVE_USERS used their smart device for tracking all the daily_activities.

LIGHT_USERS did not use their smart device for tracking heart_rate and sleep_minute.

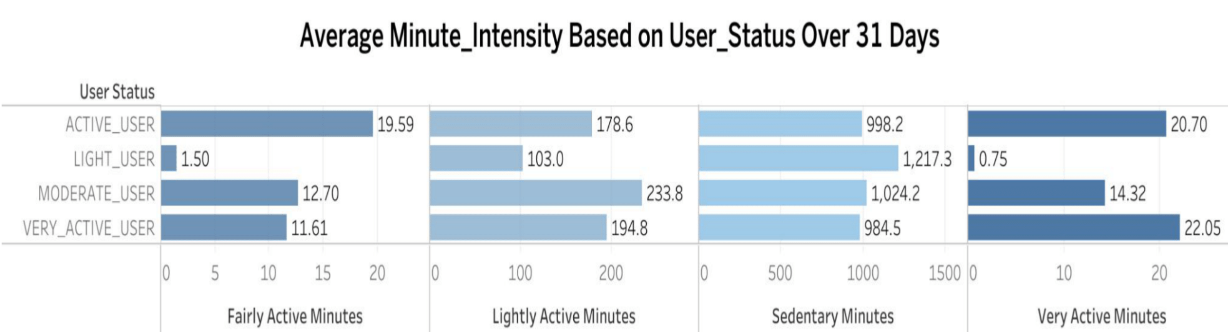


Daily-Activities



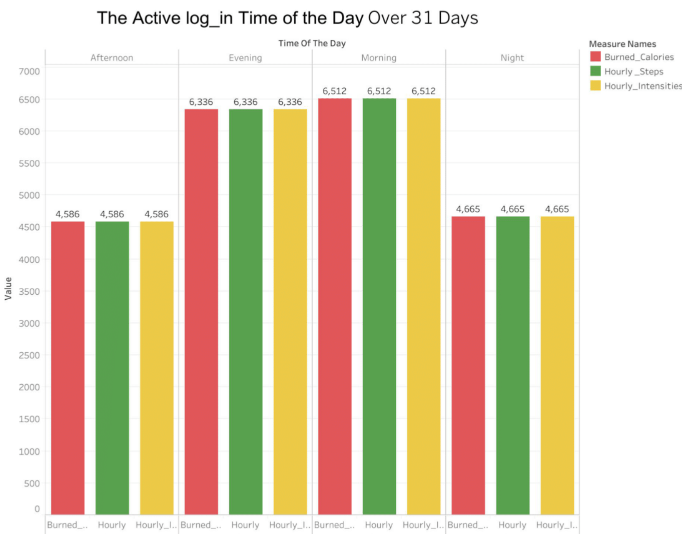
- The highest average of **Total_Steps** was recorded by VERY_ACTIVE_USERS.
- The highest average of **Burned_Calories** was recorded by ACTIVE_USERS.
- The highest average of **Heart_Rate and Sleep_Minute** was recorded by MODERATE_USERS.

Minute-Intensity



- There is a correlation between being active user and having higher active minute of the day.

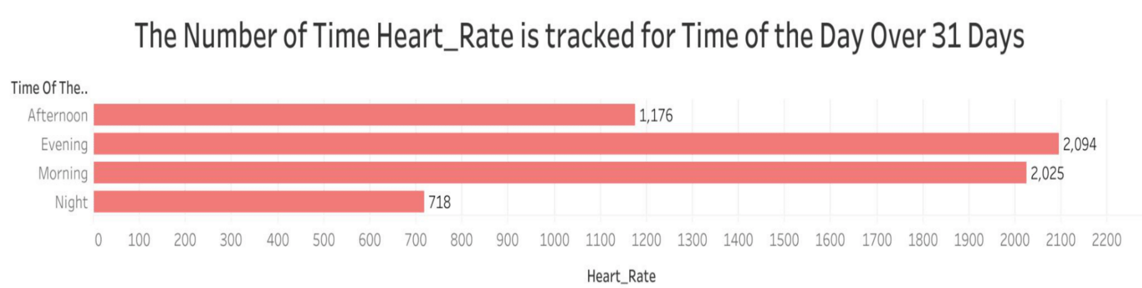
Time-of-the-Day



On average, and based on the number of times users logged_in by users over 31 days:

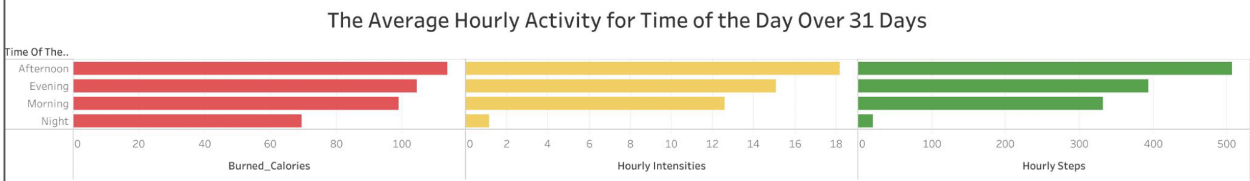
- Morning is the most active time of the day.
- Afternoon is the lowest active time of the day

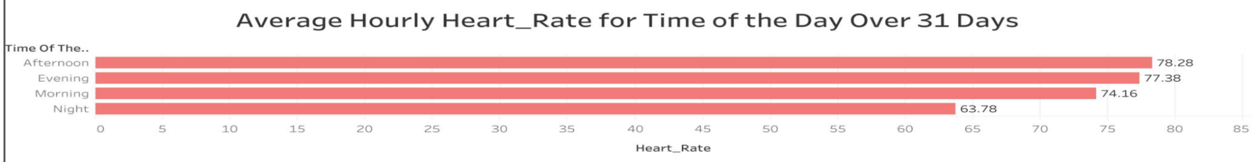
Time-of-the-Day



Evening was the most active time of the day for tracking Heart_Rate, however Night was the lowest.

Time-of-the-Day



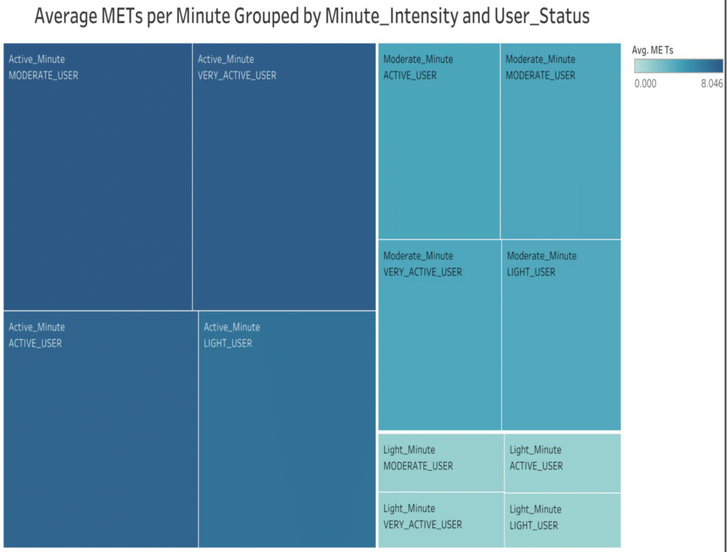


On average, and based on Hourly_Activities:

- **Afternoon** was the time of the day which recorded the highest average values of all hourly_activities.
- The Active **Time_of_the_Day** based on the number of times users logged_in (**Morning**) differs from the Active **Time_of_the_Day** based on the average of hourly activities (**Afternoon**).

METs

- **METs** stands for **Metabolic Equivalents**. Quantify exercise intensity by measuring the energy expenditure relative to resting.
- **MODERATE_USERS** have the highest METs.
- **Light_USERS** have the lowest METs



Conclusion

1

- Smart devices were used the most for tracking: **Total Steps, Burned Calories.**
- There is no correlation between being active by logging_in to smart device and using them for tracking health_activities regularly. As **LIGHT_USER** did not track sleep and heart rate at all.

3

There is no correlation between the **Active Time of the Day** based on the number of logging_in to smart device and the **Active Time of the Day** based on an average of health_activities' records.

4

Based on Minute_Intensities per day(Obtain from daily_dataset), **VERY_ACTIVE_USERS** have the highest Active_Minutes, however based on Minute_intensities per Minute (Measured by METs), **MODERATE_USERS** have the highest Active_Minute.

What is supposed to be done?

1. What are smart devices used for the most?

Focus on the activities which were untracked by users such as sleep monitoring and heart rate tracking, and prioritize the development of enhanced services and functionalities in these areas. This approach can improve user engagement with these activities while simultaneously attracting new potential users.

2. How could the company serve its customers by understanding how they use their smart devices?

By identifying customer preferences—such as which time of the day and which minute within the hour they actively use their devices and which activities they track during that time—the company can tailor its products to better meet those specific needs.

3.How effectively could the company utilize the insights on smart devices usage to develop its market strategies?

Target Light Users: Tailor marketing strategies to light users by understanding their smart device habits to grow the customer base.

Motivate Active Users: Encourage very active users to consistently track all activities each time they use their devices.
Promote Holistically: Use unbiased, comprehensive marketing that highlights multiple features—like steps, hydration, sleep, and heart rate—to encourage full usage and better health.

Thank you