Bellabeat smart device fitness data case study

Jenson Samuel

2023-09-05

### INTRODUCTION

This is a part of google data analysis capstone project, in which, Bellabeat’s smart device fitness data is used. This data is collected originally from 33 fitbit users. As we progressing further, goes through phases of ask, prepare, process, analysis share and act to successfully answer the business questions for compleating the case study.

##### i. Scenario

BellaBeat is a successful small company, but they have potential to grow bigger in global smart device market. Urška Sršen, co founder and Cheif Creative Officer of Bellabeat, believes that analyzing smart device fitness data could help unlock new growth opportunists for the company. Based on this requirement, the data analysis job is undertaken to find insights, those guide, the future marketing strategy for the company. The entire analysis is meant to present before Bellabeat executive team with all the recommendation to help the marketing strategy.

##### ii. About company

Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Collecting data on activity, sleep, stress, and reproductive health has allowed Bellabeat to empower women with knowledge about their own health and habits.Since it was founded in 2013, Bellabeat has grown rapidly and quickly positioned itself as a tech-driven wellness company for women.

#### ASK

Here junior analyst asked to analyze smart device usage data to gain insight into how consumers use Bellabeat smart devices. For selecting one Bellabeat product to apply these insights to in the presentation.

##### i. Business Tasks

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?

##### ii. Stakeholders

1. Urška Sršen: Bellabeat’s co founder and Chief Creative Officer
2. Sando Mur: Mathematician and Bellabeat’s co founder; key member of the Bellabeat executive team
3. Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat’s marketing strategy.Junior data analyst joined this team six months ago and have been busy learning about Bellabeat’’s mission and business goals as well as how a junior data analyst, can help Bellabeat achieve them.

#### PREPARE

##### i. Data source

open data is available for this case study analysis, data contains in 18 CSV files which are available both Long and wide formats. FitBit fitness tracker data from:- <https://www.kaggle.com/datasets/arashnic/fitbit>

##### ii. Reliability

Data is reliable because it generated by a survey of reliable third party Amazon Mechanical Turk.

##### iii. Originality

Data is originated from 30 original fitbit users those are consented to provide data.

##### iv. Comprehensiveness

Data provide different types of information like daily, hourly and minute activity, moving distance, step count, sleeping duration, calories burned etc.

##### v. Credibility

Data is critically examined and thoroughly checked.Data is comes from an open source, and no personal information used.

##### vi. Data drawback

Available data is 7 years old (2016), So the fitness activity pattern of users may changed over time.

#### PROCESS

##### i. Packages used

library(lubridate)  
library(tidyr)  
library(dplyr)  
library(tidyverse)  
library(here)  
library(skimr)  
library(janitor)  
library(reshape2)  
library(tinytex)

##### ii. Data tables selected

activity <- read.csv("C:/Users/Jenson/Documents/dailyActivity\_merged.csv")  
sleep <- read.csv("C:/Users/Jenson/Documents/sleepDay\_merged.csv")  
Calories <- read.csv("C:/Users/Jenson/Documents/hourlyCalories\_merged.csv")  
Intensities <- read.csv("C:/Users/Jenson/Documents/hourlyIntensities\_merged.csv")  
Steps <- read.csv("C:/Users/Jenson/Documents/hourlySteps\_merged.csv")

##### iii. Checking duplicates and cleaning

# Checking for duplicate entries in data sets  
sum(duplicated(activity))  
sum(duplicated(sleep))  
sum(duplicated(Calories))  
sum(duplicated(Intensities))  
sum(duplicated(Steps))

# Cleaning the Sleep data by removing 3 duplicate entries  
sleep1<- unique(sleep)

##### iv. Formatting data types and joining data tables

# activity date which are in \*char\* format converted to \*mdy\* date format and added a new column for week day.  
  
activity$ActivityDate <- mdy(activity$ActivityDate)  
activity$WeekDay<- weekdays.Date(activity$ActivityDate)  
  
# A new column for total active minutes created by summing up \*very active minutes\*, \*moderately active minutes\* and \*lightly active minutes\*.  
  
activity1<- activity %>% mutate(TotalActiveMinute = VeryActiveMinutes + FairlyActiveMinutes + LightlyActiveMinutes)  
  
# Sleep data file available is useful for checking the sleep pattern along with activity pattern.So to convert \*SleepDay\* into \*date-time\* format, \*SleepDay\* separated to \*date\* and \*time\* columns respectively, and then, \*date\* converted to \*mdy\* format. And then sleep data table joined with Activity table.  
  
sleep2<- separate(sleep1, SleepDay, into = c('ActivityDate', 'Time'), sep=' ')  
sleep2$ActivityDate <- mdy(sleep2$ActivityDate)  
act2<- activity1 %>% full\_join(sleep2, by=c('Id'='Id', 'ActivityDate'='ActivityDate'))  
act3<- act2 %>% select(-Time)

# Joining the Intensities, Steps and Calories  
  
hourlyAct1<- Intensities %>% full\_join(Steps, by=c('Id'='Id', 'ActivityHour'='ActivityHour'))  
hourlyAct<- hourlyAct1 %>% full\_join(Calories, by=c('Id'='Id', 'ActivityHour'='ActivityHour'))  
  
# Separating \*ActivityHour\* into \*Date\*, \*Time\* and \*UTC\*  
  
hourlyActivity<- separate(hourlyAct, ActivityHour, into = c('Date', 'TimeNew', 'UTC'), sep=' ')  
  
# Uniting \*Time\* and \*UTC\* into \*TimeNew\* and change to Time and UTC format and then \*mdy\* format  
  
hourlyActivity1 <- unite(hourlyActivity, col="Time", TimeNew, UTC, sep=" ")  
hourlyActivity1$Time <- parse\_time(hourlyActivity1$Time, "%H:%M:%S %p")  
  
# For week days analysis, add weekday column to the joined data set  
  
hourlyActivity1$Date <- mdy(hourlyActivity1$Date)  
hourlyActivity1$weekDay <- weekdays.Date(hourlyActivity1$Date)

act4<- act3 %>% drop\_na()

act4.1 <- act3 %>% arrange(TotalDistance) %>% group\_by(Id) %>% summarise(AVGSteps = mean(TotalSteps),  
AVG\_Distance = mean(TotalDistance),  
AVG\_calories = mean(Calories),   
AVG\_activeMinutes = mean(TotalActiveMinute))

##### v. Transorming data from wide to long format

act4\_wide <- act4 %>% group\_by(Id) %>% summarise(VeryActiveMinutes = sum(VeryActiveMinutes),  
FairlyActiveMinutes = sum(FairlyActiveMinutes),  
LightlyActiveMinutes = sum(LightlyActiveMinutes),  
SedentaryMinutes = sum(SedentaryMinutes))  
  
act4\_wide1 <- act4 %>% filter(Id==1503960366, ActivityDate=="2016-04-12")  
glimpse(act4\_wide1)

## Rows: 1  
## Columns: 20  
## $ Id <dbl> 1503960366  
## $ ActivityDate <date> 2016-04-12  
## $ TotalSteps <int> 13162  
## $ TotalDistance <dbl> 8.5  
## $ TrackerDistance <dbl> 8.5  
## $ LoggedActivitiesDistance <dbl> 0  
## $ VeryActiveDistance <dbl> 1.88  
## $ ModeratelyActiveDistance <dbl> 0.55  
## $ LightActiveDistance <dbl> 6.06  
## $ SedentaryActiveDistance <dbl> 0  
## $ VeryActiveMinutes <int> 25  
## $ FairlyActiveMinutes <int> 13  
## $ LightlyActiveMinutes <int> 328  
## $ SedentaryMinutes <int> 728  
## $ Calories <int> 1985  
## $ WeekDay <chr> "Tuesday"  
## $ TotalActiveMinute <int> 366  
## $ TotalSleepRecords <int> 1  
## $ TotalMinutesAsleep <int> 327  
## $ TotalTimeInBed <int> 346

act4\_long<- melt(act4\_wide1, id.vars=c("Id"), measure.vars=c("VeryActiveMinutes", "FairlyActiveMinutes",  
"LightlyActiveMinutes", "SedentaryMinutes"), variable.name="Activity\_Types", value.name="Activity\_Minutes")  
glimpse(act4\_long)

## Rows: 4  
## Columns: 3  
## $ Id <dbl> 1503960366, 1503960366, 1503960366, 1503960366  
## $ Activity\_Types <fct> VeryActiveMinutes, FairlyActiveMinutes, LightlyActive…  
## $ Activity\_Minutes <int> 25, 13, 328, 728

##### vi. Observations

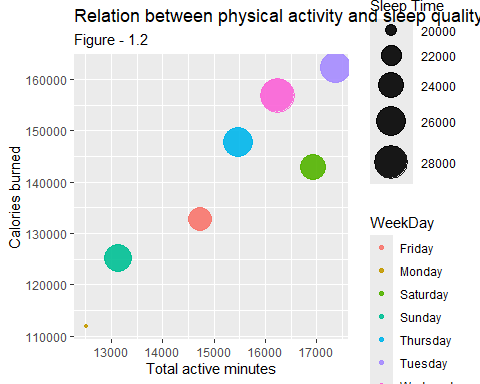
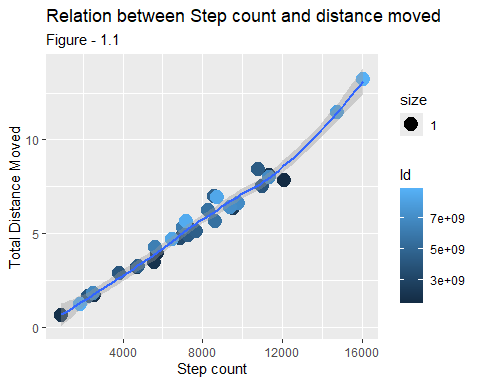
* Summary of daily physical activities of 940 observations over a month of 33 distinct people provided
* Sleep data of 24 people which consists of 413 observations of Total minutes asleep and total minutes in bed are provided
* Heart rate data for 14 people available
* A METs data for 8 people available
* 3 duplicate entries found with the Sleep data are removed for further analysis.
* Due to small sample size of heart rate and METs data, it won’t be able to provide meaningful insight. So those are dropped to take this analysis process.

#### ANALYSIS

##### i. Intensity of activity and sleeping time

act4.2 <- act4 %>% group\_by(WeekDay) %>% summarise(TotalSteps = sum(TotalSteps),   
Total\_Distance = sum(TotalDistance),  
Total\_calories = sum(Calories),  
Total\_VeryActiveMinutes = sum(VeryActiveMinutes),  
Total\_FairlyActiveMinutes = sum(FairlyActiveMinutes),  
Total\_LightlyActiveMinutes = sum(LightlyActiveMinutes),  
Total\_activeMinutes = sum(TotalActiveMinute),  
Total\_MinituesAsleep = sum(TotalMinutesAsleep),   
Total\_TimeInBed = sum(TotalTimeInBed))

ggplot(act4.1, aes(x=AVGSteps, y=AVG\_Distance)) + geom\_point(aes(color=Id, size=1)) + geom\_smooth() + labs(title="Relation between Step count and distance moved",subtitle ="Figure - 1.1", x="Step count", y="Total Distance Moved")  
ggplot(act4.2, aes(Total\_activeMinutes, Total\_calories, size=Total\_MinituesAsleep, color=WeekDay)) + geom\_point(alpha = .9) + scale\_size(range=c(1,12), name = "Sleep Time") + labs(title = "Relation between physical activity and sleep quality", subtitle ="Figure - 1.2", x="Total active minutes", y="Calories burned")



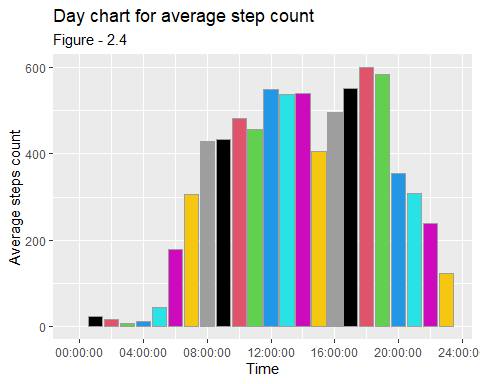
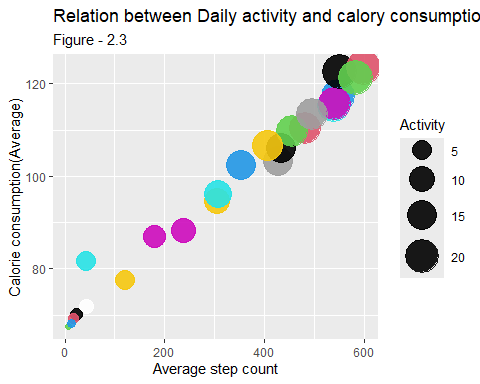
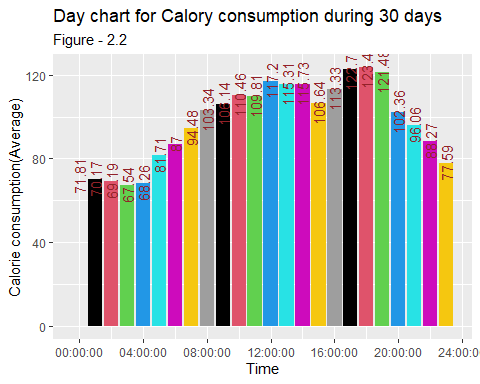
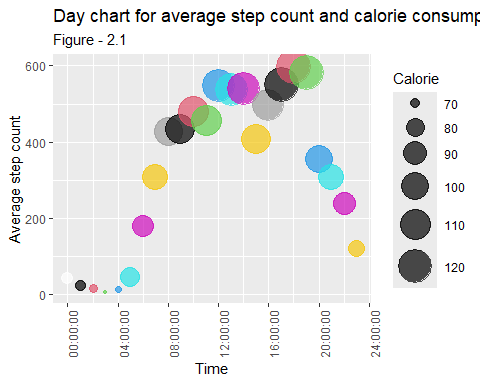
###### Observations

* Total distance and step count has a positive relationship
* Sleep data of 24 people is available from total 33 participants.
* However some disproportionate relation in weekend days, Total active minutes and calorie consumption has a direct relation ship with sleep time.

##### ii. General trend for hourly activity

hour1 <- hourlyActivity1 %>% arrange(Time) %>% group\_by(Time) %>%   
summarize(AVG\_Intensity = mean(TotalIntensity),   
AVG\_calories = mean(Calories),   
AVG\_steps = mean(StepTotal))

ggplot(hour1, aes(Time, AVG\_steps, size=AVG\_calories, color = Time)) +   
 geom\_point(alpha = 0.7) +scale\_size(range=c(1, 12), name = "Calorie") +  
 theme(axis.text.x=element\_text(angle=90, hjust=.5))+ labs(title="Day chart for average step count and calorie consumption", subtitle ="Figure - 2.1", x="Time", y="Average step count")  
ggplot(hour1, aes(x=Time, y=AVG\_calories)) + geom\_bar(stat="identity", aes(fill=Time)) + labs(title="Day chart for Calory consumption during 30 days", subtitle ="Figure - 2.2", x="Time", y="Calorie consumption(Average)") + geom\_text(aes(label = round(AVG\_calories, digits = 2)),size = 4, angle = 90, position = "identity", color = "#922229")  
ggplot(hour1, aes(AVG\_steps, AVG\_calories, size=AVG\_Intensity, color=Time)) +   
 geom\_point(alpha = .9) + scale\_size(range=c(2,12), name = "Activity") + labs(title = "Relation between Daily activity and calory consumption", subtitle ="Figure - 2.3", x="Average step count", y="Calorie consumption(Average)")  
ggplot(hour1, aes(x=Time,y=AVG\_steps, color = Time)) + geom\_bar(stat="identity", aes(fill=Time)) + labs(title="Day chart for average step count", subtitle ="Figure - 2.4", x="Time", y="Average steps count")



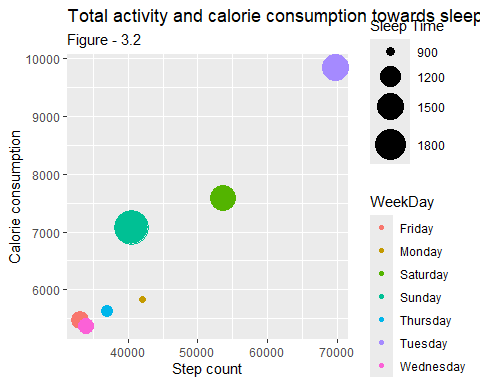
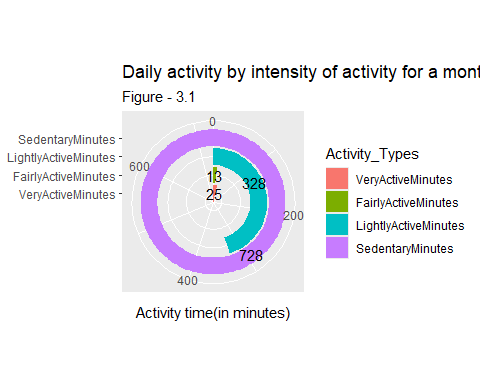
###### Observations

* As the step count increases the calorie consumption also increases, that means, total steps and calorie consumption has a positive relationship.(Fig - 2.1, Fig - 2.2)
* Average step count, Activity minutes and Calorie consumption has a very good positive relation ship (Fig - 2.3)
* Most desired workout time revealed as from 7 AM to 9 PM (Fig - 2.1, Fig - 2.4) Fig - 2.1

##### iii. Volume of intensity of activity and weekday trend

act4.3 <- act4 %>% filter(Id==1503960366) %>% group\_by(WeekDay) %>% summarise(TotalSteps = sum(TotalSteps),   
Total\_Distance = sum(TotalDistance),  
Total\_calories = sum(Calories),   
Very\_activeMinutes = sum(VeryActiveMinutes),   
Fairly\_activeMinutes = sum(FairlyActiveMinutes),   
Lightly\_activeMinutes = sum(LightlyActiveMinutes),   
Sedentary\_Minutes = sum(SedentaryMinutes),  
Total\_activeMinutes = sum(TotalActiveMinute),  
Total\_MinituesAsleep = sum(TotalMinutesAsleep),   
Total\_TimeInBed = sum(TotalTimeInBed))

ggplot(act4\_long, aes(x=Activity\_Types, y=Activity\_Minutes, fill=Activity\_Types)) + geom\_col() +   
coord\_polar("y", start=0) + geom\_text(aes(label=Activity\_Minutes), size=4,  
position = position\_stack(vjust =0.4)) +   
labs(title="Daily activity by intensity of activity for a month ",subtitle ="Figure - 3.1",   
x=NULL, y="Activity time(in minutes)")  
ggplot(act4.3, aes(TotalSteps, Total\_calories, size = Total\_MinituesAsleep)) + geom\_point(aes(color=WeekDay)) + scale\_size(range=c(2,12), name = "Sleep Time") + labs(title="Total activity and calorie consumption towards sleep time", subtitle ="Figure - 3.2", x="Step count", y="Calorie consumption")



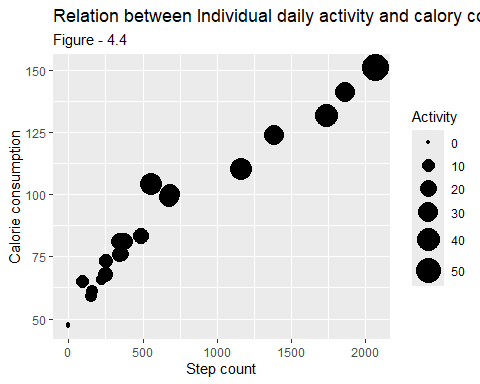
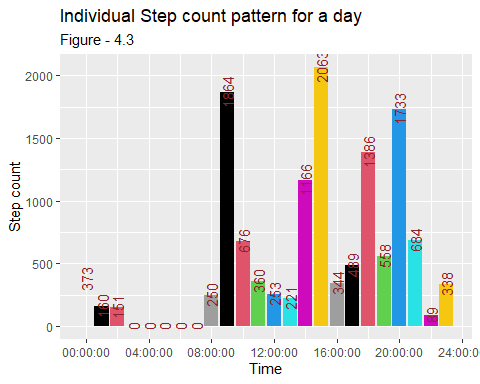
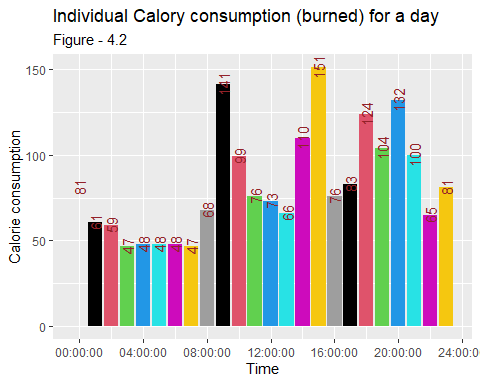
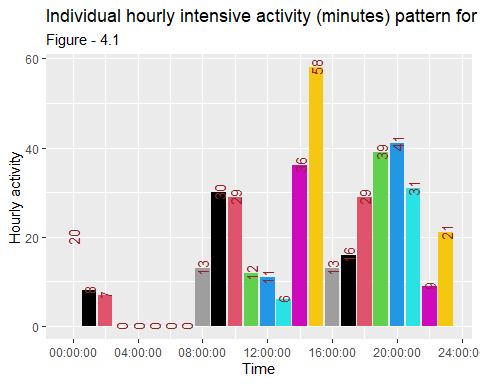
###### Observations

* How time is consumed individually, weather as very active, moderately active or lightly active shows in Fig - 3.1 (<Example:-> Id=1503960366, Date=2016-04-12)
* Weekday pattern describes the life style habits over 30 days, which show most active week days and less active weekdays(Fig - 3.2).
* Predominantly these data exhibits sedentary lifestyle pattern.

##### iv. Finding specific activity sequences and pattern

hour2 <- hourlyActivity1 %>% filter(Id==1503960366, Date=="2016-04-12") %>% group\_by(Time) %>% summarize(Total\_Intensity = sum(TotalIntensity), Total\_calories = sum(Calories), Total\_steps = sum(StepTotal))

ggplot(hour2, aes(x=Time, y=Total\_Intensity)) + geom\_bar(stat="identity", aes(fill=Time)) + labs(title = "Individual hourly intensive activity (minutes) pattern for a day", subtitle ="Figure - 4.1", x="Time", y="Hourly activity") + geom\_text(aes(label = round(Total\_Intensity,digits = 2)),size = 4, angle = 90, position = "identity", color = "#922229")  
ggplot(hour2, aes(x=Time, y=Total\_calories)) + geom\_bar(stat="identity", aes(fill=Time)) + labs(title = "Individual Calory consumption (burned) for a day ", subtitle ="Figure - 4.2", x="Time", y="Calorie consumption") + geom\_text(aes(label = round(Total\_calories,digits = 2)),size = 4, angle = 90, position = "identity", color = "#922229")  
ggplot(hour2, aes(x=Time, y=Total\_steps)) + geom\_bar(stat="identity", aes(fill=Time))+ labs(title = "Individual Step count pattern for a day", subtitle ="Figure - 4.3", x="Time", y="Step count") + geom\_text(aes(label = round(Total\_steps,digits = 2)),size = 4, angle = 90, position = "identity", color = "#922229")  
ggplot(hour2, aes(Total\_steps, Total\_calories, size=Total\_Intensity)) +   
 geom\_point(aes()) + scale\_size(range=c(1,9), name = "Activity") + labs(title = "Relation between Individual daily activity and calory consumption", subtitle ="Figure - 4.4", x="Step count", y="Calorie consumption")



###### Observations

* Daily activity pattern reveals the workout sequence, timing, idle time and sleep duration (<Example:Id==1503960366>, Date==“2016-04-12”)
* Individual hourly activity pattern shows most step count and intensive hours are from 8AM to 9PM. Sleep time can be identified as from 3AM to 7AM. (Fig - 4.1, Fig - 4.2, Fig - 4.3)
* Step count, calorie consumption and activity intensity are directly proportionate (Fig - 4.4) ######{-}

#### CONCLUSION

1. A relationship between distance moved and step count seems the available data is reliable. Among 33 individual sample data, only 24 samples have sleep time data. This shows the remaining user’s tendency of avoidance of usage of device during sleeping time. So the sleep time data is limited to 24 samples.
2. Activity minutes and calorie consumption shows a positive relationship. Here, we can see that, as the intensity of activity increases the quality of sleep (sleep duration) also increases. However, the weekend days shows a negative relation between sleep duration and physical activity minutes.
3. The intensity of activities, calorie consumption and step count shows a very positive relationship. Analyzing the hourly pattern of this factors in a day provide some meaningful insight. It describes that, most physically active time is from 7AM to 9 PM and remaining are the less active, sedentary or sleeping time. This is the general pattern of physical activity for 24 samples provided.
4. Hourly analysis helps to find summary of hours of different intensity activities for each person. Individual consumption of calorie, step count and sleep time are measurable and it is useful to disclose the weekday statistics of these activities. Helps to find individual lifestyle and useful to provide tips and notifications to improve the lifestyle.
5. Individual data for each hour of hourly activity, calorie consumption, step count, and sleep hours can be clearly analyzed. This helps to provide a clear understanding on daily activities and sleep hours, peak hours of each activity and low intensity hours. It is good to provide reminder of sleep hours of each day that improve the healthy lifestyle.

#### RECOMMENDATIONS

A. Encouraging to wear the device in sleeping time, for measuring the sleep time helps to plan a balanced fitness routine. It is useful to notify users about their sleep duration. Keeping optimal fitness activity and good quality of sleep can improve the health to lead an active life.

B. Metabolic equivalent (METs) is measure of expenditure of body energy and this using to find the exercise tolerance for individuals. Live data of heart rate identify signs of abnormal heart rhythms. Promoting to engage the users in these factors using the device is very useful for providing some precious services to plan and follow a safe fitness routine and good to keep cardiovascular health.

C. Provide app notifications of activity patterns, step count, sleeping time for each individual to get an insight on their lifestyle and tips and recommendations to improve remind users to take daily exercise.

D. Free medical checkups or free subscription for newly introduced services can be used as promotional offers to the users based on the measurement of their engagement in the device usage.