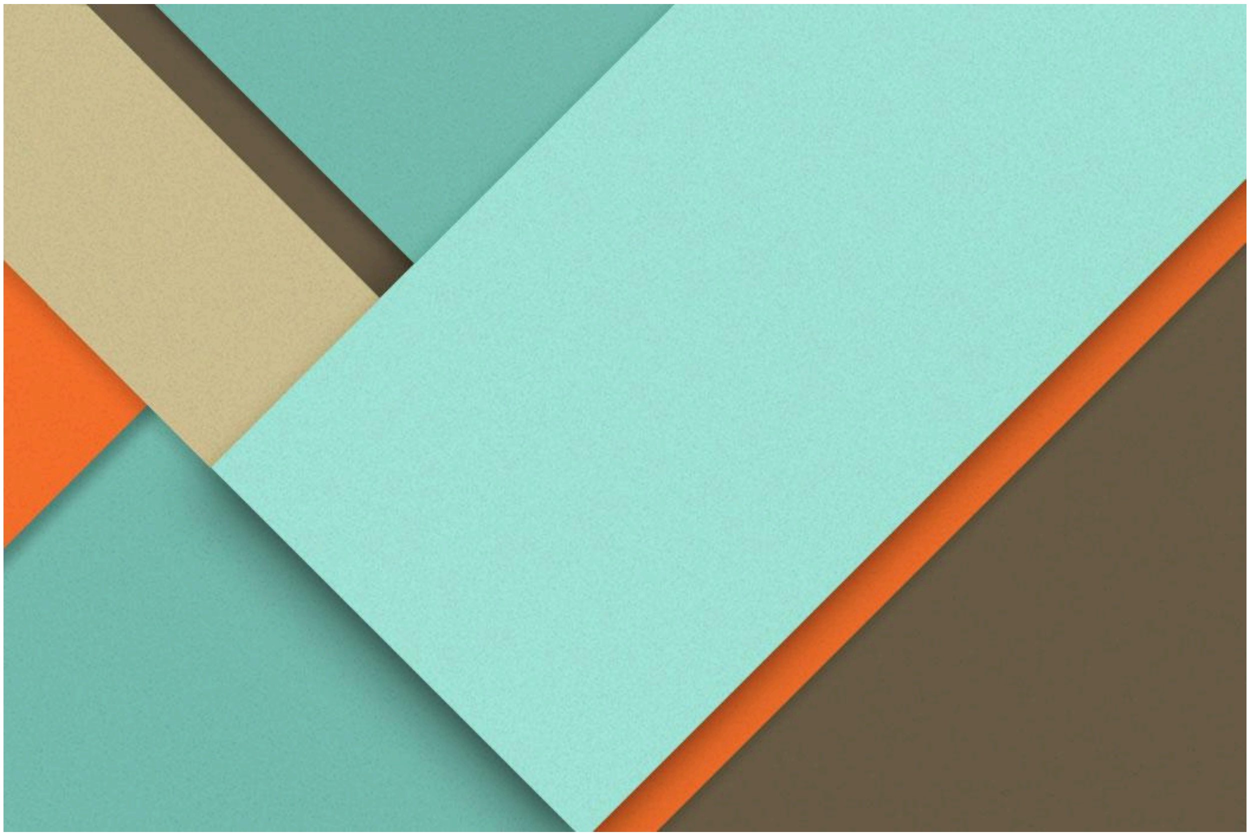


Case Study - Bellabeat

Aryan Anand

1. Summary

Bellabeat is a forward-thinking, growing company that produces cutting-edge smart devices tailored for health and wellness. These innovative gadgets capture and analyze data about various aspects such as physical activity, sleep patterns, stress levels, and reproductive health. The primary aim is to empower women by providing insightful information about their well-being and lifestyle.

The focus of this case is to explore growth opportunities for Bellabeat through the analysis of fitness data generated by smart devices. We will help unlock Bellabeat's potential to become a larger global smart device market player. In this case, our main focus will be one of Bellabeat's products: The Bellabeat App.

The Bellabeat app is a comprehensive platform that delivers health-related information that concentrates on activity levels, sleep quality, stress management, menstrual cycles, and mindfulness practices. Using this data, users gain valuable insights into their daily habits, enabling them to make informed decisions for a healthier lifestyle. The Bellabeat app seamlessly integrates with Bellabeat's multitude of intelligent wellness products.

The objective is to examine the potential avenues for growth and development from utilizing fitness data collected through the Bellabeat app.

2. Ask phase

Business Task: Identify trends in how consumers use non-Bellabeat smart devices to apply insights into Bellabeat's marketing strategy.

Stakeholders -

- Urska Srsen
- Sando Mur
- Bellabeat Marketing Analytics Team

3. Prepare phase

Dataset used -

The Kaggle dataset, FitBit Fitness Tracker Data, made available through Mobius, includes minute-level details on physical activity, heart rate, and sleep monitoring from 30, providing insights into daily habits.

Accessibility and privacy of our data -

By verifying the dataset's metadata, we can conclude it is open-source. You can use the data without permission as the owner has relinquished all rights worldwide, allowing unrestricted use, modification, distribution, and performance, even for commercial purposes.

Data credibility -

The dataset has sampling bias, as we need more demographic information from just 30 users. We cannot be sure whether or not the sample is representative of the whole population. We also have the problem of the dataset needing to be updated and not representing the current statistics, as it was done a while ago. This is why an operational reach will be used, which means we will have to use this information to improve decision-making for the daily operations of Bellabeat and shift the focus from simply understanding data to taking action on it in the tools that run the business processes.

4. Process phase

The analysis will mainly use R because of its easy accessibility and to create data visualizations so I can share the results with my stakeholders.

Package Installation -

First, I will have to install the necessary packages, which are:

- library(tidyverse)
- library(skimr)
- library(janitor)
- library(here)

After installing these packages, I proceeded to import the following data collected from Fitbit users:

- `daily_activity <- read_csv("C:/Users/shik_/Desktop/Data Case Study/Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")`

- `hourly_calories <- read_csv("C:/Users/shik_/Desktop/Data Case Study/Fitabase Data 4.12.16-5.12.16/hourlyCalories_merged.csv")`

- `sleepDay_merged <- read_csv("C:/Users/shik_/Desktop/Data Case Study/Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")`

- `hourly_steps <- read_csv("C:/Users/shik_/Desktop/Data Case Study/Fitabase Data 4.12.16-5.12.16/hourlySteps_merged.csv")`

Data Cleaning -

Data cleaning is extremely important because having clean data will ultimately increase overall productivity and allow for the highest quality information in decision-making.

Since the data is already organized, the next step is to split "mm:dd:yyyy" for month:date:year and "hh:mm:ss am/pm" for hour:minute:second am/pm to date and time by first converting it to posixct object and then format it using the format function.

This is how that process is done in R:

```
hourly_calories$activity_hour <- as.POSIXct(hourly_calories$activity_hour,
"%m/%d/%Y %I:%M:%S %p", tz = "")

hourly_calories$date <- format(hourly_calories$activity_hour, format = "%m/%d/%y")

hourly_calories$time <- format(hourly_calories$activity_hour, format = "%H:%M:%S")

---------------------------------------------------------------------------------

daily_sleep$sleep_day <- as.POSIXct(daily_sleep$sleep_day, "%m/%d/%Y %I:%M:%S %p"
, tz = "")

daily_sleep$date <- format(daily_sleep$sleep_day, format = "%m/%d/%Y")

daily_sleep$time <- format(daily_sleep$sleep_day, format = "%H:%M:%S")

---------------------------------------------------------------------------------

hourly_steps$activity_hour <- as.POSIXct(hourly_steps$activity_hour, "%m/%d/%Y
%I:%M:%S %p" , tz = "")

hourly_steps$date <- format(hourly_steps$activity_hour, format = "%m/%d/%Y")

hourly_steps$time <- format(hourly_steps$activity_hour, format = "%H:%M:%S")
```

Here, we have the Hourly Calories, Daily Sleep, and Hourly Steps. This is all quantitative data, which allows us to do this type of formatting.

We cannot do this same formatting for Daily Activity because that is a form of qualitative data.

Data Analysis -

We will look at our data and find out the average of each element.

Here, we are finding the average of each of the activities:

```
activity <- daily_activity %>%
  summarize(very_min_avg  = mean(very_active_minutes),
            fair_min_avg  = mean(fairly_active_minutes),
            light_min_avg = mean(lightly_active_minutes),
            sedentary_min_avg  = mean(sedentary_minutes))
```

Then, I convert the data frame to the long format:

```
activity_long <- activity %>%
 pivot_longer(cols = c(very_min_avg, fair_min_avg, light_min_avg, sedentary_min_avg),
names_to = "activity_type", values_to = "mean_minutes")
```

To find the average amount of calories burned in an hour, I do this:

```
hourly_cal <- hourly_calories %>%
  group_by(time) %>%
  summarize(cal_avg = mean(calories))
```

Then, to find the number of steps taken in each hour, I do this:

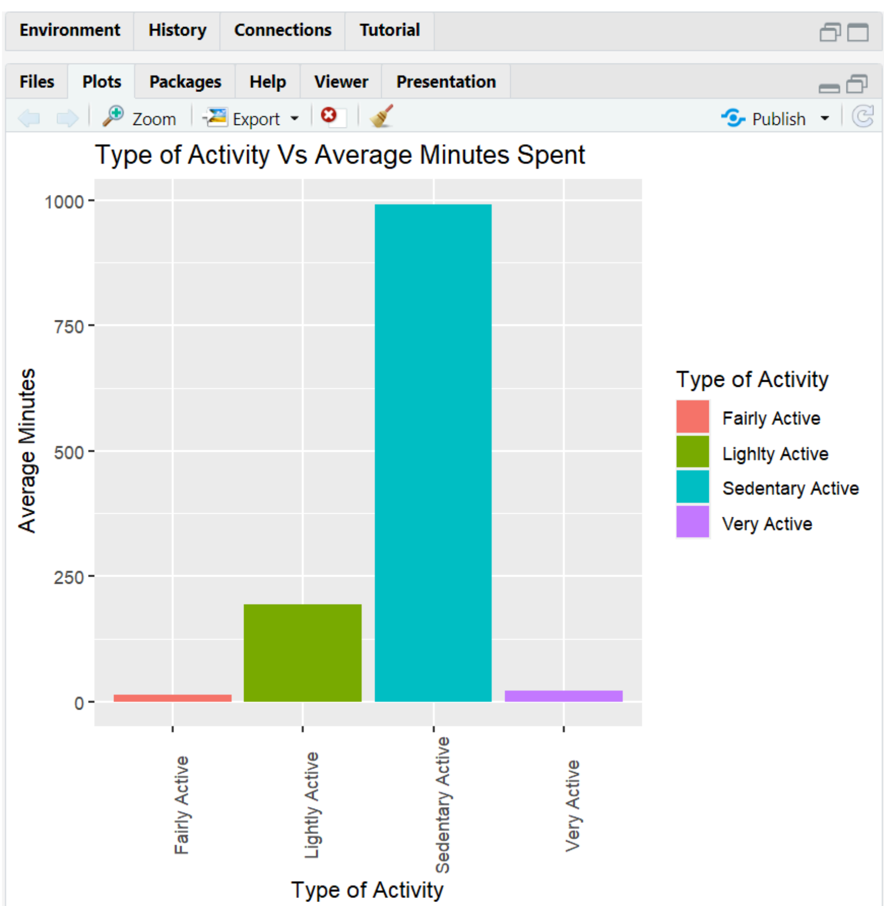
```
hourly_step <- hourly_steps %>%
  group_by(time) %>%
  summarize(step_avg = mean(total_steps_hr))
```

These findings will come in handy during the data visualizations.

Data Visualization -

The first visualization, or viz 1, represents the types of activities and the time spent doing those activities:

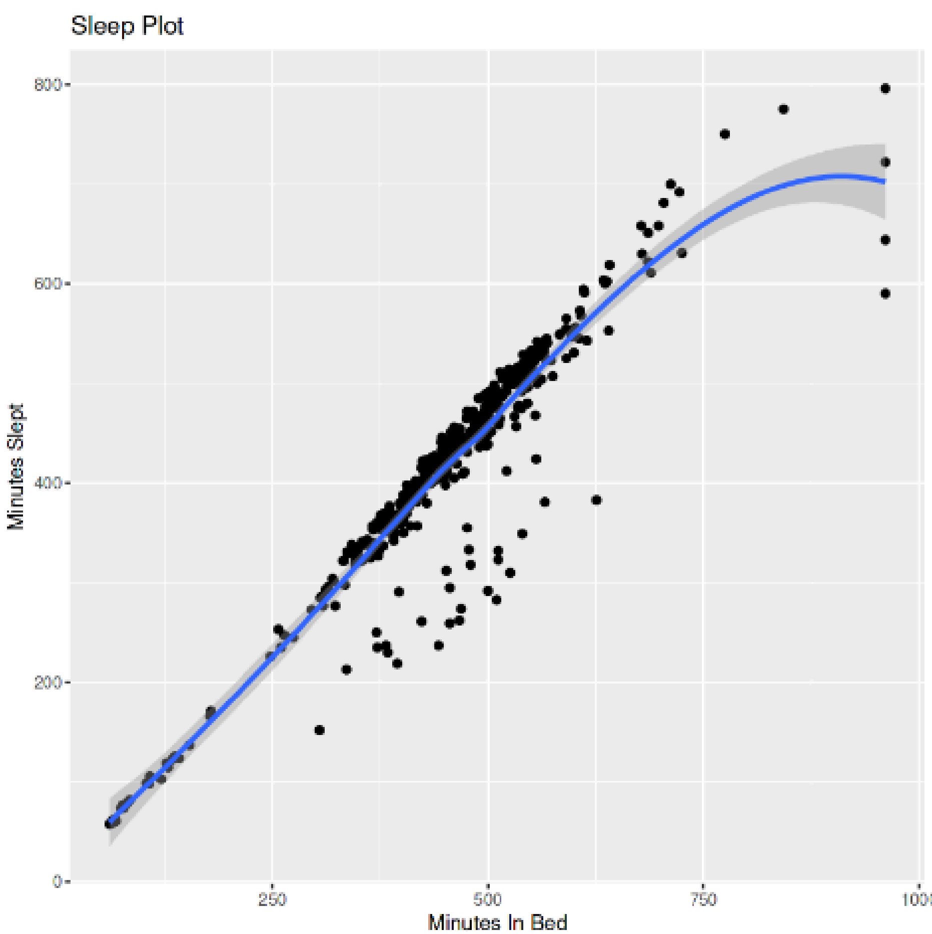
```
ggplot(data = activity_long)+
  geom_col(mapping = aes(x = activity_type, y = mean_minutes, fill = activity_type))+
  labs(title = "Type of Activity Vs Average Minutes Spent", x  = "Type of Activity", y = " Average Minutes", fill = "Type of Activity")+
  theme(axis.text.x = element_text(angle =90))+
  scale_x_discrete(labels = c("fair_min_avg" = "Fairly Active", "light_min_avg"  = "Lightly Active", "sedentary_min_avg" = "Sedentary Active",  "very_min_avg" = "Very Active"))+
  scale_fill_discrete(labels = c("Fairly Active", "Lightly Active", "Sedentary Active", "Very Active"))
```



The subsequent visualization, viz 2, is to represent the amount of time sleeping, and this is how it is done in R:

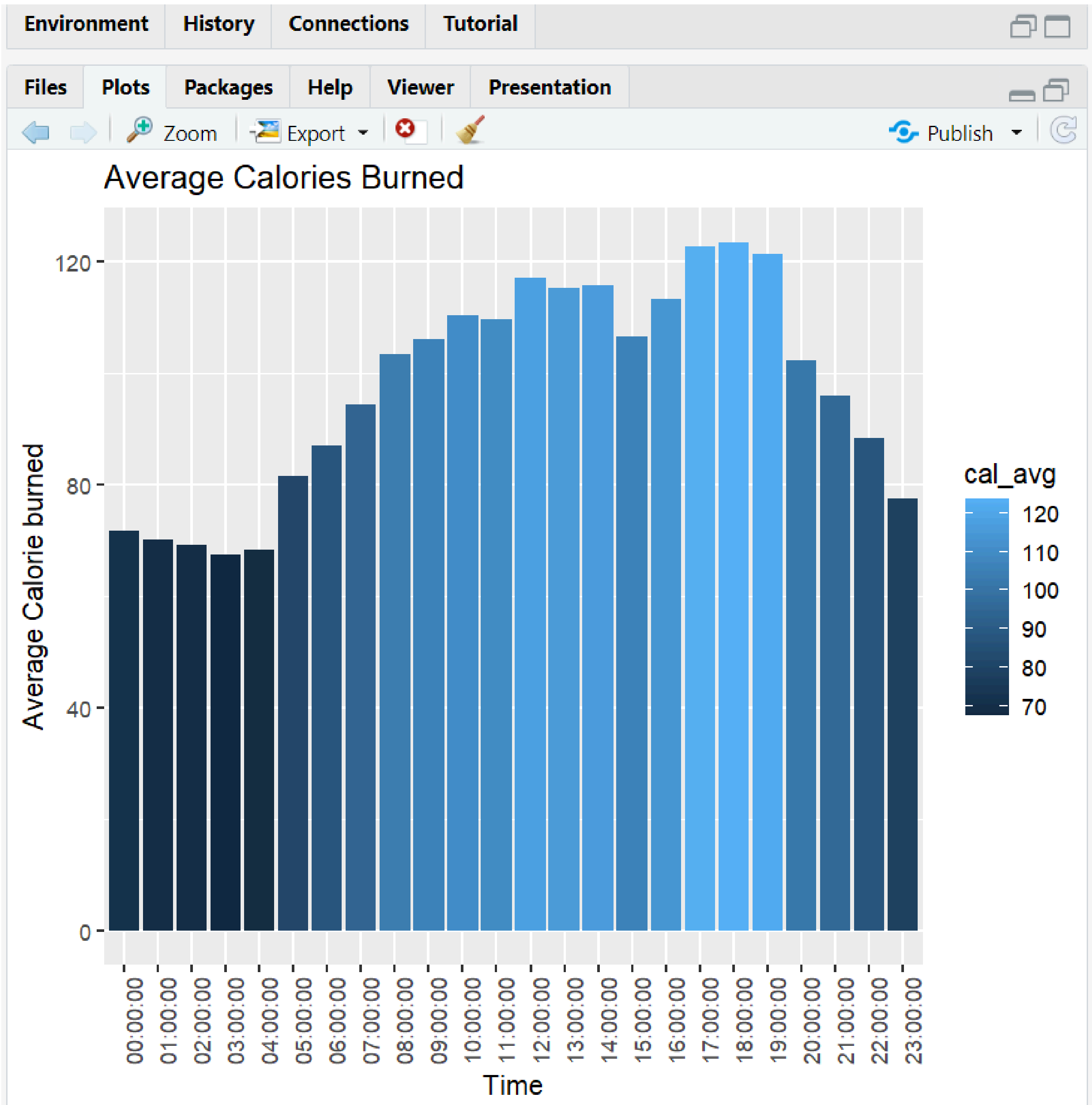
```
ggplot(data = daily_sleep)+
  geom_point(mapping = aes(x = bed_time, y = sleep_minute))+
  geom_smooth(mapping = aes(x = bed_time, y = sleep_minute))+
  labs(title = "Sleep Plot", x  = "Minutes In Bed" , y = "Minutes Slept")
```

—---



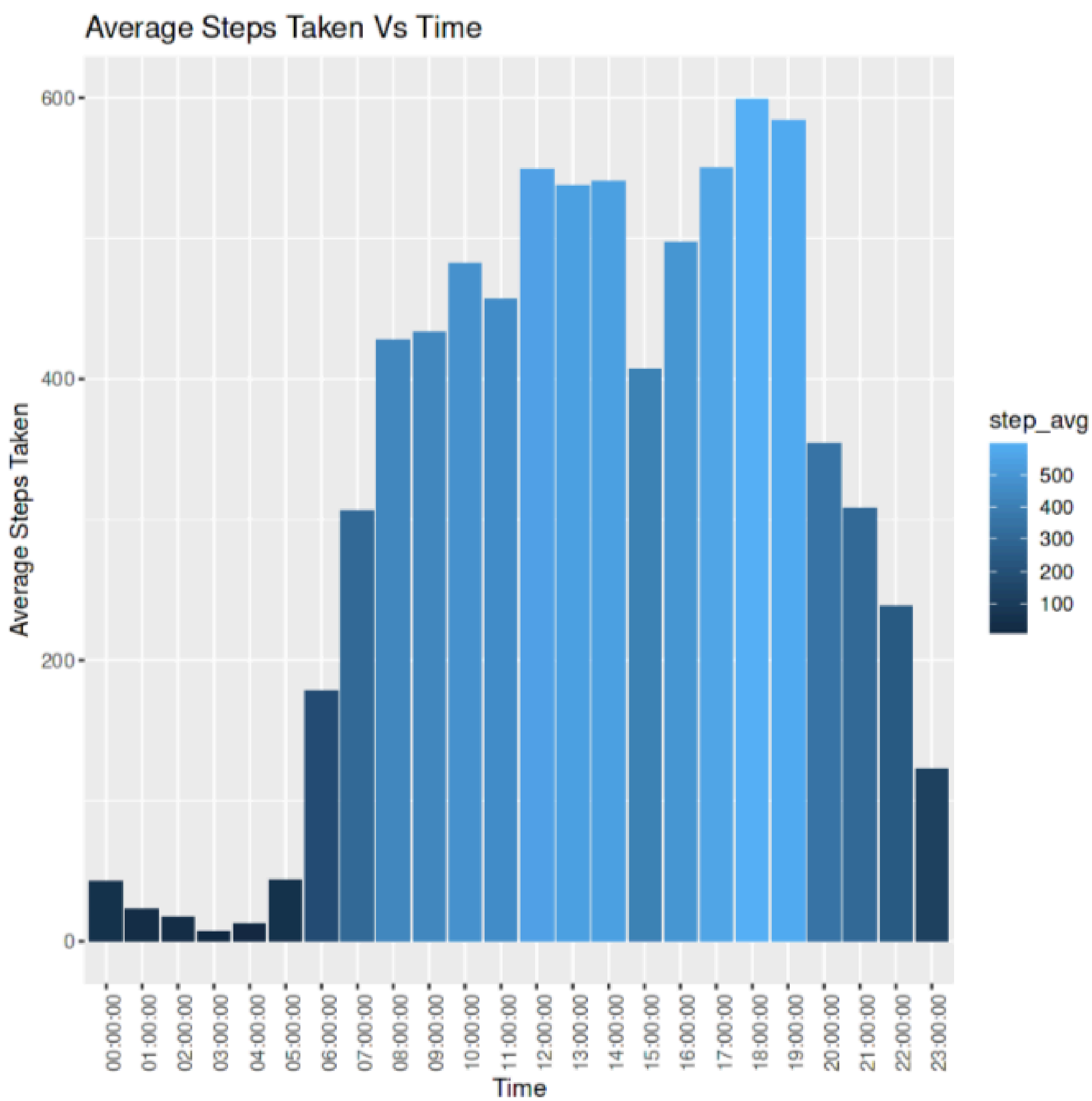
Viz 3 represents the average number of calories burned each hour:

```
ggplot(data = hourly_cal)+
 geom_col(mapping = aes(x = time, y = cal_avg,fill = cal_avg))+
 labs(title = "Average Calories Burned" , x = "Time" , y  = "Average Calorie burned")+
 theme(axis.text.x = element_text(angle = 90))
```



The final visualization, viz 4, is to represent the average steps taken in each hour:

```
ggplot(data = hourly_step) +
  geom_col(mapping  = aes(x = time, y = step_avg, fill = step_avg))+
  labs(title = "Average Steps Taken Vs Time", x = "Time", y = "Average Steps Taken")+
  theme(axis.text.x = element_text(angle = 90))
```



5. Recommendations -

These are the 3 main recommendations I would like to provide that would be effective, as these recommendations aim to use the insights gained from the data visualizations to help users improve their physical activity levels, sleep habits, and overall health and well-being...

Low Activity Levels:

From the 1st visualization, it is extremely obvious that the users spend very little time on fairly active and very active activities, which indicates a preference for inactivity or less active activities. This suggests that users may need reminders or encouragement to engage more actively. One recommendation is to send notifications through the Bellabeat app to remind users to be more active throughout their day. These reminders could be personalized based on the user's activity patterns and goals, encouraging them to incorporate more physical activity into their routines.

Sleep Duration:

The 2nd visualization shows that some users are getting less than 7 hours of sleep, below the recommended amount for adults. To address this issue, a new feature could be added to the Bellabeat app that allows users to input the time they need to wake up the next day. Based on this input, the app could calculate and recommend when users should prepare to sleep to achieve their desired wake-up time. This feature would help users track their sleep duration and make adjustments to improve their sleep quality and overall health.

Activity Patterns Throughout the Day:

The data in the 3rd and 4th visualizations show a spike in activity levels from 5:00 AM to 7:00 PM, with a dip at 3:00 PM, suggesting fatigue or decreased motivation. However, there's steady activity from 8:00 AM to 11:00 AM. Meeting the CDC's 10,000 steps recommendation is a must, so to address this, the app can send reminders during the 8:00 AM to 11:00 AM timeframe, encouraging users to increase activity levels. Suggestions like stretching or short walks can help maintain consistent activity throughout the day.