GOOGLE DATA ANALYTICS CAPSTONE PROJECT REPORT ON :

UNDERSTANDING CUSTOMER TRENDS: A DATA ANALYSIS FOR BELLABEAT

By Vibhuti Hardoneeya

Introduction

In today's fast-paced world, wellness technology is playing a crucial role in helping individuals lead healthier lives. Bellabeat, a high-growth smart wellness company, specializes in providing data-driven insights through its innovative health-tracking devices designed specifically for women. By analyzing user data, Bellabeat aims to enhance its products and expand its market presence.

This report follows a structured, data-driven approach to uncover key insights that can inform Bellabeat's business strategy. The analysis is conducted using the Google Data Analytics process, which consists of the following phases:

1. Ask: The primary objective is to understand how Bellabeat can leverage smart device usage data to optimize its products and marketing strategies. This phase involves defining key business questions and identifying areas of improvement.
2. Prepare: The dataset used in this analysis comes from a public source containing anonymized Fitbit user data. The data is reviewed for completeness, accuracy, and relevance before proceeding with further analysis.
3. Process: Cleaning and transforming the data ensures consistency and usability. This involves handling missing values, formatting date-time fields, and structuring data to facilitate meaningful analysis.
4. Analyze: Using statistical and visual analysis, patterns and trends in user activity, sleep, and fitness habits are identified. Key metrics such as daily step count, active minutes, and sleep duration are explored to uncover behavioral insights.
5. Share: Data visualizations, graphs, and dashboards are created to effectively communicate findings. These insights help in making data-driven recommendations for Bellabeat's strategic planning.
6. Act: Based on the findings, actionable recommendations are provided to enhance user engagement, improve product offerings, and refine marketing strategies to attract a wider audience.

Through this structured approach, the report aims to provide valuable insights that can help Bellabeat strengthen its position in the wellness technology market and cater more effectively to its users' needs.

About the Company

Bellabeat, a pioneering health-tech company, is on a mission to empower women through innovative smart wellness products. With a strong foundation in the market, the company has the potential to expand its reach and influence in the global smart device industry. By analyzing smart device fitness data, Bellabeat aims to uncover valuable insights into consumer behavior, identifying trends in activity levels, sleep patterns, and overall wellness habits. These insights will play a crucial role in shaping data-driven marketing strategies, allowing Bellabeat to enhance user engagement, optimize product offerings, and strengthen its competitive position. By leveraging data to better understand customer needs, Bellabeat can refine its approach and drive sustainable growth in the wellness technology sector

Products

- **Bellabeat app**: The Bellabeat app provides users with health data related to their activity, sleep, stress, menstrual cycle, and mindfulness habits. This data can help users better understand their current habits and make healthy decisions. The Bellabeat app connects to their line of smart wellness products.
- **Leaf**: Bellabeat's classic wellness tracker can be worn as a bracelet, necklace, or clip. The Leaf tracker connects to the Bellabeat app to track activity, sleep, and stress.
- **Time**: This wellness watch combines the timeless look of a classic timepiece with smart technology to track user activity, sleep, and stress. The Time watch connects to the Bellabeat app to provide you with insights into your daily wellness.
- **Spring**: This is a water bottle that tracks daily water intake using smart technology to ensure that you are appropriately hydrated throughout the day. The Spring bottle connects to the Bellabeat app to track your hydration levels.
- **Bellabeat membership**: Bellabeat also offers a subscription-based membership program for users. Membership gives users 24/7 access to fully personalized guidance on nutrition, activity, sleep, health and beauty, and mindfulness based on their lifestyle and goals.

1. ASK PHASE

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

2. PREPARE PHASE

2.1 Dataset used:

minuteSleep_merged
https://drive.google.com/file/d/1kx9MYbblwtk52AQAmg-Scj5I9OzoDBZh/view?usp=sharing

minutStepsNarrow_merged
https://drive.google.com/file/d/1r5g0RNhfdMe7vaQI7fG7JqXqDk4aHn1O/view?usp=sharing

minuteStepsWide_merged
https://drive.google.com/file/d/1qJ28-9NbsoPo-31sNsWjFR36bik-K5WK/view?usp=sharing

sleepDay_merged
https://drive.google.com/file/d/1jmEWkNIc47D6RGvfSfbWSuhB2zXema80/view?usp=sharing

weightLogInfo_merged
https://drive.google.com/file/d/1Qc7yTWEyuJ-xPP74xlBBRsXR3oErmM8q/view?usp=sharing

minuteIntensitiesNarrow_merged
https://drive.google.com/file/d/1wju_L9EPcXLQLWFWX9U5a-zI8qqOI4Aw/view?usp=sharing

minuteIntensitiesWide_merged
https://drive.google.com/file/d/1qJ28-9NbsoPo-31sNsWjFR36bik-K5WK/view?usp=sharing

minutMETsNarrow_merged
https://drive.google.com/file/d/1lRCOGrQSR0NTvkb1B1-jrqKTkpuBCo_S/view?usp=sharing

hourlyCalories_merged
https://drive.google.com/file/d/1wqN9EiQOFC2GXa8EYq0qOXtEHJw_DCjz/view?usp=sharing

hourlyIntensities_merged
https://drive.google.com/file/d/1KJUMCV6UL92-qyhwLFLoKjJtZSVxAUor/view?usp=sharing

hourlySteps_merged
https://drive.google.com/file/d/1S1CM_u-nTjQs2-TPrMYmg75EqFoIHbbe/view?usp=sharing

minuteCaloriesNarrow_merged
https://drive.google.com/file/d/1WBwPObYBKPmDZOSVvbJhvPxV-D5Nljfk/view?usp=sharing

minuteCaloriesWide_merged
https://drive.google.com/file/d/1aWrUuTo-SiS4ZYGzShwEZ3il3VSXLqUO/view?usp=sharing

dailiActivity_merged
https://drive.google.com/file/d/1rfiwO-5_f15YQMZ29HjVftxqOO3EEXtc/view?usp=sharing

dailyCalories_merged
https://drive.google.com/file/d/1U4sOsIiXwBsGHa7C4MD0IQEUVKDhWkll/view?usp=sharing

dailyIntensities_merged
https://drive.google.com/file/d/1QVVkgiERSS_-FDVtZTveq5you-IPl9B9/view?usp=sharing

dailySteps_merged
https://drive.google.com/file/d/1laRMY_ZfntbzqoHoqZ5JhxAJzXgTBwQM/view?usp=sharing

heartrate_seconds_merged
https://drive.google.com/file/d/1vNLUeBwFh-uxn73l3J5pFihEa_qy2YsV/view?usp=sharing

2.2 Information about the dataset:

These datasets were generated by respondents to a distributed survey via Amazon Mechanical Turk between 03.12.2016–05.12.2016. Thirty eligible Fitbit users consented to the submission of personal tracker data, including minute-level output for physical activity, heart rate, and sleep monitoring. Variation between output represents use of different types of Fitbit trackers and individual tracking behaviors / preferences.

2.3 Data Organization:

There are 18 CSV documents made available for the analysis. Each document represents different quantitative data tracked by Fitbit. The data is considered long since each row is one time point per subject, so each subject will have data in multiple rows.Every user has a unique ID and different rows since data is tracked by day and time.

2.4 Data Integrity and Credibility:

Due to the limitation of size (30 users) and not having any demographic information we could encounter a sampling bias. We are not sure if the sample is representative of the population as a whole. Another problem we would encounter is that the dataset is not current and also the time limitation of the survey (2 months long). That is why we will give our case study an operational approach.

3. PROCESS PHASE

I choose to use R for my analysis due to accessibility, the amount of data I will be working with, and being able to create data visualization to share my results and recommendations with stakeholders.

3.1 Setting up my environment by loading the packages

I will choose the packages that will help in working on my analysis and open them. Packages to be used for my analysis includes.

- Tidyverse
- Here
- Skimr
- Janitor
- lubridate

```
library(tidyverse)
library(here)
library(skimr)
library(janitor)
library(lubridate)
```

3.2 Importing dataset

```
> # Set the working directory (if not already set)
> setwd("C:/Users/lenovo/Downloads/COURSERA FITNESS TRACKER DATASETS/mturkfitbi$
>
> # Load all datasets
> daily_activity <- read.csv("dailyActivity_merged.csv", header = TRUE)
> daily_steps <- read.csv("dailySteps_merged.csv", header = TRUE)
> daily_sleep <- read.csv("sleepDay_merged.csv", header = TRUE)
> hourly_steps <- read.csv("hourlySteps_merged.csv", header = TRUE)
> hourly_calories <- read.csv("hourlyCalories_merged.csv", header = TRUE)
> hourly_intensities <- read.csv("hourlyIntensities_merged.csv", header = TRUE)
> minute_sleep <- read.csv("minuteSleep_merged.csv", header = TRUE)
> weight <- read.csv("weightLogInfo_merged.csv", header = TRUE)
>
```

3.3 Preview the Data

```
> head(daily_steps)
          Id ActivityDay StepTotal
1 1503960366   4/12/2016     13162
2 1503960366   4/13/2016     10735
3 1503960366   4/14/2016     10460
4 1503960366   4/15/2016      9762
5 1503960366   4/16/2016     12669
6 1503960366   4/17/2016      9705
```

3.4 Check the structure of the data sets using str()

```
> str(daily_activity)
'data.frame':   940 obs. of  15 variables:
 $ Id                      : num  1.5e+09 1.5e+09 1.5e+09 1.5e+09 1.5e+09 ...
 $ ActivityDate            : chr  "4/12/2016" "4/13/2016" "4/14/2016" "4/15/2016" ...
 $ TotalSteps              : int  13162 10735 10460 9762 12669 9705 13019 15506 10544 9819 ...
 $ TotalDistance           : num  8.5 6.97 6.74 6.28 8.16 ...
 $ TrackerDistance         : num  8.5 6.97 6.74 6.28 8.16 ...
 $ LoggedActivitiesDistance: num  0 0 0 0 0 0 0 0 0 0 ...
 $ VeryActiveDistance      : num  1.88 1.57 2.44 2.14 2.71 ...
 $ ModeratelyActiveDistance: num  0.55 0.69 0.4 1.26 0.41 ...
 $ LightActiveDistance     : num  6.06 4.71 3.91 2.83 5.04 ...
 $ SedentaryActiveDistance : num  0 0 0 0 0 0 0 0 0 0 ...
 $ VeryActiveMinutes       : int  25 21 30 29 36 38 42 50 28 19 ...
 $ FairlyActiveMinutes     : int  13 19 11 34 10 20 16 31 12 8 ...
 $ LightlyActiveMinutes    : int  328 217 181 209 221 164 233 264 205 211 ...
 $ SedentaryMinutes        : int  728 776 1218 726 773 539 1149 775 818 838 ...
 $ Calories                : int  1985 1797 1776 1745 1863 1728 1921 2035 1786 1775 ...
```

3.5 Cleaning and Formatting

3.5.1 Check the number of participants for each data set

```
> n_unique(daily_activity$Id)
[1] 33
> n_unique(daily_sleep$Id)
[1] 24
> n_unique(daily_steps$Id)
[1] 33
> n_unique(hourly_calories$Id)
[1] 33
> n_unique(hourly_intensities$Id)
[1] 33
> n_unique(hourly_steps$Id)
[1] 33
> n_unique(weight$Id)
[1] 8
```

3.5.2 Check for Duplicates

```
> sum(duplicated(daily_activity))
[1] 0
> sum(duplicated(daily_sleep))
[1] 3
> sum(duplicated(daily_steps))
[1] 0
> sum(duplicated(hourly_calories))
[1] 0
> sum(duplicated(hourly_intensities))
[1] 0
> sum(duplicated(hourly_steps))
[1] 0
```

3.5.3 Remove all Duplicates and Missing values

```
> daily_activity <- daily_activity %>%
+   distinct() %>%
+   drop_na()
> daily_sleep <- daily_sleep %>%
+   distinct() %>%
+   drop_na()
> daily_steps <- daily_steps %>%
+   distinct() %>%
+   drop_na()
> hourly_calories <- hourly_calories %>%
+   distinct() %>%
+   drop_na()
> hourly_intensities <- hourly_intensities %>%
+   distinct() %>%
+   drop_na()
> hourly_steps <- hourly_steps %>%
+   distinct() %>%
+   drop_na()
```

3.5.4 Clean and rename columns

```
clean_names(daily_activity)
daily_activity <- rename_with(daily_activity, tolower)
clean_names(daily_sleep)
daily_sleep <- rename_with(daily_sleep, tolower)
clean_names(daily_steps)
daily_steps <- rename_with(daily_steps, tolower)
clean_names(hourly_calories)
hourly_calories <- rename_with(hourly_calories, tolower)
clean_names(hourly_intensities)
hourly_intensities <- rename_with(hourly_intensities, tolower)
clean_names(hourly_steps)
hourly_steps <- rename_with(hourly_steps, tolower)
```

3.5.5 Make Date and Time columns consistent

```
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()))
hourly_calories <- hourly_calories %>%
 rename(date_time = activityhour) %>%
  mutate(date_time = as.POSIXct(date_time, format ="%m/%d/%Y %I:%M:%S
%p" , tz=Sys.timezone()))
hourly_intensities <- hourly_intensities %>%
 rename(date_time = activityhour) %>%
  mutate(date_time = as.POSIXct(date_time, format ="%m/%d/%Y %I:%M:%S
%p" , tz=Sys.timezone()))
hourly_steps<- hourly_steps %>%
 rename(date_time = activityhour) %>%
  mutate(date_time = as.POSIXct(date_time, format ="%m/%d/%Y %I:%M:%S
%p" , tz=Sys.timezone()))
```

3.6 Merging Datasets

```
> daily_activity_sleep <- merge(daily_activity, daily_sleep, by=c ("id", "date"))
```

4. ANALYZE PHASE

The analysis focuses on identifying user trends from Fitbit data to determine how these insights can inform Bellabeat's marketing strategy. By examining user activity, sleep patterns, and overall engagement with smart devices, this study aims to uncover valuable patterns that could help Bellabeat enhance its product offerings and tailor its marketing efforts to better meet consumer needs.

5. SHARE PHASE

5.1 Summarize and Explore Each Dataset

```
> daily_activity %>%
+   select(totalsteps,
+          totaldistance,
+          sedentaryminutes, calories) %>%
+   summary()
```

Discoveries from the Summary

```
  totalsteps     totaldistance    sedentaryminutes     calories
 Min.   :    0   Min.   : 0.000   Min.   :   0.0   Min.   :   0
 1st Qu.: 3790   1st Qu.: 2.620   1st Qu.: 729.8   1st Qu.:1828
 Median : 7406   Median : 5.245   Median :1057.5   Median :2134
 Mean   : 7638   Mean   : 5.490   Mean   : 991.2   Mean   :2304
 3rd Qu.:10727   3rd Qu.: 7.713   3rd Qu.:1229.5   3rd Qu.:2793
 Max.   :36019   Max.   :28.030   Max.   :1440.0   Max.   :4900
```

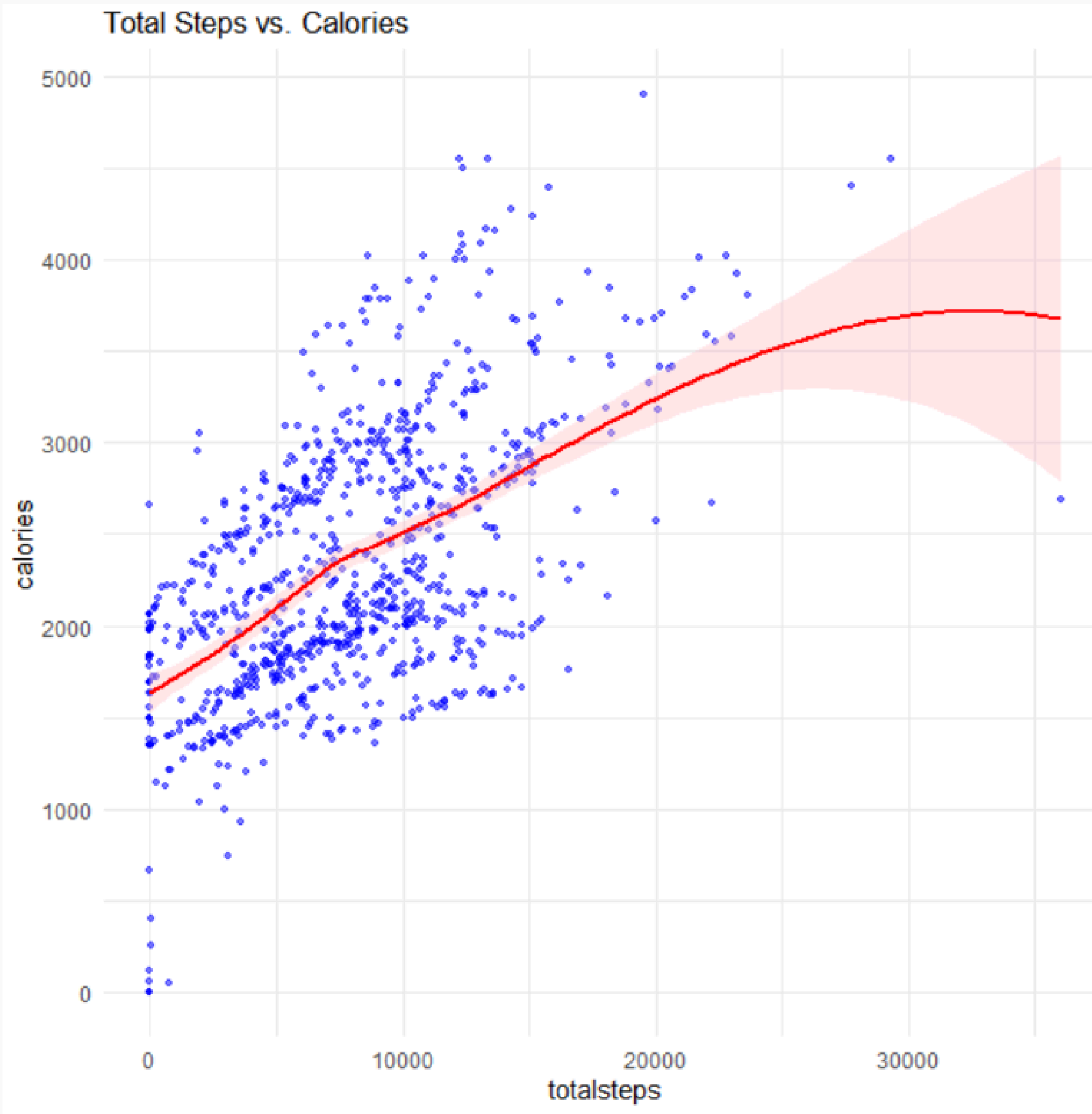
- **Average very active minutes : 21.16**
- **Average Fairly Active Minutes : 13.56**
- **Average lightly active hours : 3.21**
- **Average Sedentary hours : 16.52**
- **Distinct Id of users tracking activity : 33**
- **Average steps : 7638**
- **Average Distance : 5.48**
- **Average calories : 2303.60**
- **Distinct Id of users tracking heart rate : 14**
- **Average heartRate: 77.328**
- **Minimum heartRate : 36**
- **Maximum heartRate : 203**
- **Distinct Id of users tracking Sleep : 24**
- **Average hours sleep : 6.99**
- **Minimum hours sleep : 0.96**
- **Maximum horse sleep : 13.26**
- **Average hours in bed : 7.64**
- **Distinct Id of users tracking weight : 8**
- **Average weight : 72.05**

- **Minimum weight : 52.59**
- **Maximum weight : 133.5**
- Average total steps is 7638 in a day. The daily recommended amount of steps to be taken per day is 7500
- Sedentary minutes on an average is 991(~17 hours). This needs to be reduced.
- Majority of the participants are light users.
- Participants sleep for an average of 419 minutes (~7 hours).
- A total of 97 calories is burned per hour on average.

5.2 Steps taken and Calories burned

```
> ggplot(data = daily_activity, aes(x = totalsteps, y = calories)) +
+   geom_point(color = "blue", size = 1, alpha = 0.5) +
+   geom_smooth(color = "red", fill = "pink", method = "loess") +
+   labs(title = "Total Steps vs. Calories") +
+   theme_minimal()
```

Plot:



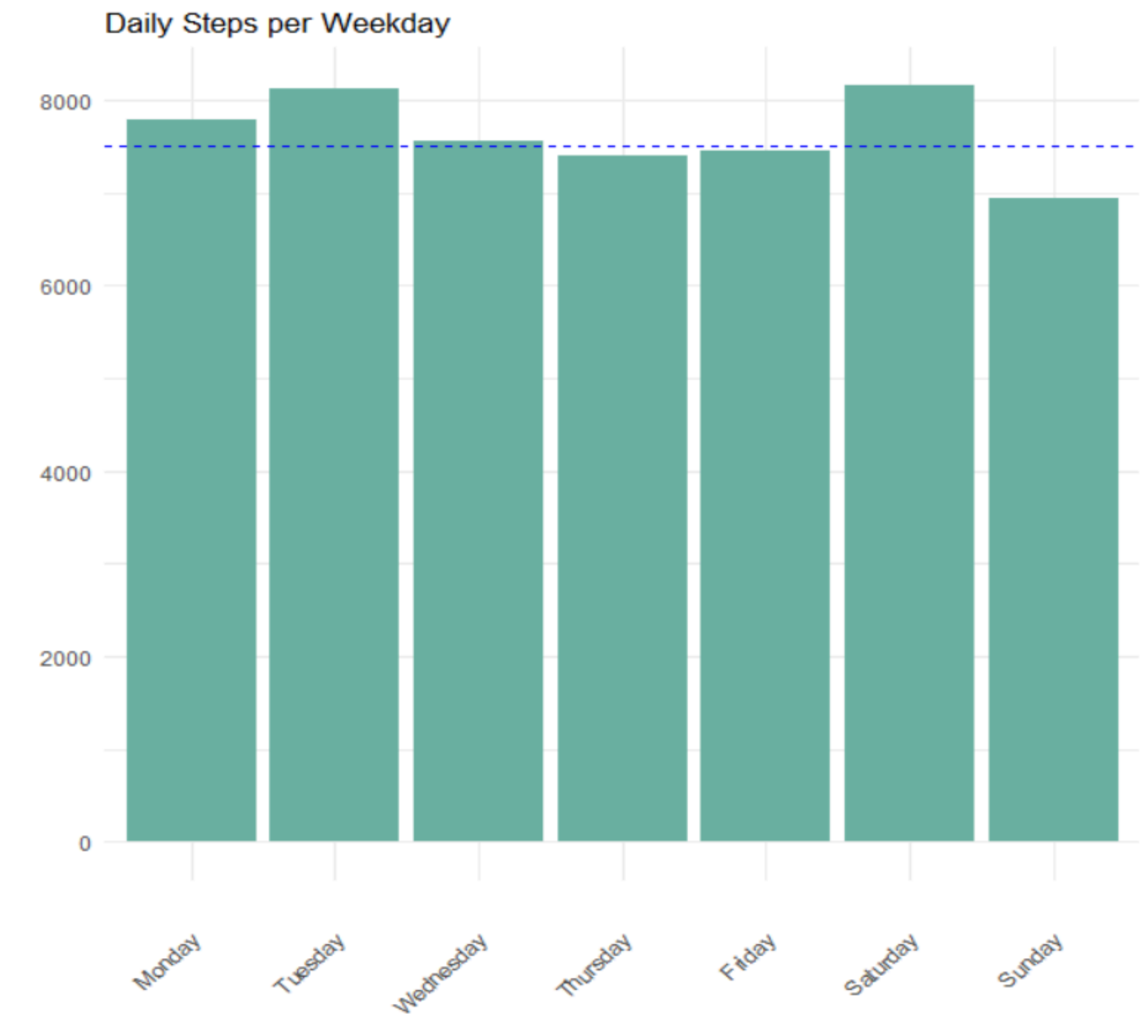
positive correlation between total steps taken and the amount of calories burned.

5.3 Steps per Weekday

```
> weekday_steps <- daily_activity %>%
+   mutate(weekday = weekdays(date))
> weekday_steps$weekday <-ordered(weekday_steps$weekday, levels=c("Monday", "Tuesday", "Wednesday"$
+                                                                 "Friday", "Saturday", "Sunday"))
> weekday_steps <-weekday_steps %>%
+   group_by(weekday) %>%
+   summarize (daily_steps = mean(totalsteps))
>
> head(weekday_steps)
# A tibble: 6 × 2
  weekday   daily_steps
  <ord>           <dbl>
1 Monday          7781.
2 Tuesday         8125.
3 Wednesday       7559.
4 Thursday        7406.
5 Friday          7448.
6 Saturday        8153.
```

```
> ggplot(weekday_steps, aes(weekday, daily_steps)) +
+   geom_col(fill = "#69b3a2") +
+   geom_hline(yintercept = 7500, color = "blue", linetype = "dashed") +
+   labs(title = "Daily Steps per Weekday", x= "", y = "") +
+   theme_minimal() +
+   theme(axis.text.x = element_text(angle = 45, vjust = 0.5, hjust = 1))
```

Plot:

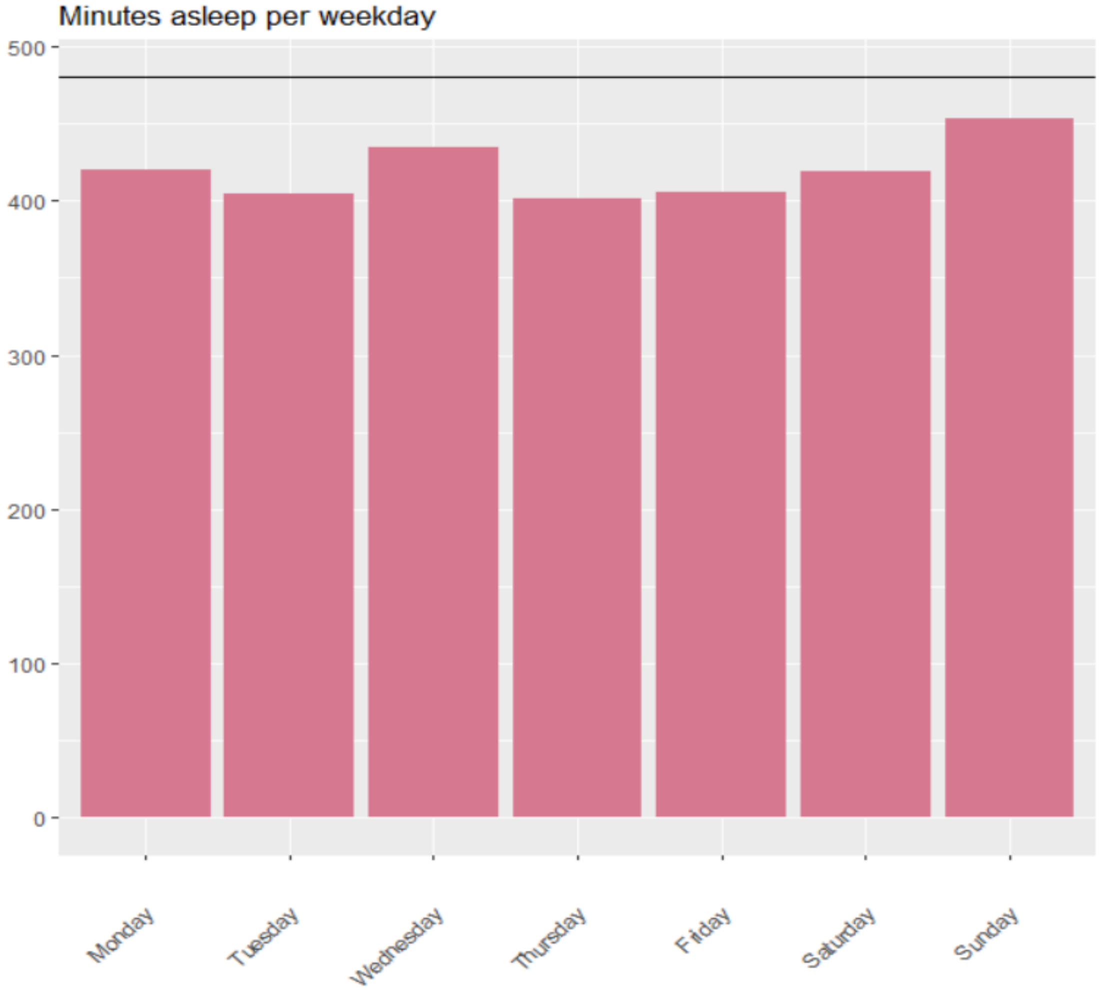


5.4 Sleeps per Weekday

```
> weekday_sleep <- daily_sleep %>%
+   mutate(weekday = weekdays(date))
> weekday_sleep$weekday <-ordered(weekday_sleep$weekday, levels=c("Monday", "Tuesday", "Wednesday"$
+                                                                  "Friday", "Saturday", "Sunday"))
>
> weekday_sleep <-weekday_sleep %>%
+   group_by(weekday) %>%
+   summarize (daily_sleep = mean(totalminutesasleep))
>
> head(weekday_sleep)
# A tibble: 6 × 2
  weekday   daily_sleep
  <ord>           <dbl>
1 Monday           420.
2 Tuesday          405.
3 Wednesday        435.
4 Thursday         401.
5 Friday           405.
6 Saturday         419.
```

```
> ggplot(weekday_sleep, aes(weekday, daily_sleep)) +
+   geom_col(fill = "#db7992") +
+   geom_hline(yintercept = 480) +
+   labs(title = "Minutes asleep per weekday", x= "", y = "") +
+   theme(axis.text.x = element_text(angle = 45,vjust = 0.5, hjust = 1))
```

Plot:



Discoveries from the above Analysis:

- In the graphs above we can deduce that users don't take the recommended amount of sleep of 8 hours
- Users take recommended number of 7500 steps a day excepts for Sundays.

5.5 Hourly intensities throughout the day

Split the datetime column into date and time columns

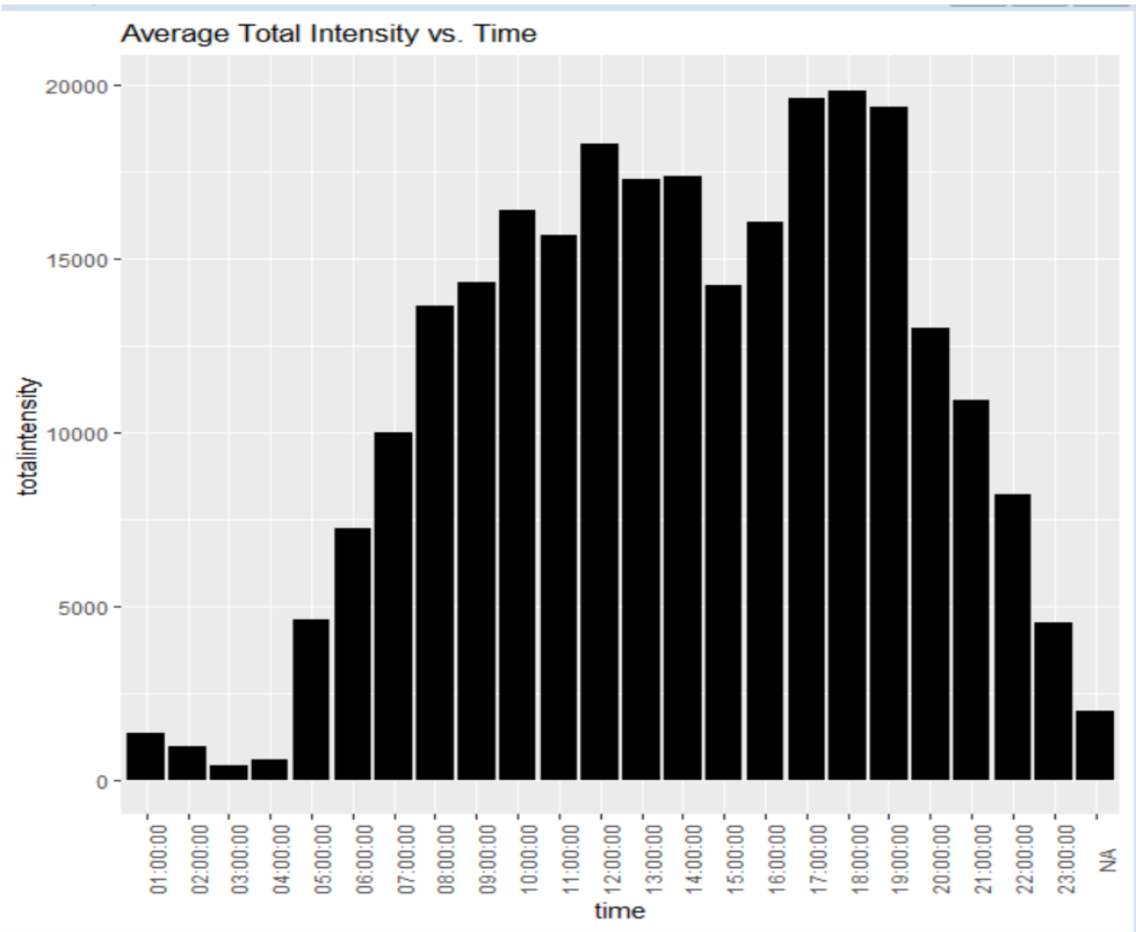
```
hourly_intensities <- hourly_intensities %>%
separate(date_time, into = c("date", "time"), sep= " ")

head(hourly_intensities)

hourly_intensities <- hourly_intensities %>%
group_by(time) %>%
drop_na() %>%
summarise(mean_total_int = mean(totalintensity)
```

```
head(hourly_intensities)
          id       date     time totalintensity averageintensity
1503960366 2016-04-12     <NA>             20         0.333333
1503960366 2016-04-12 01:00:00              8         0.133333
1503960366 2016-04-12 02:00:00              7         0.116667
1503960366 2016-04-12 03:00:00              0         0.000000
1503960366 2016-04-12 04:00:00              0         0.000000
1503960366 2016-04-12 05:00:00              0         0.000000
```

Visualization of the hourly intensities data

```
> ggplot(data = hourly_intensities, aes(x = time, y = totalintensity)) +
+   geom_col(fill = "black") +
+   theme(axis.text.x = element_text(angle = 90)) +
+   labs(title = "Average Total Intensity vs. Time")
```

Plot:



Discoveries after analyzing hourly Intensities

- Users exhibit the highest activity levels between 5 AM and 10 PM.
- Peak activity occurs between 5 PM and 7 PM, likely as individuals finish work and engage in exercise or walking.
- This time frame presents an opportunity for Bellabeat to send targeted reminders and motivationa
- l messages.
- Utilizing the Bellabeat app, users can be encouraged to stay active and maintain a consistent fitness routine.

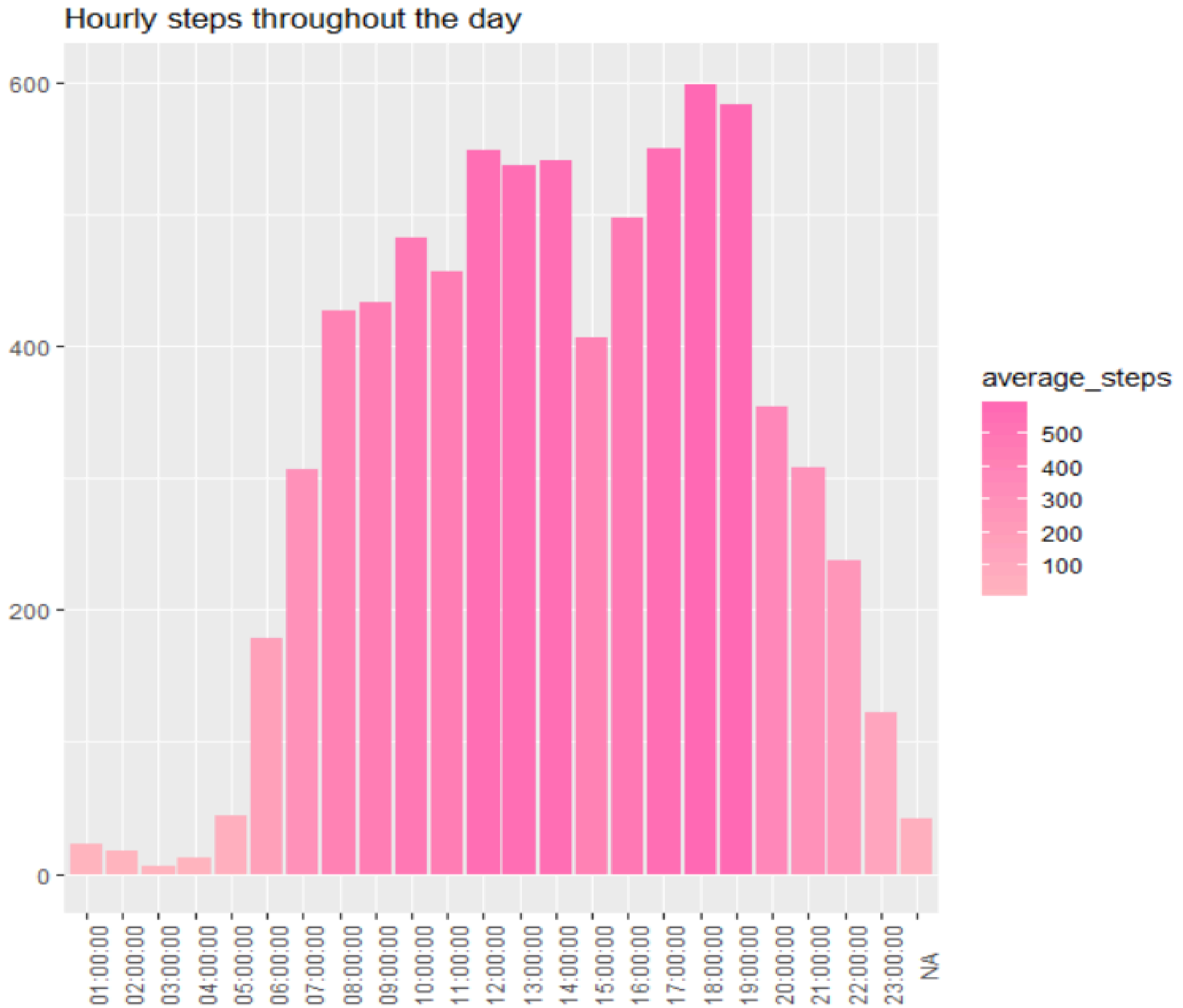
5.6 Hourly Steps throughout the day

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

> head(hourly_steps)
          id       date     time steptotal
1 1503960366 2016-04-12     <NA>       373
2 1503960366 2016-04-12 01:00:00       160
3 1503960366 2016-04-12 02:00:00       151
4 1503960366 2016-04-12 03:00:00         0
5 1503960366 2016-04-12 04:00:00         0
6 1503960366 2016-04-12 05:00:00         0
>
```

visualization of the hourly steps throughout the day

```
> hourly_steps %>%
+   group_by(time) %>%
+   summarize(average_steps = mean(steptotal)) %>%
+   ggplot() +
+   geom_col(mapping = aes(x = time, y = average_steps, fill = average_steps)) +
+   labs(title = "Hourly steps throughout the day", x = "", y = "") +
+   scale_fill_gradient(low = "#ffb6c1", high = "#ff69b4") +
+   theme(axis.text.x = element_text(angle = 90))
> |
```

Plot:



Discoveries:

- Users exhibit higher activity levels between 8 AM and 5 PM.
- The highest step counts are recorded between 12 PM to 2 PM and 5 PM to 7 PM.
- This pattern suggests that users likely take more steps during lunch breaks (12 PM - 2 PM) and after work hours (5 PM - 7 PM).
- The data indicates that most users may belong to the working-class demographic, with activity peaks aligning with typical work schedules.

5.7 Type of user based on the number of days smart device was used

Calculate the number of users that use their smart device on a daily basis, classifying our sample into three categories knowing that the duration of the survey is 31 days:

- High user — users who use their device for 21–31 days
- Moderate user — users who use their device for 10–20 days
- Low user — users who use their device for 1–10 days

```
> daily_use <- daily_activity_sleep %>%
+   group_by(id) %>%
+   summarize(days_used=sum(n())) %>%
+   mutate(user_type= case_when(
+     days_used >= 1 & days_used <= 10 ~ "low user",
+     days_used >= 11 & days_used <= 20 ~ "moderate user",
+     days_used >= 21 & days_used <= 31 ~ "high user",
+   ))
>
> head(daily_use)
# A tibble: 6 × 3
          id days_used user_type
       <dbl>     <int> <chr>
1 1503960366        25 high user
2 1644430081         4 low user
3 1844505072         3 low user
4 1927972279         5 low user
5 2026352035        28 high user
6 2320127002         1 low user
```

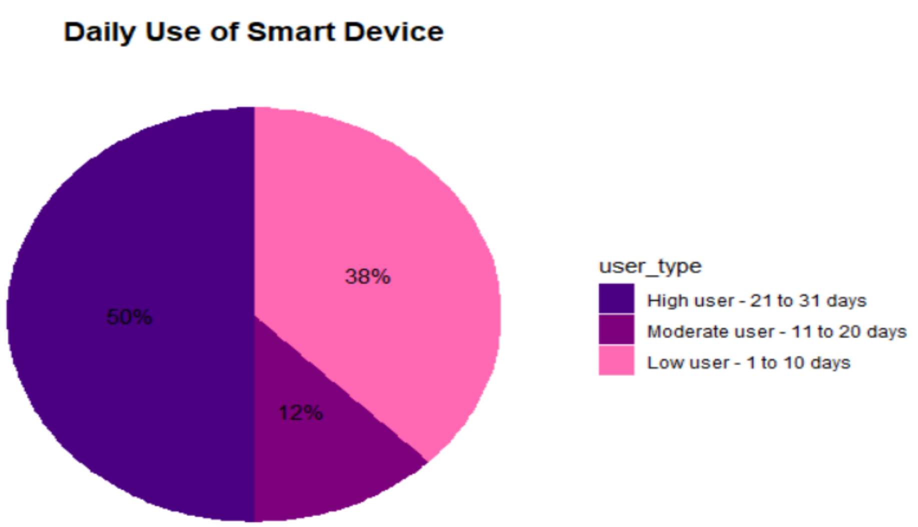
Create a percentage data frame to better visualize the results in the graph

```
> daily_use_percent <- daily_use %>%
+   group_by(user_type) %>%
+   summarise(total = n()) %>%
+   mutate(totals = sum(total)) %>%
+   group_by(user_type) %>%
+   summarise(total_percent = total / totals) %>%
+   mutate(labels = scales::percent(total_percent))
>
> daily_use_percent$user_type <- factor(daily_use_percent$user_type, levels = c("high user", "mode$
>
> head(daily_use_percent)
# A tibble: 3 × 3
  user_type     total_percent labels
  <fct>                 <dbl> <chr>
1 high user             0.5   50%
2 low user              0.375 38%
3 moderate user         0.125 12%
```

Make a visualization of the smart device usage per user

```
> daily_use_percent %>%
+   ggplot(aes(x = "", y = total_percent, fill = user_type)) +
+   geom_bar(stat = "identity", width = 1) +
+   coord_polar("y", start = 0) +
+   theme_minimal() +
+   theme(axis.title.x = element_blank(),
+         axis.title.y = element_blank(),
+         panel.border = element_blank(),
+         panel.grid = element_blank(),
+         axis.ticks = element_blank(),
+         axis.text.x = element_blank(),
+         plot.title = element_text(hjust = 0.5, size = 14, face = "bold")) +
+   geom_text(aes(label = labels),
+             position = position_stack(vjust = 0.5)) +
+   scale_fill_manual(values = c("#4B0082", "#800080", "#FF69B4"),
+                     labels = c("High user - 21 to 31 days",
+                                "Moderate user - 11 to 20 days",
+                                "Low user - 1 to 10 days")) +
+   labs(title = "Daily Use of Smart Device")
> 
```

Plot:



From the above visualization, the following insights can be deduced:

- **50%** of the users in the sample are frequent users, engaging with their device between **21 to 31 days**.
- **12%** of the users fall into the moderate category, using their device for **11 to 20 days**.
- **38%** of the users rarely use their device, with usage recorded for fewer than **10 days**.

6. ACT PHASE

- **Bellabeat: Data-Driven Insights for Marketing Strategy**

Bellabeat was founded with the mission to empower women by providing them with data-driven insights into their health and daily habits. By leveraging smart wellness technology, the company aims to help users make informed decisions about their physical activity, sleep patterns, and overall well-being.

To support this mission, an analysis was conducted on user activity, focusing on average steps taken, hours of sleep, and calories burned. Additionally, patterns related to peak activity hours, most active days, and user engagement trends were examined. The analysis was based on Fitbit Fitness data, which served as a proxy for understanding potential Bellabeat user behavior.

However, for a more precise and targeted analysis, it is recommended that future research utilize tracking data from Bellabeat's own devices. The current dataset presents limitations, as it is based on a small sample and lacks demographic details, which may introduce bias.

- ### Target Audience

Insights from the data suggest that young and adult women with full-time jobs constitute the primary target audience for Bellabeat. According to hourly intensity data, these women engage in light physical activity, often during specific time frames, likely influenced by work schedules. It is crucial to continue identifying behavioral trends to develop a marketing strategy tailored to this demographic while also promoting healthy habits.

Recommendations for the Bellabeat App

1. Daily Step Notifications

- Research from the Centers for Disease Control and Prevention (CDC) indicates that taking 8,000 steps per day is associated with a 51% lower risk of mortality, while 12,000 steps per day reduces the risk by 65% compared to 4,000 steps.
- Bellabeat can incorporate daily step notifications into the app, reminding users of their progress throughout the day.
- Personalized step goals and motivational messages can encourage users to stay active while educating them on the health benefits of walking.

2. Sleep Tracking and Alerts

- Data analysis suggests that many users do not meet the recommended daily sleep duration.
- A feature can be introduced in the app that allows users to set a preferred bedtime and receive reminders before their scheduled sleep time.
- The app could also integrate a sleep quality assessment, offering insights on how to improve rest and recovery.

3. Personalized Health and Nutrition Notifications

- Users with specific wellness goals, such as weight management, could receive tailored recommendations on daily calorie intake and nutritious meal suggestions.
- The app could leverage user input on fitness and health goals to deliver customized advice and insights.

4. Reward System for User Engagement

- A gamification approach could enhance engagement by introducing a reward system based on user activity levels.
- Users could unlock achievement levels based on their consistency in meeting activity goals.
- Accumulated reward points could be converted into discounts on Bellabeat products or wellness programs, reinforcing brand loyalty.

Conclusion

By leveraging data insights, Bellabeat can optimize its marketing strategy to align with the lifestyle patterns of its target audience. Implementing personalized app features, user engagement incentives, and a strong digital presence will not only enhance user experience but also strengthen Bellabeat's position in the global wellness technology market.