CASE STUDY: CAPSONE PROJECT ON BALLABEAT COMPANY

SUFIYANU ABUBAKAR

TABLE OF THE CONTENTS

CASE STUDY 2: CAPSONE PROJECT ON COUSERA

R <PROGRAMING LANGUAGE>

INTRODUCTION :

About the company Urška Sršen and Sando Mur founded Bellabeat, a high-tech company that manufactures health-focused smart products. Sršen used her background as an artist to develop beautifully designed technology that informs and inspires women around the world. 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. By 2016, Bellabeat had opened offices around the world and launched multiple products. Bellabeat products became available through a growing number of online retailers in addition to their own e-commerce channel on their website. The company has invested in traditional advertising media, such as radio, out-of-home billboards, print, and television, but focuses on digital marketing extensively. Bellabeat invests year-round in Google Search, maintaining active Facebook and Instagram pages, and consistently engages consumers on Twitter. Additionally, Bellabeat runs video ads on Youtube and display ads on the Google Display Network to support campaigns around key marketing dates. Sršen knows that an analysis of Bellabeat's available consumer data would reveal more opportunities for growth. She has asked the marketing analytics team to focus on a Bellabeat product and analyze smart device usage data in order to gain insight into how people are already using their smart devices. Then, using this information, she would like high-level recommendations for how these trends can inform Bellabeat marketing strategy.

İ will report the project by using the following recommendation

Ask > prepare > process > analyze > share > act phase

DESCRIPTION OF THE FOLLOWING PHASES

ASK PHASE:

In the ask phase we want to identify the problem and we want see how can solve the problem first you have re think about the strategies of the company before you continue your analysis smart device usage data in order to gain insight into how consumers use non-Bellabeat smart devices. She wants you to select one Bellabeat product to apply these insights to your presentation

What is the problem you are trying to solve?

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How could these trends help influence Bellabeat marketing strategy?

KEY TASKS

- Identify the business tasks
  These trends and analysis will let the marketing team know, how the users are using their products and what else can be done to increase their engagement level and hence increase usability for current or prospective users

.

- Consider key stakeholders
  - ✓ Urška Sršen: Bellabeat's cofounder and Chief Creative Officer
  - ✓ Sando Mur: Mathematician and Bellabeat's cofounder
  - ✓ key member of the Bellabeat executive team ○ Bellabeat marketing analytics team: A team of data analysts responsible for collecting, analyzing, and reporting data that helps guide Bellabeat's marketing strategy. You joined this team six months ago and have been busy learning about Bellabeat's mission and business goals

PRODUCT

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.

PREPARE PHASE:

encourages you to use public data that explores smart device users' daily habits. She points you to a specific data set does the data has ROCCC yes data is Reliable from the Kaggle source and is originally from the company license and is comprehensive current cited in the reliable source download the dataset and sort and filter the data

- ✓ DailyActivity_merged
- ✓ sleepDay_merged
- ✓ hourlyCalories_merged
- ✓ weigthlogInfo_merged

PROCESS PHASE

Then, process your data for analysis using the following

```
install.packages("tidyverse")
install.packages("janitor")
install.packages("skimr"
install.packages("lubridate")
library("tidyverse")
```

```
── Attaching packages ─────────────────────────────────── tidyverse 1.3.2 ──
✓ ggplot2 3.4.0      ✓ purrr   1.0.0
✓ tibble  3.1.8      ✓ dplyr   1.0.10
✓ tidyr   1.2.1      ✓ stringr 1.5.0
✓ readr   2.1.3      ✓ forcats 0.5.2
── Conflicts ────────────────────────────────────── tidyverse_conflicts() ──
✖ dplyr::filter() masks stats::filter()
✖ dplyr::lag()    masks stats::lag()
>
```

Then I use str() function in order to put data in the RStudio

```
str(dailyActivity_merged)
str(sleepDay_merged)
str(weightLogInfo_merged)
```

The output

```
str(dailyActivity_merged)
bble [857 × 18] (S3: tbl_df/tbl/data.frame)
 id                        : num [1:857] 1503960366 1503960366 1503960366
03960366 1503960366 ...
 activity_date             : Date[1:857], format: "2020-04-12" "2020-04-1
  ...
 total_steps               : num [1:857] 13162 10735 10460 9762 12669 ...
 total_distance            : num [1:857] 8.5 6.97 6.74 6.28 8.16 ...
 tracker_distance          : num [1:857] 8.5 6.97 6.74 6.28 8.16 ...
 logged_activities_distance: num [1:857] 0 0 0 0 0 0 0 0 0 0 ...
 very_active_distance      : num [1:857] 1.88 1.57 2.44 2.14 2.71 ...
 moderately_active_distance: num [1:857] 0.55 0.69 0.4 1.26 0.41 ...
 light_active_distance     : num [1:857] 6.06 4.71 3.91 2.83 5.04 ...
 sedentary_active_distance : num [1:857] 0 0 0 0 0 0 0 0 0 0 ...
 very_active_minutes       : num [1:857] 25 21 30 29 36 38 42 50 28 19
 .
 fairly_active_minutes     : num [1:857] 13 19 11 34 10 20 16 31 12 8 ...
 lightly_active_minutes    : num [1:857] 328 217 181 209 221 164 233 264
5 211 ...
 sedentary_minutes         : num [1:857] 728 776 1218 726 773 ...
 calories                  : num [1:857] 1985 1797 1776 1745 1863 ...
 day_of_week               : Ord.factor w/ 7 levels "Sun"<"Mon"<"Tue"<..:
2 3 4 5 6 7 1 2 3 ...
 total_active_hours        : num [1:857] 6.1 4.28 3.7 4.53 4.45 3.7 4.85
```

```
 str(sleepDay_merged)
pc_tbl_ [413 × 8] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ id                 : num [1:413] 1503960366 1503960366 1503960366 15039
60366 1503960366 ...
 $ sleep_day          : Date[1:413], format: "2020-04-12" "2020-04-13" ...
 $ total_sleep_records : num [1:413] 1 2 1 2 1 1 1 1 1 1 ...
 $ total_minutes_asleep: num [1:413] 327 384 412 340 700 304 360 325 361 43
 ...
 $ total_time_in_bed   : num [1:413] 346 407 442 367 712 320 377 364 384 44
 ...
 $ hours_in_bed        : num [1:413] 5.77 6.78 7.37 6.12 11.87 ...
 $ hours_asleep        : num [1:413] 5.45 6.4 6.87 5.67 11.67 ...
 $ time_taken_to_sleep : num [1:413] 19 23 30 27 12 16 17 39 23 19 ...
 - attr(*, "spec")=
  .. cols(
  ..   Id = col_double(),
  ..   SleepDay = col_character(),
  ..   TotalSleepRecords = col_double(),
  ..   TotalMinutesAsleep = col_double(),
  ..   TotalTimeInBed = col_double()
  .. )
 - attr(*, "problems")=<externalptr>
```

```
 attr(*, "problems")=<external ptr>
str(weightLogInfo_merged)
bble [67 × 8] (S3: tbl_df/tbl/data.frame)
 id              : num [1:67] 1503960366 1503960366 1927972279 2873212765
73212765 ...
 date            : POSIXct[1:67], format: "2016-05-02 23:59:59" "2016-05-
 23:59:59" ...
 weight_kg       : num [1:67] 52.6 52.6 133.5 56.7 57.3 ...
 weight_pounds   : num [1:67] 116 116 294 125 126 ...
 bmi             : num [1:67] 22.6 22.6 47.5 21.5 21.7 ...
 is_manual_report: logi [1:67] TRUE TRUE FALSE TRUE TRUE TRUE ...
 log_id          : num [1:67] 1462233599000 1462319999000 1460509732000 1
1283199000 1463097599000 ...
 bmi2            : chr [1:67] "Healthy" "Healthy" "Overweight" "Healthy"
.
```

As you can see all the file is converted to CHR instead of the date format

Let us use date () to converted the data DailyActivity_merged and also sleepDay_merged

```
#We convert the daily activity to date as  identically 12am
dailyActivity_merged$activity_date <- as.Date(dailyActivity_merged$act
sleepDay_merged$sleep_day <- as.Date(sleepDay_merged$sleep_day, "%m/%d
```

For the stake holder to understand the following project we have to use the weightlogınfo_merged so let use the paste_date_time so that we differentiated the Am and Pm of the following day

```
#We convert the daily activity to date as  identically 12am
dailyActivity_merged$activity_date <- as.Date(dailyActivity_merged$act
sleepDay_merged$sleep_day <- as.Date(sleepDay_merged$sleep_day, "%m/%d
```

This for the DailyActivity_merged

```
dailyActivity_merged$total_active_hours = round((dailyActivity_merged$very_active_minutes
+dailyActivity_merged$fairly_active_minutes + dailyActivity_merged$lightly_active_minutes)/60, digits = 2)
dailyActivity_merged$sedentary_hours = round((dailyActivity_merged$sedentary_minutes)/60, digits = 2)
```

sleepDay_merged

```
sleepDay_merged$hours_in_bed = round((sleepDay_merged$total_time_in_bed)/60, digits = 2)
sleepDay_merged$hours_asleep = round((sleepDay_merged$total_minutes_asleep)/60, digits = 2)
sleepDay_merged$time_taken_to_sleep = (sleepDay_merged$total_time_in_bed - sleepDay_merged$total_minutes_asleep)
```

ANALYSE PHASE:

Now the data is stored appropriately and has been prepared for analysis, start putting it to work

- ✓ the average of the hour sleeping in bed and also the minute taken completely total average
- ✓ the total days that we are must active sleeping
- ✓ the relationship between hour and sedentary hour
- ✓ the relationship between the hours and minute
- ✓ daily over whelmed

let u used the summary () function to identify the following strategies

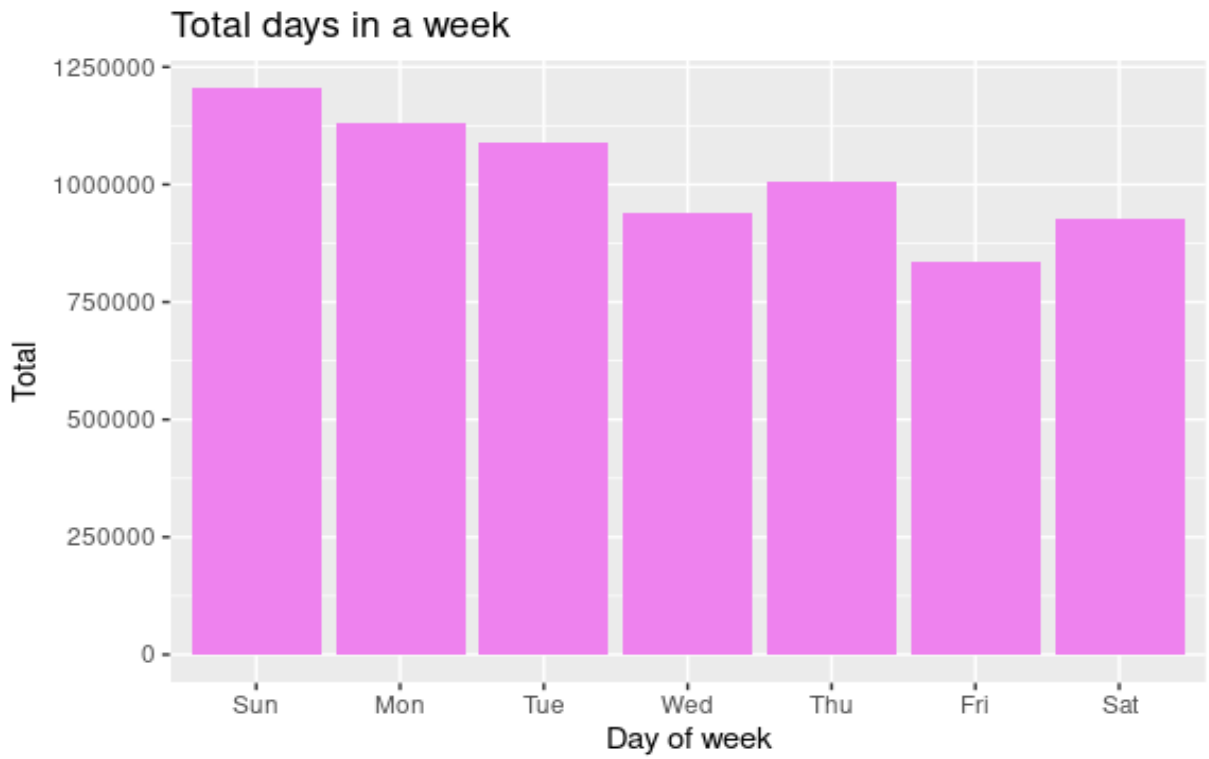
```
summary(dailyActivity_merged$total_steps)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0    4920    8053    8319   11100   36019
summary(dailyActivity_merged$sedentary_hours)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00   12.02   17.00   15.87   19.80   23.98
summary(dailyActivity_merged$very_active_minutes)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.00    0.00    7.00   23.21   36.00  210.00
summary(sleepDay_merged$hours_asleep)
 Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.970   6.020   7.220   6.992   8.170  13.270
```

This is the summary of the following data so let us move to another steps.

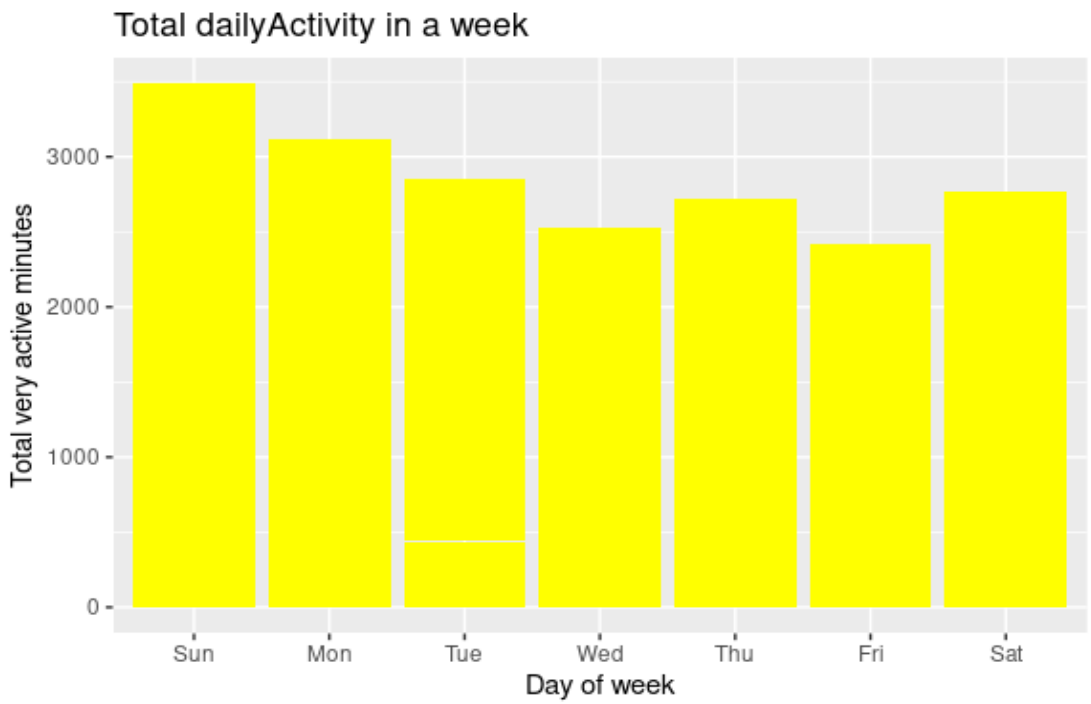
SHARE PHASE

We look at the days that we are most Actives

```
> options(scipen = 999)
> ggplot(data = dailyActivity_merged) +
+     aes(x = day_of_week, y = total_steps) +
+     geom_col(fill =  "violet") +
+     labs(x = 'Day of week', y = "Total", title = "Total days in a
week")
> ggsave('total_steps.png')
Saving 3.29 x 1.23 in image
>
```



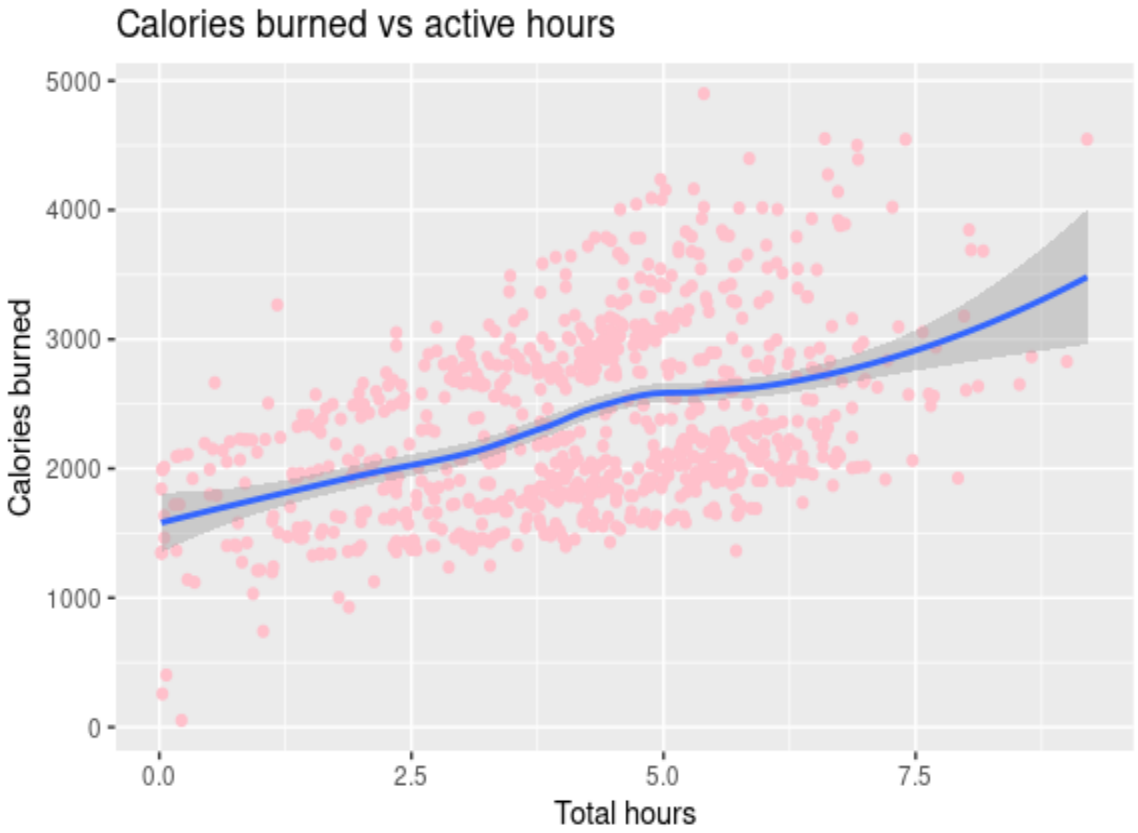

```
> ggplot(data = dailyActivity_merged) + aes(x = day_of_week, y = ve
ry_active_minutes) + geom_col(fill = "yellow") + labs(x = 'Day of w
eek', y = "Total very active minutes", title = "Total dailyActivity
in a week")
> ggsave("total_activity.png")
Saving 6.01 x 4.11 in image
>
```



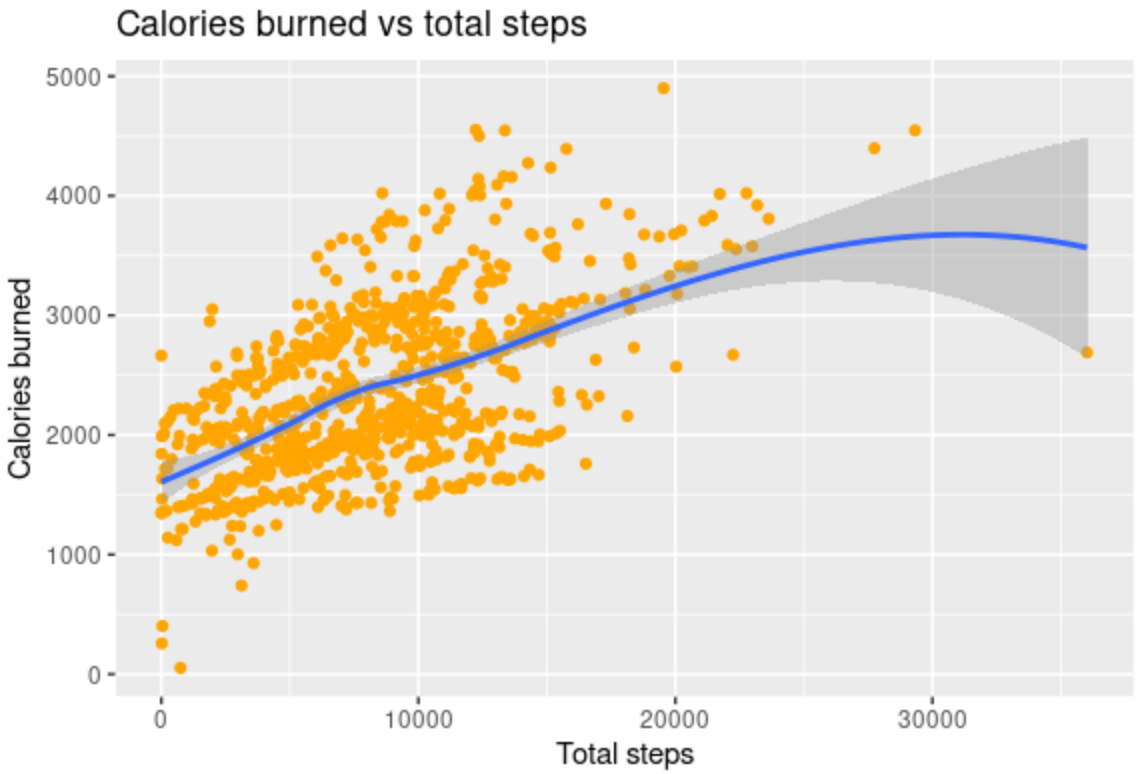

```
> ggplot(data = dailyActivity_merged) +
+     aes(x = day_of_week, y = calories) +
+     geom_col(fill =  "orange") +
+     labs(x = "Day of week", y = "Calories burned", title = "Total
calories burned in a week")
> ggsave("total_calories.png")
Saving 6.01 x 4.11 in image
>
```

✓ the relationship between between the active hour and sedentary active hours

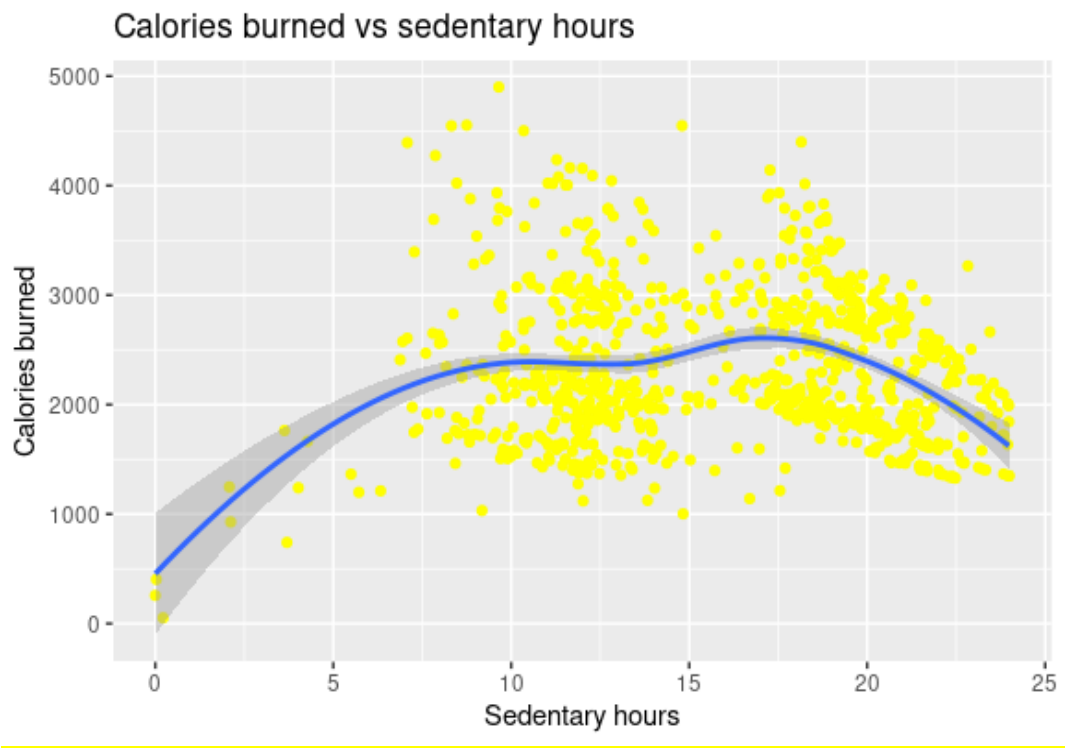
```
> ggplot(data = dailyActivity_merged) +
+     aes(x= total_active_hours, y = calories) +
+     geom_point(color = "pink") +
+     geom_smooth() +
+     labs(x = "Total hours", y = "Calories burned", title = "Calor
ies burned vs active hours")
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
> ggsave("calories_burned_vs_active_hours.png")
Saving 6.01 x 4.11 in image
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
>
```




```
 ggplot(data = dailyActivity_merged) +
     aes(x= total_steps, y = calories) +
     geom_point(color = "orange") +
     geom_smooth() +
     labs(x = "Total steps", y = "Calories burned", title = "Calor
es burned vs total steps")
geom_smooth()` using method = 'loess' and formula = 'y ~ x'
 ggsave("calories_burned_vs_total_steps.png")
aving 6.01 x 4.11 in image
geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```




```
ggplot(data = dailyActivity_merged) +
    aes(x= sedentary_hours, y = calories) +
    geom_point(color = "yellow") +
    geom_smooth() +
    labs(x = "Sedentary hours", y = "Calories burned", title = "C
lories burned vs sedentary hours")
geom_smooth()` using method = 'loess' and formula = 'y ~ x'
ggsave("sedentary_hours_vs_calories_burned.png")
aving 6.01 x 4.11 in image
geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```



ACT PHASE

Now that you have finished creating your visualizations, act on your findings. Prepare the deliverables you have been asked to create, including the high-level recommendations based on your analysis.