

Case Study 2 Analysis

DW

2022-10-05

Case Study 2

Bellabeat provided smart device fitness data and requested insights on how consumers are using the smart devices.

Data was provided for the month of April 2016.

Data was processed and cleaned based on the following workflow:

```
library(tidyverse)
library(dplyr)
library(tidyr)
library(skimr)
library(lubridate)
library(janitor)
library(dplyr)

# Extract CSV data
files = list.files(path = "C://Users//dlwan//Documents//Google_Certificate//Case2//Data")
setwd("C://Users//dlwan//Documents//Google_Certificate//Case2//Data")

file_use = list()
file_name = list()

# Check file list contains only csv
for (x in 1:length(files)) {
  ext = sapply(strsplit(files[x],split=".",fixed=TRUE),tail,1)
  name = sapply(strsplit(files[x],"_"),"[",1)
  if(ext == "csv") {
    file_use = append(file_use,files[x])
    file_name = append(file_name,name)
  }
}

id_cnt = list()
day_cnt = list()

time = list('ActivityDate','ActivityDay','Time','ActivityHour','ActivityMinute','SleepDay','Date')

# Store each csv to file name
# Convert Date to appropriate format
for (x in 1:length(file_use)) {
  temp = read_csv(file_use[x])
```

```

# Remove duplicate values
temp = distinct(temp)
for (j in 1:length(time)){
  index = which(colnames(temp) == time[j])
  # Check if time index is in column names
  if (length(index)>0){
    # Add column for date in Date format
    # Remove NA
    test = paste(time[j])
    # print(sum(duplicated(temp)))
    if (grepl(":",temp[1,index])){
      temp = temp %>%
        mutate(date_time = as.POSIXct(temp[[test]],format = "%m/%d/%Y %I:%M:%S %p", origin = '1970-01-01'))
      mutate(date_day = as.Date(date_time)) %>%
      mutate(day_of_week = wday(date_day)) %>%
      na.omit()
    } else {
      temp = temp %>%
        mutate(date_day = as.POSIXct(temp[[test]],format = "%m/%d/%Y", origin = '1970-01-01')) %>%
        mutate(day_of_week = wday(date_day)) %>%
        na.omit()
    }
  }
}
temp$Id = as.character(temp$Id)
assign(unlist(file_name[x]),temp)
id_cnt = append(id_cnt,n_distinct(get(unlist(file_name[x]))$"Id"))
day_cnt = append(day_cnt,n_distinct(get(unlist(file_name[x]))$date_day))
}

```

Daily Intensity

Daily intensity data consisted of number of steps, distance, and calories.

Summary statistics for each member, including average number of steps, average distance, average calories, and number of recorded days.

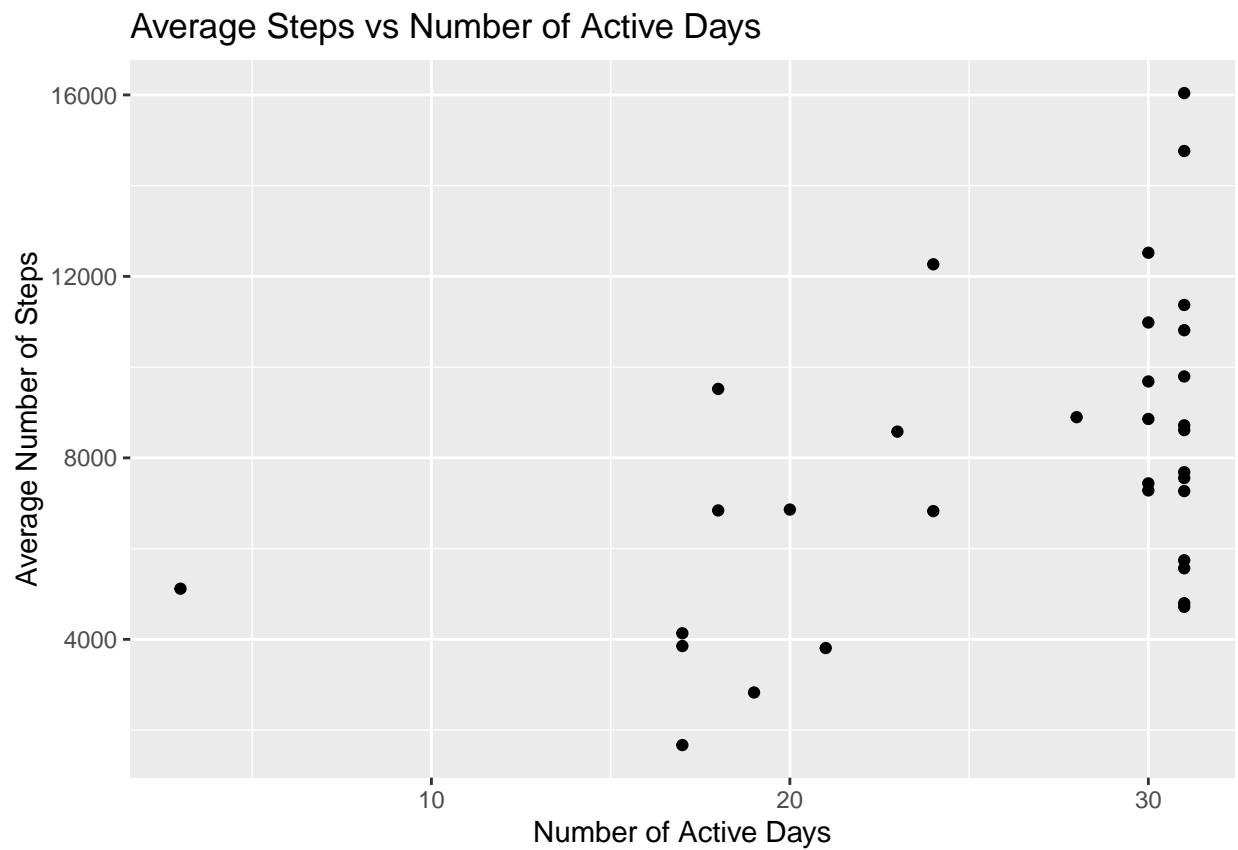
Not every member used smart device each day.

```
## # A tibble: 33 x 5
```

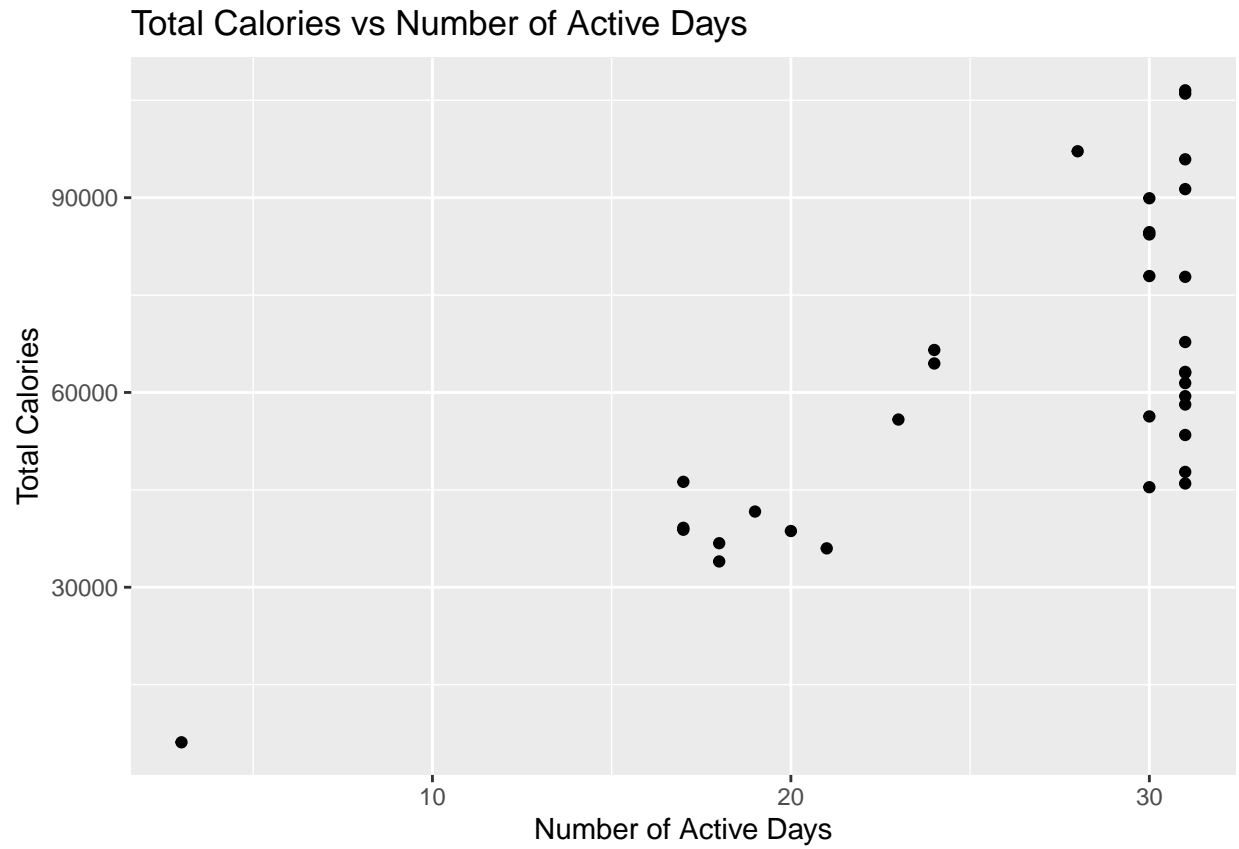
	Id	Avg_Steps	Avg_Distance	Avg_Calories	days
	<chr>	<dbl>	<dbl>	<dbl>	<int>
## 1	1503960366	12521.	8.07	1877.	30
## 2	1624580081	5744.	3.91	1483.	31
## 3	1644430081	7283.	5.30	2811.	30
## 4	1844505072	3809.	2.52	1714.	21
## 5	1927972279	1671.	1.16	2303.	17
## 6	2022484408	11371.	8.08	2510.	31
## 7	2026352035	5567.	3.45	1541.	31
## 8	2320127002	4717.	3.19	1724.	31
## 9	2347167796	9520.	6.36	2043.	18
## 10	2873212765	7556.	5.10	1917.	31
## 11	3372868164	6862.	4.71	1933.	20
## 12	3977333714	10985.	7.52	1514.	30

## 13	4020332650	4134.	2.97	2720.	17
## 14	4057192912	5117.	3.82	2040.	3
## 15	4319703577	7269.	4.89	2038.	31
## 16	4388161847	10814.	8.39	3094.	31
## 17	4445114986	4797.	3.25	2186.	31
## 18	4558609924	7685.	5.08	2033.	31
## 19	4702921684	8858.	7.19	2997.	30
## 20	5553957443	8613.	5.64	1876.	31
## 21	5577150313	8898.	6.66	3470.	28
## 22	6117666160	8579.	6.50	2427.	23
## 23	6290855005	6827.	5.16	2773.	24
## 24	6775888955	3854.	2.77	2286.	17
## 25	6962181067	9795.	6.59	1982.	31
## 26	7007744171	12267.	8.68	2686.	24
## 27	7086361926	9684.	6.60	2598.	30
## 28	8053475328	14763.	11.5	2946.	31
## 29	8253242879	6842.	4.93	1887.	18
## 30	8378563200	8718.	6.91	3437.	31
## 31	8583815059	7438.	5.80	2823.	30
## 32	8792009665	2829.	1.81	2193.	19
## 33	8877689391	16040.	13.2	3420.	31

Following visuals highlight key parameters for each member.

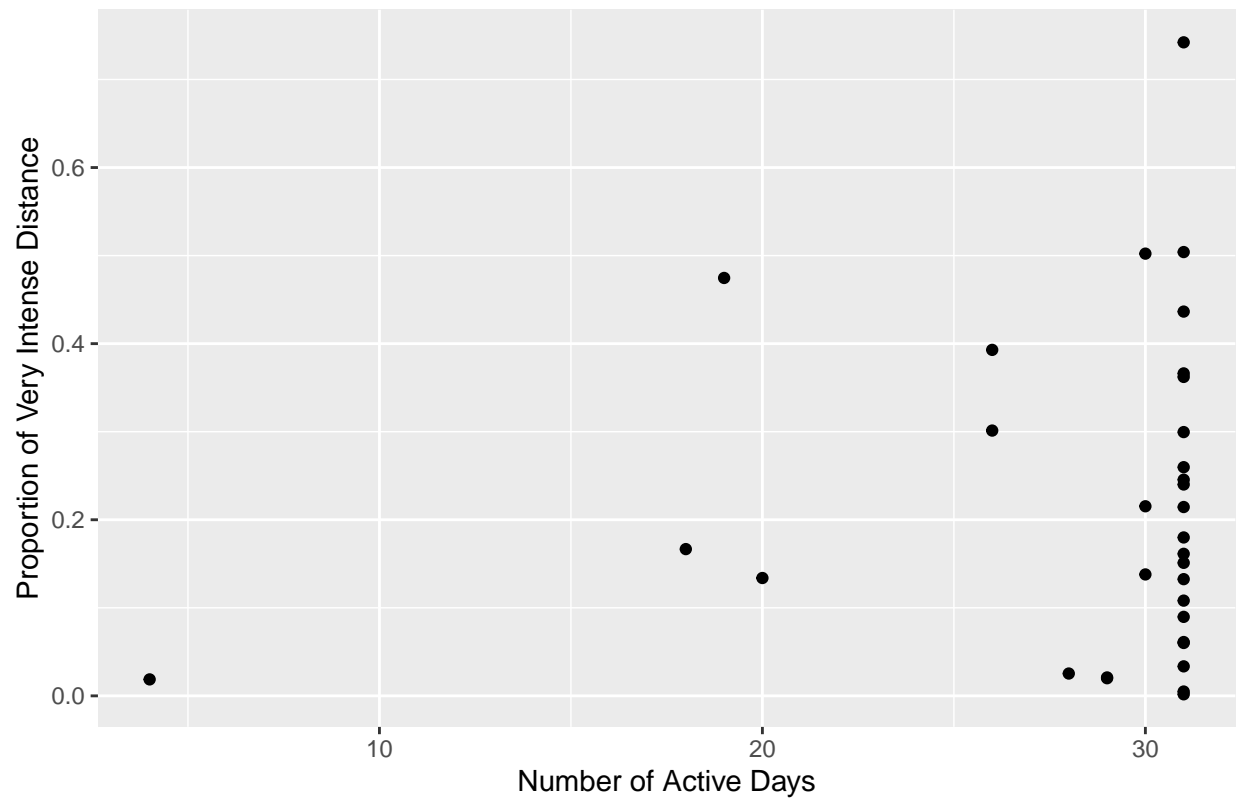


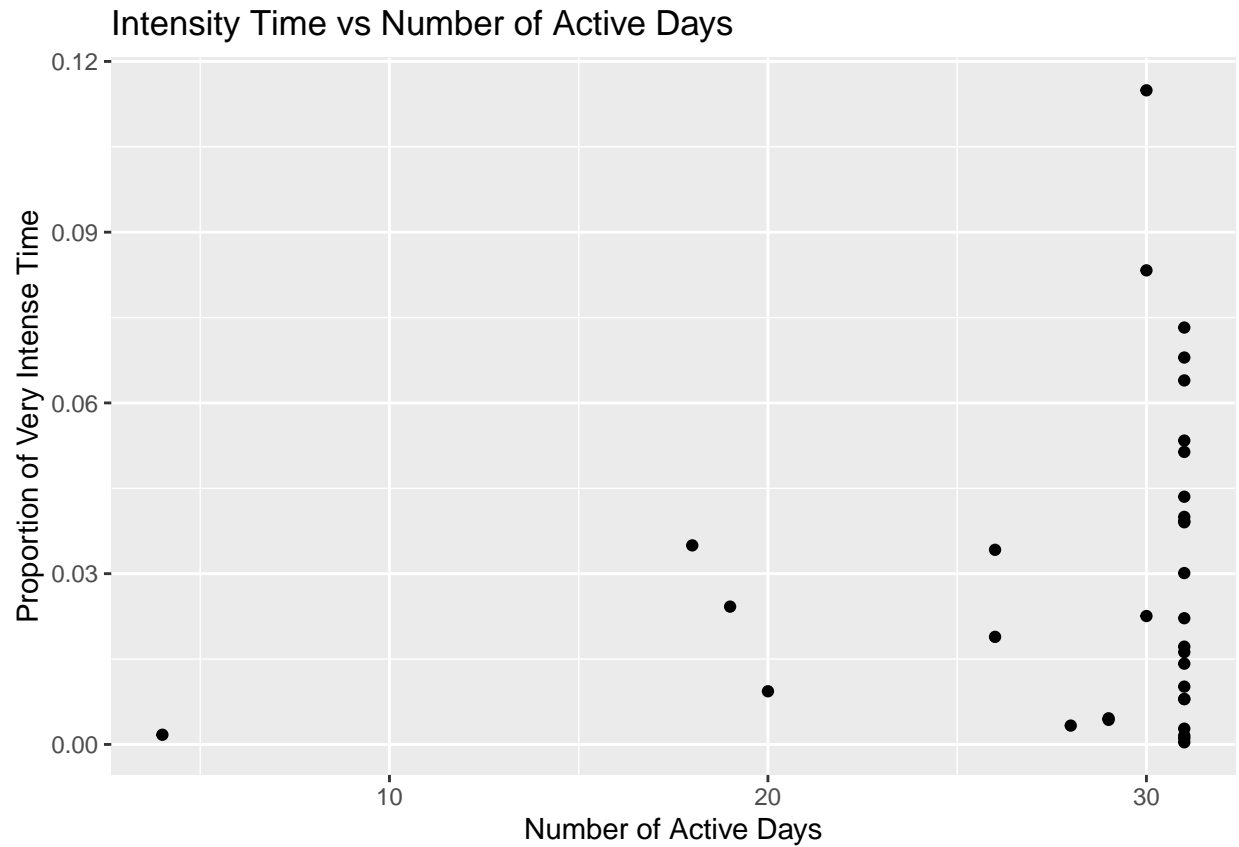
People who had an average number 6,000 steps a days were more likely to continuously use the smart device.



People who had an average monthly 45,000 calories were more likely to continuously use the smart device.

Intensity Distance vs Number of Active Days





Participants that spend 40 percent of the distance and 20 percent of active time at a high intensity would be more likely to use the smart device

Conclusions

People who are more active in terms of further distance traveled or high degree of intensity are more likely to consistently use the smart device.