title: “Bellabeat Smart Device Data Analysis” author: “Data Analyst” date: “2025-04-29” output: html_document —

Introduction

This report analyzes data collected from FitBit users to generate insights for Bellabeat, a wellness technology company. The objective is to understand user behavior and usage patterns to inform future product and marketing strategies.

Data Loading

```
activity <- read_csv("Fitabase Data 4.12.16-5.12.16/dailyActivity_merged.csv")
```

```
## Rows: 940 Columns: 15
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (1): ActivityDate
## dbl (14): Id, TotalSteps, TotalDistance, TrackerDistance, LoggedActivitiesDi...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
calories <- read_csv("Fitabase Data 4.12.16-5.12.16/hourlyCalories_merged.csv")
```

```
## Rows: 22099 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, Calories
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
steps <- read_csv("Fitabase Data 4.12.16-5.12.16/hourlySteps_merged.csv")
```

```
## Rows: 22099 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): ActivityHour
## dbl (2): Id, StepTotal
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
sleep <- read_csv("Fitabase Data 4.12.16-5.12.16/sleepDay_merged.csv")
```

```
## Rows: 413 Columns: 5
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): SleepDay
## dbl (4): Id, TotalSleepRecords, TotalMinutesAsleep, TotalTimeInBed
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Data Cleaning

```
activity <- activity %>% mutate(ActivityDate = mdy(ActivityDate))
sleep <- sleep %>% mutate(SleepDay = mdy_hms(SleepDay)) %>%
  group_by(Id, date = as.Date(SleepDay)) %>%
  summarise(TotalSleep = sum(TotalMinutesAsleep), .groups = 'drop')
```

Summary Statistics

```
activity %>%
  summarise(across(where(is.numeric), list(mean = mean, median = median), na.rm = TRUE)) %>%
  kable() %>%
  kable_styling()
```

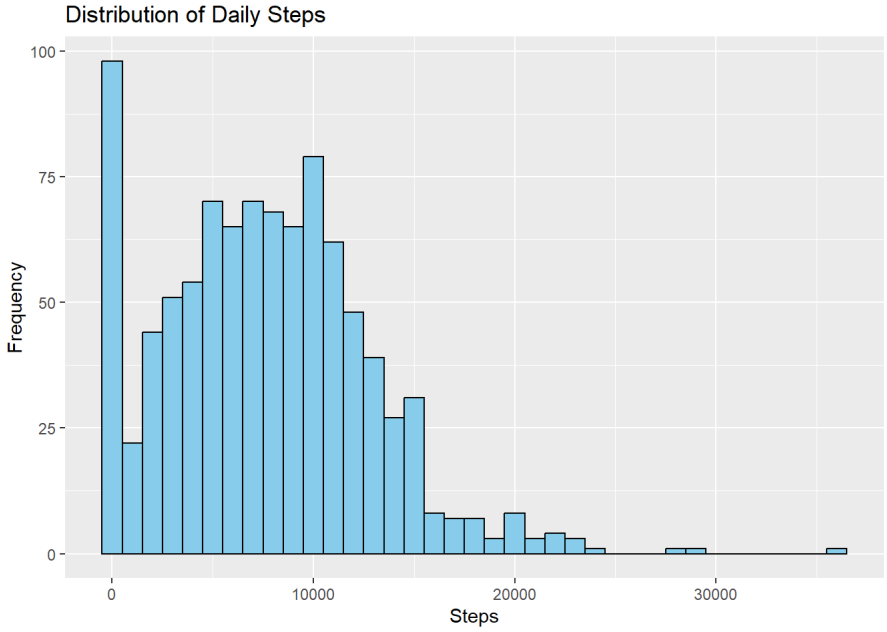
```
## Warning: There was 1 warning in `summarise()`.
## ℹ In argument: `across(...)`.
## Caused by warning:
## ! The `...` argument of `across()` is deprecated as of dplyr 1.1.0.
## Supply arguments directly to `.fns` through an anonymous function instead.
##
##   # Previously
##   across(a:b, mean, na.rm = TRUE)
##
##   # Now
##   across(a:b, \(x) mean(x, na.rm = TRUE))
```

Id_mean	Id_median	TotalSteps_mean	TotalSteps_median	TotalDistance_mean	TotalDistance_median	TrackerDistance_mean	TrackerD
4855407369	4445114986	7637.911	7405.5	5.489702	5.245	5.475351	

Visualizations

Steps Distribution

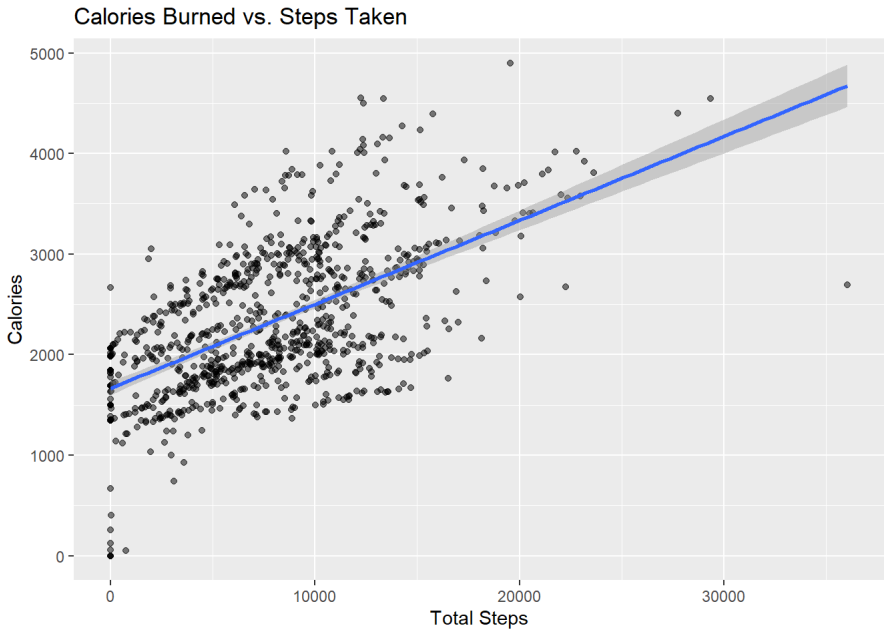
```
activity %>%
  ggplot(aes(TotalSteps)) +
  geom_histogram(binwidth = 1000, fill = "skyblue", color = "black") +
  labs(title = "Distribution of Daily Steps", x = "Steps", y = "Frequency")
```



Calories vs Steps

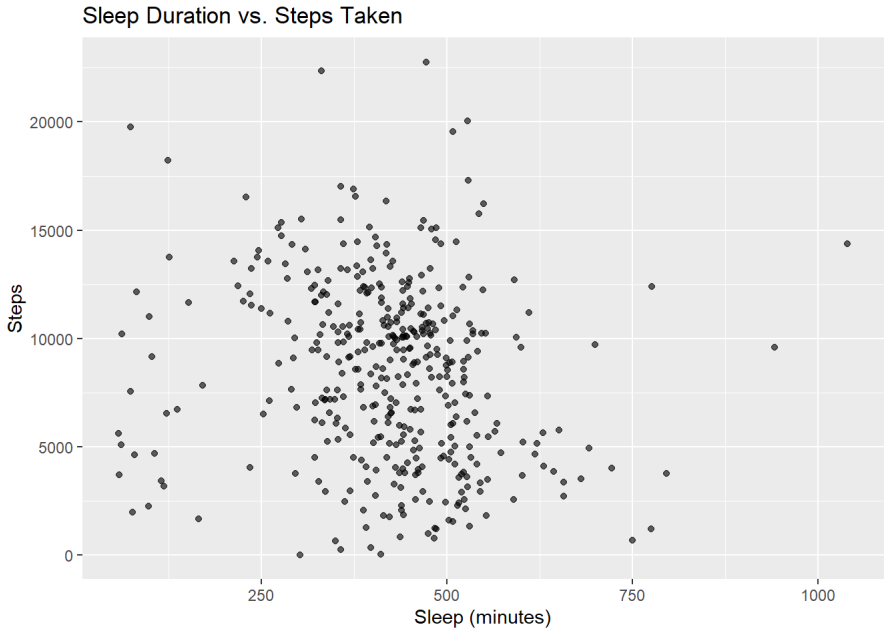
```
activity %>%
  ggplot(aes(TotalSteps, Calories)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm") +
  labs(title = "Calories Burned vs. Steps Taken", x = "Total Steps", y = "Calories")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Sleep vs Activity

```
combined <- inner_join(activity, sleep, by = c("Id" = "Id", "ActivityDate" = "date"))
combined %>%
  ggplot(aes(TotalSleep, TotalSteps)) +
  geom_point(alpha = 0.6) +
  labs(title = "Sleep Duration vs. Steps Taken", x = "Sleep (minutes)", y = "Steps")
```



Insights

- Most users average fewer than 8,000 steps per day.
- There is a positive correlation between steps taken and calories burned.
- Users with longer sleep durations tend to show slightly higher physical activity.

Recommendations

1. Promote daily step goals with app notifications.
2. Emphasize the link between sleep and activity in wellness tips.
3. Use behavior clustering to tailor in-app experiences.

Conclusion

This analysis provides a baseline understanding of user health behaviors, which Bellabeat can leverage to refine products and drive user engagement.