Bellabeat Case Study

Lidiya Arsova

2025-04-13

Ask

Bellabeat is a high-tech wellness company that manufactures health-focused smart devices for women. The goal of this project is to analyze Fitbit data and uncover user trends that Bellabeat can use to inform its marketing strategy.

Prepare

We are working with public Fitbit data collected from 30 users over two months, shared under a public domain license via Kaggle and Zenodo.

We are using two daily activity files — one from each month — and will combine them.

```
daily1 <- read_csv("dailyActivity_merged.csv")
```

```
## Rows: 457 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.
```

```
daily2 <- read_csv("dailyActivity_merged2.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.
```

Process

Here, we combine the two datasets, remove any duplicates, and convert the activity date column into proper Date format.

```
daily_full <- bind_rows(daily1, daily2) %>%
  distinct() %>%
  mutate(ActivityDate = mdy(ActivityDate))
```

Analyze

We calculate summary statistics to understand users' daily activity patterns.

```
daily_full %>%
  summarise(
    avg_steps = mean(TotalSteps, na.rm = TRUE),
    avg_calories = mean(Calories, na.rm = TRUE),
    avg_very_active = mean(VeryActiveMinutes, na.rm = TRUE),
    avg_lightly_active = mean(LightlyActiveMinutes, na.rm = TRUE),
    avg_sedentary = mean(SedentaryMinutes, na.rm = TRUE)
  )
```

```
## # A tibble: 1 x 5
##   avg_steps avg_calories avg_very_active avg_lightly_active avg_sedentary
##       <dbl>        <dbl>           <dbl>              <dbl>         <dbl>
## 1     7281.        2266.            19.7               185.          993.
```

We also explore how different activities correlate with calories burned.

```
cor_data <- daily_full %>%
  select(TotalSteps, TotalDistance, VeryActiveMinutes, FairlyActiveMinutes,
         LightlyActiveMinutes, SedentaryMinutes, Calories)

cor(cor_data, use = "complete.obs")
```

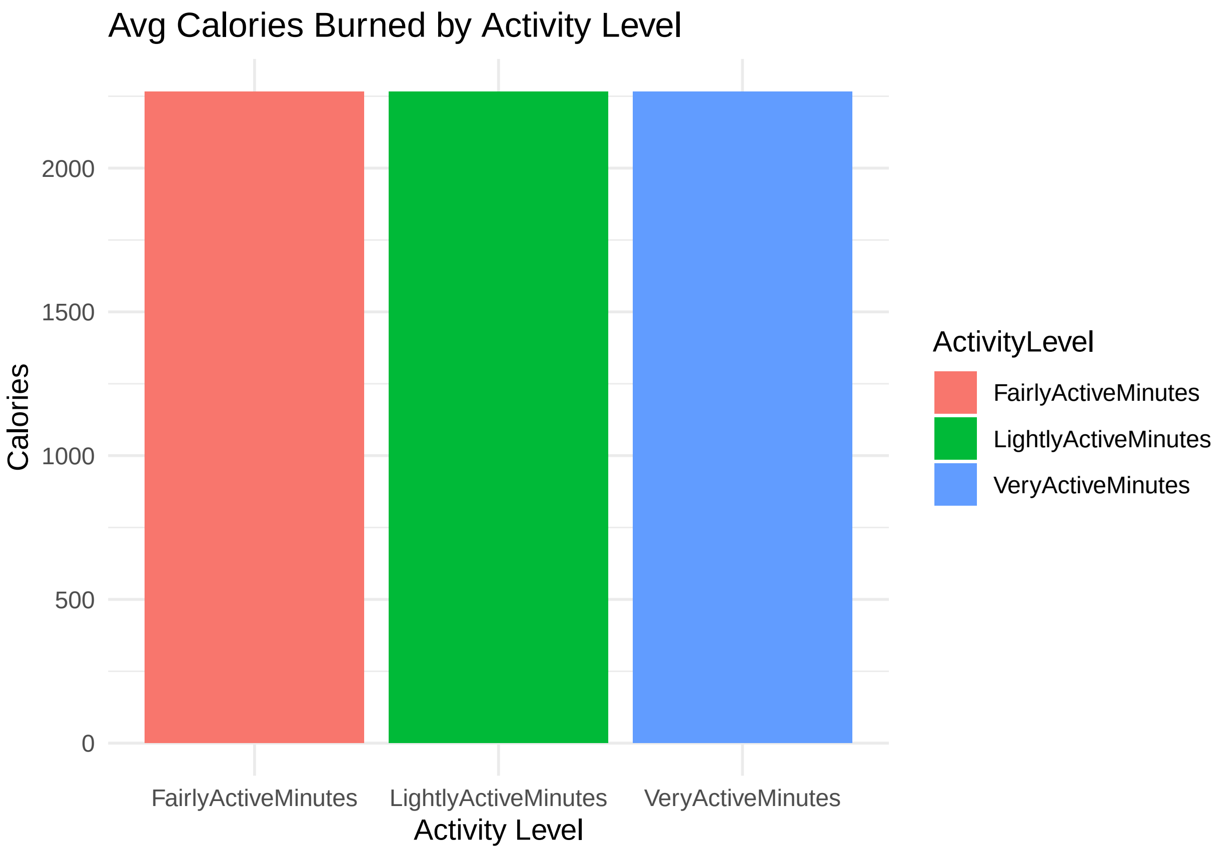
```
##                      TotalSteps TotalDistance VeryActiveMinutes
## TotalSteps            1.0000000     0.9859539         0.6765583
## TotalDistance         0.9859539     1.0000000         0.6911431
## VeryActiveMinutes     0.6765583     0.6911431         1.0000000
## FairlyActiveMinutes   0.3591856     0.3380976         0.2341731
## LightlyActiveMinutes  0.6040972     0.5553730         0.1047707
## SedentaryMinutes     -0.3110263    -0.2771641        -0.1678437
## Calories              0.5901599     0.6353040         0.5820275
##                      FairlyActiveMinutes LightlyActiveMinutes SedentaryMinutes
## TotalSteps                     0.3591856            0.6040972      -0.31102626
## TotalDistance                  0.3380976            0.5553730      -0.27716405
## VeryActiveMinutes              0.2341731            0.1047707      -0.16784374
## FairlyActiveMinutes            1.0000000            0.1191759      -0.18504495
## LightlyActiveMinutes           0.1191759            1.0000000      -0.41933197
## SedentaryMinutes              -0.1850449           -0.4193320       1.00000000
## Calories                       0.3074824            0.3257006      -0.06192441
##                         Calories
## TotalSteps            0.59015995
## TotalDistance         0.63530399
## VeryActiveMinutes     0.58202752
## FairlyActiveMinutes   0.30748240
## LightlyActiveMinutes  0.32570064
## SedentaryMinutes     -0.06192441
## Calories              1.00000000
```

Share

Average Calories Burned by Activity Level

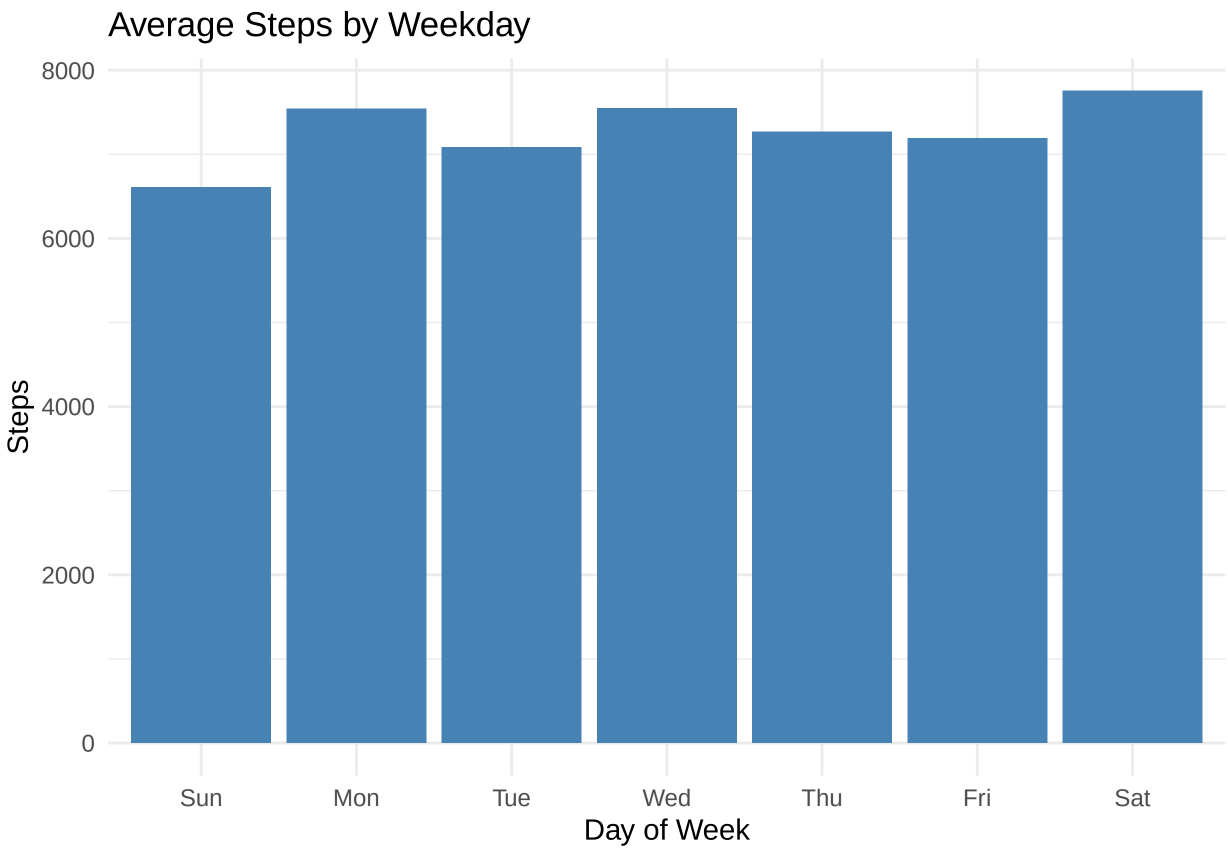
```
daily_full %>%
  pivot_longer(cols = c(VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes),
```

```r
              names_to = "ActivityLevel", values_to = "Minutes") %>%
  group_by(ActivityLevel) %>%
  summarise(AvgCalories = mean(Calories, na.rm = TRUE)) %>%
  ggplot(aes(x = ActivityLevel, y = AvgCalories, fill = ActivityLevel)) +
  geom_col() +
  labs(title = "Avg Calories Burned by Activity Level", y = "Calories", x = "Activity Level") +
  theme_minimal()
```



Average Steps by Weekday

```r
daily_full %>%
  mutate(Weekday = wday(ActivityDate, label = TRUE)) %>%
  group_by(Weekday) %>%
  summarise(AverageSteps = mean(TotalSteps, na.rm = TRUE)) %>%
  ggplot(aes(x = Weekday, y = AverageSteps)) +
  geom_col(fill = "steelblue") +
  labs(title = "Average Steps by Weekday", x = "Day of Week", y = "Steps") +
  theme_minimal()
```



Act

Based on this analysis, Bellabeat could take the following actions:

Encourage High-Intensity Movement: Very Active Minutes had a strong correlation with calories burned. Bellabeat can promote workouts that drive this metric.

Target Weekday Engagement: Most users are more active during weekdays. Bellabeat can use this insight to drive weekday-focused campaigns.

Combat Sedentary Behavior: Sedentary minutes are still high. Bellabeat can add movement reminders or gentle nudge notifications via smart devices.

Conclusion

This project explored user activity data to generate business insights for Bellabeat. Using R and tidyverse, we cleaned and analyzed the data, created visualizations, and delivered actionable recommendations for smarter engagement strategies.