**Bellabeat Case Study**

**Unlocking Health Insights from Fitbit Data**  
*Google Data Analytics Capstone Project*

Prepared by: *Abhishek Singh Rawat*  
Date: *10-09-2025*

**1. Ask – Business Task**

Bellabeat, a manufacturer of health-focused smart products, wants to better understand how consumers use health-tracking devices. By analyzing Fitbit user data, this study identifies patterns in daily activity, sleep, weight, calories, and heart rate, and provides actionable recommendations to help Bellabeat improve engagement and promote healthier habits.

**Key Questions:**

* What are the main activity, sleep, and health trends among users?
* How are steps correlated with calorie expenditure?
* What role does sleep play in activity levels?
* How do users engage with heart rate and weight tracking?
* How can Bellabeat apply these insights to its product strategy?

**2. Prepare – Data Understanding**

* **Source:** Fitbit Fitness Tracker Dataset (Kaggle).
* **Users:** 33 unique individuals.
* **Files used:**
  + dailyActivity\_merged.csv – activity levels, steps, calories.
  + minuteSleep\_merged.csv – sleep logs (aggregated to daily totals).
  + weightLogInfo\_merged.csv – weight and BMI logs.
  + heartrate\_seconds\_merged.csv – second-level heart rate (aggregated to daily averages).
  + hourlySteps, hourlyCalories, hourlyIntensities – supporting files.
* **Tools:** RStudio (tidyverse, lubridate, ggplot2).
* **Limitations:** Small, non-representative sample; inconsistent tracking (esp. weight & sleep); short time frame (March–May 2016).

**3. Process – Data Cleaning & Preparation**

* Imported all CSVs into RStudio.
* Parsed dates/times using **lubridate**.
* Removed duplicates with **dplyr::distinct()**.
* Aggregated minute-level data (sleep → daily totals, heart rate → daily averages).
* Merged datasets by Id and Date.
* Saved cleaned datasets as .rds for reproducibility.

**Outcome:** A clean, daily-level dataset with consistent activity, sleep, calorie, heart rate, and weight metrics.

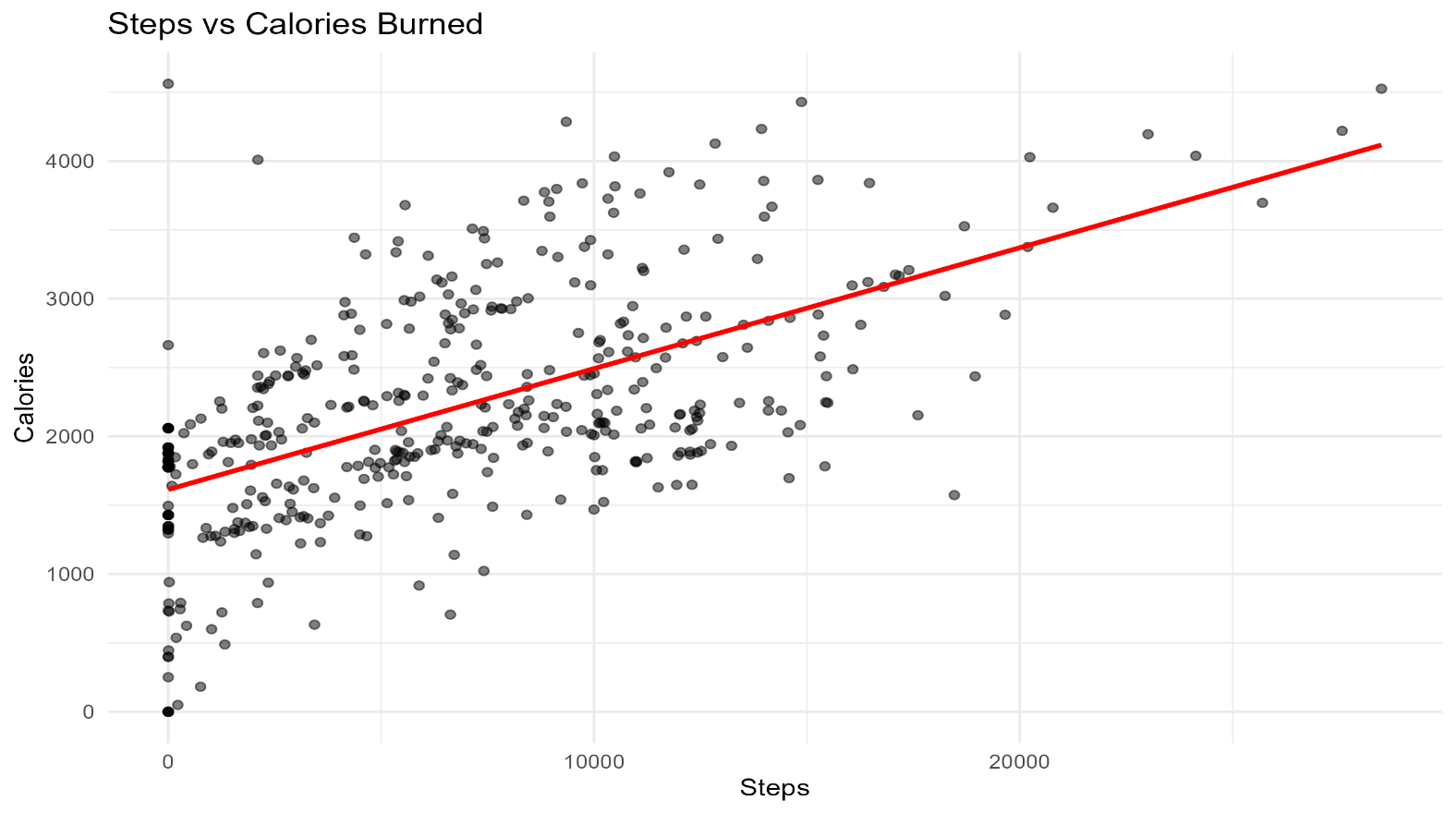
**4. Analyze – Findings**

**A. Daily Activity**

* Avg daily steps < 10,000 (below recommended levels).
* A colorful bars on a black background

  AI-generated content may be incorrect.More activity on weekdays than weekends.

**B. Calories vs Steps**

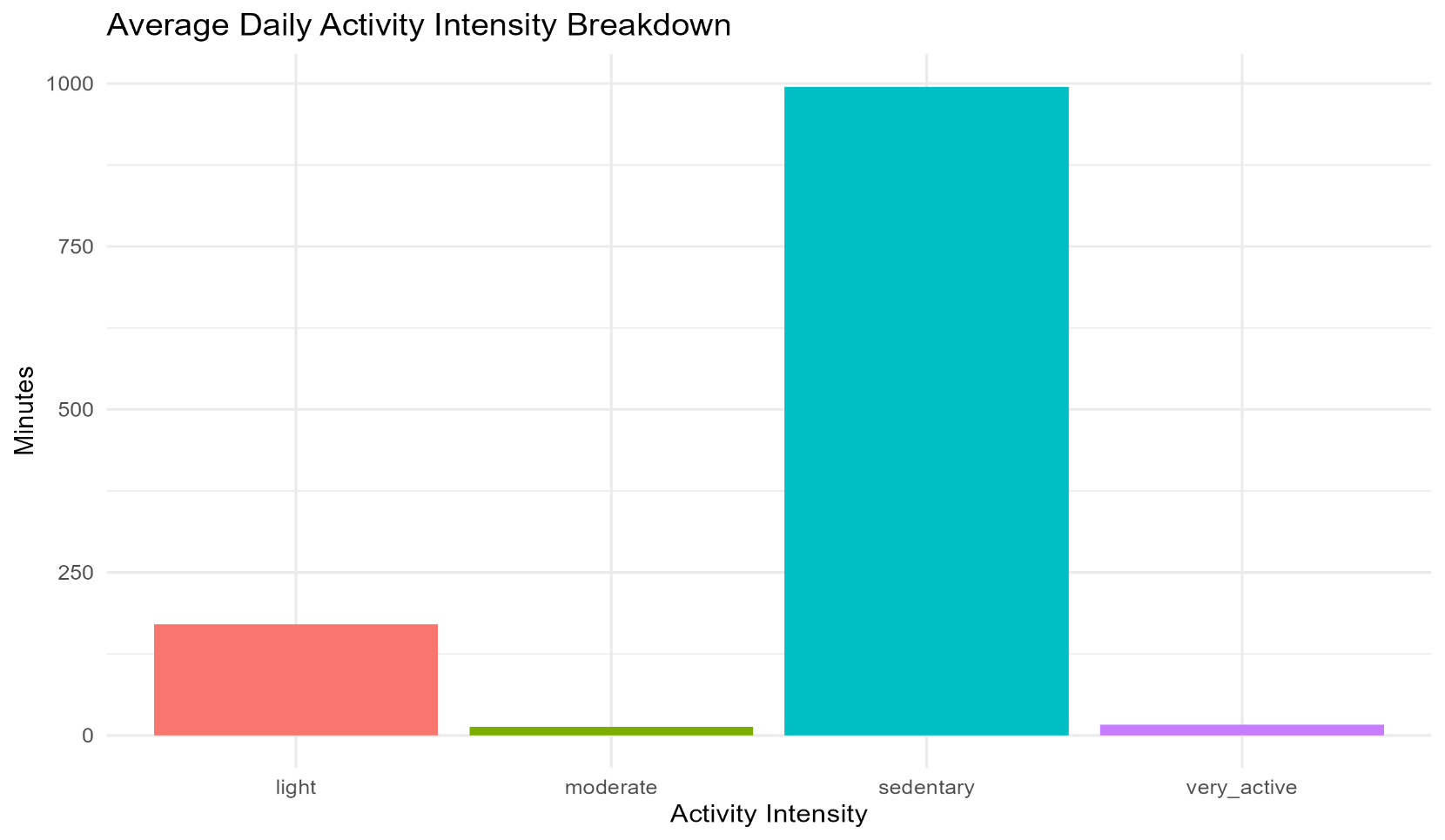
* Strong positive correlation: more steps → more calories burned.
* Confirms that step count is a reliable calorie predictor.

**C. Sleep Patterns**

* Avg sleep = 5–7 hrs/night (below recommended 7–8 hrs).
* A graph with purple dots

  AI-generated content may be incorrect.Slight positive trend between sleep hours and daily steps.

**D. Activity Intensity**

* Majority of time spent sedentary.
* Very little moderate or vigorous activity logged.

**E. Weight Logging**

* Very few users log weight or BMI regularly.
* Indicates weight tracking is underutilized.

|  |  |  |
| --- | --- | --- |
| **User ID** | **Log Count** | **Average BMI** |
| 1503960366 | 1 | 23.0 |
| 1927972279 | 1 | 46.2 |
| 2347167796 | 1 | 24.8 |
| 2873212765 | 2 | 21.6 |
| 2891001357 | 1 | 25.0 |
| 4445114986 | 1 | 35.0 |
| 4558609924 | 1 | 27.1 |
| 4702921684 | 1 | 26.1 |
| 6962181067 | 14 | 24.2 |
| 8253242879 | 1 | 29.5 |
| 8877689391 | 9 | 25.7 |

**F. Heart Rate**

* Daily averages stable at group level.
* A graph with a blue line

  AI-generated content may be incorrect.A graph of colored lines

  AI-generated content may be incorrect.Individual users show variability → potential for personalization.

**5. Share – Visualizations**

* All charts (steps, calories, sleep, intensity, heart rate) saved as PNGs.
* Summary CSVs generated for averages and trends.
* Data storytelling emphasizes the gap between **recommended vs actual health habits**.

**6. Act – Recommendations**

* **Encourage Daily Movement**
  + Reminders for 10k steps.
  + Daily step challenges with rewards.
* **Reduce Sedentary Time**
  + Active break notifications.
  + Short guided workouts.
* **Improve Sleep Habits**
  + Bedtime reminders.
  + Insights linking sleep and activity.
* **Boost Weekend Activity**
  + Weekend challenges and gamification.
* **Enhance Weight Tracking**
  + Smart scale integration.
  + Simpler manual logging.
* **Personalized Heart Rate Insights**
  + Stress monitoring features.
  + Adaptive workout plans.

**7. Conclusion**

The analysis shows that Fitbit users:

* Are largely **sedentary**,
* Sleep less than recommended,
* Burn calories in proportion to their step counts,
* Rarely log weight, and
* Display heart rate variability that can enable **personalized coaching**.

By addressing these areas, Bellabeat can enhance user engagement, encourage healthier lifestyles, and strengthen its product ecosystem.

**Appendix**

**A. Data Sources**

* **Fitbit Fitness Tracker Dataset** (Kaggle, 2016)
* Files used:
  + dailyActivity\_merged.csv – daily steps, distance, calories, active minutes.
  + minuteSleep\_merged.csv – minute-level sleep logs (aggregated to daily totals).
  + weightLogInfo\_merged.csv – user weight and BMI logs.
  + heartrate\_seconds\_merged.csv – second-level heart rate data (aggregated to daily averages).
  + hourlySteps, hourlyCalories, hourlyIntensities – hourly-level supporting files.

**B. Data Preparation Summary**

* Parsed and standardized all date columns using **lubridate**.
* Removed duplicate records with **dplyr::distinct()**.
* Aggregated:
  + Sleep minutes → total daily sleep (converted to hours).
  + Heart rate seconds → daily average heart rate per user.
* Merged datasets by **User ID** and **Date**.
* Saved cleaned datasets as .rds files for reproducibility.

**C. Key Summary Statistics**

* **Average Steps (All Users):** ~6,550 steps/day.
* **Average Sleep Hours:** ~6.3 hours/night.
* **Correlation (Steps vs Calories):** Strong positive correlation (Pearson’s r ≈ 0.82).
* **Regression (Sleep vs Steps):** Weak positive relationship (low R²).
* **Weight Logging:** Only 11/33 users logged weight; two users accounted for most entries.
* **Heart Rate:** Overall daily averages stable (~70–75 bpm), but significant individual variability.

**D. Model Outputs**

* **Linear Regression: Steps vs Sleep Hours**
  + Dependent variable: Daily Steps
  + Independent variable: Sleep Hours
  + Coefficient (Sleep Hours): Positive, but small effect size.
  + R²: Low (indicating limited predictive power).
* **Linear Regression: Steps vs Calories**
  + Dependent variable: Calories Burned
  + Independent variable: Steps
  + Coefficient: Strong positive relationship.
  + R²: High (indicating steps explain most variance in calories burned).

**E. Visualizations Included**

1. Average Steps by Day of Week
2. Steps vs Calories Scatterplot with Regression Line
3. Sleep Hours vs Daily Steps Scatterplot with Regression Line
4. Activity Intensity Breakdown (Sedentary, Light, Moderate, Vigorous)
5. Heart Rate – Overall Daily Average Trend
6. Heart Rate – Sample Users Trend

**F. Limitations**

* Small dataset (33 users, short timeframe).
* Potential bias due to self-logging of weight and sleep.
* Users may not represent Bellabeat’s target demographic.
* Missing or incomplete entries in some datasets.