Bellabeat – Smart Device Usage Analysis

1. Business Task

Objective:
Bellabeat aims to refine the marketing and product strategy for its Time watch—a wellness watch that tracks activity, sleep, and stress. The goal is to understand consumer usage patterns from non-Bellabeat Fitbit devices and identify actionable insights to convert insights into effective marketing campaigns.

Key Questions:

- What are some trends in smart device usage?
- How could these trends apply to Bellabeat customers?
- How can these trends help influence Bellabeat's marketing strategy for the Time watch?

2. Data Sources Used

The analysis is based on data collected via a distributed survey on Amazon Mechanical Turk (03.12.2016–05.12.2016). Over 30 eligible Fitbit users consented to share their minute-level personal tracker data.

Datasets Included:

- **dailyActivity_merged.csv:** Daily summary metrics (e.g., TotalSteps, TotalDistance, Calories, active and sedentary minutes).
- **heartrate_seconds_merged.csv:** Second-level heart rate readings.
- **minuteSleep_merged.csv:** Minute-level sleep data.
- **weightLogInfo_merged.csv:** Weight, BMI, and related measurements.

Data from two distinct months were concatenated to create a continuous timeline. All transformations were performed in Python before structuring the data into a star schema.

Key Columns:

- **FactActivity:** Id, ActivityDate, TotalSteps, TotalDistance, Calories, VeryActiveMinutes, FairlyActiveMinutes, LightlyActiveMinutes, SedentaryMinutes
- **FactHeartRate:** Id, Date, TimeOnly, AvgHeartRate
- **FactSleep:** Id, Date, TotalSleepMinutes
- **FactWeight:** Id, DateOnly, TimeOnly, WeightKg, BMI
- **DimUsers:** Unique UserID (with unique letter codes)
- **DimTime:** Date, Day, Month, Weekday

3. Data Cleaning & Manipulation

The following cleaning and transformation steps were performed:

- **Concatenation:** Data from both months were merged.
- **Date/Time Conversion:**
  - Daily activity data had a proper date format.
  - Datasets with combined date and time (heart rate, minute sleep, weight log) were split into separate Date and TimeOnly columns.
- **Missing Values:**
  - Rows with missing values were dropped from all datasets except for the weight dataset (only the 'Fat' column was dropped).
- **Star Schema Creation:**
  - Fact tables were created to capture daily activity, hourly heart rate, daily sleep, and weight log details.
  - Dimension tables (DimUsers and DimTime) were built to support user-level and time-based slicing.

---

4. Summary of Analysis

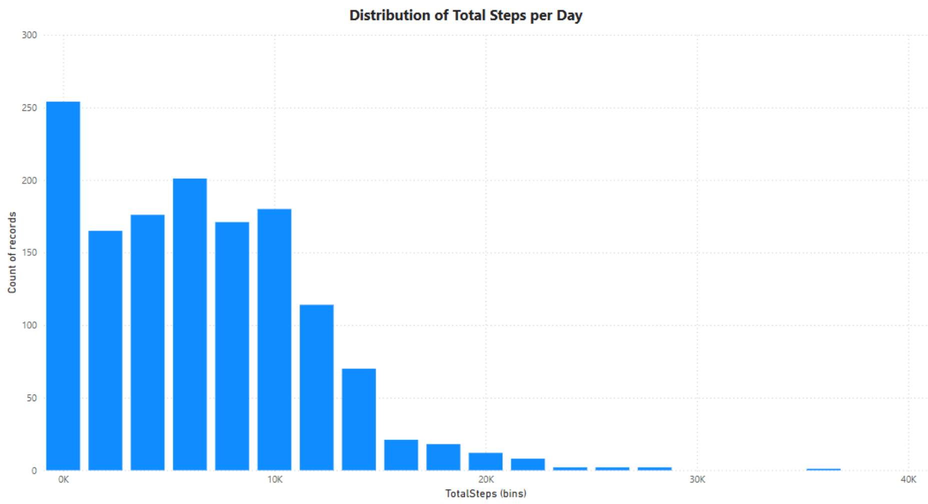
1. **Daily Activity**: Many days show low to moderate steps, suggesting room for engagement.
2. **User Segmentation**: A large portion of users are sedentary, while others are moderately or highly active.
3. **Time-Based Patterns**: Heart rate peaks in the morning and evening, indicating prime times for engagement.
4. **Sleep Behavior**: No strong linear correlation between sleep duration and average steps, but identifiable clusters suggest targeted interventions.
5. **Weight Trends**: Most users have minor fluctuations; some log weight sporadically, indicating the need for consistent tracking.
6. **Calories & Active Minutes**: Strong positive relationship—more active minutes generally lead to higher calorie burn.
7. **Weekly Trends**: Saturdays show the highest average steps, while Sundays are typically the least active day.

---

5. Supporting Visualizations & Key Findings

1. Distribution of Total Steps per Day

- Most days fall in the **0–15K** range, with a significant cluster under **5K** steps.
- A smaller portion reaches **10K+** steps, which is a commonly recommended goal.

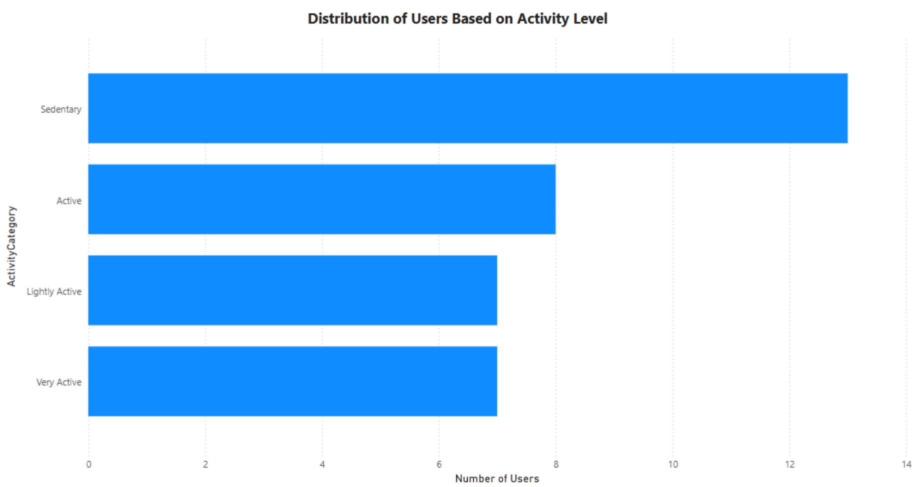


📌 **Insight:** Many days do **not** meet the recommended 10K steps, indicating a **sedentary trend** and potential for **targeted movement campaigns**.

---

2. Distribution of Users Based on Activity Level

- Majority of users are **Sedentary**, followed by **Active**, **Lightly Active**, and **Very Active**.
- Sedentary users outnumber the other categories significantly.

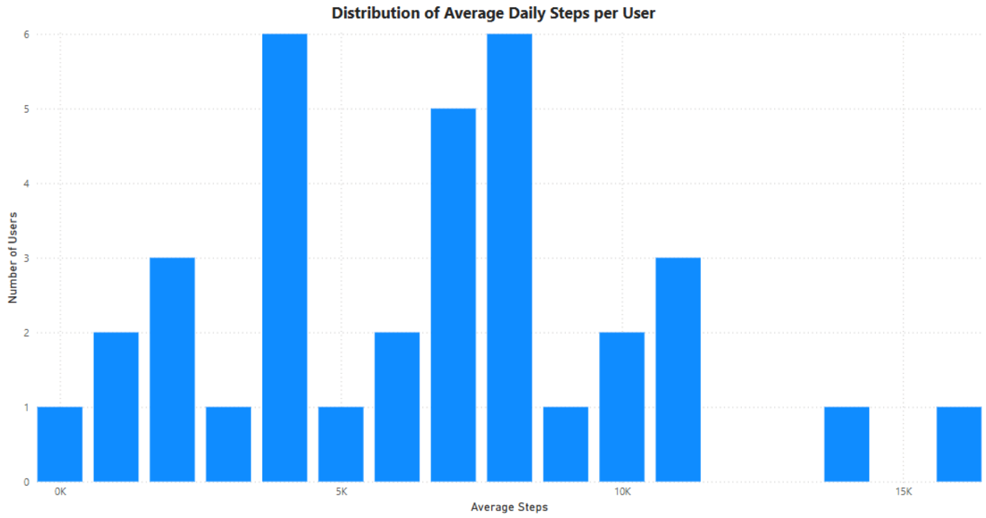


📌 **Insight:** A **large sedentary segment** presents an opportunity for **motivational interventions** to boost their activity level.

---

3. Distribution of Average Daily Steps per User

- Peaks around **5K–6K** and **9K–10K** steps, with notable spread across users.
- A few users exceed **10K** average steps, while many remain under **5K**.

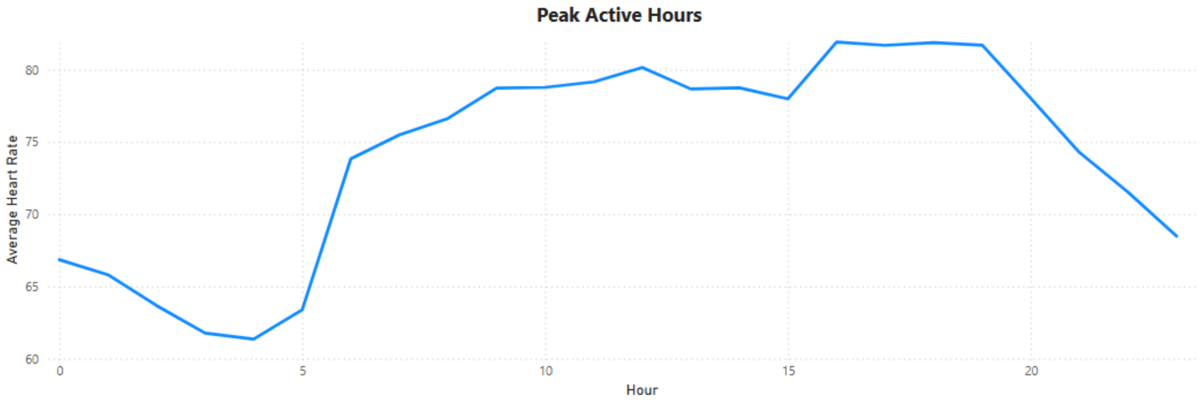


📌 **Insight: High variability** among users underscores the need for **personalized recommendations** to shift low-step users upward..

---

4. Peak Active Hours – Average Heart Rate per Hour

- Heart rate is lowest around **3–5 AM** and rises sharply by **6 AM**.
- Notable peaks around **10 AM** and **5–7 PM**.

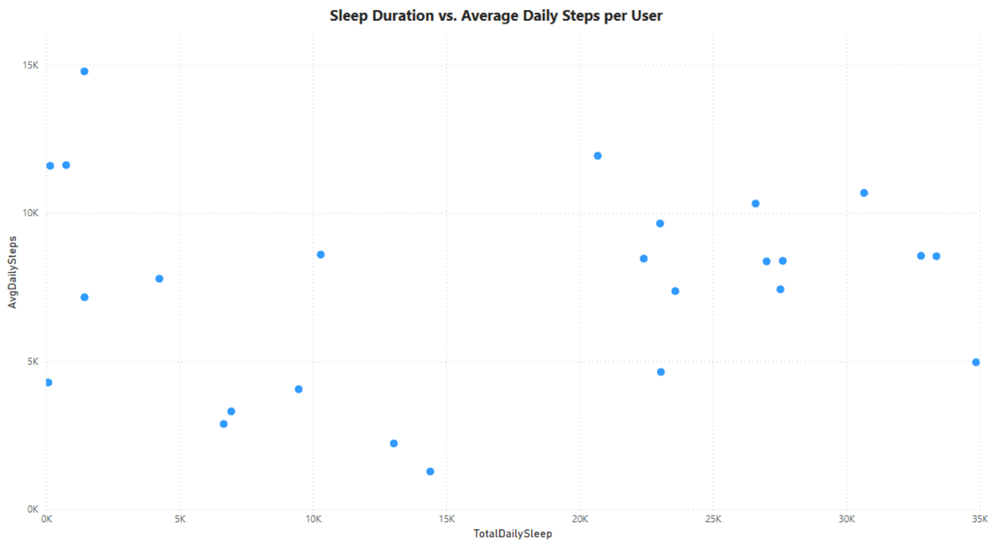


📌 **Insight:** Morning and early-evening **heart rate spikes** suggest prime times for **fitness or stress-management prompts**.

---

5. Sleep Duration vs. Average Daily Steps per User

- No strong linear correlation; users with **long sleep** can have **low steps**, and vice versa.
- Clusters show short-sleep/high-steps and long-sleep/low-steps groups.

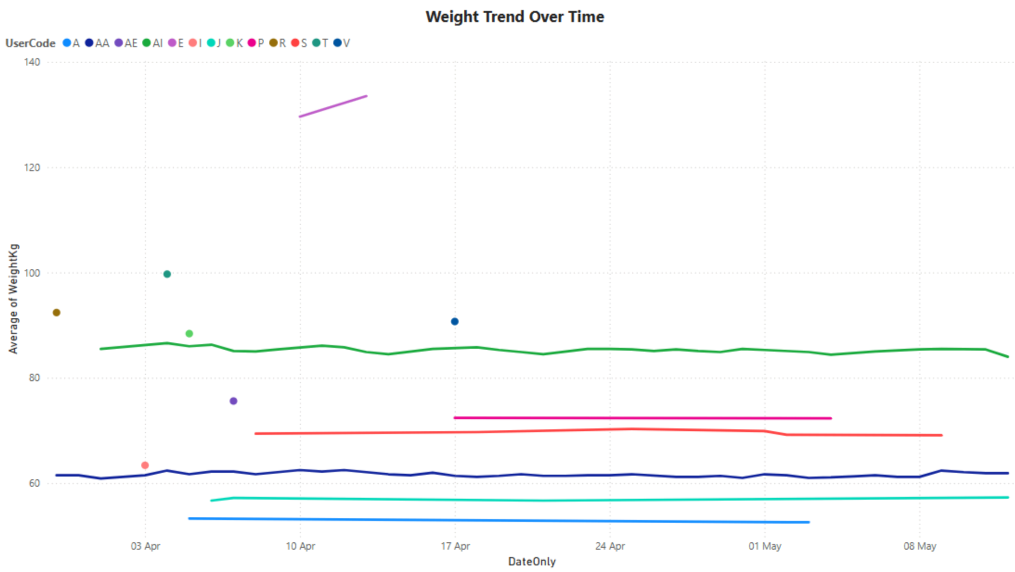


📌 **Insight:** Multiple **behavioral clusters** exist, indicating the need for **tailored sleep and activity guidance** rather than a one-size-fits-all approach.

---

6. Weight Trend Over Time

- Most users show **stable weight** with minor fluctuations.
- Some users log weight sporadically, while others maintain consistent records.

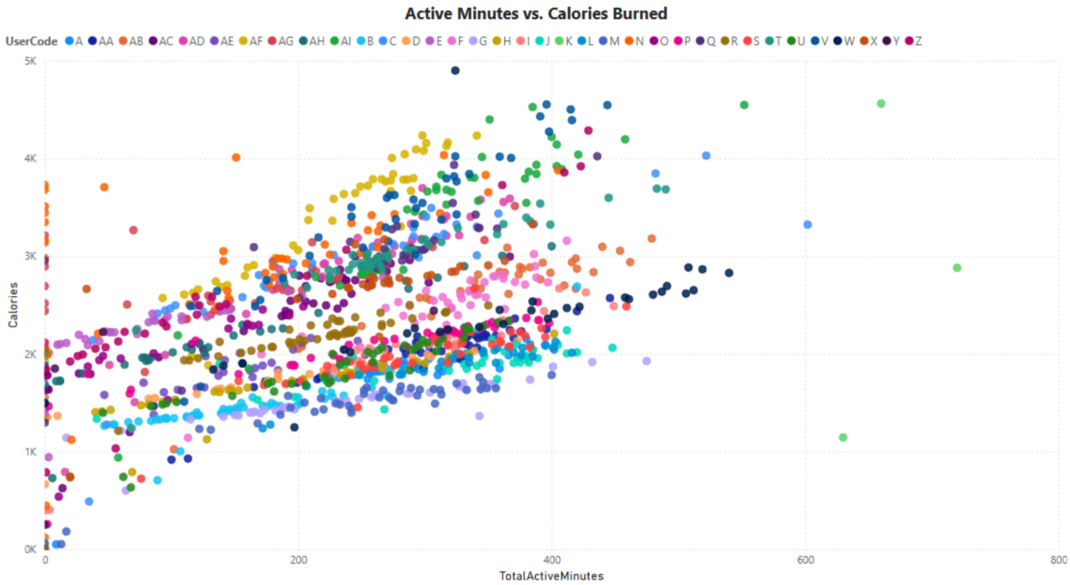


📌 **Insight: Regular weight tracking** helps detect gradual changes; sporadic loggers may benefit from **reminders** to capture accurate trends.long-term health monitoring.

---

7. Active Minutes vs. Calories Burned

- Clear upward trend: **more active minutes** generally yield **higher calories** burned.
- Outliers indicate varying intensities and individual differences.

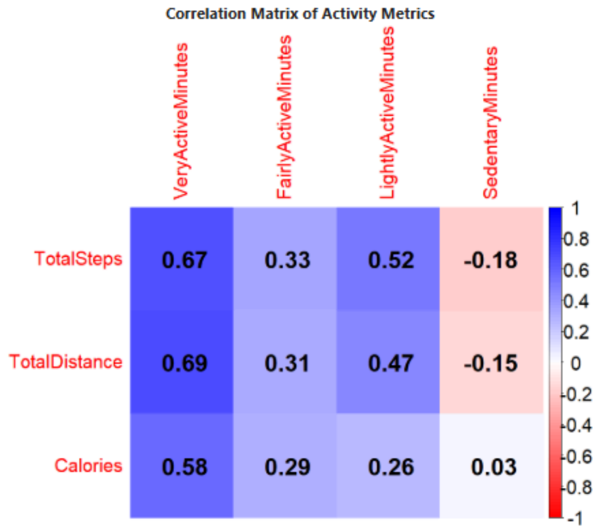


📌 **Insight: Increased activity** directly boosts **calorie expenditure**, reinforcing the value of **encouraging daily movement**.

---

8. Correlation Matrix of Activity Metrics

- **Strong positive** correlation between **VeryActiveMinutes** and (Steps, Distance, Calories).
- **SedentaryMinutes** is negatively correlated with Steps/Distance, slightly neutral with Calories.

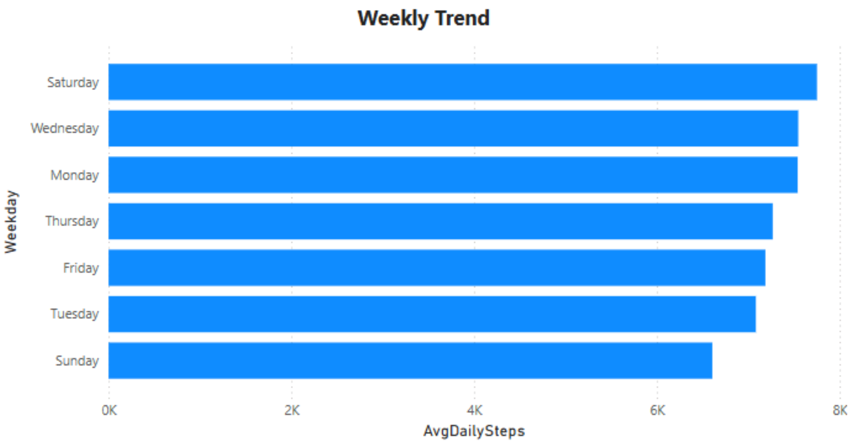


📌 **Insight: High-intensity activities** (very active minutes) are most impactful on steps and calorie burn, while **sedentary** behavior reduces overall movement.

---

9. Weekly Trend – Average Steps by Day of Week

- **Saturday** leads in average steps, while **Sunday** shows the lowest.
- Weekdays vary, with moderate to high steps on certain days.



📌 **Insight: Weekend peaks** suggest leveraging **Saturday fitness challenges**, while **Sunday** may need gentle reminders or rest-day content.

---

6. Recommendations

1. **Personalized Engagement Strategies**
   - Target **morning and evening** for app notifications and workout prompts.
   - Segment users by **activity level** (especially sedentary) to offer relevant challenges.

2. **Sleep & Activity Integration**
   - Use **sleep data** to provide **customized suggestions** for better rest and subsequent activity.
   - Offer **stress-reduction features** (guided breathing) during peak heart rate hours.

3. **Marketing Focus**
   - Emphasize **weekend campaigns** to leverage Saturday's natural activity spike.
   - Create promotions that **encourage short bursts of active minutes** for higher calorie burn.

4. **Weight Tracking Support**
   - Prompt **consistent weight logging** to capture meaningful trends.
   - Provide **insights on minor weight fluctuations** and tie them to daily steps or calorie data.

---

7. Final Conclusion

The analysis of non-Bellabeat Fitbit data highlights:

- **Many days under 5K steps** → Indicating sedentary lifestyles.
- **Large sedentary user segment** → Potential for motivational campaigns.
- **Peak heart rate hours in morning/evening** → Opportunity for well-timed reminders.
- **No strict correlation between sleep and steps** → Suggesting multiple user behavior clusters.
- **Active minutes strongly drive calorie burn** → Reinforcing daily movement as a key strategy.
- **Saturday** as the most active day → Ideal for weekend fitness challenges.

By **leveraging these insights**, Bellabeat can tailor its Time watch experience—sending timely notifications, crafting personalized workout suggestions, and marketing specialized content to users based on their activity segments. This data-driven approach positions Bellabeat to **enhance user engagement** and **support healthier lifestyles** for its Time watch customers.

Appendix:
The full Python notebook detailing data cleaning, aggregation, and star schema creation, along with the Power BI dashboard file for interactive exploration are added to the repository.
