Bellabeat Consumer Data Analysis

Data Analysis Project

By: Gwendolyne Madrid

Project Overview & Business Task

Understanding User Behavior to drive growth

- Identify ways consumers use smart devices to inform a targeted marketing strategy.

Key stakeholders: Bellabeat marketing analytics team.

Data Preparation

Primary source: FitBit Fitness Tracker Data (Kaggle)

Data Scope: March 12 - May 12, 2016

Key Data Points Used for Analysis:

Daily Activity: Steps, distance, active minutes, calories

Sleep Monitoring: Sleep duration, time in bed, sleep efficiency

Note: Heart rate and weight data available for future analysis

Data Quality: High - Reliable, Original, Comprehensive, Current, Cited (ROCCC)

Data Cleaning & Manipulation (1/2)

Tools & Workflow:

-The data processing was performed using R in Posit Cloud.
-Key Packages utilized:
-Tidyverse (specifically dplyr and lubridate) for data manipulation and data functions.

-Both datasets CSV files were imported and combined into a single data frame for an unified analysis.

-The str(), summary(), and head() functions were used to understand the data structure and identify initial issues like missing values or incorrect data types.

Data Cleaning & Manipulation (2/2)

Data Cleaning & Transformation:

Combined Datasets: Merged the two time periods (March-April and April-May) into a unified dataset.

Fixed Date Format: Converted ActivityDate from text to proper date format for time-based analysis.

Data Integrity Checks: Removed duplicate entries and verified data consistency across the combined dataset.

Analysis Preparation:

Created User Summary: Generated a separate analysis table with calculated metrics per user:

Average daily steps, calories, and sedentary minutes

Total number of usage days

Segmentation Ready: Structured data to enable user segmentation by activity level.

Outcome: The resulting datasets are clean, validated, and optimized for exploratory analysis and visualization.

Analysis Summary

Methodology:

-Very active users (10,000+ steps/day) use their devices most consistently (43.3 days)

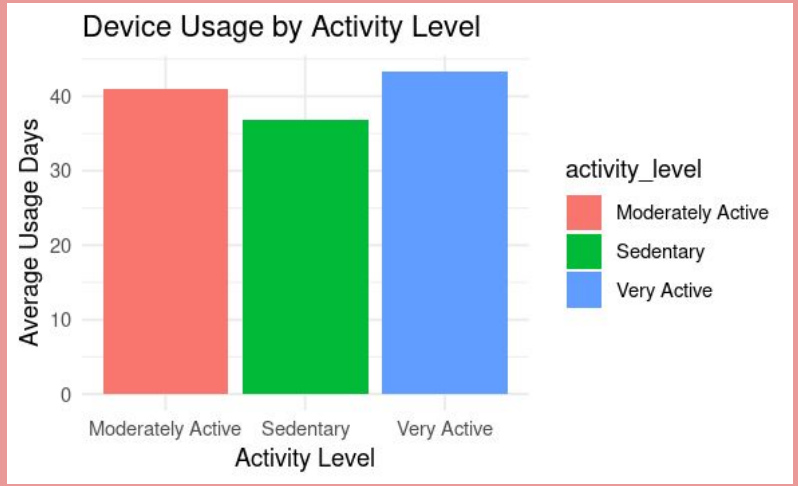
-The cleaned dataset was analyzed using R to calculate descriptive statistics and aggregate metrics.
-Key calculations include:
-Mean activity level, sleep data.

Activity Level Summary

-Very active users (10,000+ steps/day) use their devices most consistently (43.3 days)

-Sedentary users (<5,000 steps/day) use their devices least consistently (36.9 days)

-More active = more consistent usage.



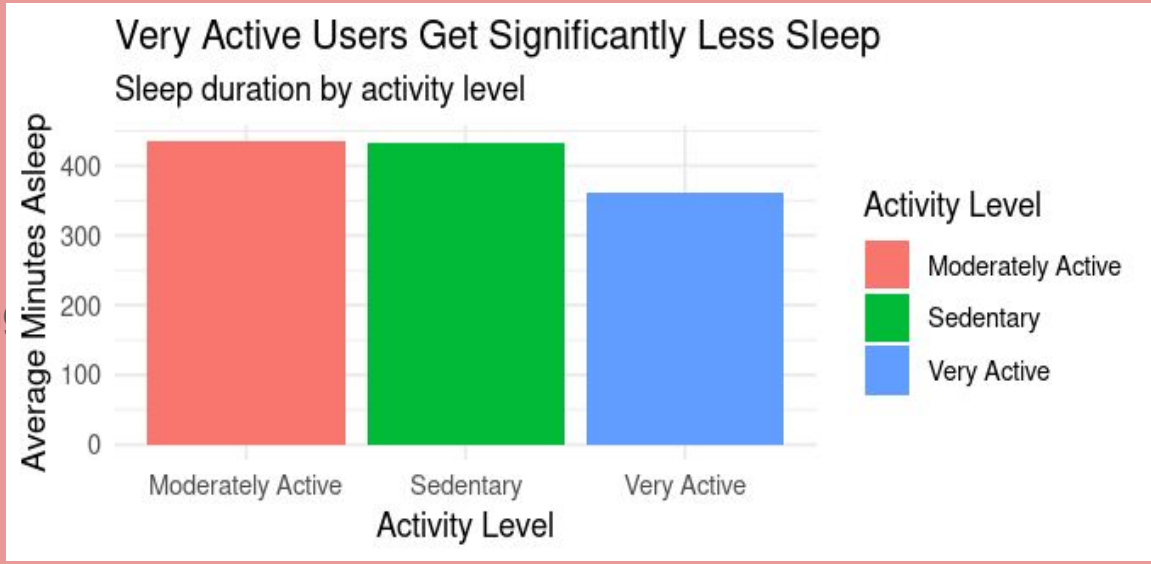
Sleep Data Summary

Very Active Users:

-Sleep 1.2 hours LESS than other groups (361 vs 434 min)

-Have significantly worse sleep efficiency (85% vs 93%)

-Spend more time tossing and turnin in bed.



Moderately Active Users:

-Get the optimal sleep (7.3 hours)

-Have the best sleep efficiency (94%)

-Represent the "sweet spot" for balanced health

Recommendations

1. Target the "Balance Seekers":

-Focus marketing on moderately active users who value consistent, sustainable health habits.

-Develop features that help very active users optimize sleep and recovery.

2. Leverage the Engagement-Activity Connection:

-Create gamification that rewards consistent daily use, not just high activity spikes.

-Develop reminders that encourage sedentary users to build gradual habits.

3. Position as a Holistic Health Partner:

-Emphasize sleep and recovery tracking as much as activity tracking.

-Create content about balancing intense workouts with quality rest.


thank
you