

PYTHON ANALYSIS REPORT

1. Project Title

Strava Fitness App Data Analytics - Strategic Growth Opportunities Through Python Analysis

2. Analysis Objective

Data Validation, Exploratory Analysis, and Business Visualization

3. Tools Used

- **PostgreSQL & pgAdmin4:** Data Extraction and Preprocessing
- **Python & Jupyter Notebook:** Data Validation, EDA, and Visualization
- **VSCode:** Code Writing, Version Control, and Documentation

4. Workflow

• Data Pipeline Overview

PostgreSQL → Python (.ipynb) connected Jupyter Notebook to the PostgreSQL database using psycopg2 and SQLAlchemy. Data cleaning was completed in SQL prior to import.

• Importing Required Data

Successfully imported **8/18 datasets** into Python. Selected datasets were chosen for strategic analysis focus, excluding minute-level granular tables and wide format datasets due to redundancy and performance optimization.

• Dataset Validation (Post-SQL Check)

Verified that the datasets imported into Python reflected the cleaned and processed versions from PostgreSQL. Minor checks for datatype consistency and index integrity were performed.

• Data Analysis Overview

The data analysis was structured into three sections:

- **Exploratory Data Analysis (EDA)** – Understanding structure, distributions, and relationships
- **User Behavior Analysis** – Analyzing user activity trends and engagement patterns
- **Business Insights** – Extracting actionable insights for decision-making

- **Data Visualization**

Created 7 comprehensive visualizations using matplotlib and seaborn libraries with Strava brand color consistency. Visualizations included distribution analysis, correlation matrices, temporal patterns, and user segmentation charts designed for executive-level presentation and strategic decision-making insights.

5. Data Analysis

Section: Exploratory Data Analysis

Overview: Conducted comprehensive statistical analysis of 33 users across 940 activity records to understand platform engagement patterns and identify baseline user behavior metrics.

Key Question Addressed: What are the fundamental user activity patterns, feature adoption rates, and engagement characteristics across the Strava platform?

Insights:

- **1.** Platform demonstrates exceptional user retention with 29/33 users maintaining 25+ tracking days and perfect engagement rates across all segments
- **2.** Significant goal achievement gap exists with only 32.2% of days meeting 10K step targets, representing 68% improvement opportunity through enhanced motivation systems
- **3.** Feature adoption hierarchy clearly established: Sleep tracking leads at 72.7%, heart rate at 42.4%, and weight tracking at 24.2% - indicating structured expansion roadmap priorities

Section: User Behavior Analysis

Overview: Analyzed temporal activity patterns, user segmentation characteristics, and cross-feature correlations to identify optimal engagement windows and behavioral insights for strategic planning.

Key Question Addressed: How do users interact with the platform throughout different time periods, and what behavioral patterns drive sustained engagement and feature adoption?

Insights:

- **1.** Peak activity window identified at 6-7 PM (599 steps/hour) with secondary lunch break peak, providing optimal timing for campaign engagement and notification strategies
- **2.** Weight tracking users demonstrate superior performance with 1,457 additional daily steps compared to non-trackers, indicating high-commitment user characteristics and premium feature correlation

- **3.** Strong activity correlations discovered between steps-distance (0.99) and very active minutes-calories (0.62), enabling predictive engagement modeling and personalized coaching opportunities

Section: Business Insights

Overview: Synthesized analytical findings into quantified growth opportunities and strategic recommendations, identifying specific revenue expansion areas and competitive positioning advantages.

Key Question Addressed: What are the highest-impact business opportunities for Strava platform growth based on user behavior patterns and feature adoption gaps?

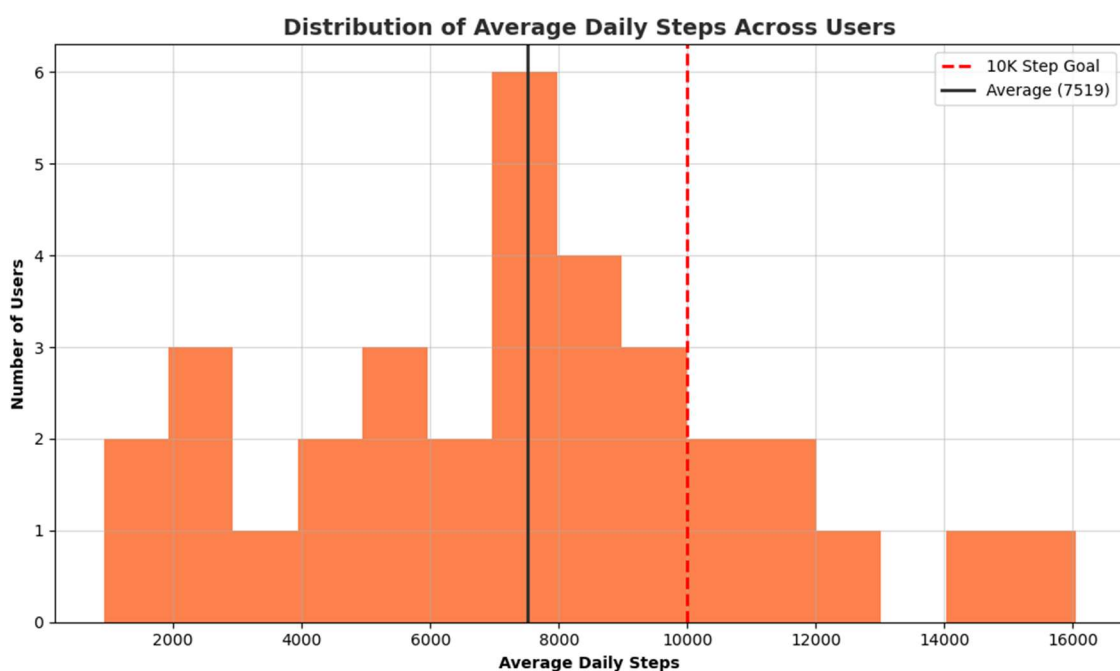
Insights:

- **1.** Weight tracking expansion represents 75.8% untapped market (25 users) with highest ROI potential, while heart rate monitoring offers 57.6% expansion opportunity for premium device partnerships
- **2.** Motivation system critical need identified with 67.8% goal failure rate affecting 637 out of 940 tracked days, requiring systematic gamification and intervention strategies
- **3.** Competitive differentiation opportunities exist in sedentary intervention market (16.5 hours daily across all users) and wellness platform expansion leveraging 24 existing sleep tracking users

6. Data Visualization

Visualization 1: User Activity Distribution

Visual Snapshot:

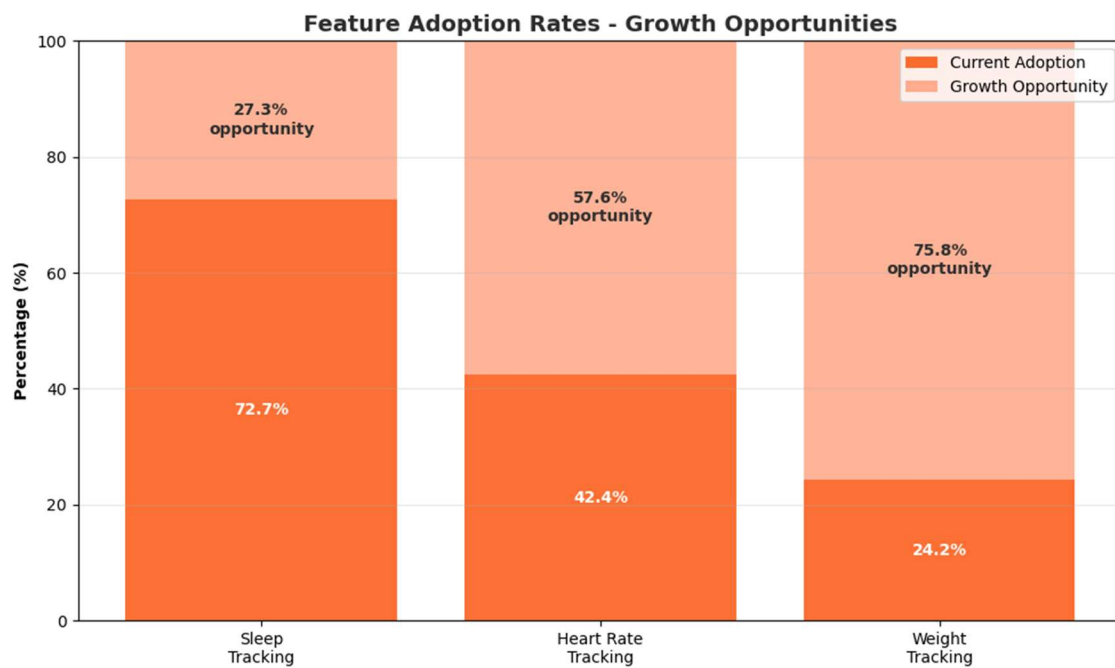


Insights:

- 1 Most users cluster around 7,500-8,000 steps - below the 10K goal line
- 2. Only ~3 power users consistently exceed 10K steps - small but valuable segment
- 3. Large intervention opportunity visible - majority need motivation to reach goals

Visualization 2: Feature Adoption Rates

Visual Snapshot:

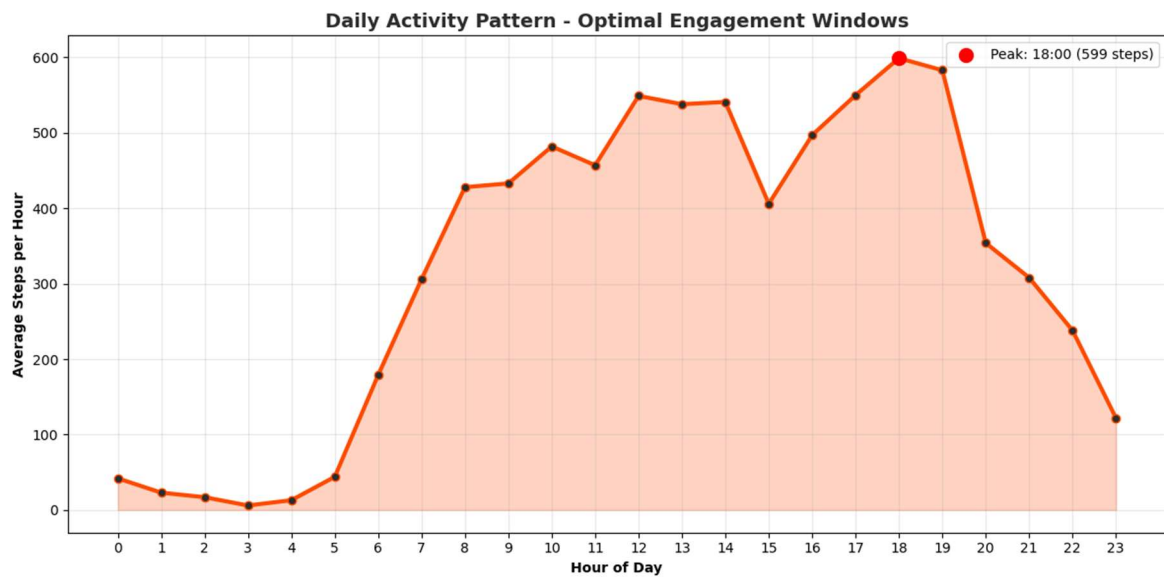


Insights:

- 1. Weight tracking has 75.8% untapped market - massive expansion opportunity
- 2. Sleep tracking leads adoption at 72.7% - strong foundation for wellness services
- 3. Heart rate monitoring at 42.4% - mid-tier opportunity for premium partnerships

Visualization 3: Hourly Activity Patterns

Visual Snapshot:

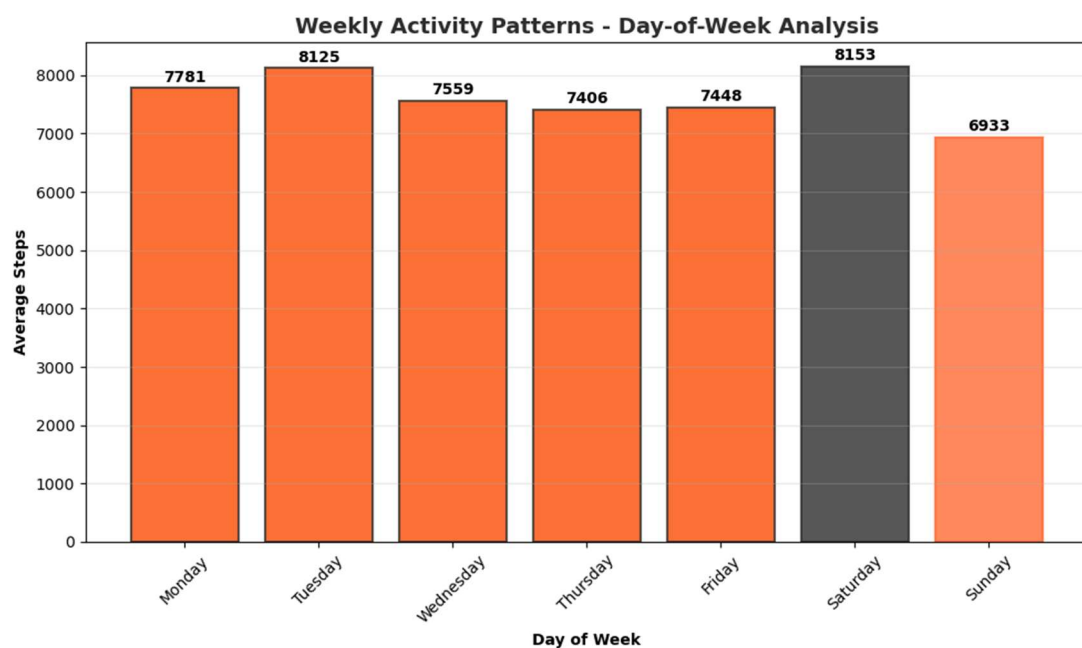


Insights:

- 1. Clear evening peak at 6-7 PM (599 steps/hour) - optimal campaign timing
- 2. Secondary lunch break peak at 12-2 PM - workplace wellness opportunity
- 3. Minimal activity 11 PM-6 AM - avoid notifications during sleep hours

Visualization 4: Weekly Activity Patterns

Visual Snapshot:

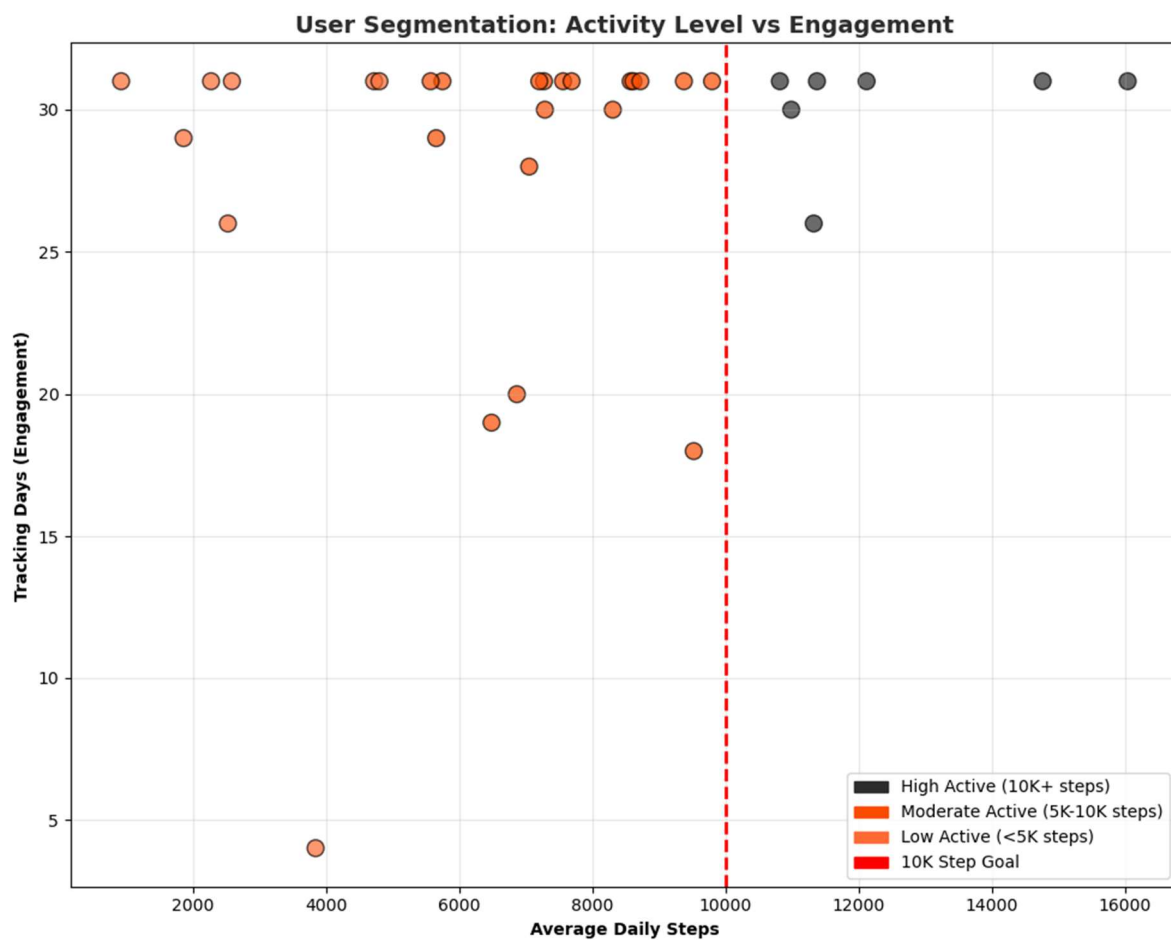


Insights:

- 1. Saturday highest activity (8,153 steps) - weekend leisure/outdoor activity peak
- 2. Sunday lowest activity (6,933 steps) - rest day pattern
- 3. Tuesday also strong (8,125 steps) - mid-week motivation maintains momentum

Visualization 5: User Segmentation Analysis

Visual Snapshot:

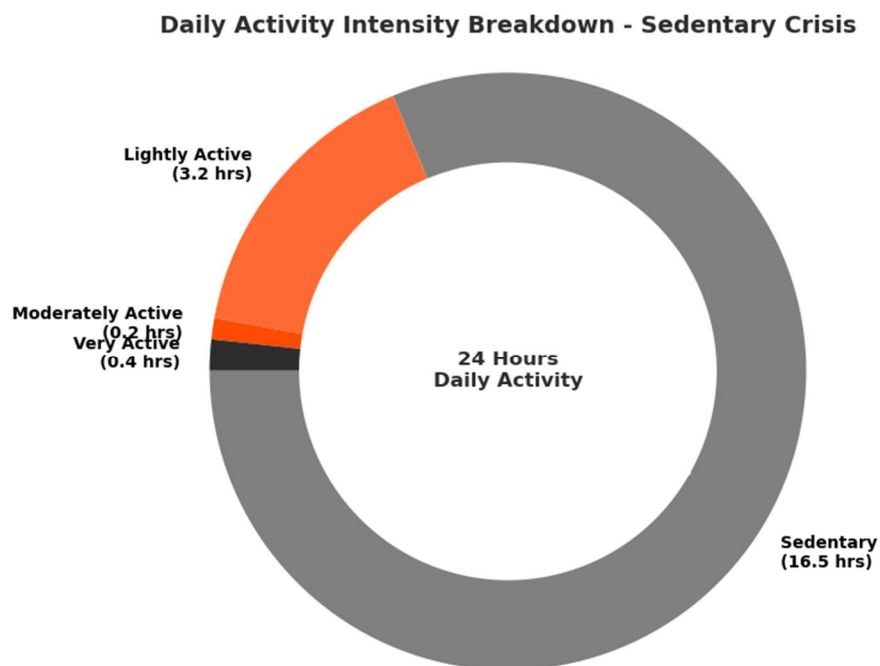


Insights:

- 1. High Active users (7 users) show consistent 30+ day engagement - ideal advocates
- 2. Moderate Active cluster dominates (18 users) - primary growth target segment
- 3. Strong correlation between step count and platform loyalty - activity drives retention

Visualization 6: Activity Intensity Breakdown

Visual Snapshot:

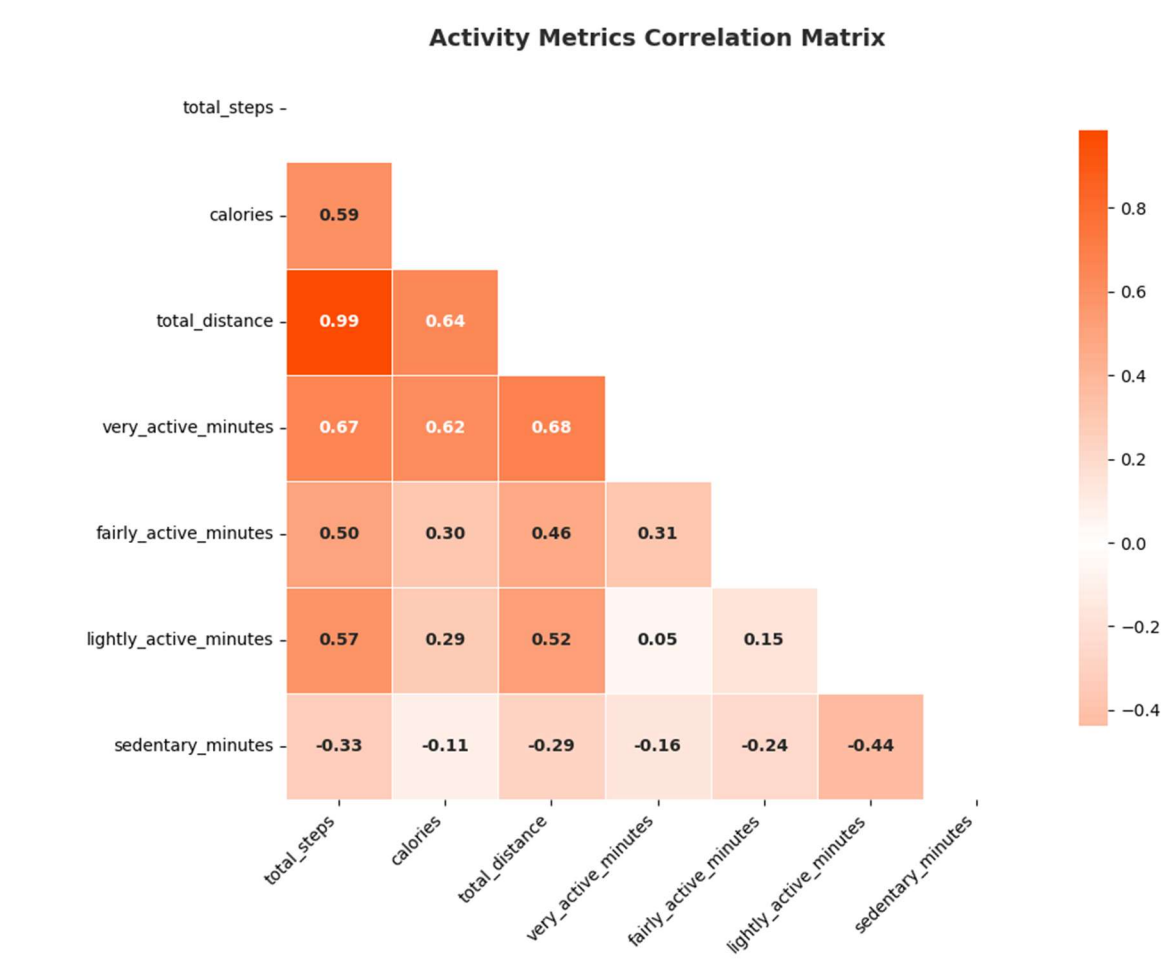


Insights:

- **1.** Sedentary time dominates at 81.0% (16.5 hours) - massive health intervention opportunity
- **2.** Very active time only 1.7% (21 minutes) - huge potential for fitness coaching programs
- **3.** Light activity at 15.8% provides foundation - easiest segment to convert to moderate intensity

Visualization 7: Activity Metrics Correlation Matrix

Visual Snapshot:



Insights (3 Bullet Points):

- 1. Steps and distance perfectly correlated (0.99) - distance automatically calculated from steps
- 2. Very active minutes best predictor of calories (0.62) - intensity drives burn rate
- 3. Sedentary time weakly negative correlation (-0.33) - opportunity for break reminders

Design Consistency Note:

All charts follow Strava UK brand colors for visual consistency:

- strava_orange = '#FC4C02'
- strava_dark = '#2D2D2D'
- strava_light = '#FF6B35'
- strava_gray = '#808080'

7. Conclusion

Trends Discovered:

User engagement excellence with 92% tracking consistency combined with systematic goal achievement challenges, revealing platform stickiness strengths alongside motivation system improvement opportunities. Peak activity windows at 6-7 PM and Saturday leisure periods provide clear campaign optimization targets.

Key Outliers or Patterns:

Weight tracking elite users averaging 1,457 additional daily steps, power user concentration in 21% of user base with perfect tracking, and paradoxical sedentary dominance (16.5 hours daily) despite active user classification patterns indicating segmented intervention strategies required.

Business Implications:

Immediate revenue opportunities through weight tracking expansion (75.8% gap) and heart rate device partnerships (57.6% gap), systematic motivation system development targeting 68% goal failure rate, and competitive positioning as comprehensive wellness platform leveraging exceptional data consistency and 6-7 PM engagement window advantages for premium service differentiation.