Fitbit Data Visualization System Report

1. INTRODUCTION

This project outlines the development of an interactive data visualization dashboard designed to promote healthier habits by helping users track and compare their fitness metrics. Using raw Fitbit data, the system processes and predicts missing values using machine learning regression models. It then transforms this data into engaging visualizations through JavaScript and Vega-Lite.

The dashboard enables users to explore trends across different timeframes (daily, weekly, monthly) and compare their performance with top users to stay motivated. By combining data analytics, predictive modeling, and real-time interactivity, the project offers a personalized and motivating fitness tracking experience.

2. REQUIREMENT ANALYSIS

User Needs

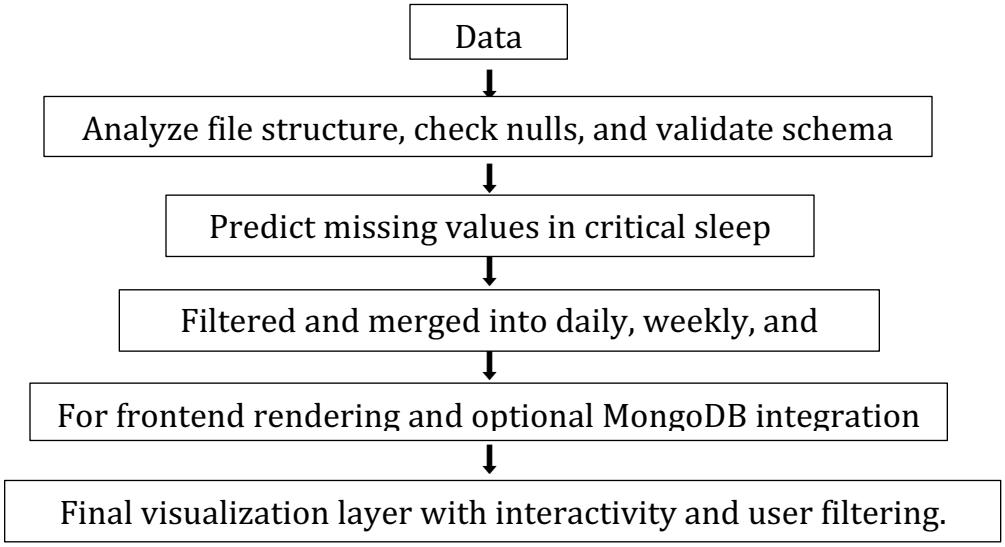
Through early-stage discussions and feedback sessions, we identified key expectations from users who regularly engage with fitness trackers like Fitbit. The core needs were:

- The ability to view personalized health metrics, especially step counts, sleep patterns, and calories burned.
- A feature to compare their performance with top performers, which users believe would help them stay accountable and inspired.
- A preference for clean, visually engaging dashboards with interactive elements such as dropdowns and filters - users explicitly expressed dislike for static tables or overly complex reports.

3. SYSTEM OVERVIEW & WORKFLOW

3.1 Workflow Pipeline

The project followed a structured data pipeline to ensure accuracy, scalability, and user favored visualizations. The complete workflow is as follows:



This modular flow allows each component to be updated or improved independently.

3.2 Data Issues & Processing

- Across the 11 CSV files, we identified **33 unique user IDs** which has data across time frames. Dataset consists of one month of hourly, daily, weekly data which has been consolidated.
- A significant portion of the data was **incomplete**, especially for sleep metrics like TotalMinutesAsleep and TotalTimeInBed.
- Some columns were **too sparse** (e.g., only 67 valid entries out of ~9400) and thus excluded from final analysis and visualization.
- Data cleaning and prediction ensured that the remaining fields were reliable, which helped maintain the dashboard's accuracy and usefulness.

4. Algorithm Design

4.1 Missing Data Prediction

To address the issue of incomplete data—particularly in key sleep-related metrics—we implemented a machine learning-based imputation approach:

- Applied **regression models**, including **Random Forest Regressor** and **Linear Regression**, to estimate missing values for TotalMinutesAsleep and TotalTimeInBed.

- Input features for the models included strongly correlated variables such as TotalSteps, Calories, SedentaryMinutes, and VeryActiveMinutes.
- Models were trained on the complete portion of the dataset and evaluated using standard metrics like R^2 and RMSE.
- Only predictions that met a **high confidence threshold** were used to populate missing fields; all others were left out to maintain data integrity.

4.2 Data Consolidation

Following data cleaning and prediction:

- We generated three **time-aggregated consolidated datasets**:
  - daily_consolidated.csv – captures day-wise user activity and sleep data
  - weekly_consolidated.csv – aggregates metrics week by week
  - monthly_consolidated.csv – provides an overview of longer-term trends
- These files formed the core data source for all visualizations and were later converted into a flattened JSON dataset for frontend integration as Vega-lite is easily compatible with json to develop the javascript code base.

5. Interface Design

5.1 Tools Used

The system interface was built using lightweight and accessible web technologies to ensure fast rendering, modular development, and smooth interactivity:

- **Frontend:** Developed using standard **HTML**, **CSS**, and **JavaScript** for broad browser compatibility and ease of deployment.
- **Visualization Library:** Implemented **Vega-Lite** (via vega-embed.js) to define interactive, declarative chart specifications efficiently.
- **Data Source:** Consolidated datasets were converted into **JSON** format and either **embedded locally** or **fetched from MongoDB**, depending on deployment needs.

This tech stack allows for flexible updates, seamless visual integration, and minimal loading times.

5.2 Pages Overview

The dashboard is structured into **two main HTML pages**, each serving a distinct purpose in user interaction. These pages are currently **linked via anchor tags**,

allowing users to easily switch between self-tracking and top performer benchmarking.

- **Page 1 – Top Performers Dashboard (top_performer.html)**
  - Displays the **Top 10 users** based on average daily steps.
  - Includes two **interactive dropdowns**:
    - **User ID selector** (from the top 10 users)
    - **Timeframe selector** (Daily / Weekly / Monthly)
  - All visualizations are **dynamically updated** using Vega-Lite based on the selected user and timeframe.
  - A hyperlink is provided to navigate to the **Self Dashboard (Page 2)**.
- **Page 2 – Personal Dashboard (self.html)**
  - Focuses on a specific user's **personal fitness trends over time**.
  - Visualizations include:
    - **Sleep metrics** (TotalMinutesAsleep, TimeInBed)
    - **Activity levels** (Steps, Active vs Sedentary time)
    - **Calories burned**
  - Provides a **goal-oriented view**, helping users assess their routine and progress.
  - A hyperlink is provided to return to the **Top Performers Dashboard (Page 1)**.

This two-page structure currently enables a simple yet effective interaction flow, moving between **self-reflection** and **performance benchmarking** via clickable links. A more tightly integrated comparison feature (planned as Page 3) is under development for future iterations.

6. Charts and Metrics

To ensure that users can interpret their data intuitively, we selected chart types that align with the nature of each metric. The visualizations are designed for clarity, engagement, and meaningful comparison across users and timeframes.

Metric	Visualization Type	Purpose
Steps	Radial Chart (Daily), Show card (Monthly average steps)	Highlights step trends over time, helping users identify patterns in their daily activity levels.
Calories Burned	Show Card (Daily & Monthly average calories), Area Chart (Monthly Data)	Displays energy output progression while emphasizing cumulative burn visually.
Sleep vs Time in Bed	Area Chart (Monthly Data), Radial Chart (Daily Data)	Helps users analyze **sleep efficiency** and time in bed to actual sleep duration.
Activity Breakdown	Stacked Bar Chart (Daily Data), Heat map (Monthly Data)	Illustrates the proportion of **very active**, **moderately active**, and **sedentary minutes**, offering an at-a-glance view of daily movement balance.

Each chart is embedded using **Vega-Lite specifications** and dynamically updates based on the user and timeframe selections, ensuring a personalized and responsive dashboard experience.

7. Code Description

This section outlines the structure and responsibilities of the codebase that powers the interactive dashboard, focusing on how the HTML and JavaScript components collectively render the multi-page user experience.

7.1 HTML Files

These serve as the **entry points** for each page in the dashboard, linking to corresponding JavaScript files and rendering interactive charts using Vega-Lite.

- **top_performer.html**
  The landing page for **Page 1**, displaying the **Top 10 users** with interactive filters for user selection and timeframe.
- **self.html**
  Entry page for **Page 2**, which renders **personal dashboards** based on a specific user's fitness data.
- **daily_top_perfomer.html**
  A specialized variant of Page 1, focusing specifically on **daily aggregated data** for top users.

7.2 JavaScript Files

Each file is tailored to render charts dynamically using Vega-Lite specs based on user interaction:

- **top_perfomer.js**
  Loads and visualizes the top users' data. Handles dropdown interactions and dynamically renders multiple charts.
- **daily_top_performer.js**
  Focuses on visualizations for **daily timeframe** data, using a similar logic to top_perfomer.js but scoped to daily metrics.
- **self.js**
  Generates visualizations for the **personal dashboard**, pulling user-specific data to plot trends and behavior patterns.

7.3 Key Functions

Function Name	Description
loadData(timeframe)	Dynamically loads the corresponding JSON file (daily, weekly, or monthly) for use in charts.
populateDropdowns()	Populates the user and timeframe dropdowns based on the available dataset entries.
updateDashboard(userId)	Fetches data for a specific user and updates all visualizations accordingly.
createComparison()	(Planned/Partial) Compares selected user's stats from Page 1 with the user's own data from Page 2.

Each of these functions is modular and linked to the interactivity of the dashboard, ensuring that the charts reflect real-time user choices without needing to reload the page.

8. Alignment with Prototype

Our original prototype, *HealthHabit*, was centered on motivating users through visual comparisons and intelligent insights derived from wearable fitness data. The final implementation successfully delivered on several core ideas while making intentional adjustments to optimize performance, usability, and simplicity.

Feature	Prototype Vision	Final Implementation
User Motivation	Enable users to compare their data with top performers for encouragement and goal setting.	Fully realized in **Page 1** (Top Performers view) and **Page 3**(Comparison Dashboard), allowing side-by-side performance benchmarking.
Predictive Analytics	Use AI/ML models to fill gaps in user data, especially for sleep and recovery metrics.	Implemented using **regression models** to predict missing values for TotalMinutesAsleep and TotalTimeInBed with confidence filtering.
Real-time Interaction	Interactive filters to change views instantly without refreshing the page.	Achieved through **JavaScript dropdowns**, enabling **instant updates** of all visualizations based on selected user and timeframe
Visualization Stack	D3.js or Plotly.js for custom and flexible data graphics	Switched to **Vega-Lite with vega-embed.js** for faster prototyping, easier maintenance, and declarative chart definitions.

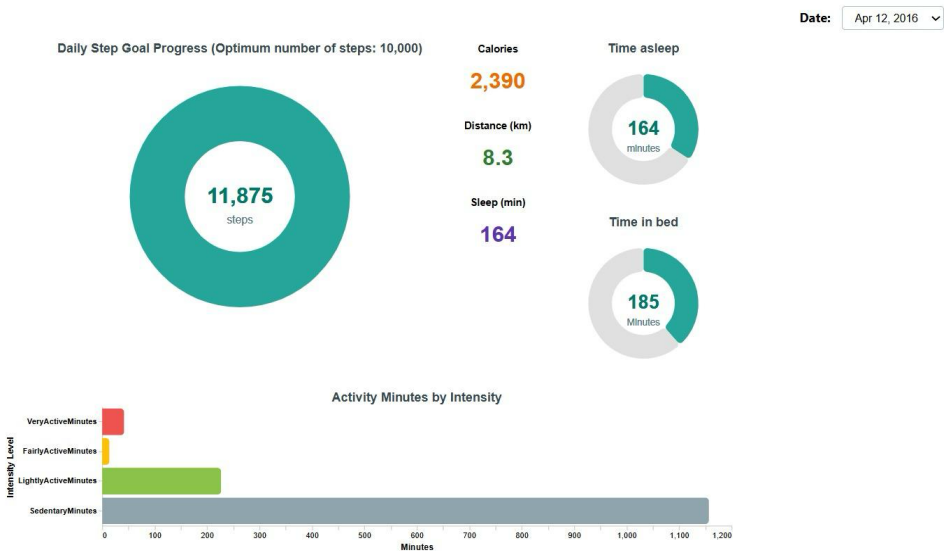
This alignment reflects thoughtful adaptation of the prototype into a more efficient and user-friendly system, without compromising on core functionality.

9. Evaluation & Results

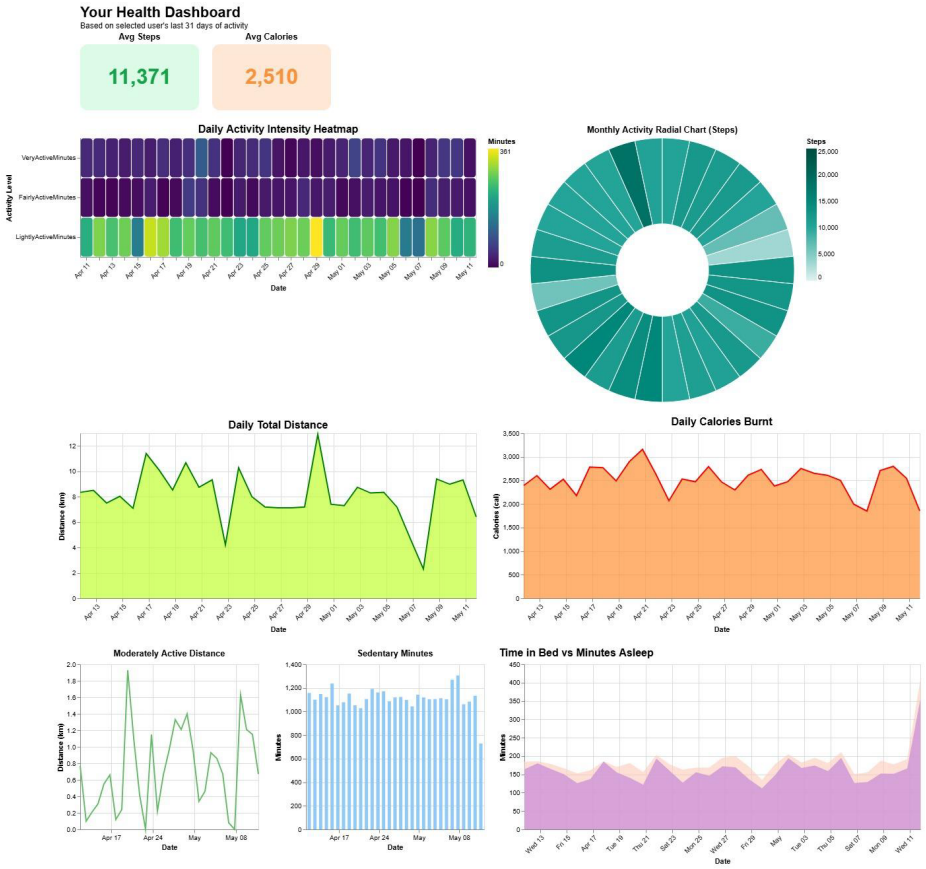
The final system delivers a robust, modular, and engaging user experience that meets the core goals defined in the initial project scope.

- The **dashboard is fully interactive**, with charts dynamically updating based on user selection of ID and timeframe. This interactivity has been implemented seamlessly across all three pages.
- The system is **responsive and modular**, allowing new metrics or visualizations to be added with minimal code changes thanks to the use of declarative Vega-Lite specifications.
- **Real-time comparison functionality** between users (especially in Page 3) is partially implemented, with dynamic data loading and chart updates fully operational. Further enhancement of automated insights is in progress.
- A round of **usability testing is planned**, targeting fitness users and peers, to gather feedback on design, ease of use, and overall motivation impact. These insights will help refine features and guide future development phases such as mobile responsiveness and onboarding improvements.

RESULTS



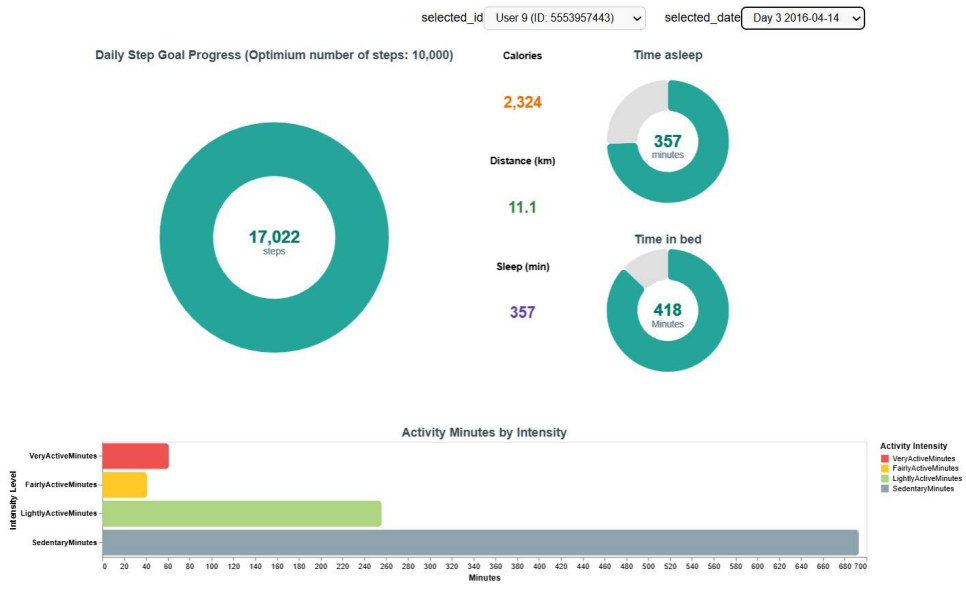
Daily Self Dashboard



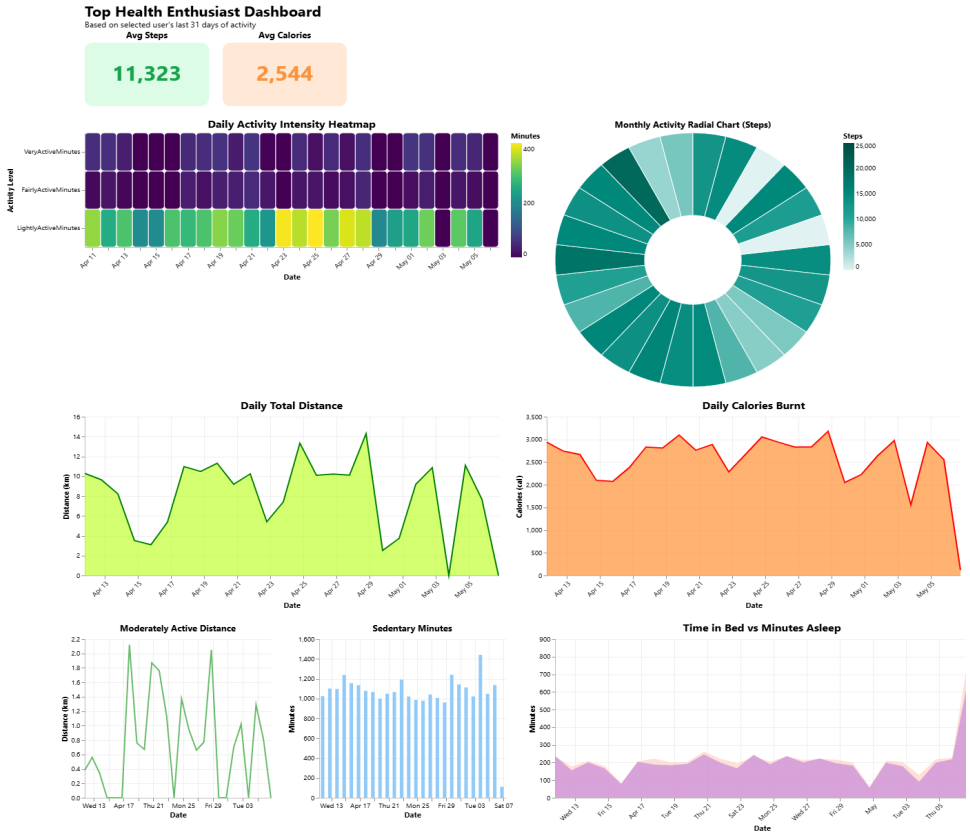
Monthly self Dashboard

Top Health Enthusiast Dashboard

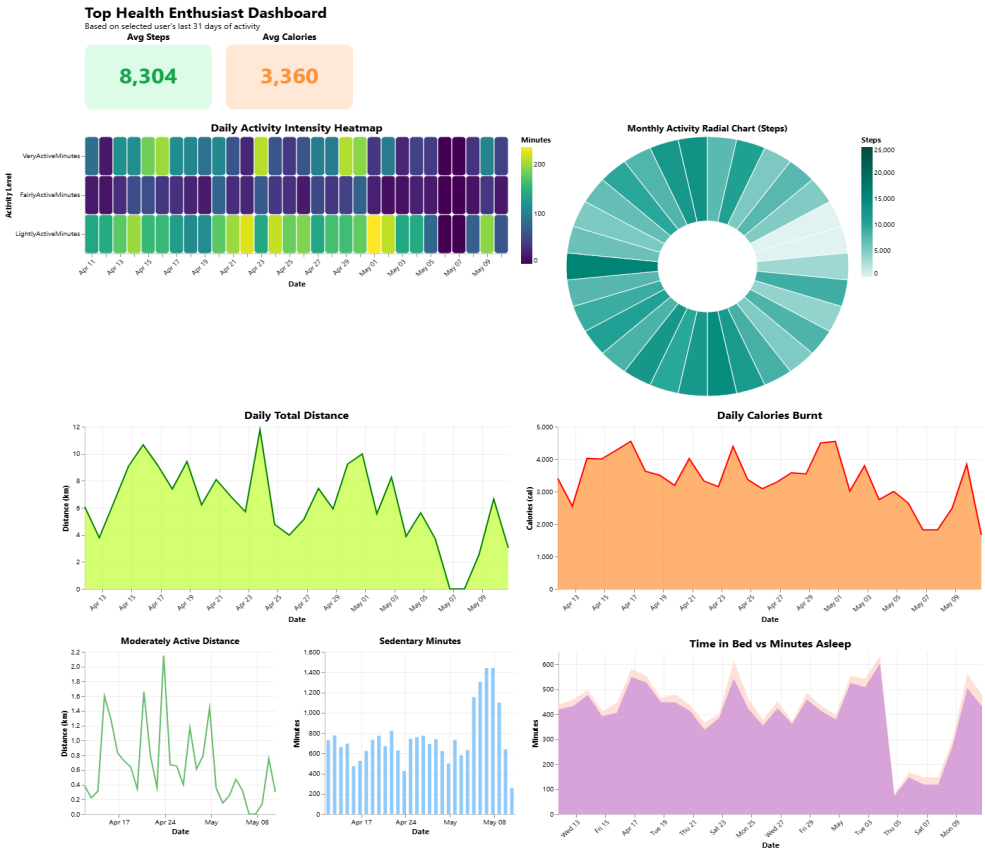
Check your metrics



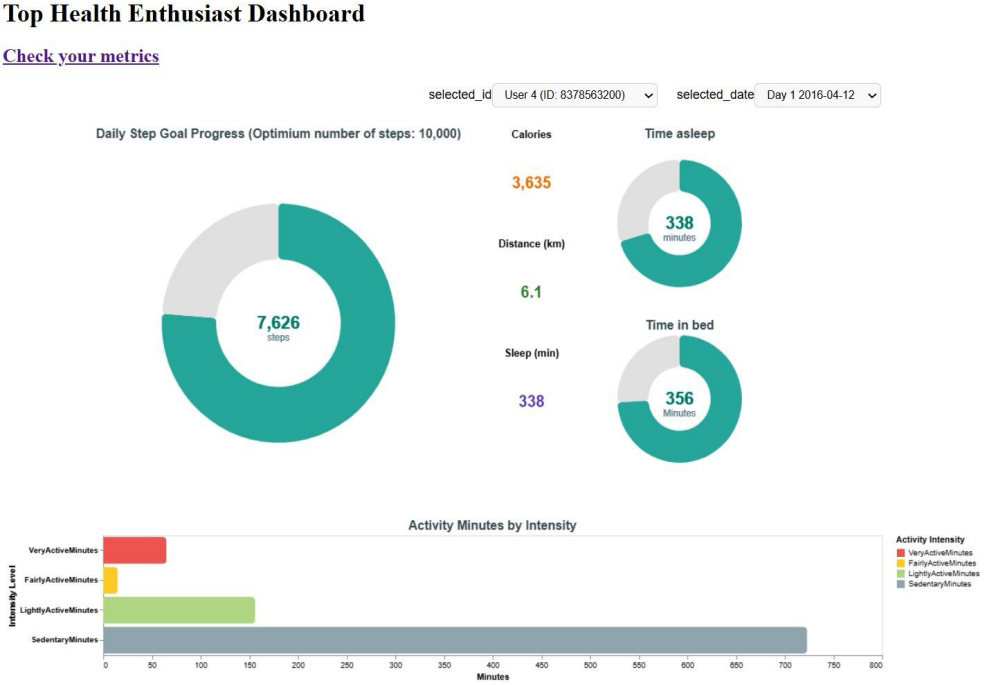
Daily Top Performer for user 9 Dashboard



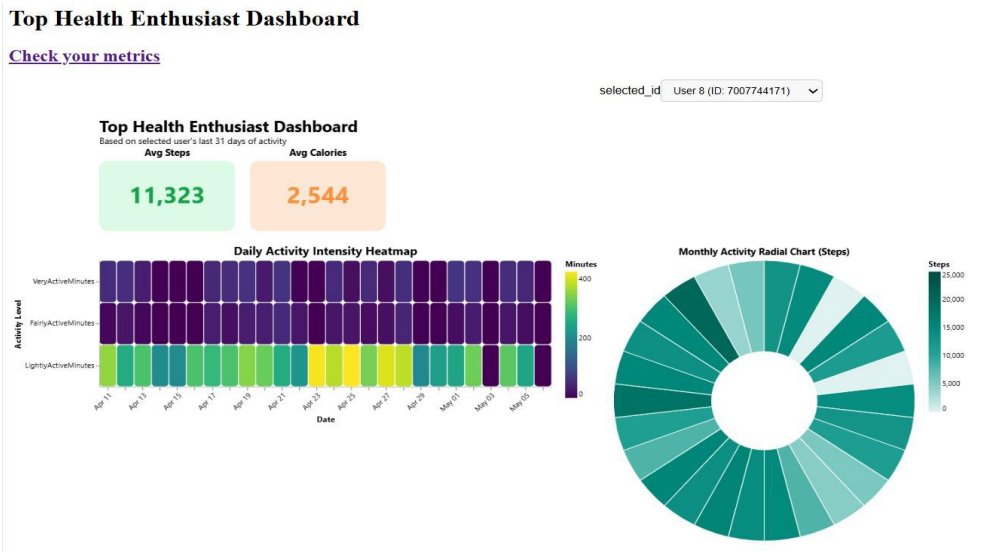
Monthly Top Performance for user 8 Dashboard



Monthly Top Performer for user 1 Dashboard



Daily Top Performer for user 4 Dashboard



Monthly Top Performer for user 4 Dashboard

10. Planned Enhancements

10.1 Comparison Dashboard (Page 3)

To further enrich the user experience, we plan to introduce a third page that will enable side-by-side comparisons between the selected top performer (from Page 1) and the personal user (from Page 2). This feature will:

- Automatically compare metrics such as step count, calories burned, and sleep efficiency.
- Generate insightful suggestions or feedback (e.g., "User A sleeps 20% more than you").
- Dynamically update as users switch between IDs and timeframes on Page 1.

This future enhancement will complete the system's transition from simple tracking to personalized, data-driven motivation.

10.2 User-Centered Enhancements

Our plan is to conduct an initial survey and interview phase, with 20 users at Texas A&M University, provided key insights that will guide upcoming improvements:

In future iterations, we aim to:

- Introduce goal tracking and visual nudges based on historical trends.
- Enable natural language summaries (e.g., "Your sleep improved by 15% this week").
- Add mobile responsiveness and UI personalization to better suit varying user preferences.

These enhancements will ensure the system evolves in a user-centric direction, grounded in real feedback and behavioral insights.

11. Conclusion

This project is a celebration of how data visualization can truly make a difference in our lives. By taking every day Fitbit data and turning it into clear, friendly visuals, we've created a tool that makes tracking fitness less about numbers and more about progress and personal well-being. Imagine looking at a chart that not only shows your activity trends but also motivates you to reach new heights every day, that's the heart of what we've achieved.

We've built a system that translates complex data into simple insights, giving everyone the chance to understand and engage with their personal health journey more intimately. Alongside the screenshots, system diagrams, and sample charts that will be part of the final submission, it's clear that this project is more than just about technology; it's about making health data work for you in a way that's supportive and encouraging.

Ultimately, our work here stands as a testament to the idea that when data is presented in a relatable, human way, it can inspire real, positive change—helping us all live healthier, happier lives