

Prompt Context:

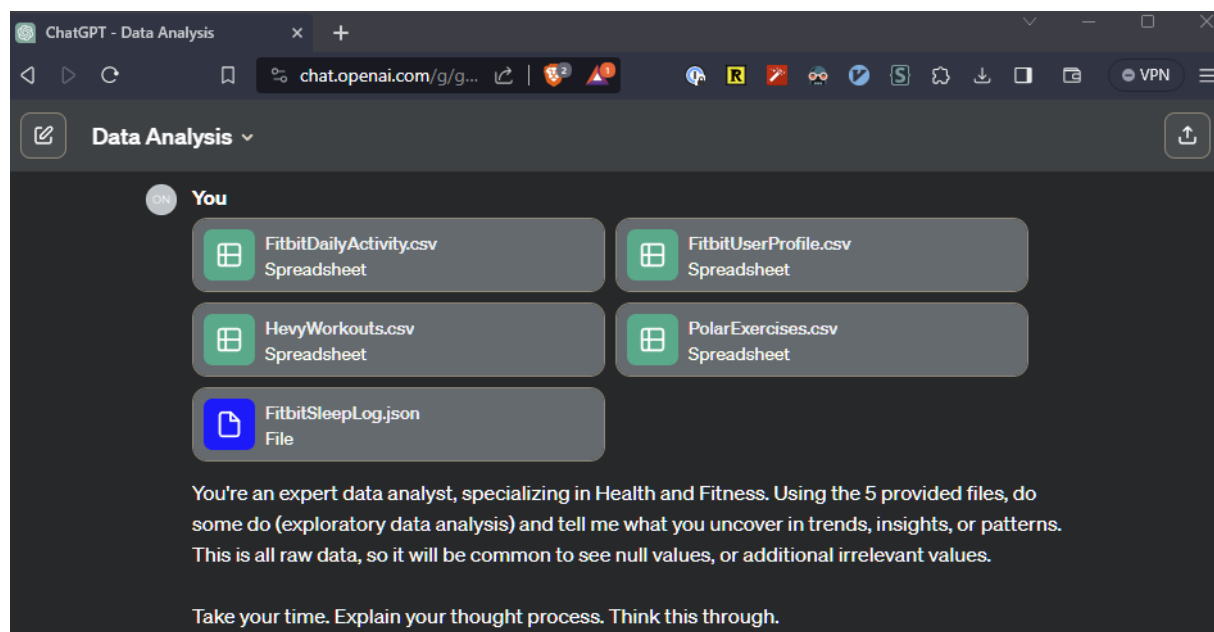
I provided ChatGPTv4 these 5 files to analyze

- [FitbitDailyActivity.csv](#)
- [FitbitUserProfile.csv](#)
- [PolarExercises.csv](#)
- [HeavyWorkouts.csv](#)
- [FitbitSleepLog.json](#)

Then used this prompt:

You're an expert data analyst, specializing in Health and Fitness. Using the 5 provided files, do some do (exploratory data analysis) and tell me what you uncover in trends, insights, or patterns. This is all raw data, so it will be common to see null values, or additional irrelevant values.

Take your time. Explain your thought process. Think this through.



ChatGPT insights/observations:

- ChatGPT gave great descriptive data of the files (see highlights on full response below)
- Didn't extrapolate much meaning in the data. I expected this, being < 1 month total data

My own insights, top takeaways from exploring the data:

- Learning is messy, this data is messy. Going to be a challenge to parse if too much is incomplete. It's crucial that I stick to the core top-level metrics (sleep, total volume lifted, HRMAX intensity training zones) and build off that. A "win" for the project is seeing the relationship of those key variables over time, and having access to all the historical data. Next steps will include transformation/serving, and using Power BI to explore trends.

ChatGPT (v4) Full Response:

Data Analysis

Great, let's start the exploratory data analysis (EDA) on the provided health and fitness data. Our goal is to uncover trends, insights, or patterns that may exist in the data. Given the nature of this analysis, we'll proceed in stages:

1. Data Loading and Initial Assessment:

We'll load each file and briefly assess its contents and structure. This will help us understand what kind of data we are dealing with.

2. Data Cleaning and Preprocessing: Based on the initial assessment, we will clean and preprocess the data. This may include handling missing values, irrelevant data, and ensuring the data is in a format suitable for analysis.

3. Exploratory Data Analysis: We will analyze each dataset to uncover trends, patterns, and insights. This will involve looking at distributions, correlations, and other statistical measures.

4. Synthesis and Reporting: We'll try to integrate findings from each dataset to form a holistic view of the health and fitness data.

Let's begin by loading and initially assessing each file.

1. Data Loading and Initial Assessment

1.1 Fitbit Daily Activity

Let's start by loading the FitbitDailyActivity.csv file and examining its contents.

The FitbitDailyActivity.csv file contains the following columns:

- **activeZoneMinutes:** Likely represents the number of minutes spent in active heart rate zones.
- **activityLevels:** Contains detailed activity level data, possibly in a nested JSON format.
- **calories:** Nested data showing basal metabolic rate (bmr) and total calories burned.
- **customHeartRateZones:** Custom heart rate zones, if any.
- **distance:** Total distance covered.
- **elevation:** Elevation gain.
- **floors:** Number of floors climbed.
- **heartRateZones:** Detailed data on time spent in different heart rate zones.
- **steps:** Number of steps taken.

This dataset seems to track daily physical activity metrics. It includes nested JSON structures that may require unpacking for a detailed analysis.

1.2 Fitbit User Profile

Next, let's load the FitbitUserProfile.csv file and take a look at its contents.

The FitbitUserProfile.csv file contains a wide array of columns, mainly focusing on user profile information. Key columns include:

- **age**: Age of the user.
- **ambassador**: Indicates if the user is an ambassador (likely a program or status within Fitbit).
- **autoStrideEnabled, strideLengthWalking, strideLengthRunning**: Related to stride measurements.
- **avatar, avatar150, avatar640**: URLs to user's avatar images at different resolutions.
- **averageDailySteps**: Average number of steps taken daily.
- **clockTimeDisplayFormat**: Format of time display (e.g., 24-hour).
- **corporate**: Indicates if the account is corporate.
- **dateOfBirth**: User's date of birth.
- **displayName**: User's display name.
- **gender**: User's gender.
- **height**: User's height.
- **memberSince**: Date since the user has been a member.
- **strideLengthWalkingType**: Type of stride length for walking.
- **timezone**: User's timezone.
- **topBadges**: Information about top badges earned.
- **visibleUser**: Whether the user profile is public or private.
- **waterUnit, waterUnitName**: Unit for water tracking.
- **weight, weightUnit**: User's weight and its unit.

This dataset appears to provide comprehensive profile information about Fitbit users, which could be useful for demographic analyses and understanding user characteristics.

1.3 Hevy Workouts

Now, let's load the HevyWorkouts.csv file to understand the workout data it contains.

The HevyWorkouts.csv file contains detailed workout data with the following key columns:

- **id, short_id, index**: Unique identifiers for each workout.
- **name**: Name of the workout.

- **description:** Description of the workout, potentially including perceived exertion or notes.
- **start_time, end_time:** Start and end times of the workout.
- **created_at, updated_at:** Timestamps for when the workout was created and last updated.
- **routine_id:** Identifier for the workout routine.
- **workout_type:** Type of the workout (e.g., strength training, cardio).
- **duration:** Duration of the workout.
- **like_count, is_liked_by_user:** Social features, such as likes.
- **is_private:** Whether the workout is private.
- **comments, comment_count:** User comments and their count.
- **media, image_urls:** Media associated with the workout.
- **exercises:** Detailed information about the exercises performed, likely in a nested JSON format.
- **estimated_volume_kg:** Estimated total volume of the workout in kilograms.

This dataset appears to track individual workout sessions, including types of exercises, durations, and user feedback. The detailed exercise data may provide insights into workout patterns and preferences.

1.4 Polar Exercises

Next, let's load and examine the PolarExercises.csv file.

The PolarExercises.csv file contains data related to exercise sessions with these key columns:

- **id:** Unique identifier for each exercise session.
- **upload_time:** Timestamp of when the exercise data was uploaded.
- **polar_user:** Link to the user's Polar profile.
- **device:** The device used for the exercise session (e.g., Polar Flow app, Polar BEAT).
- **start_time, start_time_utc_offset:** Start time of the exercise and its UTC offset.
- **duration:** Duration of the exercise.
- **heart_rate:** Average and maximum heart rate during the exercise.
- **sport:** The type of sport or exercise.
- **has_route:** Indicates if the exercise session included a route (e.g., for running or cycling).
- **detailed_sport_info:** More detailed information about the type of sport or exercise.
- **training_load_pro:** Data related to the training load, possibly in a nested JSON format.
- **distance:** Distance covered during the exercise (if applicable).
- **calories:** Calories burned during the exercise.

This dataset appears to focus on detailed exercise sessions, including heart rate data, type of exercise, and other metrics like distance and calories. It could be valuable for analyzing exercise patterns and intensity.

1.5 Fitbit Sleep Log

Finally, we'll load the FitbitSleepLog.json file to explore the sleep data it contains.

The FitbitSleepLog.json file, converted to a DataFrame, contains detailed sleep data with the following key columns:

- **dateOfSleep:** The date of the sleep log.
- **duration:** Total duration of the sleep session.
- **efficiency:** A metric indicating sleep efficiency.
- **endTime, startTime:** End and start times of the sleep session.
- **isMainSleep:** Indicates if this is the main sleep session (as opposed to naps).
- **levels.data:** Nested data detailing the sleep stages (light, deep, REM, awake) and their timestamps.
- **levels.shortData:** Additional data about short periods of sleep.
- **levels.summary:** Summary data for each sleep stage, including total minutes and counts.
- **logId, logType:** Identifiers for the sleep log.
- **minutesAfterWakeup, minutesAsleep, minutesAwake, minutesToFallAsleep:** Various metrics about the sleep session.
- **timeInBed:** Total time spent in bed.
- **type:** Type of sleep (e.g., stages).

This dataset provides a comprehensive view of sleep patterns, including duration, efficiency, and stages of sleep.

2. Data Cleaning and Preprocessing

Given the large number of columns and the presence of missing values, here's a streamlined approach:

1. **Handling Missing Values:** We'll focus on columns that are essential for analysis and check if they have significant missing values. If yes, we'll decide whether to impute or drop them.
2. **Unpacking Nested JSON Structures:** Columns with nested JSON (like activityLevels, heartRateZones in Fitbit data, exercises in Hevy Workouts, and levels.data in Fitbit Sleep Log) will be unpacked for a more detailed analysis.
3. **Standardizing Date and Time Formats:** We'll ensure that all date and time columns are in a consistent format to facilitate temporal analysis.

4. **Merging Datasets:** Where relevant, we might merge datasets, like linking user profiles to their activity or sleep data.

Fitbit Daily Activity

- Focus on key metrics: steps, distance, calories, active minutes.
- Unpack activityLevels, heartRateZones.

Fitbit User Profile

- Essential columns: age, gender, height, weight.
- Check for missing values in these columns and decide on imputation or exclusion.

Hevy Workouts

- Key focus: workout type, duration, exercises.
- Unpack exercises for detailed analysis.

Polar Exercises

- Focus on exercise type, duration, heart rate, distance.
- Handle missing values in heart_rate, distance.

Fitbit Sleep Log

- Main focus: sleep duration, efficiency, sleep stages.
- Unpack levels.data for sleep stage analysis.

3. Exploratory Data Analysis

We'll conduct a high-level analysis to uncover trends, insights, or patterns in each dataset.

3.1 Fitbit Daily Activity

- **Activity Patterns:** Analyze steps, distance, and floors data to understand daily activity patterns.
- **Calorie Burn:** Investigate the relationship between different types of physical activities and calorie burn.
- **Active Zone Minutes:** Explore how users spend their time in different heart rate zones.

3.2 Fitbit User Profile

- **Demographic Analysis:** Understand user demographics like age, gender, height, and weight distributions.
- **Engagement Metrics:** Look at average daily steps and membership duration to gauge user engagement.

3.3 Hevy Workouts

- **Workout Types and Durations:** Identify common types of workouts and their average durations.
- **Exercise Preferences:** Examine the types of exercises and their frequencies.

3.4 Polar Exercises

- **Exercise Intensity:** Analyze heart rate data to understand exercise intensity.
- **Sport Types:** Look at the variety of sports and exercise activities and their popularity.

3.5 Fitbit Sleep Log

- **Sleep Patterns:** Analyze sleep duration, efficiency, and stages.
- **Correlation with Activity:** Investigate if there's a correlation between daily activity levels and sleep patterns.

4. Synthesis and Reporting

- **Integrated Health Profile:** Combine insights from activity, workout, and sleep data to create a comprehensive health profile.
- **Behavioral Insights:** Identify correlations between lifestyle choices (e.g., exercise routines, sleep habits) and health outcomes.
- **Trends Over Time:** Look for trends and patterns over longer periods to assess the impact of lifestyle changes on health metrics.
- **User-Specific Recommendations:** Based on the combined analysis, generate personalized recommendations for improving health and fitness.

This analysis provides a holistic view of health and fitness, offering valuable insights into lifestyle habits and their impacts on well-being. By integrating data from multiple sources, we can better understand the interplay between physical activity, exercise habits, and sleep quality in the context of overall health.