

Fitness Tracker Coach

This dataset is taken from Kaggle.com here is the link:

<https://www.kaggle.com/datasets/arnavsmayan/fitness-tracker-dataset>

The dataset provided contains information from a **fitness tracker** with 1,000,000 entries by various individuals. It records key metrics of physical activity, sleep patterns, and mood, captured across different dates and environments. The data comprises a range of health-related parameters, including steps taken, calories burned, distance traveled, active minutes, and sleep hours, as well as average heart rates during various workouts. Additionally, it captures contextual information such as weather conditions, user location, and mood during activities.

This dataset is ideal for analyzing user behavior and fitness trends, building personalized fitness recommendations, or evaluating the impact of environmental conditions on physical activity and well-being.

The columns include:

- **user_id**: A unique identifier for each user.
- **date**: The date when the data was recorded.
- **steps**: The number of steps taken by the user.
- **calories_burned**: Calories expended during activities.
- **distance_km**: Distance covered in kilometers.
- **active_minutes**: Duration of activity in minutes.
- **sleep_hours**: Hours of sleep logged.
- **heart_rate_avg**: Average heart rate during the activity.
- **workout_type**: Type of workout (e.g., walking, running).
- **weather_conditions**: Weather during the activity (e.g., clear, rain).
- **location**: The place where the activity occurred (e.g., park, office).
- **mood**: The user's self-reported mood (e.g., happy, tired).

This dataset provides a comprehensive look at daily fitness and health-related behavior, enabling deep analysis into the factors influencing physical activity, wellness, and emotional states.

Some questions from dataset:

1. How does physical activity (e.g., steps, workout type) affect mood?

Ans: **Variables Involved:**

Steps
Workout type
Mood

Hypothesis: Higher levels of physical activity (more steps, intensive workouts) may lead to more positive moods (e.g., "happy"), while sedentary behavior or lighter workouts might correlate with neutral or negative moods (e.g., "tired", "stressed").

Example: We could calculate average mood scores (assign numerical values to moods) for different ranges of steps (low, medium, high) and analyze if there's a positive correlation. Look at specific workout types, such as "Running" or "Yoga," and see if they tend to correlate with happier moods.

2. Is there a relationship between sleep hours and heart rate average?

Ans: **Variables Involved:**

Sleep hours
Heart rate average

Hypothesis: More sleep might result in a lower average heart rate the next day, as better rest generally leads to lower stress and better cardiovascular health.

Example: We could group the data by ranges of sleep hours (e.g., less than 5 hours, 5-7 hours, 7+ hours) and calculate the mean heart rate for each group.

3. What is the impact of weather conditions on physical activity levels?

Ans: **Variables Involved:**

Weather conditions
Steps
Active minutes
Workout type

Hypothesis: People may be less active during bad weather conditions like "rain" or "snow," while they might engage in more outdoor activities during clear weather.

Example:

Compare the average number of steps and active minutes across different weather conditions to see if there's a significant drop in activity during inclement weather.

Analyze if indoor workouts (e.g., "Gym Workout") are more common during bad weather conditions.

4. How does the location of a workout affect mood?

Ans: **Variables Involved:**

Location

Mood

Hypothesis: Workouts done in nature (e.g., "Park") might be more beneficial for mental well-being, resulting in happier moods, while workouts in places like "Office" or "Gym" might correlate with more neutral or stressed moods.

Example: We could calculate the average mood score for each location and compare them to see if certain locations, such as outdoor spaces, have a higher average mood rating.

5. How does stress (as indicated by mood) affect heart rate during exercise?

Ans: **Variables Involved:**

Mood (especially "Stressed")

Heart rate average

Workout type

Hypothesis: People who are stressed may have a higher average heart rate during exercise due to elevated cortisol levels.

Example: We could compare the heart rates of individuals who reported feeling "Stressed" with those who felt "Happy" or "Neutral" during similar workout types.

Ethical Data Collection and Management Plan

Introduction to Ethical Data Collection and Use

The fitness tracker dataset contains sensitive information about users, including their physical activities, health metrics, and emotional states. To ethically collect and manage this data, we adhere to guidelines that ensure privacy, informed consent, data security, and fairness. The goal is to develop a solution that leverages the data for personalized fitness recommendations while safeguarding user rights.

Data Collection Considerations

1. Informed Consent:

- Users are fully informed about the type of data being collected, why it is being collected, and how it will be used. This includes health metrics (steps, heart rate), location data, and mood information.
- Consent is granular, allowing users to decide which data they are comfortable sharing (e.g., some users prefer to share steps but not heart rate or mood).

2. Anonymization:

- Data is anonymized so that users cannot be identified from the dataset. Even though user_id is present, it is de-identified by either masking or removing any personally identifiable information.
- Anonymization techniques such as **k-anonymity** or **differential privacy** can be employed to ensure that individuals cannot be re-identified through reverse engineering.

3. Minimization of Data:

- Collect only the data necessary for the specific objectives of the product.
- Avoid over-collection, especially when dealing with sensitive health and location data.

4. **Transparency:**

- The data collection process is transparent, providing users with easy-to-understand explanations of how their data will be processed, stored, and analyzed.
- Regular updates about the data usage, security practices, and possible third-party data sharing should be communicated to users.

5. **Compliance with Regulations:**

- Data collection comply with privacy regulations like **GDPR** (General Data Protection Regulation) or **CCPA** (California Consumer Privacy Act), which mandate user rights to access, delete, and control their personal data.

Selection of Appropriate Tools, Models, and Methods

1. Tools for Data Analysis and Preprocessing:

- **Pandas:** For data manipulation and cleaning, essential for handling structured data.
- **NumPy:** Efficient numerical computations, useful for manipulating fitness data.
- **Matplotlib/Seaborn:** Visualization tools for exploratory data analysis (EDA).

2. Machine Learning Models:

- **K-Means Clustering:** Segments users based on fitness behavior; suitable when the number of clusters is known.
- **DBSCAN:** Clusters based on data density, useful for discovering natural user groups and handling noise.
- **Random Forest/XGBoost:** High-accuracy models for predicting outcomes (e.g., mood, activity); useful for structured data.
- **Logistic Regression:** Simple classification model for predicting binary or multiclass outcomes (e.g., mood levels).
- **ARIMA:** Time-series forecasting for predicting trends like daily steps; best for linear time-series data.
- **LSTM:** Deep learning model for long-term trend predictions, suitable for sequential data with complex dependencies.

3. Bias Detection and Mitigation:

- **Aequitas:** Tool for auditing bias in model predictions, ensuring fairness across different user groups.
- **Fairness Indicators:** Measures fairness metrics across demographic groups, ensuring no segment is disproportionately impacted.

4. Explainability Tools:

- **SHAP:** Interprets complex model outputs, explaining feature importance and model behavior.

Crew member:

- Ajaypal Singh (0855054)
- Rohit (0860615)
- Gagandeep Singh (0858500)
- Tamanna (0857860)