**Tarefa da aula 02 - Coleta de Dados**

**Aluno:** Marco Leone Merini  
**Data:** 12/08/2025  
**Professor:** Paulo Manseira

**1. Nome do Dataset**

**FitBit Fitness Tracker Data**

**2. Fonte**

Kaggle – Disponível em: <https://www.kaggle.com/datasets/arashnic/fitbit>

**3. Justificativa**

Este dataset foi escolhido por conter registros diários de usuários de dispositivos Fitbit, incluindo dados sobre passos, calorias queimadas, minutos ativos e sono. Essas variáveis estão diretamente relacionadas à **pergunta de negócio formulada na Aula 01**, que busca identificar quais métricas de uso mais influenciam a continuidade do uso do aplicativo após as primeiras 4 semanas.  
A granularidade e a variedade das informações permitem analisar padrões de comportamento e correlacioná-los com a retenção de usuários, possibilitando a criação de estratégias personalizadas para aumentar o engajamento.

**4. Dicionário de Dados Preliminar**

| **Coluna** | **Descrição** | **Tipo de Dado** |
| --- | --- | --- |
| Id | Identificador único do usuário | Numérico (Inteiro) |
| ActivityDate | Data do registro da atividade | Data |
| TotalSteps | Total de passos dados no dia | Numérico (Inteiro) |
| TotalDistance | Distância total percorrida (milhas) | Numérico (Decimal) |
| TrackerDistance | Distância registrada pelo dispositivo (milhas) | Numérico (Decimal) |
| LoggedActivitiesDistance | Distância registrada manualmente (milhas) | Numérico (Decimal) |
| VeryActiveMinutes | Minutos em atividades de alta intensidade | Numérico (Inteiro) |
| FairlyActiveMinutes | Minutos em atividades moderadas | Numérico (Inteiro) |
| LightlyActiveMinutes | Minutos em atividades leves | Numérico (Inteiro) |
| SedentaryMinutes | Minutos em comportamento sedentário | Numérico (Inteiro) |
| Calories | Total de calorias queimadas no dia | Numérico (Inteiro) |
| TotalMinutesAsleep | Total de minutos dormindo | Numérico (Inteiro) |
| TotalTimeInBed | Tempo total na cama (inclui períodos acordado) | Numérico (Inteiro) |

**5. Desafios Previstos**

* **Valores ausentes:** Algumas colunas podem não ter registros para todos os usuários em todos os dias, especialmente as de sono.
* **Formato de datas:** A coluna *ActivityDate* pode precisar de conversão para o formato de data padrão para facilitar filtragens e análises.
* **Usuários com dados insuficientes:** Pode haver participantes com poucos dias de registro, dificultando a análise de retenção.
* **Unidades diferentes:** Distância em milhas pode precisar ser convertida para quilômetros, dependendo do público-alvo da análise.