

Missão Prática – RPG0033: Tratando a Imensidão dos Dados

Aluno: Josthoon Miguel Alejandro Rivas Jaramillo

Disciplina: RPG0033 – Tratando a Imensidão dos Dados

Tema: Limpeza e tratamento de dados com Pandas (Python)

Ambiente utilizado: VS Code + Python + Pandas

Data: 14/05/25

Objetivo

O objetivo principal desta missão prática foi aplicar conhecimentos básicos da biblioteca **Pandas** em **Python** para tratar um conjunto de dados brutos, corrigir problemas e deixá-los prontos para análises futuras.

Ferramentas e Tecnologias

- Python 3.11
 - Pandas
 - Visual Studio Code (VS Code)
 - Sistema de arquivos local (Windows)
-

Fonte dos Dados

Foi utilizado um arquivo chamado `dados.csv`, contendo registros de monitoramento de exercícios físicos. O arquivo possui as seguintes colunas:

- ID
- Duration
- Date
- Pulse

- Maxpulse
- Calories

Etapas Realizadas

1. Leitura do CSV

O arquivo foi lido usando `pd.read_csv()`, com o separador `;`.

2. Exibição de dados

Foram exibidas as primeiras e últimas linhas e informações gerais do DataFrame.

3. Cópia do DataFrame original

Foi criada uma cópia de segurança dos dados com `df.copy()`.

4. Tratamento de valores ausentes

- Valores nulos da coluna `Calories` foram substituídos por `0`.
- Valores nulos da coluna `Date` foram temporariamente substituídos por `'1900/01/01'`, mas posteriormente corrigidos para `NaN`.

5. Conversão de datas

- A coluna `Date` foi convertida para o tipo `datetime`, após tratar valores mal formatados como `20201226`.

6. Remoção de dados inválidos

Foi removida a linha com data ausente definitiva (`NaN`).

Resultados Obtidos

Ao final do processo, obteve-se um DataFrame totalmente limpo, sem valores nulos, com datas padronizadas no formato `datetime`, pronto para ser utilizado em tarefas de análise de dados ou visualização.

Conclusão

A atividade permitiu aplicar na prática conceitos essenciais de tratamento de dados, como:

- Leitura de arquivos CSV;
- Manipulação de valores ausentes;
- Conversão de tipos;
- Limpeza de dados inconsistentes.

O uso do **Pandas** se mostrou eficiente para lidar com dados em larga escala e preparar conjuntos para etapas posteriores de análise.