



Campus: Polo Duque de Caxias - RJ

Curso: Desenvolvimento full stack

Disciplina: RPG0033 - TRATANDO A IMENSIDÃO DOS DADOS

Semestre letivo: Mundo 5

Aluno: Lucas Pietro Santos de Souza

TRATANDO A IMENSIDÃO DOS DADOS (pico web)

Introdução

Neste projeto, embarcamos na jornada de preparar um conjunto de dados para análise, utilizando Python e a poderosa biblioteca Pandas. Nosso objetivo foi limpar os dados, tornando-os aptos para serem explorados e interpretados com confiança. O conjunto de dados inicial é composto por colunas importantes: ID, Duration, Date, Pulse, Maxpulse e Calories.

Componentes do Projeto

- Importação de Bibliotecas

Começamos importando o Pandas, nossa ferramenta principal para manipulação e análise de dados

- Função para Conversão de Datas

Criamos uma função personalizada, `try_convert_date`, para lidar com a variedade de formatos de data presentes nos nossos dados

- Leitura do Arquivo CSV

Carregamos os dados de um arquivo CSV, ajustando os parâmetros conforme necessário

- Verificação Inicial dos Dados

Para entender melhor os dados que recebemos, exibimos informações gerais e visualizamos algumas amostras

- Tratamento de Dados

Fazemos uma cópia do DataFrame original para preservar os dados brutos e trabalhar com segurança

- Remoção de Registros Nulos

Finalizamos removendo quaisquer registros que ainda possuam valores nulos, assegurando a integridade dos dados finais

Conclusão

Este projeto não apenas demonstrou técnicas fundamentais de manipulação de dados com Pandas, mas também ilustrou a importância de garantir a qualidade dos dados para análises subsequentes. Ao seguir este roteiro, adquirimos habilidades essenciais para preparar, limpar e validar conjuntos de dados complexos, preparando-os para serem utilizados eficazmente em projetos de análise e descoberta de conhecimento.

Essa abordagem prática não só fortalece nosso entendimento de Python e Pandas, mas também nos capacita a enfrentar desafios reais em ambientes de análise de dados.