



Centro Universitário Estácio

Curso: Desenvolvimento Full Stack

Disciplina: Tratando a Imensidão Dos Dados

Polo: São Lourenço da Mata

Aluno: Jonnatha Walben Saldanha da Silva

Relatório da Atividade

O objetivo da missão prática foi capacitar o aluno a realizar o **tratamento de dados brutos**, preparando-os para tarefas de **mineração de dados e análise estatística**, utilizando a linguagem Python e a biblioteca Pandas. O foco esteve na **limpeza, padronização e transformação** de um conjunto de dados que continha valores ausentes e formatações inconsistentes.

Conjunto de Dados Utilizado

O dataset fornecido continha informações sobre sessões de exercícios físicos, com as seguintes colunas:

- ID: Identificador do registro
- Duration: Duração do exercício (minutos)
- Date: Data da atividade
- Pulse: Frequência cardíaca média
- Maxpulse: Frequência cardíaca máxima
- Calories: Calorias gastas

Ferramentas Utilizadas

- Linguagem: **Python**
- Biblioteca: **Pandas**
- Ambiente: **Google Colab** (ou Jupyter Notebook)

Etapas Realizadas

1. Leitura do Arquivo

Foi feita a leitura do arquivo CSV com o separador ; (ponto e vírgula), utilizando o comando `pd.read_csv()`. Isso permitiu a correta separação das colunas no DataFrame.

2. Verificação Inicial dos Dados

Utilizou-se os métodos `df.info()`, `df.head()` e `df.tail()` para verificar:

- Quantidade de registros
- Tipos de dados
- Presença de valores nulos ou duplicados
- Visualização de amostras dos dados

3. Cópia de Segurança

Criou-se uma **cópia do DataFrame original** para preservar os dados crus, garantindo a possibilidade de comparação e reprocessamento, se necessário.

4. Tratamento da Coluna Calories

- Substituição de valores nulos por 0 com `fillna(0)`
- Conversão para tipo numérico

5. Tratamento da Coluna Date

- Substituição temporária de valores nulos por '1900/01/01'
- Tentativa de conversão para o tipo `datetime`, o que resultou em erro para datas mal formatadas
- Identificação de data incorreta '20201226' → substituída por '2020/12/26'
- Substituição do placeholder '1900/01/01' por NaN
- Conversão completa da coluna para o formato `datetime` usando `pd.to_datetime()`

6. Remoção de Registros com Valores Nulos

- Após a conversão da coluna Date, removeu-se o registro com valor nulo restante (linha 22) utilizando dropna().

Resultados Obtidos

Ao final do processo, obteve-se um DataFrame:

- Sem valores nulos
- Com as datas padronizadas no formato YYYY-MM-DD
- Com os dados prontos para análise, visualização ou modelagem

Aprendizados e Competências Desenvolvidas

- Leitura e interpretação de arquivos CSV
- Identificação e tratamento de dados nulos
- Conversão de tipos de dados (string para datetime e numéricos)
- Manipulação de dados com Pandas
- Criação de cópias seguras de dados
- Limpeza e padronização de dados