

Missão Prática - RPG0033

Tratando a Imensidão dos Dados

Aluno: Felipe de Souza Komatsu

Matrícula: 202302883443

1. Introdução

Este documento descreve o passo a passo da execução da missão prática da disciplina RPG0033, seguindo todas as orientações do roteiro fornecido. O objetivo foi realizar o tratamento de dados em um arquivo CSV utilizando Python e a biblioteca Pandas, preparando o conjunto de dados para futuras análises e mineração de conhecimento.

2. Etapas Realizadas

1. Leitura do arquivo 'dados.csv' com separador ';', utilizando a engine 'python' e encoding 'utf-8'.
 2. Exibição das informações gerais do conjunto de dados (info), primeiras e últimas linhas (head, tail).
 3. Criação de uma cópia do DataFrame original para tratamento.
 4. Substituição de valores nulos da coluna 'Calories' por 0.
 5. Substituição de valores nulos na coluna 'Date' por '1900/01/01'.
 6. Tentativa de conversão da coluna 'Date' para datetime (gera erro esperado).
 7. Substituição de '1900/01/01' por valor nulo (NaT).
 8. Nova tentativa de conversão para datetime (erro ainda ocorre por conta do valor 20201226).
 9. Correção do valor 20201226 na coluna 'Date' para '2020/12/26'.
 10. Conversão completa da coluna 'Date' para datetime, com tratamento de erros.
 11. Remoção de todas as linhas com valores nulos (incluindo a linha 22).
 12. Exibição final dos dados tratados (DataFrame e info).
 13. Exportação do DataFrame final para 'dados_tratados.csv'.
-

3. Conclusão

Com a execução das etapas descritas acima, foi possível realizar com sucesso a limpeza e o preparo do conjunto de dados, garantindo que ele esteja pronto para aplicações futuras de análise ou mineração de dados. O processo utilizou práticas comuns em ciência de dados e demonstrou domínio da biblioteca Pandas.