

Universidade Estácio de Sá

Aluno: Luan Augusto Vieira Bandeira

Nível 3: Tratando a imensidão dos dados

```
#Importar as bibliotecas necessárias:
import pandas as pd

#Ler o arquivo CSV fornecido:
df = pd.read_csv('dados.csv', sep=';', engine='python', encoding='latin1')

#Exibir informações gerais sobre o conjunto de dados:
print('Exibir informações gerais sobre o conjunto de dados')
df.info()

#Exibir as primeiras e últimas linhas do arquivo:
print('\nExibir as primeiras e últimas linhas do arquivo ')
print(df.head())
print(df.tail())

#Criar uma nova variável contendo uma cópia dos dados originais
print('\nCriar uma nova variável contendo uma cópia dos dados originais')
df_copy = df.copy()

#Substituir valores nulos da coluna 'Calories' por 0:
print('\nSubstituir valores nulos da coluna 'Calories' por 0')
df_copy['Calories'].fillna(0, inplace=True)
print(df_copy)

#Substituir valores nulos da coluna 'Date' por '1900/01/01'
print('\nSubstituir valores nulos da coluna 'Date' por '1900/01/01' e
transformá-la em datetime')
df_copy['Date'].fillna('1900/01/01', inplace=True)
df_copy['Date'] = pd.to_datetime(df_copy['Date'], format='%Y/%m/%d',
errors='coerce')
print(df_copy)

#Corrigir valores de data incorretos
df_copy['Date'] = df_copy['Date'].replace('20201226', '2020-12-26')
df_copy['Date'] = pd.to_datetime(df_copy['Date'], format='%Y/%m/%d',
errors='coerce')
```

```
print(df_copy)

#Remover registros que ainda contenham valores nulos:
print('\nRemover registros que ainda contenham valores nulos')
df_clean = df_copy.dropna()
print(df_clean)

#Verificar todas as transformações e finalizar
print('\nVerificar todas as transformações e finalizar')
print(df_clean.info())
```