Logotipo, nome da empresa

Descrição gerada automaticamente

**Campus:** Rua Manoel João Gonçalves, 410/412 – Alcântara CEP: 24711-080

**Curso:** Desenvolvimento Full-Stack

**Disciplina:**  RPG0033 - TRATANDO A IMENSIDÃO DOS DADOS

**Turma:** 9001

**Semestre letivo**: 2024.4 FLEX

**Integrante:**

Nome:Samir Campos LimaMatrícula: 2022.11.47141-1

**Link do repositório no GIT:** [**samircamposlima/Miss-o-Pr-tica-N-vel-3-Mundo-5**](https://github.com/samircamposlima/Miss-o-Pr-tica-N-vel-3-Mundo-5)

**Micro atividade 1: Descrever como ler um arquivo**

**CSV usando a biblioteca Pandas (Python)**

**- Procedimentos**

1. Salve o conjunto de dados em formato CSV que utilizará num local acessível pela ferramenta de escrita de código que utilizará;
   1. Salve o conjunto de dados em formato CSV que utilizará num local acessível pela ferramenta de escrita de código que utilizará;
2. Crie um novo arquivo e:
   1. Importe a biblioteca pandas;
   2. Cria uma variável;
   3. Leia o conteúdo do arquivo CSV, passando como parâmetros o separador de

colunas, a engine – com o valor ‘python’ e o enconding relativo aos dados

constantes no arquivo lido (esse último parâmetro pode ser opcional,

* 1. Leia o conteúdo do arquivo CSV, passando como parâmetros o separador de
  2. colunas, a engine – com o valor ‘python’ e o enconding relativo aos dados
  3. constantes no arquivo lido (esse último parâmetro pode ser opcional,

dependendo do enconding existente);

* 1. Atribua os dados lidos do CSV à variável criada anteriormente; Salve as

alterações;

* 1. Imprima/exiba em tela os dados da variável.

Código:

import pandas as pd

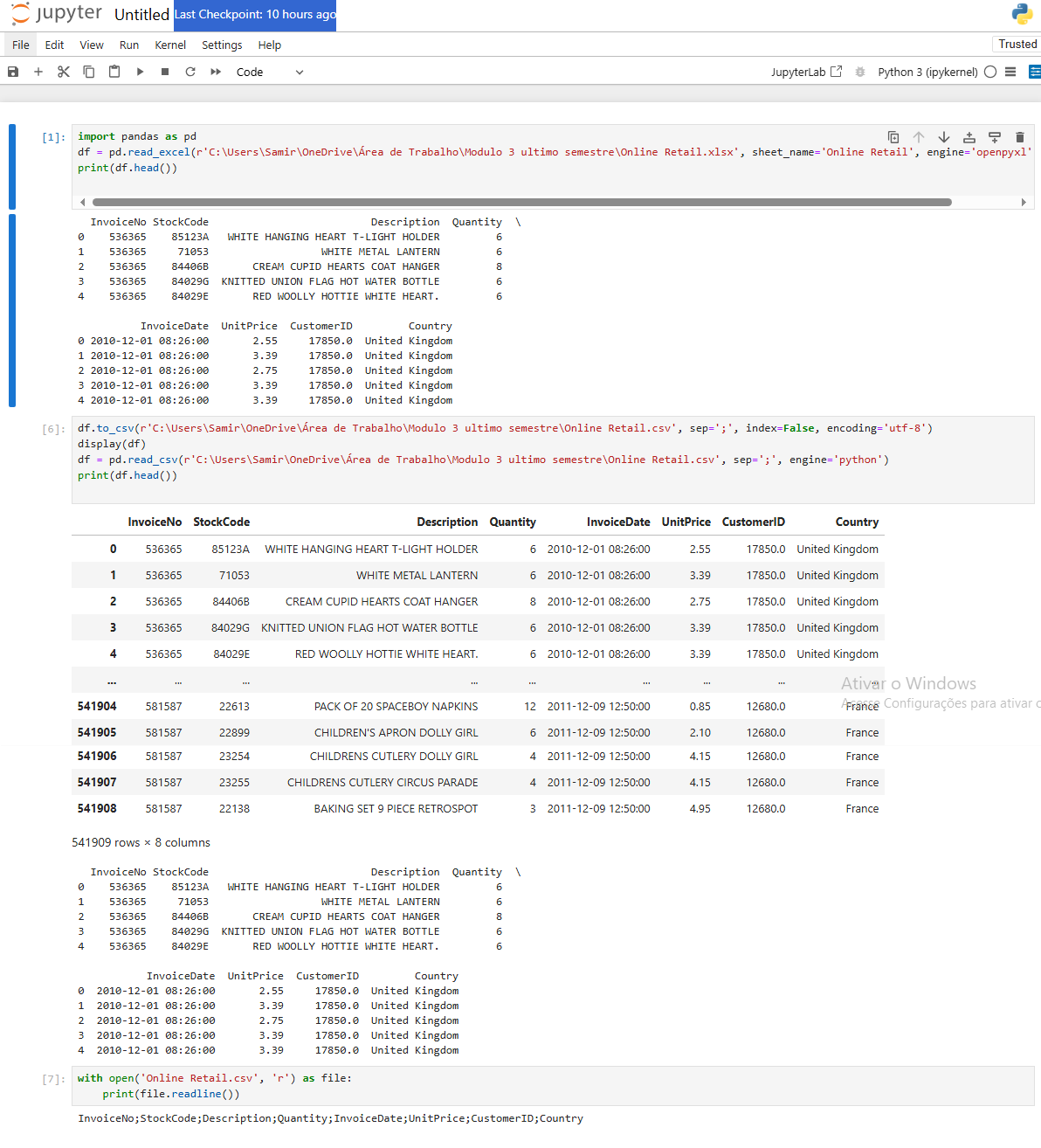
df = pd.read\_excel(r'C:\Users\Samir\OneDrive\Área de Trabalho\Modulo 3 ultimo semestre\Online Retail.xlsx', sheet\_name='Online Retail', engine='openpyxl', header=0)  
print(df.head())

df.to\_csv(r'C:\Users\Samir\OneDrive\Área de Trabalho\Modulo 3 ultimo semestre\Online Retail.csv', sep=';', index=False, encoding='utf-8')  
display(df)

df = pd.read\_csv(r'C:\Users\Samir\OneDrive\Área de Trabalho\Modulo 3 ultimo semestre\Online Retail.csv', sep=';', engine='python')  
print(df.head())

with open('Online Retail.csv', 'r') as file:  
 print(file.readline())

imagem:

****

* 1. Atribua os dados lidos do CSV à variável criada anteriormente; Salve as

**Micro atividade 2: Descrever como criar um**

**subconjunto de dados a partir de um conjunto**

**existente usando a biblioteca Pandas (Python)**

**- Procedimentos:**

1. No mesmo arquivo/script utilizado na microatividade 1, crie uma nova variável;
2. Atribua, a essa nova variável, um subconjunto de dados contendo apenas parte

das colunas (recomenda-se a utilização de 3 colunas) disponíveis no conjunto de

dados original;

1. Salve as alterações realizadas;
2. Imprima/exiba em tela os dados da nova variável (que contém o subconjunto de

dados).

1. Atribua, a essa nova variável, um subconjunto de dados contendo apenas parte
2. das colunas (recomenda-se a utilização de 3 colunas) disponíveis no conjunto de

Código:

arquivo\_csv = 'Online Retail.csv'   
colunas\_selecionadas = ['InvoiceNo', 'Description', 'UnitPrice']  
df\_selecionado = df[colunas\_selecionadas]  
  
print(df\_selecionado.head())  
display(df\_selecionado)

imagem:

**Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente**

**Micro atividade 3: Descrever como configurar o**

**número máximo de linhas a serem exibidas na**

**visualização de um conjunto de dados usando a**

**biblioteca Pandas (Python)**

**- Procedimentos**

1. Abra o arquivo/script utilizado nas microatividades anteriores;
2. Usando as opções de configuração da biblioteca pandas, defina um novo valor

para a propriedade “max\_rows”, definindo o novo valor para 9999;

1. Salve as alterações;
2. Imprima na tela o conjunto de dados original (criado na microatividade 1) usando

o método “to\_string()”.

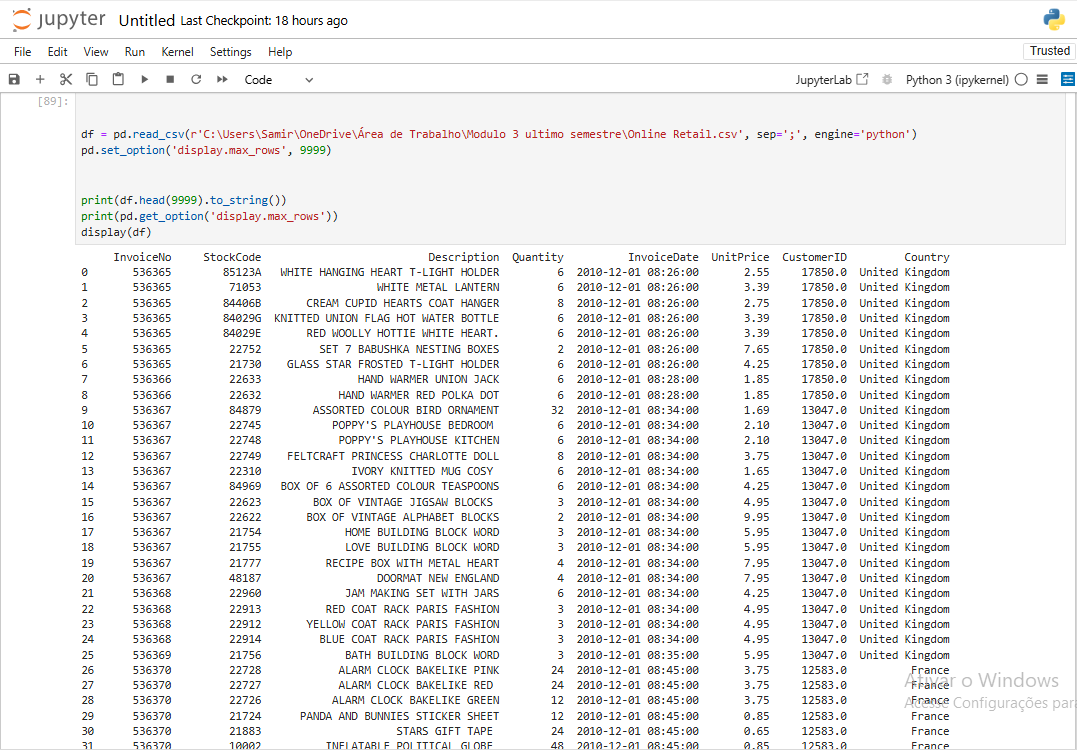
Código:

df = pd.read\_csv(r'C:\Users\Samir\OneDrive\Área de Trabalho\Modulo 3 ultimo semestre\Online Retail.csv', sep=';', engine='python')

pd.set\_option('display.max\_rows', 9999)

print(df.head(9999).to\_string())  
print(pd.get\_option('display.max\_rows'))  
display(df)

imagem:

****

**Interface gráfica do usuário, Texto, Aplicativo, Tabela

Descrição gerada automaticamente**

**Micro atividade 4: Descrever como exibir as**

**primeiras e últimas “N” linhas de um conjunto de**

**dados usando a biblioteca Pandas (Python)**

**- Procedimentos**

1. Abra o arquivo/script utilizado nas microatividades anteriores;
2. Imprima na tela as apenas as primeiras 10 linhas do conjunto de dados original

(criado na microatividade 1);

1. Imprima na tela as apenas as últimas 10 linhas do conjunto de dados original

(criado na microatividade 1).

Código:

print(df.head(10))   
  
  
print(df.tail(10))

imagem:

**Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente**

**Micro atividade 5: Descrever como exibir**

**informações gerais sobre as colunas, linhas e**

**dados de um conjunto de dados usando a**

**biblioteca Pandas (Python)**

**- Procedimentos**

1. Abra o arquivo/script utilizado nas microatividades anteriores;
2. Tendo como base o conjunto de dados original:
   1. Imprima as informações gerais sobre o conjunto – suas colunas, linhas e dados;
   2. Descubra a partir do comando acima:
      1. O total de linhas;
      2. O total de colunas;
      3. A quantidade de dados nulos, caso existam;
      4. O tipo de dado de cada coluna;
      5. A quantidade de memória utilizada pelo conjunto de dados.

Código:

print("Informações gerais:")  
print(df.info())  
  
  
print(f"\nQuantidade de linhas: {df.shape[0]}")  
  
  
print(f"Quantidade de colunas: {df.shape[1]}")  
  
  
print("\nQuantidade de dados nulos por coluna:")  
print(df.isnull().sum())  
  
print("\nTipos de dados de cada coluna:")  
print(df.dtypes)  
  
print(f"\nQuantidade de memória utilizada: {df.memory\_usage(deep=**True**).sum() / 1024:.2f} KB")

imagem:

**Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente**

**Missão Prática | Tratando a imensidão dos**

**dados 💻**

**- Procedimentos**

1. Para essa atividade você deverá, obrigatoriamente, utilizar o conjunto de dados

(fornecido anteriormente, na seção “Contextualização”) composto pelas colunas

ID;Duration;Date;Pulse;Maxpulse;Calories

\***OBS: Utilizei a planilha do link que não me direcionou para planilha correta\***

1. Crie um novo arquivo/script;
2. Leia o conteúdo do CSV fornecido, atentando-se para a necessidade ou não de

incluir parâmetros adicionais como os relativos ao separador dos dados, a engine e o enconding;

1. Atribua os dados lidos a uma variável;

Código:

**import** pandas **as** pd  
  
df = pd.read\_csv(r'C:\Users\Samir\OneDrive\Área de Trabalho\Modulo 3 ultimo semestre\Online Retail1.csv', sep=';', engine='python')  
display(df)

imagem:

Interface gráfica do usuário, Texto, Aplicativo, Email

Descrição gerada automaticamente

1. Verifique se os dados foram importados adequadamente:
2. Imprima as informações gerais sobre o conjunto de dados;
3. Imprima as primeiras e últimas N linhas do arquivo.

Código:

print("Informações gerais sobre o conjunto de dados:")  
print(df.info())   
  
print("\nQuantidade de dados nulos por coluna:")  
print(df.isnull().sum())  
  
print("\nPrimeiras 5 linhas do arquivo:")  
print(df.head())   
  
print("\nÚltimas 5 linhas do arquivo:")  
print(df.tail())

imagem:

Tabela

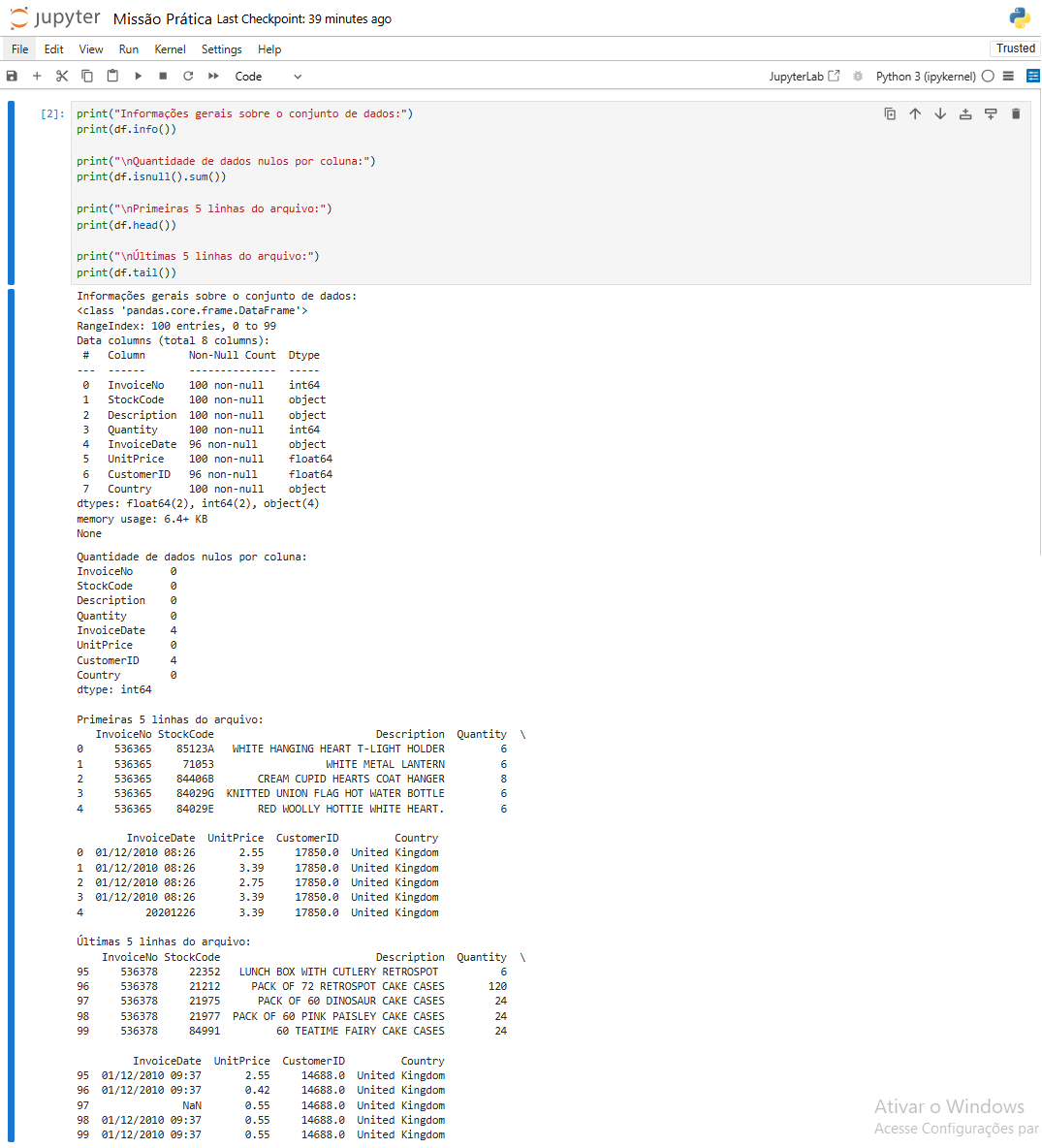
Descrição gerada automaticamente com confiança baixa

1. Verifique se os dados foram importados adequadamente:
2. Imprima as informações gerais sobre o conjunto de dados;
3. Imprima as primeiras e últimas N linhas do arquivo.

Código:

print("Informações gerais sobre o conjunto de dados:")  
print(df.info())   
  
print("\nQuantidade de dados nulos por coluna:")  
print(df.isnull().sum())  
  
print("\nPrimeiras 5 linhas do arquivo:")  
print(df.head())   
  
print("\nÚltimas 5 linhas do arquivo:")  
print(df.tail())

imagem:

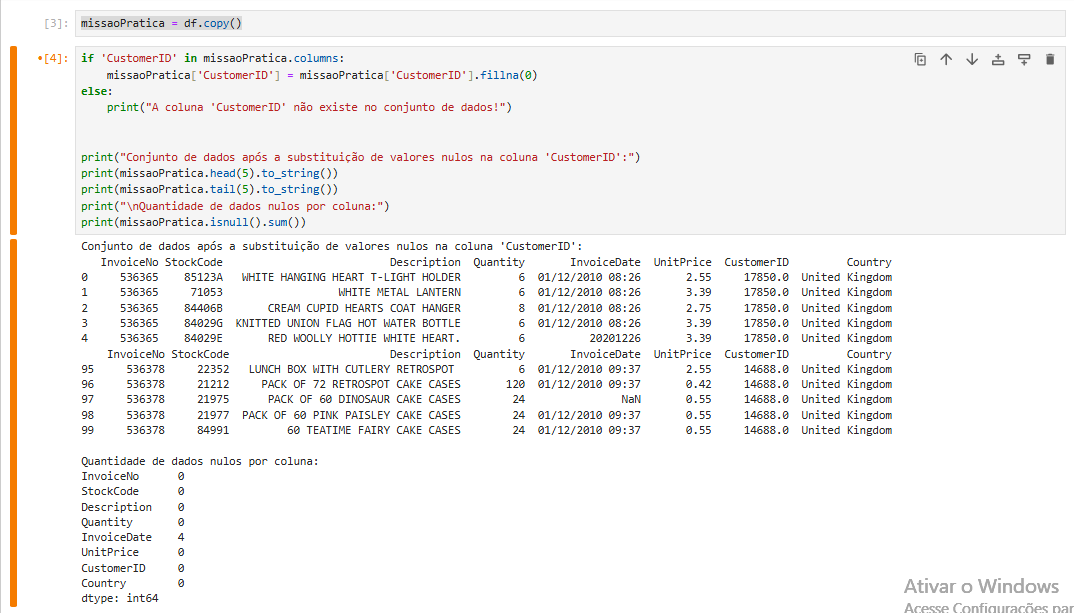


1. Crie uma nova variável e atribua a ela uma cópia do conjunto de dados original (variável criada no passo 4);
2. Nessa nova variável, contendo uma cópia dos dados:
3. Substitua todos os valores nulos da coluna ‘CustomerID’ por 0;
4. Imprima o conjunto de dados para verificar se a mudança acima foi aplicada com sucesso;

Código:

missaoPratica = df.copy()  
  
**if** 'CustomerID' **in** missaoPratica.columns:  
 missaoPratica['CustomerID'] = missaoPratica['CustomerID'].fillna(0)  
**else**:  
 print("A coluna 'CustomerID' não existe no conjunto de dados!")  
  
  
print("Conjunto de dados após a substituição de valores nulos na coluna 'CustomerID':")  
print(missaoPratica.head(5).to\_string())  
print(missaoPratica.tail(5).to\_string())  
print("\nQuantidade de dados nulos por coluna:")  
print(missaoPratica.isnull().sum())

imagem:



1. Transforme os dados da coluna ‘Date’ em datetime usando o método ‘to\_datetime’;

Código:

missaoPratica['InvoiceDate'] = pd.to\_datetime(missaoPratica['InvoiceDate'], format='%d/%m/%Y %H:%M')

imagem:

Texto

Descrição gerada automaticamente

1. Nesse ponto, você deverá ter esbarrado em um erro, informando agora que o valor "20201226" não corresponde ao formato "'%Y/%m/%d'" . Você precisará, agora, na coluna ‘Date”, transformar especificamente esse valor, atualmente uma string, para o formato datetime. Para isso você deverá combinar os métodos ‘replace’ e ‘to\_datetime’;

Código:

**def** **adicionar\_hora\_se\_necessario**(data):  
 **if** isinstance(data, str) **and** len(data) == 8: # Verifica se é uma string com 8 caracteres  
 **return** data + ' 00:00'  
 **else**:  
 **return** data  
  
missaoPratica['InvoiceDate'] = missaoPratica['InvoiceDate'].apply(adicionar\_hora\_se\_necessario)  
missaoPratica['InvoiceDate'] = missaoPratica['InvoiceDate'].fillna('120201226')  
missaoPratica['InvoiceDate'] = missaoPratica['InvoiceDate'].replace('20201226 00:00', '26/12/2020 00:00')  
  
  
  
print(missaoPratica)

imagem:

Interface gráfica do usuário, Texto

Descrição gerada automaticamente

1. Ainda na nova variável:
2. Substitua os valores nulos da coluna ‘Date’ por ‘1900/01/01’;
3. Imprima o conjunto de dados e confira se a mudança foi aplicada com sucesso;

Código:

missaoPratica['InvoiceDate'] = missaoPratica['InvoiceDate'].fillna('1900/01/01 00:00')  
print(missaoPratica)  
print(missaoPratica.isnull().sum())

imagem:

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

1. Transforme os dados da coluna ‘Date’ em datetime usando o método ‘to\_datetime’;
2. Tendo seguido todas as instruções anteriores, ao executar o passo anterior você deverá ter encontrado outro erro informando que o valor ‘1900/01/01’ não corresponde ao formato ‘%Y/%m/%d’. Para resolver esse problema:

Código:

missaoPratica['InvoiceDate'] = pd.to\_datetime(missaoPratica['InvoiceDate'], format='%d/%m/%Y %H:%M')

imagem:

Texto

Descrição gerada automaticamente

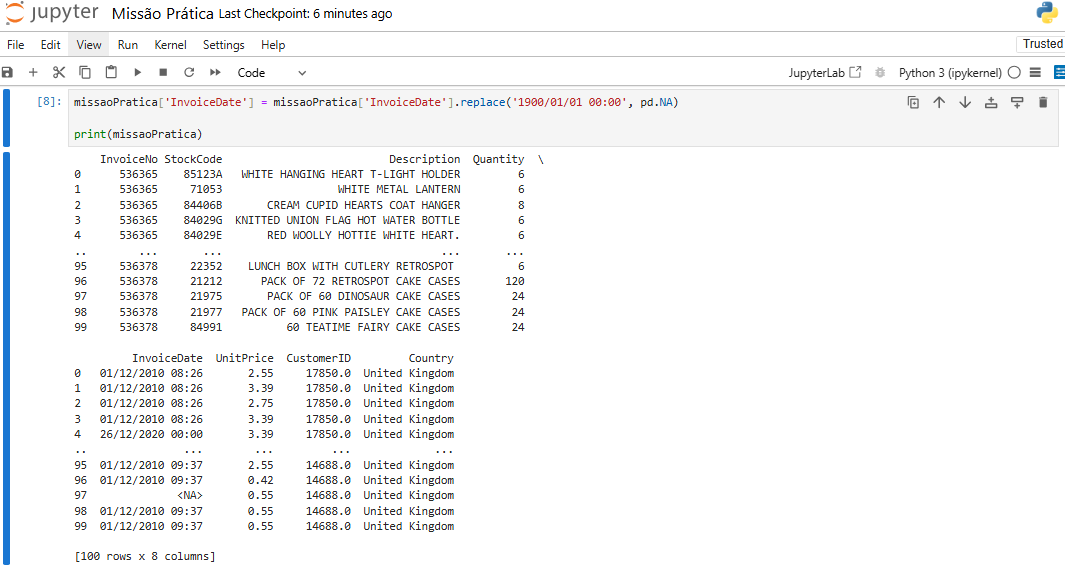
1. Substitua, na coluna ‘Date’, o valor ‘1900/01/01’ por ‘NaN’;
2. Utilizando o método ‘to\_datetime’, repita o passo de transformação dos dados da coluna ‘Date’ para datetime;
3. Imprima o conjunto de dados para verificar se as mudanças acima foram

aplicadas com sucesso;

Código:

missaoPratica['InvoiceDate'] = missaoPratica['InvoiceDate'].replace('1900/01/01 00:00', pd.NA)  
  
print(missaoPratica)

imagem:



1. Após o passo anterior, execute novamente a transformação de todos os dados dacoluna ‘Date’ para o formato datetime (usando o to\_datetime). Imprima o conjunto de dados atual para verificar se todas as transformações foram executadas com sucesso;

Código:

missaoPratica['InvoiceDate'] = pd.to\_datetime(missaoPratica['InvoiceDate'], format='%d/%m/%Y %H:%M')  
print(missaoPratica)

imagem:

Interface gráfica do usuário, Texto, Aplicativo

Descrição gerada automaticamente

1. Por fim, remova os registros contendo valores nulos. Nesse ponto, apenas a coluna ‘Date’ possui um registro que atende a essa premissa (em 4 linha). Logo,

utilize-a como base para realizar a transformação solicitada;

Código:

missaoPratica = missaoPratica.dropna(subset=['InvoiceDate'])  
print(missaoPratica)

imagem:

Interface gráfica do usuário, Texto

Descrição gerada automaticamente