

Documentação do Script Python para tratamento de dados

Script Python utilizado para o tratamento dos dados:

```
import pandas as pd

# Carregar a base de dados (substitua 'caminho/do/arquivo.csv' pelo seu caminho)
df = pd.read_excel('/content/Qlik Sense - COMPRADORES (SEMESTRE) - 10 de junho de 2024.xlsx', skiprows = 2)

#Removendo linhas duplicadas
df.drop_duplicates(subset=['CÓDIGO'], inplace=True)

# Removendo a coluna 'coluna_a_remove'
df.drop(columns=['Unnamed: 10'], inplace=True)
df.drop(columns=['jun'], inplace=True)

# Renomeando colunas do dataframe
df = df.rename(columns={'FUINACIONÁRIO': 'FUNCIONÁRIO'})
df = df.rename(columns={'RAZAOSOCIAL': 'RAZAO_SOCIAL'})
df = df.rename(columns={'NOMEFANTASIA': 'NOME_FANTASIA'})
df = df.rename(columns={'mai': 'fat_mai'})
df = df.rename(columns={'abr': 'fat_abr'})
df = df.rename(columns={'mar': 'fat_mar'})
df = df.rename(columns={'fev': 'fat_fev'})
df = df.rename(columns={'jan': 'fat_jan'})
df = df.rename(columns={'dez': 'fat_dez'})

# Preencher valores faltantes (se houver) com valores padrão, como 'N/A'
```

```
df.fillna('N/A', inplace=True)
```

```
# Converter colunas para tipos de dados adequados, se necessário
```

```
Exemplo: df['CÓDIGO'] = df['CÓDIGO'].astype(int)
```

```
#Substituição de valores no dataframe
```

```
df['SEGMENTO'] = df['SEGMENTO'].replace('-', 'Não Informado')
```

```
df['SEGMENTO'] = df['SEGMENTO'].replace('# NAO UTILIZAR #', 'Não Informado')
```

```
df['BAIRRO'] = df['BAIRRO'].replace('10', 'Não Informado')
```

```
df['BAIRRO'] = df['BAIRRO'].replace('*****', 'Não Informado')
```

```
df['BAIRRO'] = df['BAIRRO'].replace('***', 'Não Informado')
```

```
df['BAIRRO'] = df['BAIRRO'].replace(' ', 'Não Informado')
```

```
#Substitui '-' por '0' em colunas numéricas
```

```
df['fat_mai'] = df['fat_mai'].replace('-', '0')
```

```
df['fat_abr'] = df['fat_abr'].replace('-', '0')
```

```
df['fat_mar'] = df['fat_mar'].replace('-', '0')
```

```
df['fat_fev'] = df['fat_fev'].replace('-', '0')
```

```
df['fat_jan'] = df['fat_jan'].replace('-', '0')
```

```
df['fat_dez'] = df['fat_dez'].replace('-', '0')
```

```
# Salvar a base de dados limpa e corrigida (substitua  
'caminho/do/novo_arquivo.csv' pelo seu caminho)
```

```
df.to_excel('/content/Compadores.xlsx', index=False)
```

O script utiliza a biblioteca pandas para manipulação e limpeza de dados em um arquivo Excel. Primeiramente, importa a biblioteca pandas com o alias 'pd' para facilitar a referência a funções da biblioteca. Em seguida, carrega o arquivo Excel especificado, pulando as duas primeiras linhas (skiprows=2), e armazena os dados em um DataFrame chamado 'df'.

No DataFrame 'df', são realizadas as seguintes operações:

1. Remoção de linhas de clientes duplicados que dificulta a contagem real do número total dos clientes com base na coluna 'CÓDIGO'.
2. Remoção das colunas 'Unnamed: 10' e 'jun', que não são necessárias, já que o relatório é no período de dezembro a maio.
3. Renomeação das colunas conforme especificado no script, corrigindo possíveis erros de digitação ou nomes inadequados.
4. Preenchimento de valores faltantes (NaN) com o valor 'N/A' para manter a consistência nos dados.
5. Substituição de valores específicos nas colunas 'SEGMENTO' e 'BAIRRO' por 'Não Informado' para padronização e melhor interpretação dos dados.
6. Substituição do caractere '-' por '0' em colunas numéricas relacionadas às métricas de faturamento para garantir a coerência nos cálculos.

Por fim, o DataFrame 'df' limpo e corrigido é salvo em um novo arquivo Excel chamado 'Compadores.xlsx' no caminho especificado, permitindo o uso dos dados tratados em outras análises ou aplicações.

Este script é útil para pré-processamento e correção de dados em arquivos Excel antes de sua análise ou utilização em outros contextos, garantindo a qualidade e integridade dos dados manipulados.