

```
In [306... import pandas as pd
import numpy as np
```

```
In [319... #Leitura das planilhas de excel

vendas = pd.read_excel('Vendas.xlsx', skiprows=5)
cliente = pd.read_excel('Cliente.xlsx')
file_path = 'Metas.xlsx'

#Leitura de arquivos csv

local = pd.read_csv('Localizacao.csv', delimiter=';', skiprows=2)
produtos = pd.read_csv('Produto.csv', delimiter=';', skiprows=2)

#Leitura de arquivos JSON

file_path_json = 'Subcategoria.json'
```

```
In [323... vendas.to_excel('VENDAS[TRATADA].xlsx', index=False)
```

```
In [321... # TRANSFORM TABLE[SUBCATEGORIA]

sub_categoria = pd.read_json(file_path_json)

sub_categoria.to_excel('SUBCATEGORIA[TRATADA].xlsx', index=False)
```

```
In [310... # TRANSFORM TABLE[CLIENTE]

cliente['Data Nascimento'] = pd.to_datetime(cliente['Data Nascimento'], errors='coerce')
cliente['Data Nascimento'].fillna(pd.to_datetime('1800-01-01'), inplace=True)
cliente.loc[:, cliente.columns != 'Data Nascimento'] = cliente.loc[:, cliente.columns != 'Data Nascimento'].fillna('Não Informado') #trata nulo de todas as colunas,

cliente.to_excel('CLIENTE[TRATADA].xlsx', index=False)
```

```
In [322... # TRANSFORM TABLE[METAS]

dfs = []

sheets = pd.ExcelFile(file_path).sheet_names

for sheet in sheets:
    metas = pd.read_excel(file_path, sheet_name=sheet, header=None)

    metas = metas[~metas[0].str.contains('Total', na=False)]
    metas = metas.iloc[1:]

    metas.columns = ['Categoria'] + list(metas.iloc[0, 1:])
    metas = metas[1:]

    for continent in metas.columns[1:]:
        continent_data = metas[['Categoria', continent]].copy()
```

```

continent_data['Ano'] = sheet
continent_data['Continente'] = continent
continent_data = continent_data.rename(columns={continent: 'Meta'})
dfs.append(continent_data[['Ano', 'Continente', 'Categoria', 'Meta']]) # Adiciona o DataFrame à lista

metas = pd.concat(dfs, ignore_index=True)

metas.to_excel('METAS[TRATADA].xlsx', index=False)

```

```

In [ ]: # TRANSFORM TABLE[LOCALIZACAO]

linha_continente = local[local['ID Localização'].str.contains('Continente:', na=False)]

indices = linha_continente.index.tolist()

local['Continente'] = np.nan

for i in indices:
    local.at[i, 'Continente'] = local.at[i, 'Tipo Localização']
    local.iloc[i]

local['Continente'] = local['Continente'].ffill()

local = local[~(local['ID Localização'].str.contains('Continente:', na=False) | local['ID Localização'].isna() | (local['ID Localização'] == ''))]
local['Cidade'].fillna('Não Informado', inplace=True)
local['Estado'].fillna('Não Informado', inplace=True)

local.to_excel('LOCALIZACAO[TRATADA].xlsx', index=False)

```

```

In [ ]: # TRANSFORM TABLE[PRODUTOS]

linha_marca = produtos[produtos['Produto'].str.contains('Marca:', na=False)]

indices = linha_marca.index.tolist()

produtos['Marca'] = np.nan

for i in indices:
    produtos.at[i, 'Marca'] = produtos.at[i, 'Subcategoria']
    produtos.iloc[i]

produtos['Marca'] = produtos['Marca'].ffill()

produtos = produtos[~(produtos['Produto'].str.contains('Marca:', na=False) | produtos['Produto'].isna() | (produtos['Produto'] == ''))]

produtos['ID do Produto'] = produtos['Produto'].str.split('-').str[0].str.split('#').str[1].str.strip()
produtos['Produtos'] = produtos['Produto'].str.split('-').str[1].str.strip()
produtos['ID Subcategoria'] = produtos['Subcategoria'].str.split('-').str[0].str.split('#').str[1].str.strip()
produtos['Subcategorias'] = produtos['Subcategoria'].str.split('-').str[1].str.strip()

produtos = produtos.drop(columns=['Produto', 'Subcategoria'])
produtos = produtos[['ID do Produto', 'Produtos', 'ID Subcategoria', 'Subcategorias', 'Marca']]

```

```
produtos.to_excel('PRODUTOS[TRATADA].xlsx', index=False)
```

```
In [ ]: %%shell  
jupyter nbconvert --to html /content/OUT.ipynb
```