### **Introdução**

O texto aborda a estrutura e organização dos dados relacionados a transações de vendas. Ele detalha as categorias de produtos e os preços associados a diferentes segmentos, como Varejo, Atacado e Farmácia. Além disso, o texto descreve as cidades e regiões envolvidas nas vendas e fornece informações sobre os vendedores e suas respectivas identidades. O objetivo é gerar e analisar dados de vendas com base em categorias de produtos, regiões e vendedores, para criar um conjunto de dados robusto que pode ser utilizado em análises e relatórios.

### **Estrutura dos Dados**

* **Código**: Código identificador da transação
* **Razão Social**: Nome da empresa ou razão social
* **Categoria**: Tipo de venda
* **Cidade**: Localidade
* **Região**: Região da cidade
* **Ano**: Ano da venda
* **Mês**: Mês da venda
* **Vendedor\_ID**: Identificação do vendedor
* **Preço dos Produtos**: Preços dos produtos em diferentes categorias
* **Quantidade de Produtos**: Quantidade vendida de cada produto
* **Vendas Totais**: Cálculo das vendas totais para cada produto e para o total

### **Dados**

**Categorias**:

* Varejo
* Atacado
* 4 SKU
* 10 SKU
* Farmácia

**Cidades**:

* Plano Piloto
* Gama
* Taguatinga
* Brazlândia
* Sobradinho
* Planaltina
* Paranoá
* Núcleo Bandeirante
* Ceilândia
* Guará
* Cruzeiro
* Samambaia
* Santa Maria
* São Sebastião
* Recanto das Emas
* Lago Sul
* Riacho Fundo
* Lago Norte
* Candangolândia
* Águas Claras
* Riacho Fundo II
* Sudoeste/Octogonal
* Varjão
* Park Way
* SCIA
* Sobradinho II
* Jardim Botânico
* Itapoã
* SIA
* Vicente Pires
* Fercal
* Sol Nascente/Pôr do Sol
* Arniqueira
* Arapoanga
* Água Quente

**Regiões**:

* **Região Central**: Plano Piloto, Sudoeste/Octogonal, Cruzeiro, SCIA, SIA
* **Região Sul**: Gama, Santa Maria, Núcleo Bandeirante, Candangolândia, Park Way, Riacho Fundo, Riacho Fundo II
* **Região Oeste**: Taguatinga, Ceilândia, Samambaia, Recanto das Emas, Águas Claras, Vicente Pires, Sol Nascente/Pôr do Sol, Arniqueira, Brazlândia
* **Região Norte**: Sobradinho, Sobradinho II, Planaltina, Fercal, Lago Norte, Varjão
* **Região Leste**: Paranoá, São Sebastião, Lago Sul, Jardim Botânico, Itapoã, Arapoanga, Água Quente

**Vendedores**:

* **Varejo**: IDs de 101 a 110
* **Atacado**: IDs de 201 a 205
* **4 SKU**: IDs de 301 a 310
* **10 SKU**: IDs de 401 a 405
* **Farmácia**: IDs de 501 a 510

**Preços dos Produtos**:

* **Água**:
  + Varejo: R$ 2,50
  + Atacado: R$ 1,80
  + 4 SKU: R$ 2,40
  + 10 SKU: R$ 2,20
  + Farmácia: R$ 2,50
* **Café**:
  + Varejo: R$ 8,00
  + Atacado: R$ 6,50
  + 4 SKU: R$ 7,70
  + 10 SKU: R$ 7,20
  + Farmácia: R$ 8,50
* **Cerveja**:
  + Varejo: R$ 6,00
  + Atacado: R$ 4,80
  + 4 SKU: R$ 5,70
  + 10 SKU: R$ 5,30
  + Farmácia: R$ 6,00
* **Refrigerante**:
  + Varejo: R$ 4,50
  + Atacado: R$ 3,80
  + 4 SKU: R$ 4,40
  + 10 SKU: R$ 4,10
  + Farmácia: R$ 4,50
* **Suco**:
  + Varejo: R$ 5,00
  + Atacado: R$ 4,20
  + 4 SKU: R$ 4,80
  + 10 SKU: R$ 4,60
  + Farmácia: R$ 5,00
* **Vinho**:
  + Varejo: R$ 35,00
  + Atacado: R$ 28,00
  + 4 SKU: R$ 33,00
  + 10 SKU: R$ 31,00
  + Farmácia: R$ 36,00

### **Código Python**

Esse código cria um DataFrame com 10000 registros de vendas e salva em um arquivo Excel chamado vendas\_dados.xlsx. O código a seguir gera os dados conforme especificado e salva o DataFrame em um arquivo Excel:

import pandas as pd

import numpy as np

# Definir dados

categorias = ['Varejo', 'Atacado', '4 SKU', '10 SKU', 'Farmácia']

cidades = ['Plano Piloto', 'Gama', 'Taguatinga', 'Brazlândia', 'Sobradinho', 'Planaltina', 'Paranoá',

'Núcleo Bandeirante', 'Ceilândia', 'Guará', 'Cruzeiro', 'Samambaia', 'Santa Maria',

'São Sebastião', 'Recanto das Emas', 'Lago Sul', 'Riacho Fundo', 'Lago Norte',

'Candangolândia', 'Águas Claras', 'Riacho Fundo II', 'Sudoeste/Octogonal', 'Varjão',

'Park Way', 'SCIA', 'Sobradinho II', 'Jardim Botânico', 'Itapoã', 'SIA',

'Vicente Pires', 'Fercal', 'Sol Nascente/Pôr do Sol', 'Arniqueira', 'Arapoanga',

'Água Quente']

regioes = {

'Região Central': ['Plano Piloto', 'Sudoeste/Octogonal', 'Cruzeiro', 'SCIA', 'SIA'],

'Região Sul': ['Gama', 'Santa Maria', 'Núcleo Bandeirante', 'Candangolândia', 'Park Way',

'Riacho Fundo', 'Riacho Fundo II'],

'Região Oeste': ['Taguatinga', 'Ceilândia', 'Samambaia', 'Recanto das Emas', 'Águas Claras',

'Vicente Pires', 'Sol Nascente/Pôr do Sol', 'Arniqueira', 'Brazlândia'],

'Região Norte': ['Sobradinho', 'Sobradinho II', 'Planaltina', 'Fercal', 'Lago Norte', 'Varjão'],

'Região Leste': ['Paranoá', 'São Sebastião', 'Lago Sul', 'Jardim Botânico', 'Itapoã',

'Arapoanga', 'Água Quente']

}

vendedores = {

'Varejo': list(range(101, 111)),

'Atacado': list(range(201, 206)),

'4 SKU': list(range(301, 311)),

'10 SKU': list(range(401, 406)),

'Farmácia': list(range(501, 511))

}

precos = {

'Água': {'Varejo': 2.50, 'Atacado': 1.80, '4 SKU': 2.40, '10 SKU': 2.20, 'Farmácia': 2.50},

'Café': {'Varejo': 8.00, 'Atacado': 6.50, '4 SKU': 7.70, '10 SKU': 7.20, 'Farmácia': 8.50},

'Cerveja': {'Varejo': 6.00, 'Atacado': 4.80, '4 SKU': 5.70, '10 SKU': 5.30, 'Farmácia': 6.00},

'Refrigerante': {'Varejo': 4.50, 'Atacado': 3.80, '4 SKU': 4.40, '10 SKU': 4.10, 'Farmácia': 4.50},

'Suco': {'Varejo': 5.00, 'Atacado': 4.20, '4 SKU': 4.80, '10 SKU': 4.60, 'Farmácia': 5.00},

'Vinho': {'Varejo': 35.00, 'Atacado': 28.00, '4 SKU': 33.00, '10 SKU': 31.00, 'Farmácia': 36.00}

}

anos = [2021, 2022, 2023]

meses = ['Janeiro', 'Fevereiro', 'Março', 'Abril', 'Maio', 'Junho', 'Julho', 'Agosto',

'Setembro', 'Outubro', 'Novembro', 'Dezembro']

# Função para criar os dados

def criar\_dados(num\_linhas):

np.random.seed(0)

dados = []

for \_ in range(num\_linhas):

cidade = np.random.choice(cidades)

categoria = np.random.choice(categorias)

regiao = [regiao for regiao, cidades\_regiao in regioes.items() if cidade in cidades\_regiao][0]

ano = np.random.choice(anos)

mes = np.random.choice(meses)

vendedor\_id = np.random.choice(vendedores[categoria])

produto = np.random.choice(list(precos.keys()))

preco = precos[produto][categoria]

quantidade = np.random.randint(1, 50)

venda\_total = preco \* quantidade

dados.append([cidade, regiao, ano, mes, vendedor\_id, produto, preco, quantidade, venda\_total])

df = pd.DataFrame(dados, columns=['Cidade', 'Região', 'Ano', 'Mês', 'Vendedor\_ID',

'Produto', 'Preço\_Produto', 'Quantidade\_Produto', 'Venda\_Total'])

return df

# Gerar e salvar os dados

df\_dados = criar\_dados(1000)

df\_dados.to\_excel('vendas\_dados.xlsx', index=False)