



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Analytics e Business Intelligence

UMA ANÁLISE DO *E-COMMERCE* DE UMA SUPER LOJA

Marília Figueiredo Santos

Belo Horizonte
2023

SUMÁRIO

| | |
|---------------------------------------------------------------------|-----------|
| 1. Introdução | 3 |
| 1.1. O problema proposto | 4 |
| 2. Coleta de Dados | 6 |
| 3. Processamento/Tratamento de Dados..... | 8 |
| 3.1. Ferramentas utilizadas..... | 8 |
| 3.1.2 Bibliotecas utilizadas..... | 8 |
| 3. Integração, Tratamento e Carga de Dados | 9 |
| 3.1. Fontes de Dados..... | 9 |
| 3.2. Obtendo dados | 11 |
| 3.2.1 Dataset: EU Superstore.xlsx e WhatsgoodlyData-6.xlsx | 11 |
| 4. Análise e Exploração dos Dados | 20 |
| 4.1 Exploração dos dados | 20 |
| 4.2 Análises com novas colunas..... | 21 |
| 4.2.1 Mesclando as Bases da Super Loja, de Ano, Mês e de Dias | 25 |
| 4.2.2 A coluna Preco_unit_produto..... | 26 |
| 4.2.3 Coluna Desconto_euro | 27 |
| 4.2.4 Coluna Lucro_prod_unit..... | 28 |
| 4.3 Explorações dos dados correlacionando colunas | 29 |
| 5. Modelo de Dados | 42 |
| 5.2. Fatos e Dimensões | 42 |
| 5.3. Dashboard | 43 |
| 5.3.1 Painel Estratégico | 43 |
| 5.3.2 Painel Tático | 44 |
| 5.3.3 Painel Operacional..... | 45 |
| 5.4 Modelo Canvas | 47 |
| 6. Links | 48 |
| REFERÊNCIAS | 49 |

1. Introdução

A presente pesquisa aborda o tema de um e-commerce de uma Super Loja do seguimento mobiliário corporativo e consumidor que atua em toda Europa nas regiões Sul, Norte e Central, consiste nas vendas de eletrodomésticos para escritório, móveis e tecnologia nas quais os atores se relacionam por um meio eletrônico. Temos produtos desde cadeiras, estantes de livros, telefones e copiadoras. As mudanças representadas pelas transações eletrônicas impactaram em crescimento econômico, na geração de empregos e criaram relações singulares entre o setor produtivo e seus consumidores.

O comércio por intermédio eletrônico advém da evolução das tecnologias de comunicação e informação e está acessível a praticamente todo tipo de empresa, dos mais variados tamanhos e setores, em qualquer parte do mundo. A integração proposta pela globalização e pelas novas tecnologias da informação causou a tendência de aumento desse tipo de comércio no ambiente interno e externo.

Nakamura (2011) define o termo comércio eletrônico como os processos realizados e executados unicamente por um meio eletrônico em um ambiente eletrônico que possam envolver uma determinada cadeia de valor. Esses processos empregam ferramentas que contém grande tecnologia de comunicação e de informação, com o intuito principal de atender as necessidades e as mais variadas exigências dos negócios.

A internet tem se tornado parte da vida e do cotidiano de grande parte dos cidadãos. Há um entendimento global que ela deveria ser alçada a um direito humano fundamental e universal. As organizações, por sua vez, utilizam cada vez mais a tecnologia de informação tanto no ambiente organizacional, quanto na relação com seus clientes. As mudanças de paradigmas causadas pela globalização e pela quebra de barreiras físicas no mundo eletrônico justificam o estudo e a aplicação do comércio eletrônico como ferramenta fundamental de aquisição e sustentação de novos mercados pelas empresas. A mineração de dados é uma das principais ferramentas para obter dados relevantes e realizar um mapeamento mais aguçado para ajudar na tomada de decisões.

Para Silva (2009), o aparecimento do Intercâmbio Eletrônico de Dados (Electronic Data Interchange), na década de 1960 gerou grandes evoluções que favoreceram as empresas nas trocas de mensagens, documentos, ordens de pagamentos, solicitações de material e outras operações, o que tornou a negociação mais democrática e rápida ao transformar os processos obsoletos de negociação utilizados na era pré-internet.

Esse trabalho objetiva aplicar técnicas de Mineração de Dados e Modelos Preditivos em vendas a fim de classificar e encontrar padrões extrair algumas informações e gerar insights que podem influenciar em estratégias de vendas e vendas.

A estratégia metodológica utilizada no presente trabalho foi um estudo descritivo e exploratório, com análise dos dados através de uma abordagem quantitativa. Primeiramente buscou-se analisar as informações das vendas da Super Loja. Essas informações foram obtidas no site data.world. As palavras-chave utilizadas nas pesquisas foram: Super Lojas, comércio eletrônico, lojas virtuais.

Posteriormente à coleta das informações, foi realizada a leitura e seleção do material e em seguida foi efetuada uma análise com o objetivo de compreender e entender sobre as vendas da Super Loja e, assim, elaborar o referencial teórico da investigação, onde o trabalho foi feito utilizando o Google Colab.

1.1. O problema proposto

Neste trabalho será utilizado a Análise Exploratória e Modelagem Preditiva para extração de informações importantes do conjunto de dados da Super Loja com o objetivo de realizar uma análise visando verificar o quão bem ou mal as vendas estão indo na empresa. Para isso todos os atributos serão classificados com um grau de importância, desta maneira conseguimos analisar os resultados e utilizá-los em previsões futuras.

Para isso, serão analisados o conjunto de dados da Super Loja, disponibilizados no site da data.world. Os principais objetivos dessa análise são:

- Realizar uma análise nos dados das vendas da Super Loja para tomar decisões estratégicas na venda desses produtos.
- Verificar quais são os produtos que mais vendem?
- Verificar quais os produtos que menos vendem?
- Verificar quais os maiores compradores?
- Analisar em quais regiões se concentram as maiores e menores taxas de vendas?
- Apresentar quais os melhores e piores meses de venda no ano.

Os dados que serão analisados, foram coletados do site da data.world. Foi necessário coletar algumas informações separadamente, são elas:

- 1 Data set da Super Loja: neste data set é apresentado informações sobre as vendas dos produtos por categorias, contendo informações como: seguimento, nome do cliente, cidade, região, quantidade, lucro, vendas, entre outras.
- 2 Data set plataforma de compras: neste data set é apresentado qual plataforma mais influenciou as compras online

- As análises realizadas, têm como objetivo encontrar padrões, métricas e tendências que auxiliarão no entendimento das bases trabalhadas. E assim poderemos indicar quais características da Super Loja que a tornam uma empresa com crescente receita e um negócio durável.
- Em relação a aspectos geográficos, os resultados obtidos destinam-se exclusivamente a vendas online ocorridas no site da Super Loja.
- Os dados se estendem de 2020 a 2023 e incluem tanto os totais de vendas quanto detalhes de subcategorias de vendas no varejo.

¹ <https://data.world/vizwiz/superstore-europe/workspace/file?filename=Sample+-+EU+Superstore.xls>

² <https://data.world/ahalps/social-influence-on-shopping/workspace/file?filename=WhatsgoodlyData->

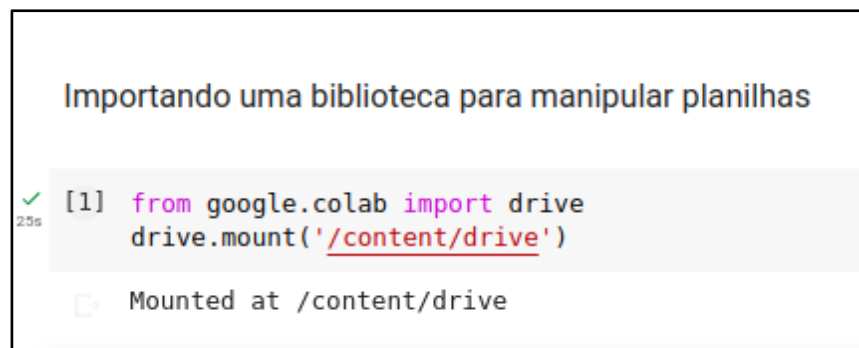
2. Coleta de Dados

O Trabalho foi desenvolvido no Google Colab, onde a coleta de dados foi separada em duas partes: a primeira e a segunda com dados do data.world com dados da Super Loja e a plataforma de compras. Como observação, o primeiro dataset veio formatado em xls, e o segundo como arquivo csv por este motivo foi necessário a conversão de formato para.xlsx.

No Google Colab, onde o projeto foi executado, a importação foi feita conforme abaixo:

Importar o dataset do data.world, o qual foi feito download, para isso foi necessário utilizar uma importação do google, selecionar o arquivo (fazendo upload para o Google Drive) e preparando ele para o pandas fazer a leitura.

Figura 1: Importando Drive

A screenshot of a Google Colab code cell. At the top, the text 'Importando uma biblioteca para manipular planilhas' is displayed. Below it, a code cell shows the execution of the following Python code:

```
[1] from google.colab import drive
drive.mount('/content/drive')
```

 To the left of the code, there is a green checkmark and the text '23s'. Below the code, a message indicates the drive is mounted:

```
Mounted at /content/drive
```

Fonte: Autor

2.1 Listagem descritiva das colunas:

A tabela abaixo mostra uma descrição da estrutura encontrada no Dataset EU Superstore.xls exibindo o nome do atributo, tipo e sua descrição.

Tabela 1: Estrutura Dataset EU Superstore.xlsx

| Nome da coluna | Descrição | Tipo |
|----------------|--------------|----------------|
| Row ID | ID_pedido | int64 |
| Order ID | Tipo_cliente | object |
| Order Date | Data_pedido | datetime64[ns] |
| Dispatch Date | Data_envio | datetime64[ns] |
| Delivery Mode | Modo_envio | object |
| Customer ID | ID_cliente | object |
| Customer Name | Nome_cliente | object |
| Segment | Seguimento | object |
| City | Cidade | object |
| State/Province | Estado | object |
| Country/Region | País | object |
| Region | Região | object |
| Product ID | ID_produto | object |
| Category | Categoria | object |
| Sub-Category | Subcategoria | object |
| Product Name | Nome_produto | object |
| Sales | Vendas | float64 |
| Quantity | Quantidade | int64 |
| Discount | Desconto | float64 |
| Profit | Lucro | float64 |

Fonte: Autor

O Dataset WhatsgoodlyData-6.xlsx contém informações referentes a plataforma de compras, contém a seguinte estrutura:

Tabela 2: Estrutura Dataset WhatsgoodlyData-6.xlsx

| Nome da coluna | Descrição | Tipo |
|---------------------|-----------------------|---------|
| Question | Pergunta | object |
| Segment Type | Tipo_segmento | object |
| Segment Description | Descrição_seguimento | object |
| Answer | Respostas | object |
| Count | Contagem_respostas | int64 |
| Percentage | Porcentagem_respostas | float64 |

Fonte: Autor

Dessa forma os dados estão preparados para serem trabalhados e posteriormente analisados.

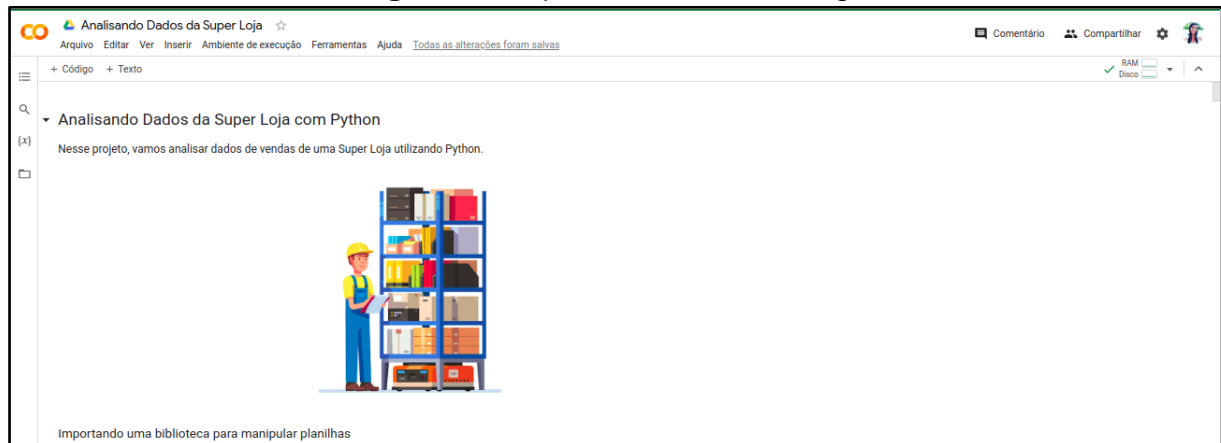
3. Processamento/Tratamento de Dados

Nessa seção serão apresentadas todas as ferramentas e bibliotecas utilizadas para o processamento e o tratamento dos dados.

3.1. Ferramentas utilizadas

Como ferramenta para desenvolvimento dos scripts em python, foi escolhido a ferramenta em nuvem do Google Colaboratory, (figura 2), disponível em <https://colab.research.google.com/>

Figura 2: Captura da tela do Google Colab



Fonte: Autor

Foi escolhida a ferramenta Google Colab, pois possui as principais ferramentas e bibliotecas para realizarmos todas as codificações necessárias para a análise e tratamento dos dados. Esta ferramenta nos permite unir código e texto, facilitando nossa organização no projeto com notebook Jupyter.

3.1.2 Bibliotecas utilizadas

Para realizar o processamento e o tratamento dos dados, foi necessário importar algumas bibliotecas conforme a Figura 3 abaixo.

Figura 3: Captura da tela da importação das bibliotecas

```
[182] import pandas as pd
import numpy as np
%matplotlib inline
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.graph_objects as go
import plotly.express as px
import matplotlib.image as mpimg
from datetime import datetime
import bar_chart_race as bcr
from itertools import combinations
from collections import Counter
import chart_studio.plotly as py
import chart_studio
import plotly.graph_objs as pg
from plotly.subplots import make_subplots

[183] import chart_studio
chart_studio.tools.set_credentials_file(username='mariliafigueiredo007', api_key='uIWj5SxRpkIPamIjRV50')
```

Fonte: Autor

3. Integração, Tratamento e Carga de Dados

3.1. Fontes de Dados

Descrição das bases de dados ou arquivos utilizados pelo projeto como fonte, apresentando possíveis diagramas dos bancos de dados relacionais.

A tabela 2 abaixo mostra de forma detalhada as bibliotecas importadas.

Tabela 3: Bibliotecas utilizadas

| Bibliotecas | Descrição | Comando(s) Utilizados |
|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Pandas | Pacote de ferramentas para análise de dados e manipulação, construída sobre a base da linguagem python. Informações: https://pandas.pydata.org/ | Import pandas as pd |
| Numpy | Pacote de ferramentas utilizada para realizar cálculos em arrays multidimensionais, construída sobre a base da linguagem de programação python. Informações: https://numpy.org/ | Import numpy as np |

| | | |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------|
| Matplotlib | Pacote de ferramentas utilizadas para criação de gráficos e visualização de dados, construída sobre a base da linguagem python. Informações: https://matplotlib.org/ | Import matplotlib.pyplot as plt %matplotlib inline Import matplotlib.image as mpimg |
| Seaborn | Pacote de ferramentas utilizadas para criação de gráficos e visualização de dados de alto nível baseada na lib Matplotlib. Informações: https://docs.python.org/pt-br/3/library/datetime.html | Import seaborn as sns |
| Plotly graph | Pacote de ferramentas utilizadas para a criação de gráficos. https://plotly.com/python/graph-objects/ | Import plotly.graph_objects as go Import plotly.graph_objs as pg from plotly.subplots import make_subplots |
| Plotly.express | Pacote de ferramentas utilizadas para a criação de figuras. https://plotly.com/python/plotly-ex | Import plotly.express as px |
| Bar_chart | Pacote python para criar corridas de gráfico de barras. Informações: https://matplotlib.org/stable/gallery/lines_bars_and_markers/barchart.html | import bar_chart_race as bcr |
| Itertools | Funções que criam iteradores para um loop eficiente. Informações: https://docs.python.org/3/library/itertools.html | from itertools import combinations |
| Collections | A classe Python Counter faz parte do módulo Collections . Counter é uma subclasse de Dictionary e usado para controlar os elementos e sua contagem. Informações: https://docs.python.org/3/library/collections.html | from collections import Counter |
| Chart studio | O Chart Studio fornece um serviço da Web para hospedar gráficos. Informações: https://plotly.com/python/getting-started-with-chart-studio/ | import chart_studio.plotly as py |

Fonte: Autor

3.2. Obtendo dados

Nessa seção será apresentado como foi coletado os dados após a execução do notebook Analisando Dados da Super Loja, salvo no google drive arquivo.xlsx.

3.2.1 Dataset: EU Superstore.xlsx e WhatsgoodlyData-6.xlsx

Na Figura abaixo é apresentado a parte do código utilizada para obter os dados salvos em formato.xlsx e transformá-los em um formato Data Frame para iniciarmos a nossa análise.

Os dados utilizados neste trabalho tiveram origem em duas fontes. A primeira foi a base das vendas fornecida pela site data.world EU Superstore.xlsx. A segunda foi o a plataforma de vendas, onde essa base foi fornecida pelo site data.world WhatsgoodlyData-6.xlsx.

Esses dois datasets foram unificados, onde foram realizadas as análises diante do cenário da Super Loja, ou seja, verificar a correlação com as vendas e a plataforma de vendas.

Figura 4: Armazenando dados df1

```
#Armazenando os dados na variável 'df1'  
df1 = pd.read_excel('https://query.data.world/s/33svkaypxczmjvnyvxsb7fjlx4dx2?dws=00000')
```

Fonte: Autor

Para esse caso, foi criado a variável df1 para trazer os dados da super loja e armazenar no dataframe.

O resultado em df1 dos dados coletados, selecionando apenas os 5 primeiros registros encontrados através do comando df1.head().

Figura 5: Exibindo dados dataset do df1

df1.head()

| | Row ID | Order ID | Order Date | Dispatch Date | Delivery Mode | Customer ID | Customer Name | Segment | City | State/Province | Country/Region | Region | Product ID | Category | Sub-Category | Product Name | Sales | Quantity | Discount | Profit |
|---|--------|-----------------|------------|---------------|----------------|-------------|----------------|-----------|-------|----------------|----------------|--------|-----------------|-----------------|--------------|--------------------------------------------|--------|----------|----------|--------|
| 0 | 1 | ES-2023-1311038 | 2023-02-07 | 2023-02-11 | Standard Class | AS-10045 | Aaron Smayling | Corporate | Leeds | England | United Kingdom | North | OFF-ST-10000988 | Office Supplies | Storage | Fellowes Folders, Blue | 79.20 | 3 | 0.0 | 39.60 |
| 1 | 2 | ES-2023-1311038 | 2023-02-07 | 2023-02-11 | Standard Class | AS-10045 | Aaron Smayling | Corporate | Leeds | England | United Kingdom | North | TEC-AC-10004144 | Technology | Accessories | SanDisk Numeric Keypad, Bluetooth | 388.92 | 7 | 0.0 | 0.00 |
| 2 | 3 | ES-2023-1311038 | 2023-02-07 | 2023-02-11 | Standard Class | AS-10045 | Aaron Smayling | Corporate | Leeds | England | United Kingdom | North | OFF-LA-10001915 | Office Supplies | Labels | Avery Legal Exhibit Labels, 5000 Label Set | 35.19 | 3 | 0.0 | 16.11 |
| 3 | 4 | ES-2023-1311038 | 2023-02-07 | 2023-02-11 | Standard Class | AS-10045 | Aaron Smayling | Corporate | Leeds | England | United Kingdom | North | OFF-ST-10004550 | Office Supplies | Storage | Fellowes Folders, Wire Frame | 50.94 | 2 | 0.0 | 13.20 |
| 4 | 5 | ES-2023-1311038 | 2023-02-07 | 2023-02-11 | Standard Class | AS-10045 | Aaron Smayling | Corporate | Leeds | England | United Kingdom | North | TEC-AC-10004068 | Technology | Accessories | Memorex Memory Card, USB | 307.44 | 3 | 0.0 | 73.71 |

Fonte: Autor

A extração apenas do Orçamento de vendas é executada conforme o script abaixo.

Figura 6: Armazenando dados df2

```
df2 = pd.read_csv('https://query.data.world/s/famzmneo2z6vlce5zj7evr5o4obyqp?dws=000000')
df2.head()
```

Fonte: Autor

O resultado em df2 dos dados coletados, selecionando apenas os 5 primeiros registros encontrados através do comando df2.head().

Figura 7: Exibindo dados dataset do df2

| | Question | Segment Type | Segment Description | Answer | Count | Percentage |
|---|---------------------------------------------------|--------------|---------------------|-----------|-------|------------|
| 0 | What social platform has influenced your onlin... | Mobile | Global results | Facebook | 548 | 0.205 |
| 1 | What social platform has influenced your onlin... | Mobile | Global results | Instagram | 916 | 0.342 |
| 2 | What social platform has influenced your onlin... | Mobile | Global results | Snapchat | 86 | 0.032 |
| 3 | What social platform has influenced your onlin... | Mobile | Global results | Twitter | 179 | 0.067 |
| 4 | What social platform has influenced your onlin... | Mobile | Global results | None | 947 | 0.354 |

Fonte: Autor

3.3 Informações dos DataFrames

Para cada objeto do Dataset foi utilizado a função info() para visualizarmos algumas informações como a quantidade de registros, quantidade de colunas,

informações de cada coluna e o tipo dela. Podemos observar na figura abaixo, que no Datasete df1 foram encontrados 10000 registros com um total de 20 colunas.

Figura 8: Exibindo informações Dataset df1

```
#Vesualizando informações sobre os dados antes da formatação
df1.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Row ID                 10000 non-null  int64
1   Order ID               10000 non-null  object
2   Order Date             10000 non-null  datetime64[ns]
3   Dispatch Date          10000 non-null  datetime64[ns]
4   Delivery Mode           10000 non-null  object
5   Customer ID            10000 non-null  object
6   Customer Name          10000 non-null  object
7   Segment                10000 non-null  object
8   City                   10000 non-null  object
9   State/Province         10000 non-null  object
10  Country/Region         10000 non-null  object
11  Region                 10000 non-null  object
12  Product ID             10000 non-null  object
13  Category               10000 non-null  object
14  Sub-Category           10000 non-null  object
15  Product Name           10000 non-null  object
16  Sales                  10000 non-null  float64
17  Quantity               10000 non-null  int64
18  Discount               10000 non-null  float64
19  Profit                 10000 non-null  float64
dtypes: datetime64[ns](2), float64(3), int64(2), object(13)
memory usage: 1.5+ MB
```

Fonte: Autor

Para o Dataset com os valores da tabela orçamento de vendas, temos o seguinte resultado:

Figura 9: Exibindo informações DataFrame df2

```
#Visualizando informações sobre os dados antes da formatação
df2.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1450 entries, 0 to 1449
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Question              1450 non-null  object
1   Segment Type          1450 non-null  object
2   Segment Description    1450 non-null  object
3   Answer                1450 non-null  object
4   Count                 1450 non-null  int64
5   Percentage            1450 non-null  float64
dtypes: float64(1), int64(1), object(4)
memory usage: 68.1+ KB
```

Fonte: Autor

Logo podemos ver que foram encontrados 1450 registros a quantidade do Dataset da super loja, porém somente com 6 colunas.

3.4 Tratamento dos Dados

Nessa seção será detalhado cada passo realizado no tratamento dos dados antes de começarmos as análises.

3.4.1 Unindo os Datasets

Antes de realizarmos o tratamento em cada Dataset, para facilitar foi feita a união dos Datasets. Representado na Figura 10 abaixo:

Figura 10: Unindo os Datasets: superLoja

```
#Unindo os Datasets
superLoja = pd.concat([df1, df2], axis=1, ignore_index=False)
superLoja.head()
```

Fonte: Autor

3.4.2 Removendo as colunas

Após realizado análise em cima dos Datasets df1 e df2, verificou-se que existem colunas que não serão necessárias e não terão impacto em análises futuras. Desta forma essas colunas foram removidas conforme a Figura 11.

Figura 11: Remoção colunas DataFrame superLoja

```
#Excluindo colunas
del superLoja['Order ID']
del superLoja['Question']
del superLoja['Segment Description']
del superLoja['Segment Type']
```

Fonte: Autor

3.4.3 Renomeando as colunas

Para deixar nossos dados formatados e padronizados, algumas colunas foram renomeadas utilizando a função `rename()`, conforme a Figura 12 abaixo.

Figura 12: Renomeando as colunas do DataFrame: superLoja

```
#Traduzindo as colunas do Dataset
superLoja.rename(columns={'Row ID': 'ID_pedido',
                          'Order ID': 'Tipo_cliente',
                          'Customer Name': 'Nome_cliente',
                          'Order Date': 'Data_pedido',
                          'Dispatch Date': 'Data_envio',
                          'Delivery Mode': 'Modo_envio',
                          'Customer ID': 'ID_cliente',
                          'Customer Name': 'Nome_cliente',
                          'Segment': 'Seguimento',
                          'Country/Region': 'Pais',
                          'City': 'Cidade',
                          'State/Province': 'Estado',
                          'Region': 'Regiao',
                          'Product ID': 'ID_produto',
                          'Category': 'Categoria',
                          'Sub-Category': 'SubCategoria',
                          'Product Name': 'Nome_produto',
                          'Sales': 'Vendas',
                          'Quantity': 'Quantidade',
                          'Discount': 'Desconto',
                          'Profit': 'Lucro',
                          'Question': 'Pergunta',
                          'Segment Type': 'Tipo_segmento',
                          'Segment Description': 'Descricao_segmento',
                          'Answer': 'Respostas',
                          'Count': 'Contagem_respostas',
                          'Percentage': 'Porcentagem_respostas'}, inplace = True)
```

Fonte: Autor

3.4.4 Traduzindo as linhas do dataframe

No entanto, também para deixar os dados formatados e padronizados, algumas linhas foram renomeadas, conforme a Figura 13 abaixo.

Figura 13: Traduzindo as colunas DataFrame: superLoja

```
#Traduzindo as linhas do dataframe
superLoja['Categoria'].replace({'Technology':'Tecnologia', 'Office Supplies':'Material de Escritório', 'Furniture':'Móveis'}, inplace = True)
superLoja['Seguimento'].replace({'Consumer':'Consumidor', 'Corporate':'Corporativo',
                                 'Home Office':'Escritório em casa'}, inplace = True)
superLoja['SubCategoria'].replace({'Accessories':'Acessórios', 'Appliances':'Eletrodomésticos', 'Art':'Art', 'Binders':'Fichários',
                                   'Bookcases':'Estantes', 'Chairs':'Cadeiras', 'Copiers':'Copiadoras', 'Envelopes':'Envelopes',
                                   'Fasteners':'Fechos', 'Furnishings':'Mobiliária', 'Labels':'Etiquetas', 'Machines':'Máquinas',
                                   'Paper':'Papel', 'Phones':'Telefones', 'Storage':'Armazenar', 'Supplies':'Suprimentos',
                                   'Tables':'Tabelas' }, inplace = True)
superLoja['Regiao'].replace({'Central':'Central', 'North':'Norte', 'South':'Sul'}, inplace = True)
superLoja['Pais'].replace({'Austria':'Áustria', 'Belgium':'Bélgica', 'Denmark':'Dinamarca', 'Finland':'Finlândia', 'France':'França',
                          'Germany':'Alemanha', 'Ireland':'Irlanda', 'Italy':'Itália', 'Netherlands':'Holanda', 'Norway':'Noruega',
                          'Portugal':'Portugal', 'Spain':'Espanha', 'Sweden':'Suécia', 'Switzerland':'Suíça',
                          'United Kingdom':'Reino Unido'}, inplace = True)
```

Fonte: Autor

3.4.4 Padronizando os tipos das colunas

Para deixar os tipos de cada coluna padronizados, foi realizada a conversão dos tipos em algumas colunas utilizando a função `astype()` e outros em `int64`, conforme Figura 14.

Figura 14: Padronizando colunas do DataFrame: superLoja

```
# Convertendo dados que estão como object para string e inteiro
superLoja['ID_pedido'] = superLoja['ID_pedido'].astype('int64')
superLoja['Quantidade'] = superLoja['Quantidade'].astype('int64')
superLoja['Nome_cliente'] = superLoja['Nome_cliente'].astype('string')
superLoja['Seguimento'] = superLoja['Seguimento'].astype('string')
superLoja['Cidade'] = superLoja['Cidade'].astype('string')
superLoja['Pais'] = superLoja['Pais'].astype('string')
superLoja['Estado'] = superLoja['Estado'].astype('string')
superLoja['Regiao'] = superLoja['Regiao'].astype('string')
superLoja['Categoria'] = superLoja['Categoria'].astype('string')
superLoja['SubCategoria'] = superLoja['SubCategoria'].astype('string')
```

Fonte: Autor

As colunas `Nome_cliente`, `Seguimento`, `Cidade`, `Estado`, `Região`, `País`, `Categoria` e `Subcategoria` foram convertidos para `string`. O valor da coluna `ID_pedido` e `Quantidade` `object`, foi convertido para um formato `int64`, correspondendo ao ID do pedido e a quantidade de vendas.

3.4.5 Informações dos DataFrames

Para certificar que os dados estão padronizados e que não existem valores

nulos no DataFrame, foi executado a função `isnull()` onde recupera todos os registros que estão com valores nulo e logo em seguida aplicado a função `sum()`, pois somamos a quantidade de registros nulos da coluna correspondente, conforme a Figura 15.

Figura 15: Verificando valores nulos

```
superLoja.isnull().sum()
```

| | |
|---------------------|-------|
| Row ID | 0 |
| Order ID | 0 |
| Order Date | 0 |
| Dispatch Date | 0 |
| Delivery Mode | 0 |
| Customer ID | 0 |
| Customer Name | 0 |
| Segment | 0 |
| City | 0 |
| State/Province | 0 |
| Country/Region | 0 |
| Region | 0 |
| Product ID | 0 |
| Category | 0 |
| Sub-Category | 0 |
| Product Name | 0 |
| Sales | 0 |
| Quantity | 0 |
| Discount | 0 |
| Profit | 0 |
| Question | 8550 |
| Segment Type | 8550 |
| Segment Description | 8550 |
| Answer | 8550 |
| Count | 8550 |
| Percentage | 8550 |
| dtype: | int64 |

Fonte: Autor

O resultado obtido apresenta a listagem de todas as colunas e o valor da soma de valores nulos obtidos. Neste caso todas as colunas que eram do dataset `df2`, possuem registro com valores nulos, por isso os valores aparecem com o total de registros.

Por conta desse concat, os dados que não têm dados correspondentes acabaram por ficarem como NaN em todos os campos. Para isso, foi utilizado a função `fillna()` para preencher com a média e none esses casos.

Figura 16: Substituindo valores NaN

| superLoja.sample(5) | | | | | | | | | | | | | |
|---------------------|---------|-----------------|------------------------|--------------|-------------------------------------|---------|------------|----------|---------|-----------|--------------------|-----------------------|--|
| Pais | Regiao | ID_produto | Categoria | SubCategoria | Nome_produto | Vendas | Quantidade | Desconto | Lucro | Respostas | Contagem_respostas | Porcentagem_respostas | |
| França | Central | FUR-BO-10003738 | Móveis | Estantes | Dania Library with Doors, Mobile | 985.284 | 3 | 0.1 | 142.254 | None | 35.013793 | 0.199313 | |
| Alemanha | Central | FUR-CH-10002647 | Móveis | Cadeiras | Novimex Rocking Chair, Red | 354.861 | 3 | 0.1 | 31.491 | None | 35.013793 | 0.199313 | |
| Alemanha | Central | OFF-AR-10003377 | Material de Escritório | Art | Boston Pencil Sharpener, Easy-Erase | 241.680 | 8 | 0.0 | 38.640 | None | 6.000000 | 0.261000 | |
| Espanha | Sul | OFF-AR-10002165 | Material de Escritório | Art | Boston Sketch Pad, Fluorescent | 349.020 | 7 | 0.0 | 0.000 | None | 35.013793 | 0.199313 | |
| Alemanha | Central | TEC-MA-10002118 | Tecnologia | Máquinas | Konica Printer, Red | 529.980 | 2 | 0.0 | 238.440 | None | 35.013793 | 0.199313 | |

Fonte: Autor

Dessa forma, agora os dados estão preparados para uma melhor exploração, e nesse caso, verificar a correlação entre as vendas da super Loja e a plataforma de vendas.

Figura 17: Verificando valores nulos após tratamento

| #Conferindo dados faltantes | |
|-------------------------------------------------------------------------------------|-----|
| faltantes_percentual = (superLoja.isnull().sum() / len(superLoja['ID_pedido']))*100 | |
| print(faltantes_percentual) | |
| ID_pedido | 0.0 |
| Data_pedido | 0.0 |
| Data_envio | 0.0 |
| Modo_envio | 0.0 |
| ID_cliente | 0.0 |
| Nome_cliente | 0.0 |
| Seguimento | 0.0 |
| Cidade | 0.0 |
| Estado | 0.0 |
| Pais | 0.0 |
| Regiao | 0.0 |
| ID_produto | 0.0 |
| Categoria | 0.0 |
| SubCategoria | 0.0 |
| Nome_produto | 0.0 |
| Vendas | 0.0 |
| Quantidade | 0.0 |
| Desconto | 0.0 |
| Lucro | 0.0 |
| Respostas | 0.0 |
| Contagem_respostas | 0.0 |
| Porcentagem_respostas | 0.0 |
| dtype: float64 | |

Fonte: Autor

Para exibir as informações do DataFrame de forma detalhada, utilizamos a função info(), ela retorna informações importantes de cada coluna, tais como: nome da coluna, tipo da coluna e se a coluna aceita valores nulos, conforme a Figura 20 abaixo.

Figura 18: Verificando valores nulos após tratamento

```
#informações sobre o dataframe criado
superLoja.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 22 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   ID_pedido             10000 non-null  int64
 1   Data_pedido           10000 non-null  datetime64[ns]
 2   Data_envio            10000 non-null  datetime64[ns]
 3   Modo_envio            10000 non-null  object
 4   ID_cliente            10000 non-null  object
 5   Nome_cliente          10000 non-null  object
 6   Seguimento            10000 non-null  object
 7   Cidade                10000 non-null  object
 8   Estado                10000 non-null  object
 9   Pais                  10000 non-null  object
10   Regiao                10000 non-null  object
11   ID_produto            10000 non-null  object
12   Categoria              10000 non-null  object
13   SubCategoria          10000 non-null  object
14   Nome_produto          10000 non-null  object
15   Vendas                 10000 non-null  float64
16   Quantidade            10000 non-null  int64
17   Desconto              10000 non-null  float64
18   Lucro                 10000 non-null  float64
19   Respostas             10000 non-null  object
20   Contagem_respostas    10000 non-null  float64
21   Porcentagem_respostas 10000 non-null  float64
dtypes: datetime64[ns](2), float64(5), int64(2), object(13)
memory usage: 1.7+ MB
```

Fonte: Autor

Foi utilizado a função `shape()`, onde podemos visualizar as dimensões do DataFrame.

Figura 19: Informações das dimensões do DataFrame

```
#Verificar o número de linhas e colunas, após exclusão de colunas
superLoja.shape

(10000, 22)
```

Fonte: Autor

Com isso obtemos como resultado 10,000 linhas e 22 colunas. Além disso, foi verificado a ocorrência de valores duplicados como linhas inteiras com os mesmos dados exatos.

Figura 20: Verificando valores duplicados

```
#Verificando dados duplicados
superLoja.duplicated().sum()

0
```

Fonte: Autor

Usando a função duplicada do pandas confirmamos que o dataframe não possui linhas duplicadas.

4. Análise e Exploração dos Dados

Nessa seção será mostrado todas as análises e exploração dos dados tratados anteriormente. Analisaremos as ocorrências, padrões e informações importantes que levantamos do dataframe.

4.1 Exploração dos dados

Para iniciar a exploração dos dados foi feita, inicialmente a descrição estatística deles:

Figura 21: Descrição estatística do DataFrame

```
#Descrição estatística da Super Loja
superLoja.describe().round(2)
```

| | ID_pedido | Vendas | Quantidade | Desconto | Lucro | Contagem_respostas | Porcentagem_respostas |
|-------|-----------|----------|------------|----------|----------|--------------------|-----------------------|
| count | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.00 | 10000.0 |
| mean | 5000.50 | 293.81 | 3.78 | 0.10 | 37.28 | 35.01 | 0.2 |
| std | 2886.90 | 486.06 | 2.20 | 0.17 | 178.10 | 36.19 | 0.1 |
| min | 1.00 | 2.96 | 1.00 | 0.00 | -3059.82 | 0.00 | 0.0 |
| 25% | 2500.75 | 49.46 | 2.00 | 0.00 | 1.32 | 35.01 | 0.2 |
| 50% | 5000.50 | 119.36 | 3.00 | 0.00 | 14.22 | 35.01 | 0.2 |
| 75% | 7500.25 | 320.71 | 5.00 | 0.10 | 48.51 | 35.01 | 0.2 |
| max | 10000.00 | 7958.58 | 14.00 | 0.85 | 3979.08 | 947.00 | 1.0 |

Fonte: Autor

Nessa descrição podemos ver que a mínima e máxima modificaram bastante, o que já indica uma grande movimentação e crescimento da receita. Por conseguinte, foi verificada as 3 maiores vendas e as 3 menores.

Figura 22: Três maiores e Três menores vendas

| # Descobrindo as 3 maiores vendas superLoja.nlargest(3, "Vendas").round(2) | | | | | | | | | | | | | | | | |
|--------------------------------------------------------------------------------|-------------|------------|------------|----------------|--------------|-------------------|-------------|--------------|-------------|-------------|------------|-----------------|------------------------|------------------|--------------------------------------------|------------|
| ID_pedido | Data_pedido | Data_envio | Modo_envio | ID_cliente | Nome_cliente | Seguimento | Cidade | Estado | Pais | Regiao | ID_produto | Categoria | SubCategoria | Nome_produto | Vendas | Quantidade |
| 1774 | 1775 | 2023-09-08 | 2023-09-14 | Standard Class | PJ-18835 | Patrick Jones | Corporativo | Prato | Tuscany | Itália | Sul | OFF-AP-10004512 | Material de Escritório | Eletrodomesticos | Hoover Stove, Red | 7958.58 |
| 3265 | 3266 | 2020-09-09 | 2020-09-12 | Second Class | ER-13855 | Elpida Rittenbach | Corporativo | Lugo | Galicia | Espanha | Sul | OFF-AP-10002330 | Material de Escritório | Eletrodomesticos | Hamilton Beach Stove, Silver | 6517.08 |
| 6386 | 6387 | 2021-12-14 | 2021-12-18 | Standard Class | GT-14710 | Greg Tran | Consumidor | Huddersfield | England | Reino Unido | Norte | TEC-PH-10004583 | Tecnologia | Telefones | Motorola Smart Phone, Cordless | 5785.02 |
| # Descobrindo as 3 menores vendas superLoja.nsmallest(3, "Vendas").round(2) | | | | | | | | | | | | | | | | |
| ID_pedido | Data_pedido | Data_envio | Modo_envio | ID_cliente | Nome_cliente | Seguimento | Cidade | Estado | Pais | Regiao | ID_produto | Categoria | SubCategoria | Nome_produto | Vendas | Quantidade |
| 9304 | 9305 | 2023-04-18 | 2023-04-25 | Standard Class | JS-16030 | Joy Smith | Consumidor | Copenhagen | Hovedstaden | Dinamarca | Norte | OFF-BI-10001119 | Material de Escritório | Fichários | Wilson Jones Index Tab, Clear | 2.96 |
| 5019 | 5020 | 2021-06-30 | 2021-07-05 | Standard Class | MV-18190 | Mike Vittorini | Consumidor | Solna | Stockholm | Suécia | Norte | OFF-LA-10000553 | Material de Escritório | Etiquetas | Hon Round Labels, Laser Printer Compatible | 3.33 |
| 5045 | 5046 | 2022-01-15 | 2022-01-15 | Same Day | TG-21640 | Trudy Glocke | Consumidor | Vienna | Vienna | Áustria | Central | OFF-BI-10001544 | Material de Escritório | Fichários | Wilson Jones Hole Reinforcements, Clear | 3.99 |

Fonte: Autor

Analizando as vendas é possível verificar que as tendências e expectativas de vendas para os períodos são muito boas.

4.2 Análises com novas colunas

Após as análises realizadas, concluímos que o tempo de envio também pode ser considerado um atributo relevante para tomar decisões estratégicas na venda desses produtos. Dessa forma, para a análise diminuimos os valores da coluna Data_envio pelos valores da coluna Data_pedido e assim foi criada a nova coluna chamada Tempo_envio que fornecerá em dias o tempo entre o pedido feito e o envio do produto.

Figura 23: Tempo de Envio

```
#Verificando em dias o tempo do envio do pedido e o envio do produto
superLoja["Tempo_envio"] = superLoja["Data_envio"] - superLoja["Data_pedido"]
superLoja.head()
```

| Seguimento | Cidade | Estado | Pais | Regiao | ID_produto | Categoria | SubCategoria | Nome_produto | Vendas | Quantidade | Desconto | Lucro | Respostas | Contagem_respostas | Porcentagem_respostas | Tempo_envio |
|-------------|--------|---------|-------------|--------|-----------------|------------------------|--------------|--------------------------------------------|--------|------------|----------|-------|-----------|--------------------|-----------------------|-------------|
| Corporativo | Leeds | England | Reino Unido | Norte | OFF-ST-10000988 | Material de Escritório | Armazenar | Fellowes Folders, Blue | 79.20 | 3 | 0.0 | 39.60 | Facebook | 548.0 | 0.205 | 4 days |
| Corporativo | Leeds | England | Reino Unido | Norte | TEC-AC-10004144 | Tecnologia | Acessórios | SanDisk Numeric Keypad, Bluetooth | 388.92 | 7 | 0.0 | 0.00 | Instagram | 916.0 | 0.342 | 4 days |
| Corporativo | Leeds | England | Reino Unido | Norte | OFF-LA-10001915 | Material de Escritório | Etiquetas | Avery Legal Exhibit Labels, 5000 Label Set | 35.19 | 3 | 0.0 | 16.11 | Snapchat | 86.0 | 0.032 | 4 days |
| Corporativo | Leeds | England | Reino Unido | Norte | OFF-ST-10004550 | Material de Escritório | Armazenar | Fellowes Folders, Wire Frame | 50.94 | 2 | 0.0 | 13.20 | Twitter | 179.0 | 0.067 | 4 days |
| Corporativo | Leeds | England | Reino Unido | Norte | TEC-AC-10004068 | Tecnologia | Acessórios | Memorex Memory Card, USB | 307.44 | 3 | 0.0 | 73.71 | None | 947.0 | 0.354 | 4 days |

Fonte: Autor

Sendo assim, com a coluna pronta é possível verificar o valor máximo, mínimo e a média de dias do envio do pedido.

Figura 24: Tempo envio max, min e média

```
#verificando a quantidade máxima, mínima e a média de tempo de envio em dias
print(superLoja["Tempo_envio"].max())
print(superLoja["Tempo_envio"].min())
print(superLoja["Tempo_envio"].mean())

8 days 00:00:00
-364 days +00:00:00
1 days 16:56:55.680000
```

Fonte: Autor

A mediana do tempo de envio está em 1 dia. Porém, também podemos ver que temos envios, no mesmo dia e até 8 dias depois do pedido. É possível verificar que temos um valor negativo o que indica que o tempo de envio tem ocorrido em menos de um dia, ou seja em horas é realizado o envio. Concluímos então que os dados das colunas de dados de pedido e dados de envio estão corretos e dentro da normalidade.

Mormente, é necessário verificar quantos valores unicos temos de ID_produto e de Nome_produto.

Figura 25: Verificando valores únicos

```
print(f'Quantidade de ID de produtos: {len(superLoja["ID_produto"].unique())}')
print(f'Quantidade de produtos únicos: {len(superLoja["Nome_produto"].unique())}')

Quantidade de ID de produtos: 1915
Quantidade de produtos únicos: 1856
```

Fonte: Autor

No entanto, é possível perceber que temos mais ID's de produtos do que nomes de produtos, quer dizer, que provavelmente temos mais de um ID para um produto com o mesmo nome. Sendo assim, talvez não seja um erro, pois podemos ter produtos com o mesmo nome e ID diferente por conta de características diferentes que não estão presentes no nosso DataFrame. Dessa maneira, o mesmo procedimento foi feito os ID's dos clientes.

Figura 26: Agrupando por Nome_produto e ID_produto

```
Agrupando_nome_prod=superLoja.groupby(superLoja["Nome_produto"], as_index=False)["ID_produto"].nunique()
Agrupando_nome_prod
```

| | Nome_produto | ID_produto |
|------|-----------------------------------|------------|
| 0 | Acco 3-Hole Punch, Clear | 1 |
| 1 | Acco 3-Hole Punch, Durable | 1 |
| 2 | Acco 3-Hole Punch, Economy | 1 |
| 3 | Acco 3-Hole Punch, Recycled | 1 |
| 4 | Acco Binder Covers, Clear | 1 |
| ... | ... | ... |
| 1851 | Xerox Note Cards, Recycled | 1 |
| 1852 | Xerox Parchment Paper, 8.5 x 11 | 1 |
| 1853 | Xerox Parchment Paper, Multicolor | 1 |
| 1854 | Xerox Parchment Paper, Premium | 1 |
| 1855 | Xerox Parchment Paper, Recycled | 1 |

1856 rows x 2 columns

Fonte: Autor

Logo, possuímos uma lista dos produtos com o mesmo nome e que tem mais de um ID. Assim sendo, verificou-se no DataFrame se o contrário também ocorre, ou seja, se temos nomes diferentes com ID's iguais, de acordo com a figura abaixo.

Figura 27: Agrupando por ID_producto e Nome_producto

```
Agupando_id_prod=superLoja.groupby(superLoja["ID_producto"], as_index=False)["Nome_producto"].nunique()
```

```
Agupando_id_prod
```

| | ID_producto | Nome_producto | |
|------|-----------------|---------------|--|
| 0 | FUR-BO-10000002 | 1 | |
| 1 | FUR-BO-10000022 | 1 | |
| 2 | FUR-BO-10000034 | 1 | |
| 3 | FUR-BO-10000120 | 1 | |
| 4 | FUR-BO-10000155 | 1 | |
| ... | ... | ... | |
| 1910 | TEC-PH-10004879 | 1 | |
| 1911 | TEC-PH-10004882 | 1 | |
| 1912 | TEC-PH-10004910 | 1 | |
| 1913 | TEC-PH-10004915 | 1 | |
| 1914 | TEC-PH-10004992 | 1 | |

1915 rows x 2 columns

Fonte: Autor

Nesse sentido, é possível verificar que o contrário também ocorre, com isso, para testar utilizaremos um produto de cada lista para analisarmos. Filtrando pelo ID do “OFF-BI-10002570” vemos que para o mesmo ID temos o produto “Acco Binder Covers, Clear” que são capas para fichários e o produto “Cardinal 3-Hole Punch, Clear” que é perfurador de 3 furos. Produtos completamente diferentes um do outro, mas que seguem o mesmo ID. Além disso, em questão de preço podemos ver que o primeiro produto custa 1 unidade 12.78 euros enquanto 3 unidades do segundo produto custam 84.58 euros.

Figura 28: Filtrando pelo ID do “OFF-BI-10002570”

```
superLoja.loc[superLoja["ID_producto"] == "OFF-BI-10002570"]
```

| | Cidade | Estado | Pais | Regiao | ID_producto | Categoria | SubCategoria | Nome_producto | Vendas | Quantidade | De |
|--|------------|---------------|-------------|---------|-----------------|------------------------|--------------|------------------------------|---------|------------|----|
| | Leicester | England | Reino Unido | Norte | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 63.900 | 5 | |
| | Saint-Malo | Brittany | França | Central | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 12.780 | 1 | |
| | Exeter | England | Reino Unido | Norte | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 25.560 | 2 | |
| | Watford | England | Reino Unido | Norte | OFF-BI-10002570 | Material de Escritório | Fichários | Cardinal 3-Hole Punch, Clear | 84.780 | 3 | |
| | Taranto | Apulia | Itália | Sul | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 63.900 | 5 | |
| | Oslo | Oslo | Noruega | Norte | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 12.780 | 1 | |
| | Sceaux | Ile-de-France | França | Central | OFF-BI-10002570 | Material de Escritório | Fichários | Acco Binder Covers, Clear | 12.780 | 1 | |
| | Turin | Piedmont | Itália | Sul | OFF-BI-10002570 | Material de Escritório | Fichários | Cardinal 3-Hole Punch, Clear | 84.780 | 3 | |
| | Berlin | Berlin | Alemanha | Central | OFF-BI-10002570 | Material de Escritório | Fichários | Cardinal 3-Hole Punch, Clear | 228.906 | 9 | |

Fonte: Autor

Por conseguinte, pelo nome do produto “Harbour Creations Removable Labels, Adjustable” que são etiquetas removíveis. Podemos ver que o produto de mesmo nome tem pelo menos dois ID's diferentes, isso ocorre pelo fato da variabilidade de cor ou modelo do produto.

Figura 29: Filtrando pelo nome do produto

| #Analisando pelo Nome do produto superLoja.loc[superLoja["Nome_produto"] == "Harbour Creations Removable Labels, Adjustable"] | | | | | | | | | | | | | | | | |
|----------------------------------------------------------------------------------------------------------------------------------|-------------|------------|------------|----------------|--------------|------------------|--------------------|----------------|------------------------|-------------|------------|-----------------|------------------------|--------------|------------------------------------------------|--------|
| ID_pedido | Data_pedido | Data_envio | Modo_envio | ID_cliente | Nome_cliente | Seguimento | Cidade | Estado | Pais | Regiao | ID_produto | Categoria | SubCategoria | Nome_produto | Vendas | Qu |
| 273 | 274 | 2021-12-25 | 2021-12-30 | Standard Class | KH-16690 | Kristen Hastings | Corporativo | Chilly-Mazarin | Ile-de-France | França | Central | OFF-LA-10003283 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 41.250 |
| 426 | 427 | 2022-08-21 | 2022-08-26 | Second Class | RB-19570 | Rob Beeghly | Consumidor | Rudge | England | Reino Unido | Norte | OFF-LA-10001549 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 53.700 |
| 3079 | 3080 | 2023-01-21 | 2023-01-25 | Standard Class | GM-14695 | Greg Maxwell | Corporativo | Letchworth | England | Reino Unido | Norte | OFF-LA-10003283 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 16.500 |
| 3677 | 4774 | 2023-03-17 | 2023-03-21 | Second Class | MA-17995 | Michelle Arnett | Escritório em casa | Duisburg | North Rhine-Westphalia | Alemanha | Central | OFF-LA-10003283 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 41.250 |
| 4866 | 4867 | 2021-06-07 | 2021-06-12 | Second Class | RP-19270 | Rachel Payne | Corporativo | Berlin | Berlin | Alemanha | Central | OFF-LA-10003283 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 37.125 |
| 4983 | 4984 | 2020-06-28 | 2020-07-03 | Standard Class | JM-15580 | Jill Matthias | Consumidor | Roenne | Auvergne-Rhône-Alpes | França | Central | OFF-LA-10003283 | Material de Escritório | Etiquetas | Harbour Creations Removable Labels, Adjustable | 16.500 |

Fonte: Autor

Contudo, para a análise vamos utilizar o Nome do produto.

4.2.1 Mesclando as Bases da Super Loja, de Ano, Mês e de Dias

Nesta etapa do trabalho será feito a transformação dos dados de forma a ter um uso mais fácil. Para tal tarefa foi usado a função split do pandas para separar as colunas Data_pedido e Data_envio em colunas com apenas o ano, mês e dia de cada uma delas para facilitar as análises.

Figura 30: Colunas Ano, Mês e Dia

```
# Extraíndo o ano, mês e o dia do pedido
superLoja["Mes_envio", superLoja["Dia_envio"]] = (superLoja["Data_envio"].dt.month, superLoja["Data_envio"].dt.day)
superLoja.sample(5)
```

| categoria | Nome_produto | Vendas | Quantidade | Desconto | Lucro | Respostas | Contagem_respostas | Porcentagem_respostas | Tempo_envio | Ano_pedido | Mes_pedido | Dia_pedido | Ano_envio | Mes_envio | Dia_envio |
|-----------|--------------------------------------------------|----------|------------|----------|---------|-----------|--------------------|-----------------------|-------------|------------|------------|------------|-----------|-----------|-----------|
| Cadeiras | Hon Executive Leather Armchair, Red | 826.902 | 2 | 0.1 | -91.878 | None | 35.013793 | 0.199313 | 5 days | 2022 | 8 | 15 | 2022 | 8 | 20 |
| elefones | Apple Headset, with Caller ID | 329.670 | 5 | 0.1 | 18.270 | None | 35.013793 | 0.199313 | 6 days | 2021 | 7 | 19 | 2021 | 7 | 25 |
| Art | Sanford Highlighters, Easy-Erase | 48.060 | 3 | 0.0 | 1.890 | None | 35.013793 | 0.199313 | 3 days | 2023 | 9 | 5 | 2023 | 9 | 8 |
| envelopes | GlobeWeis Clasp Envelope, with clear poly window | 13.440 | 1 | 0.0 | 4.950 | None | 35.013793 | 0.199313 | 4 days | 2020 | 9 | 29 | 2020 | 10 | 3 |
| láminas | StarTech Inkjet, White | 1494.600 | 5 | 0.0 | 657.600 | None | 35.013793 | 0.199313 | 4 days | 2021 | 10 | 31 | 2021 | 11 | 4 |

Fonte: Autor

4.2.2 A coluna Preço_unit_produto

A coluna preço unitário contém as informações do preço unitário de cada produto, visto que, no DataFrame superLoja, já possui uma coluna para o valor total do pedido com o desconto aplicado, a quantidade de produtos neste pedido, o desconto oferecido e o lucro sobre este pedido. Dessa forma vamos criar a coluna Preço Unitário Produto e esta coluna se dará pela fórmula:

Figura 31: Fórmula preço unitário

$$PU = \frac{(V / (1 - D))}{Q}$$

Fonte: Autor

Onde:

PU = Preço Unitário

V = Venda com desconto

D = Valor do desconto (em decimal)

Q = Quantidade do Pedido

Figura 32: Coluna Preço Unitário

```
#Criando coluna de preço em euro
superLoja["Preco_unit_prod"] = preco_unitario(superLoja["Vendas"], superLoja["Quantidade"], superLoja["Desconto"])
superLoja.head()
```

| Respostas | Contagem_respostas | Porcentagem_respostas | Receita | Tempo_envio | Ano_pedido | Mes_pedido | Dia_pedido | Ano_envio | Mes_envio | Dia_envio | diferenca_dias | trimestre_vendas | Preco_unit_prod |
|-----------|--------------------|-----------------------|---------|-------------|------------|------------|------------|-----------|-----------|-----------|----------------|------------------|-----------------|
| Facebook | 548.0 | 0.205 | 237.60 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 1 | 26.40 |
| Instagram | 916.0 | 0.342 | 2722.44 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 1 | 55.56 |
| Snapchat | 86.0 | 0.032 | 105.57 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 1 | 11.73 |
| Twitter | 179.0 | 0.067 | 101.88 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 1 | 25.47 |
| None | 947.0 | 0.354 | 922.32 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 1 | 102.48 |

Fonte: Autor

4.2.3 Coluna Desconto_euro

A coluna Desconto em euro, foi criada para também facilitar a visualização de valores.

Figura 33: Fórmula Desconto em Euro

$$DE = (PU \times Q) \times D$$

Fonte: Autor

Onde:

DE = Desconto em Euro

PU = Preço Unitário

Q = Quantidade

D = Desconto em decimais

Figura 34: Coluna Desconto em Euro

```
superLoja["Desconto_euro"] = desconto_euro(superLoja["Preco_unit_prod"], superLoja["Quantidade"],superLoja["Desconto"])
superLoja.sample(5)
```

| Id | Lucro | Respostas | Contagem_respostas | Porcentagem_respostas | Tempo_envio | Ano_pedido | Mes_pedido | Dia_pedido | Ano_envio | Mes_envio | Dia_envio | diferenca_dias | Preco_unit_prod | Desconto_euro |
|------|---------|-----------|--------------------|-----------------------|-------------|------------|------------|------------|-----------|-----------|-----------|----------------|-----------------|---------------|
| 1.15 | 321.552 | None | 35.013793 | 0.199313 | 4 days | 2023 | 1 | 31 | 2023 | 2 | 4 | 4 days | 309.24 | 371.088 |
| 1.10 | 24.114 | None | 35.013793 | 0.199313 | 4 days | 2020 | 6 | 15 | 2020 | 6 | 19 | 4 days | 134.13 | 26.826 |
| 1.00 | 38.880 | Instagram | 0.000000 | 0.000000 | 4 days | 2023 | 11 | 12 | 2023 | 11 | 16 | 4 days | 24.03 | 0.000 |
| 1.10 | 94.518 | None | 35.013793 | 0.199313 | 4 days | 2022 | 2 | 2 | 2022 | 2 | 6 | 4 days | 168.81 | 33.762 |
| 1.00 | 13.200 | None | 35.013793 | 0.199313 | 4 days | 2023 | 1 | 16 | 2023 | 1 | 20 | 4 days | 50.79 | 0.000 |

Fonte: Autor

4.2.4 Coluna Lucro_prod_unit

A coluna Lucro sobre valor unitário, também foi criada para ajudar a visualização de valores, onde apresenta o lucro por cada produto.

Figura 35: Coluna Desconto em Euro

$$\text{LPU} = \frac{(\text{DE} + \text{LAD})}{Q}$$

Fonte: Autor

Onde:

LPU = Lucro sobre produto unitário

DE = Desconto em Euro

LAD = Lucro após desconto

Q = Quantidade

Figura 36: Lucro após desconto

```
superLoja["Lucro_prod_unit"] = lucro_unitario(superLoja["Desconto_euro"], superLoja["Lucro"], superLoja["Quantidade"])
superLoja
```

| Porcentagem_respostas | Tempo_envio | Ano_pedido | Mes_pedido | Dia_pedido | Ano_envio | Mes_envio | Dia_envio | diferenca_dias | Preco_unit_prod | Desconto_euro | Lucro_prod_unit | Lucro_apos_desc |
|-----------------------|-------------|------------|------------|------------|-----------|-----------|-----------|----------------|-----------------|---------------|-----------------|-----------------|
| 0.205000 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 26.40 | 0.0 | 13.20 | 39.60 |
| 0.342000 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 55.56 | 0.0 | 0.00 | 0.00 |
| 0.032000 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 11.73 | 0.0 | 5.37 | 16.11 |
| 0.067000 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 25.47 | 0.0 | 6.60 | 13.20 |

Fonte: Autor

Para finalizar também foi criada a coluna de lucro por produto unitário em porcentagem, onde é feita a divisão dos valores da coluna Lucro_prod_unit pelos valores presentes na coluna Preco_unit_prod e multiplicar por 100, com isso resulta na figura 36, abaixo:

Figura 37: Lucro por produto unitário em (%)

```
#Coluna de lucro por produto unitário em porcentagem
superLoja["Lucro_prod_unit(%)"] = round(superLoja["Lucro_prod_unit"]/superLoja["Preco_unit_prod"],2)*100
superLoja.head()
```

| _respostas | Tempo_envio | Ano_pedido | Mes_pedido | Dia_pedido | Ano_envio | Mes_envio | Dia_envio | diferenca_dias | Preco_unit_prod | Desconto_euro | Lucro_prod_unit | Lucro_apos_desc | Lucro_prod_unit(%) |
|------------|-------------|------------|------------|------------|-----------|-----------|-----------|----------------|-----------------|---------------|-----------------|-----------------|--------------------|
| 0.205 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 26.40 | 0.0 | 13.20 | 39.60 | 50.0 |
| 0.342 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 55.56 | 0.0 | 0.00 | 0.00 | 0.0 |
| 0.032 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 11.73 | 0.0 | 5.37 | 16.11 | 46.0 |
| 0.067 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 25.47 | 0.0 | 6.60 | 13.20 | 26.0 |
| 0.354 | 4 days | 2023 | 2 | 7 | 2023 | 2 | 11 | 4 days | 102.48 | 0.0 | 24.57 | 73.71 | 24.0 |

Fonte: Autor

Com isso, obtemos o seguinte DataFrame com as colunas novas inseridas.

4.3 Explorações dos dados correlacionando colunas

Para encontrar informações relacionadas ao tema do estudo nos dados da base apresentada na figura 36, se faz necessário correlacionar suas colunas e perceber como uma variável se comporta com a variação da outra, logo essa seção se dedica a explorar os dados, criar visualizações e procurar responder as questões de negócio de uma maneira que seja possível a partir dessa análise descritiva para tomar decisões futuras.

4.3.1 Quais os produtos apresentam o maior número de vendas?

A primeira correlação feita foi da variável Nome_produto e a variável Quantidade, onde o objetivo foi concentrar na quantidade do produto vendido independente de lucro ou valor. Produtos de alta demanda e liquidez têm alta rotatividade de estoque e devem ter uma atenção especial quanto a isso.

Figura 38: Produtos mais vendidos em todos os anos

| TOP 10 PRODUTOS MAIS VENDIDOS | | |
|-------------------------------|---------------------------------------|------------|
| | Nome_produto | Quantidade |
| 1405 | Rogers File Cart, Single Width | 119 |
| 596 | Eldon File Cart, Single Width | 115 |
| 299 | Binney & Smith Sketch Pad, Blue | 97 |
| 1689 | Stanley Pencil Sharpener, Water Color | 96 |
| 1575 | Sanford Pencil Sharpener, Water Color | 93 |
| 1573 | Sanford Pencil Sharpener, Easy-Erase | 93 |
| 187 | Avery Index Tab, Clear | 90 |
| 1019 | Ibico Index Tab, Clear | 89 |
| 317 | Boston Pencil Sharpener, Water Color | 87 |
| 305 | Boston Canvas, Fluorescent | 86 |

Fonte: Autor

Dessa forma, temos os produtos mais vendidos em todos os anos registrados e que devem ter uma alta rotação de estoque. Para uma melhor visualização foi realizada a divisão para ver qual foi o produto mais vendido em cada ano e se a demanda foi se alterando.

Figura 39: Produtos mais vendidos no ano de 2020

| TOP 10 PRODUTOS MAIS VENDIDOS EM 2020 | | |
|---------------------------------------|---------------------------------------|------------|
| | Nome_produto | Quantidade |
| 106 | Avery Binding Machine, Durable | 31 |
| 728 | Rogers Shelving, Industrial | 29 |
| 188 | Boston Canvas, Fluorescent | 29 |
| 270 | Cardinal Binder, Clear | 28 |
| 137 | BIC Markers, Water Color | 28 |
| 164 | Binney & Smith Canvas, Blue | 27 |
| 198 | Boston Pencil Sharpener, Water Color | 25 |
| 133 | BIC Highlighters, Water Color | 24 |
| 182 | Binney & Smith Sketch Pad, Blue | 23 |
| 881 | Stanley Pencil Sharpener, Water Color | 23 |

Fonte: Autor

Figura 40: Produtos mais vendidos no ano de 2021

| TOP 10 PRODUTOS MAIS VENDIDOS EM 2021 | | |
|---------------------------------------|----------------------------------------|------------|
| | Nome_produto | Quantidade |
| 973 | Sanford Pencil Sharpener, Easy-Erase | 38 |
| 382 | Eldon File Cart, Single Width | 36 |
| 192 | Binney & Smith Sketch Pad, Water Color | 33 |
| 205 | Boston Pencil Sharpener, Easy-Erase | 32 |
| 138 | BIC Canvas, Blue | 31 |
| 449 | Fellowes Box, Industrial | 30 |
| 1118 | Stockwell Staples, 12 Pack | 30 |
| 975 | Sanford Pencil Sharpener, Water Color | 29 |
| 282 | Cardinal 3-Hole Punch, Economy | 29 |
| 281 | Cardinal 3-Hole Punch, Durable | 29 |

Fonte: Autor

Figura 41: Produtos mais vendidos no ano de 2022

| TOP 10 PRODUTOS MAIS VENDIDOS EM 2022 | | |
|---------------------------------------|------------------------------------------|------------|
| | Nome_produto | Quantidade |
| 968 | Rogers File Cart, Single Width | 46 |
| 156 | BIC Markers, Fluorescent | 46 |
| 168 | BIC Sketch Pad, Water Color | 40 |
| 135 | Avery Index Tab, Clear | 39 |
| 1134 | Smead Lockers, Blue | 38 |
| 1076 | Sanford Pencil Sharpener, Water Color | 33 |
| 160 | BIC Pencil Sharpener, Fluorescent | 32 |
| 16 | Acco Hole Reinforcements, Durable | 31 |
| 206 | Binney & Smith Highlighters, Fluorescent | 31 |
| 218 | Binney & Smith Sketch Pad, Blue | 30 |

Fonte: Autor

Figura 42: Produtos mais vendidos no ano de 2023

| TOP 10 PRODUTOS MAIS VENDIDOS EM 2023 | | |
|---------------------------------------|---------------------------------------|------------|
| | Nome_produto | Quantidade |
| 1318 | Smead Shelving, Blue | 49 |
| 254 | Boston Pencil Sharpener, Water Color | 44 |
| 1383 | Stockwell Paper Clips, Assorted Sizes | 41 |
| 473 | Eldon File Cart, Single Width | 41 |
| 1339 | Stanley Markers, Water Color | 39 |
| 1396 | Stockwell Staples, Metal | 38 |
| 134 | Avery Binder, Clear | 36 |
| 4 | Acco Binder Covers, Clear | 36 |
| 1130 | Rubbermaid Frame, Duo Pack | 36 |
| 1238 | Sanford Markers, Fluorescent | 35 |

Fonte: Autor

Conforme, as informações das tabelas acima têm os produtos com alta demanda. Estes produtos são mais sujeitos a não processar parados em estoque. São produtos de alta liquidez e que darão um rápido retorno a um investimento, deste modo é uma posição que tem alto capital de giro.

4.3.1 Quais os produtos apresentam o menor número de vendas?

A seguinte correlação feita foi da variável Nome_produto e a variável Quantidade, onde o objetivo foi concentrar na quantidade do produto vendido de acordo com o seu lucro ou valor. Produtos com baixa participação de mercado e baixas taxas de crescimento, possuem baixa rotatividade de estoque e são produtos que geram pouco lucro ou mesmo prejuízos para a empresa.

Figura 43: Produtos menos vendidos em todos os anos

```
#Produtos com menos vendas
Menos_vendidos_por_ano = superLoja.groupby(superLoja['Nome_produto'], as_index = False)[['Quantidade']].sum().sort_values(by = 'Quantidade', ascending=True)
Menos_vendidos_por_ano[:30].style.background_gradient(cmap='Greys', subset=['Quantidade']).set_properties(**{'text-align': 'center'}, subset=['Quantidade']).set_properties(**{'text-align': 'left'}, subset=['Nome_produto'])
```

| TOP 10 PRODUTOS COM MENOS VENDAS | |
|--------------------------------------------------------|------------|
| Nome_produto | Quantidade |
| 123 Ames Interoffice Envelope, Set of 50 | 1 |
| 775 GlobeWeis Interoffice Envelope, Security-Tint | 1 |
| 1166 Lesro Training Table, Adjustable Height | 1 |
| 1060 Jiffy Manila Envelope, Recycled | 1 |
| 1383 Panasonic Card Printer, Wireless | 1 |
| 838 Hamilton Beach Blender, White | 1 |
| 546 Deflect-O Frame, Duo Pack | 1 |
| 499 Cuisinart Microwave, Black | 1 |
| 340 Breville Stove, Red | 1 |
| 204 Avery Shipping Labels, Alphabetical | 1 |
| 1546 SanDisk Mouse, USB | 1 |
| 48 Acme Box Cutter, Steel | 1 |
| 1063 Jiffy Interoffice Envelope, Security-Tint | 1 |
| 235 Barricks Conference Table, Rectangular | 1 |
| 205 Avery Shipping Labels, Laser Printer Compatible | 1 |
| 1845 Xerox Message Books, Multicolor | 1 |
| 326 Breville Blender, Black | 1 |
| 978 Hoover Blender, Silver | 2 |
| 751 Fiskars Ruler, High Speed | 2 |
| 875 Harbour Creations File Folder Labels, Alphabetical | 2 |
| 844 Hamilton Beach Microwave, Silver | 2 |
| 841 Hamilton Beach Coffee Grinder, Silver | 2 |

Fonte: Autor

Como de se esperar, são várias a quantidade de ocorrências de produtos com uma ou duas vendas, além de muitos outros com apenas três vendas. Com a análise do Top dez não é possível visualizar uma informação mais precisa, com isso podemos então fazer uma análise de quais são os produtos que tiveram a última venda mais antiga e desde então estão sem vender.

Figura 44: Produtos com data de venda mais antiga

```
Menos_vendidos_por_date = Menos_vendidos_por_date.groupby(Menos_vendidos_por_date["Nome_produto"]).max().sort_values(by= "Data_pedido", ascending = True)
```

```
Menos_vendidos_por_date[:20].style.set_properties(**{'text-align': 'center'},subset=['Data_pedido']).set_table_styles([dict(selector='th',props=[('text-align', 'left')])])
```

| | Data_pedido |
|----------------------------------------------|---------------------|
| Nome_produto | |
| Bevis Training Table, with Bottom Storage | 2020-01-11 00:00:00 |
| Barricks Coffee Table, Fully Assembled | 2020-02-15 00:00:00 |
| Brother Copy Machine, High-Speed | 2020-02-24 00:00:00 |
| Sauder Classic Bookcase, Pine | 2020-04-01 00:00:00 |
| Barricks Training Table, with Bottom Storage | 2020-04-07 00:00:00 |
| OIC Rubber Bands, 12 Pack | 2020-05-12 00:00:00 |
| Bevis Coffee Table, with Bottom Storage | 2020-05-28 00:00:00 |
| Hon Computer Table, Adjustable Height | 2020-05-31 00:00:00 |
| Green Bar Cards & Envelopes, Premium | 2020-06-08 00:00:00 |
| Safco Library with Doors, Pine | 2020-06-27 00:00:00 |
| KitchenAid Refrigerator, Silver | 2020-07-10 00:00:00 |
| Eldon Stacking Tray, Black | 2020-07-13 00:00:00 |
| Ames Interoffice Envelope, Set of 50 | 2020-07-25 00:00:00 |

Fonte: Autor

Dessa maneira, podemos analisar os vinte produtos com vendas mais antigas, com isso entende-se que esses produtos que quase não foram vendidos, não precisam ser repostos. Caso ainda haja itens no estoque é possível fazer uma promoção para tentar acabar com a mercadoria parada antes que fiquem obsoletos ou danificados. Como o DataFrame não possui informações sobre estoque de produtos essa análise fica um pouco prejudicada. Em uma situação real a melhor solução seria bater essas informações com informações de estoque.

4.3.2 Elasticidade de preço

Figura 45: Margem de lucro sobre os produtos

```
lista_desc_lucro = Margem_lucro.groupby(Margem_lucro["Nome_produto"],as_index = False)\
[["Preco_unit_prod", "Lucro_prod_unit(%)", "Lucro_prod_unit"]].mean().round(2)
```

```
lista_desc_lucro
```

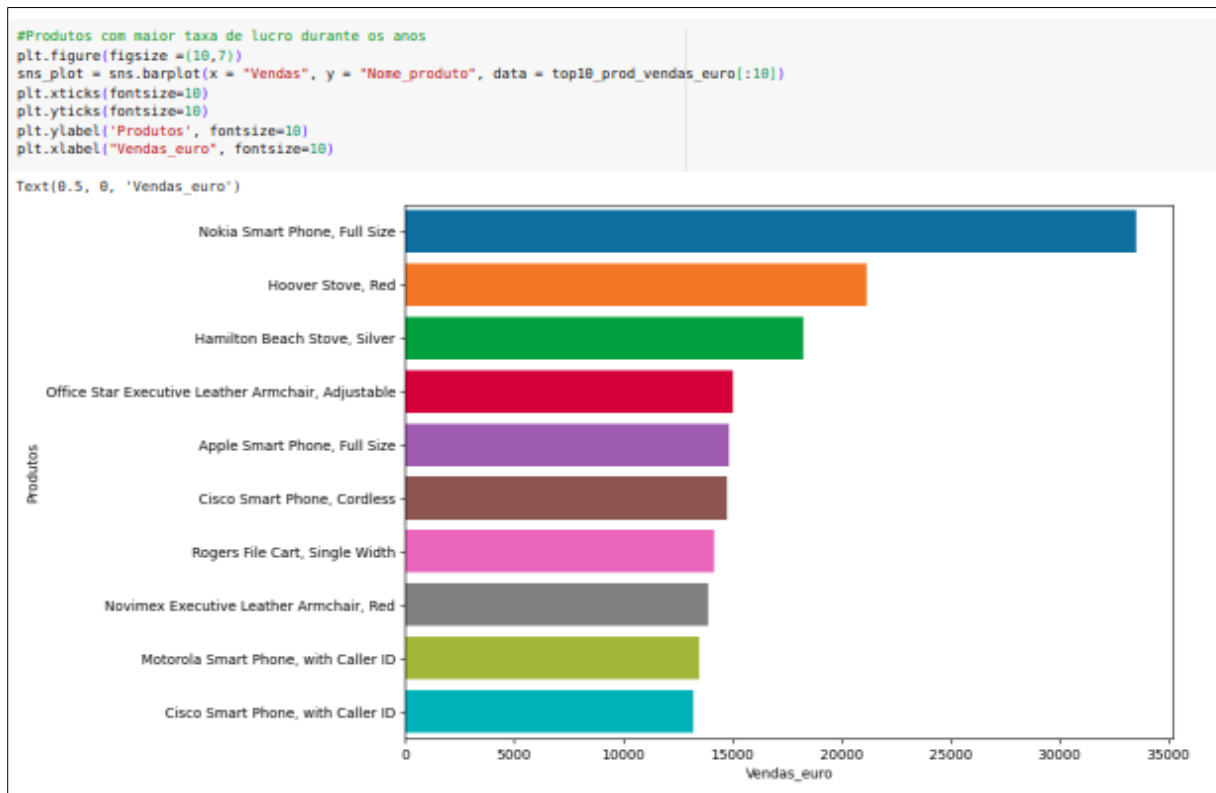
| | Nome_produto | Preco_unit_prod | Lucro_prod_unit(%) | Lucro_prod_unit |
|------|-----------------------------------|-----------------|--------------------|-----------------|
| 0 | Acco 3-Hole Punch, Clear | 29.58 | 17.0 | 5.01 |
| 1 | Acco 3-Hole Punch, Durable | 31.62 | 41.0 | 12.96 |
| 2 | Acco 3-Hole Punch, Economy | 30.12 | 19.0 | 5.70 |
| 3 | Acco 3-Hole Punch, Recycled | 30.48 | 9.0 | 2.73 |
| 4 | Acco Binder Covers, Clear | 12.78 | 43.0 | 5.49 |
| ... | ... | ... | ... | ... |
| 1851 | Xerox Note Cards, Recycled | 26.28 | 20.0 | 5.25 |
| 1852 | Xerox Parchment Paper, 8.5 x 11 | 17.40 | 13.0 | 2.25 |
| 1853 | Xerox Parchment Paper, Multicolor | 14.67 | 38.0 | 5.55 |
| 1854 | Xerox Parchment Paper, Premium | 13.26 | 32.0 | 4.23 |
| 1855 | Xerox Parchment Paper, Recycled | 13.17 | 45.0 | 5.91 |

1856 rows x 4 columns

Fonte: Autor

Conforme a tabela acima, podemos perceber a lista de produtos com seu Lucro por produto unitário, assim sendo, a porcentagem de lucro é bem relevante se comparado a preço do produto. Na afigura abaixo, podemos visualizar os produtos com maior taxa de lucro durante os anos.

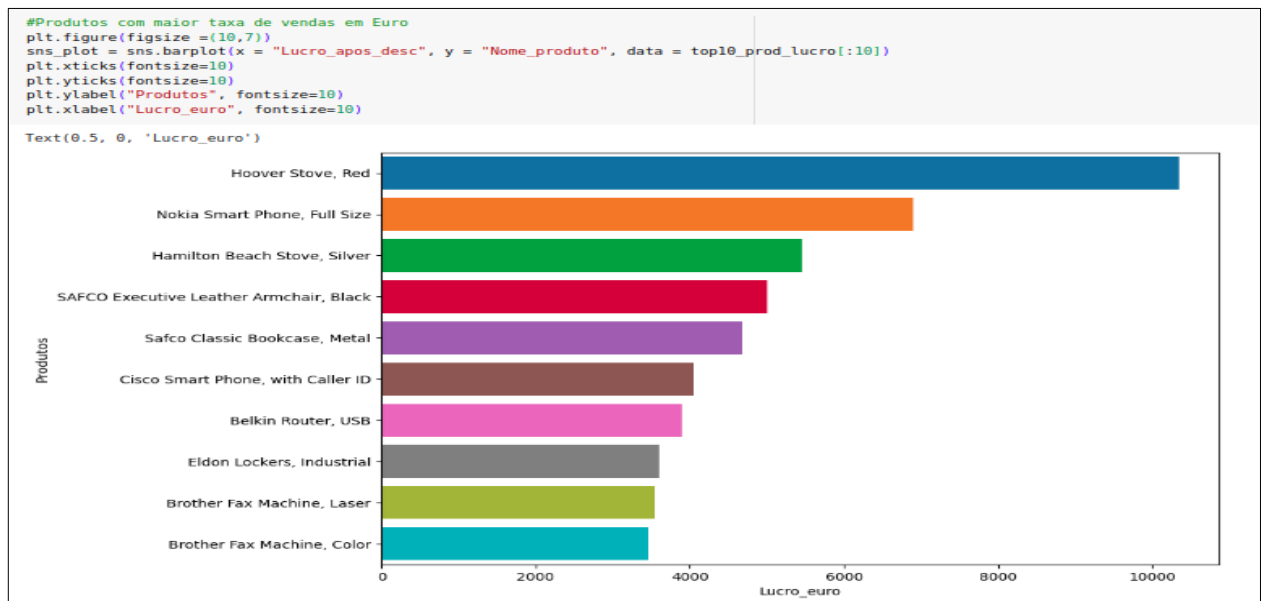
Figura 46: Produtos com maior taxa de lucro



Fonte: Autor

Ao visualizar a o Gráfico, temos que o produto “Nokia Smart Phone, Full Size” é o que possui maior taxa de lucro durante o ano, em seguida do “Hoover Stove, Red”.

Figura 47: Produtos com maior taxa de vendas



Fonte: Autor

Conforme, os gráficos podemos concluir que nem sempre o produto que nos dá o maior valor bruto em vendas trás o maior lucro.

4.4 Insights sobre compradores

Nesta seção analisaremos o padrão de clientes, onde é considerado um atributo importante para nossa análise, pois identifica os maiores compradores. Para melhor análise dos dados e aplicarmos os gráficos, consideramos o top 10 de clientes por valor de compras bruto. Em vermelho temos o valor total comprado pelo cliente e em roxo temos o lucro gerado por essas compras. Apesar do cliente Harry Greence ter comprado pouco mais de 2 mil euros em produtos ele não dá lucro para a loja, pelo contrário, dá prejuízo.

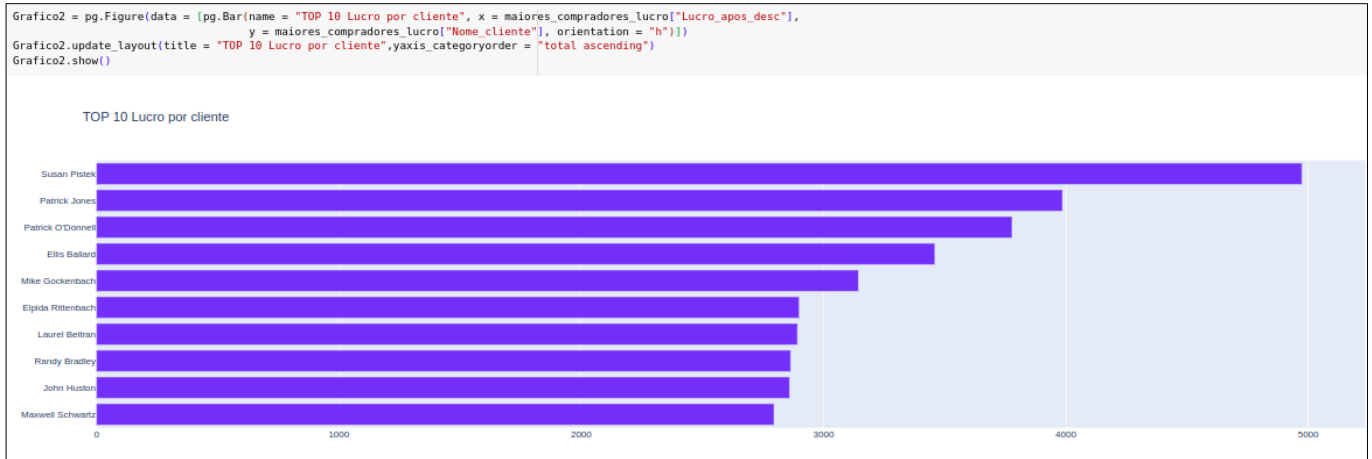
Figura 48: Top 10 de clientes por valor de compras bruto



Fonte: Autor

Agrupamos os registros dos 10 clientes que mais geraram lucro para a Super Loja. Pois desta forma conseguimos visualizar melhor os resultados, uma vez que na nossa base consta mais de 30 clientes.

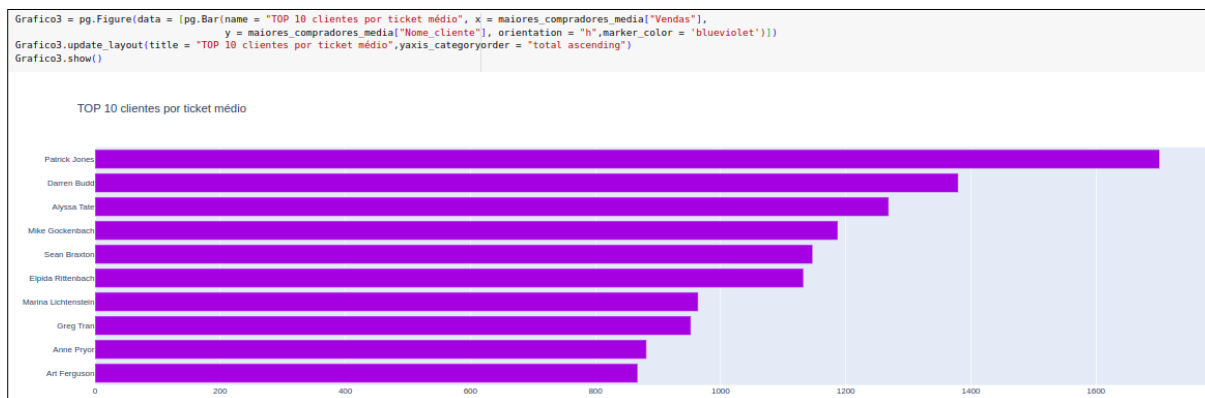
Figura 49: Top 10 Lucro por cliente



Fonte: Autor

Agrupamos também os registros dos 10 clientes por ticket médio, para visualizar o valor médio de vendas por cliente.

Figura 50: Ticket Médio



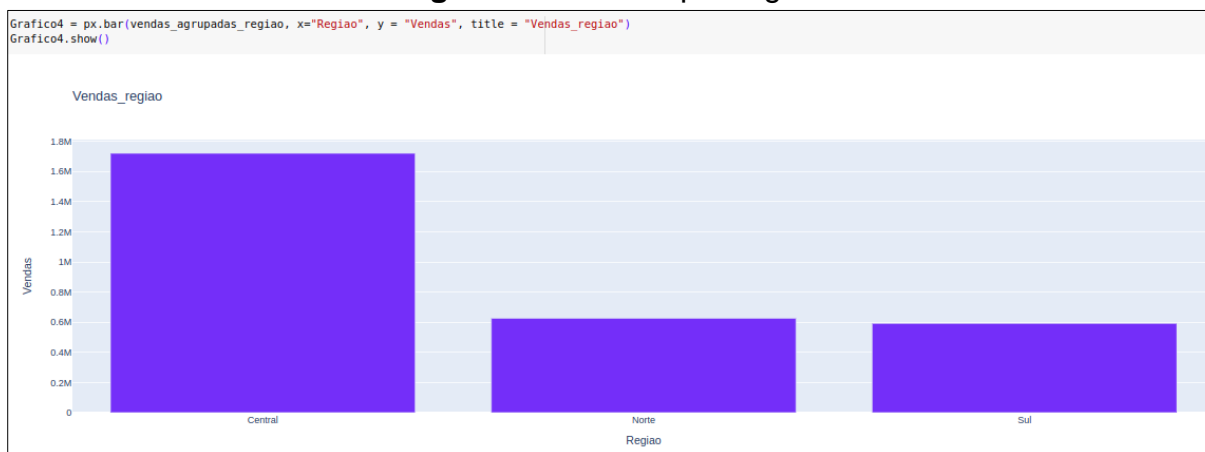
Fonte: Autor

Portanto, podemos verificar segundo o gráfico acima, que o cliente Patrick Jones é o nosso melhor cliente.

4.5 Quais regiões possuem as maiores e menores taxas de vendas?

Para melhorar nossa análise, construímos também um gráfico do agrupamento mostrando apenas as vendas por região e outro segmentando por Estado.

Figura 51: Vendas por região



Fonte: Autor

Figura 52: Vendas por região e Estado



No entanto, a realização dessa análise é de suma importância, para sabermos qual área concentra a maioria de nossas vendas e a região e estados que estão defasados. Com essa informação é possível trabalhar com possibilidade de campanhas de marketing e promoções para áreas deficitárias com objetivo de atrair novos clientes e aumentar a presença da empresa nessas áreas. Também é possível investir em análises para saber por que as regiões com maiores vendas estão nessa posição e tentar replicar características de venda e marketing dessas áreas para áreas de menor expressividade .

4.6 Sazonalidades

Para darmos continuidade na análise dos dados, também vimos a necessidade de verificar quais os melhores e piores meses de venda no ano. Para responder esta pergunta plotamos o gráfico abaixo. Este gráfico mostra a soma de todos os anos de dados e o faturamento por mês. Como podemos ver o mês de Fevereiro e Março são os mais fracos do ano. Após isso há uma grande alta em Junho e em seguida no mês de julho uma baixa brusca e uma taxa de faturamento sem muita variabilidade até o mês de Agosto onde as vendas aumentam muito. Por esse gráfico é possível ver que entre o mês 6 e 12 são os meses de maior faturamento. É perceptível que algo ocorre nesses meses para gerar essa alta, cisto que possível tomar decisões a partir desta informação para diversas estratégias.

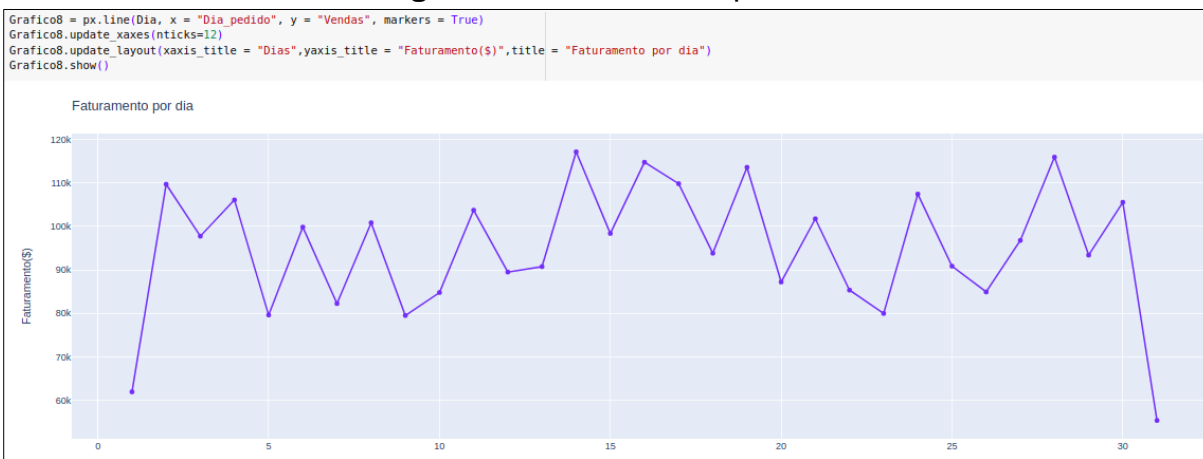
Figura 53: Faturamento por Mês



Fonte: Autor

Para avaliarmos melhor os dados obtidos do agrupamento da seção anterior, criamos um gráfico mostrando quais os melhores e piores dias de venda no mês.

Figura 54: Faturamento por Dia

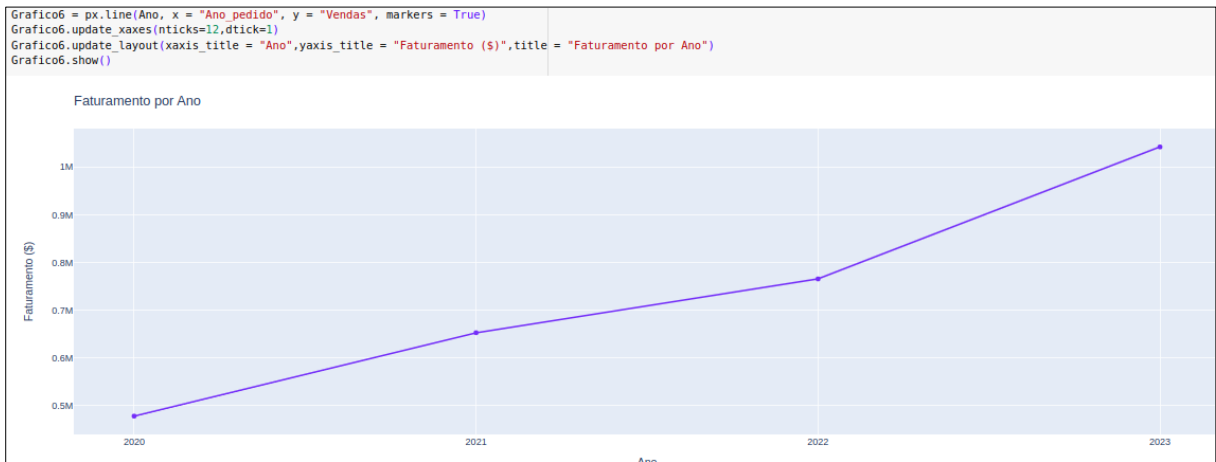


Fonte: Autor

Este gráfico permite uma tomadas de decisão em escalas menores como indicar bons dias para promoção ou melhores dias para enviar sugestões de compra para os clientes por e-mail, mensagens de celular e etc.

Para visualizarmos o faturamento anual e suas tendências, com o gráfico de faturamento por ano temos uma visão de como a empresa performou nos anos anteriores, mas também podemos ver uma tendência futura.

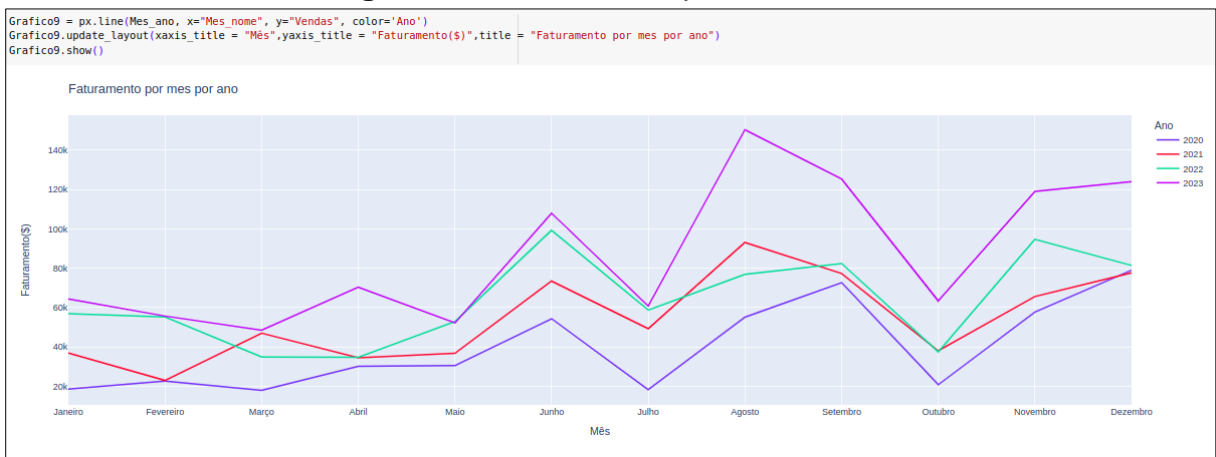
Figura 55: Faturamento por Ano



Fonte: Autor

Este Concluímos então que segundo o gráfico acima a tendência é de subida para os próximos anos. Agora ao invés de mostrar a somatória dos anos podemos ter uma visão mais comparativa entre a performance de cada mês de cada ano analisado.

Figura 56: Faturamento por Mês e Ano



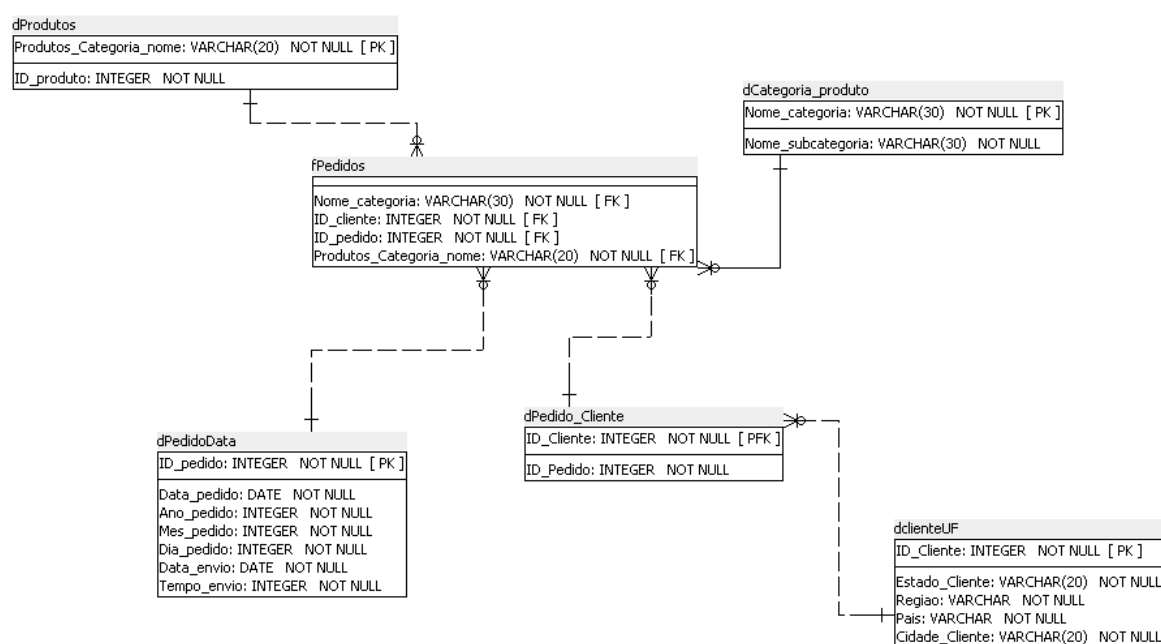
Fonte: Autor

5. Modelo de Dados

5.1. Modelo Dimensional

Para este trabalho, o modelo dimensional utilizado foi o *star schema* (uma tabela fato ligada a várias tabelas dimensão).

Figura 57: Diagrama dimensional, elaborado com a ferramenta SQL Power Architect



Fonte: Autor

5.2. Fatos e Dimensões

Para este trabalho foram utilizadas as seguintes tabelas:

Tabela 1 - Descrição das tabelas fato e dimensão.

| Tabela | Descrição | Tipo |
|----------------|-----------------------------------------------------------------------------|----------|
| dPedidoCliente | Contém o id do pedido (PK) e o id do cliente. | Dimensão |
| dClienteUF | Contém o id do cliente (PK) e a sigla do seu estado, cidade, país e região. | Dimensão |

| | | |
|-------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| dPedidoData | Contém o id do pedido (PK), data do pedido, data de envio, mês, dia e ano. Também contém as colunas calculadas, criadas diretamente no Power BI: tempo_envio. | Dimensão |
| dProduto | Contém o id do produto (PK) e sua categoria. | Dimensão |
| dCategoriaProduto | Contém o id da categoria (PK) e a subcategoria. | Dimensão |
| fPedido | Contém o id do pedido (FK), do produto e da categoria. | Fato |

5.3. Dashboard

5.3.1 Painel Estratégico

O painel estratégico apresenta uma visão gerencial dos dados. Na concepção de Chiavenato (2014), o planejamento estratégico procura especificar como fazer para alcançar os objetivos empresariais, visto que, o planejamento estratégico é um processo amplo, abrange toda a organização e está orientado para o futuro.

Nesse painel é possível filtrar a visualização por anos considerando a data do pedido, onde o filtro localiza-se no canto superior direito.

Logo em seguida temos a tabela que destaca os números de faturamentos, o percentual de vendas e o gráfico de barras na própria tabela que mostra a representatividade do maior e menor número de vendas. Ainda na tabela, temos um gráfico de linhas variando pelo período, que mostra quais os meses que venderam mais e menos ou seja a sazonalidade.

Temos também, o gráfico de funil que exibe o seguimento que teve o maior número de vendas até os menores números de vendas. Logo depois temos três cards:

- Faturamento: trata-se da soma do valor de todas as vendas;
- Quantidade: é o total da quantidade de pedidos;
- Ticket médio: valor médio das vendas, esse item é visto no painel tático sob perspectiva mensal;

$$TicketMédio = \frac{\sum Faturamento}{Quantidade de vendas}$$

- Vendas por ano: mostra o desempenho das vendas por ano.
- Faturamento por mês: O Gráfico de barras quais os melhores e piores meses de vendas.

Figura 57: Painel Estratégico



Fonte: Autor

5.3.2 Painel Tático

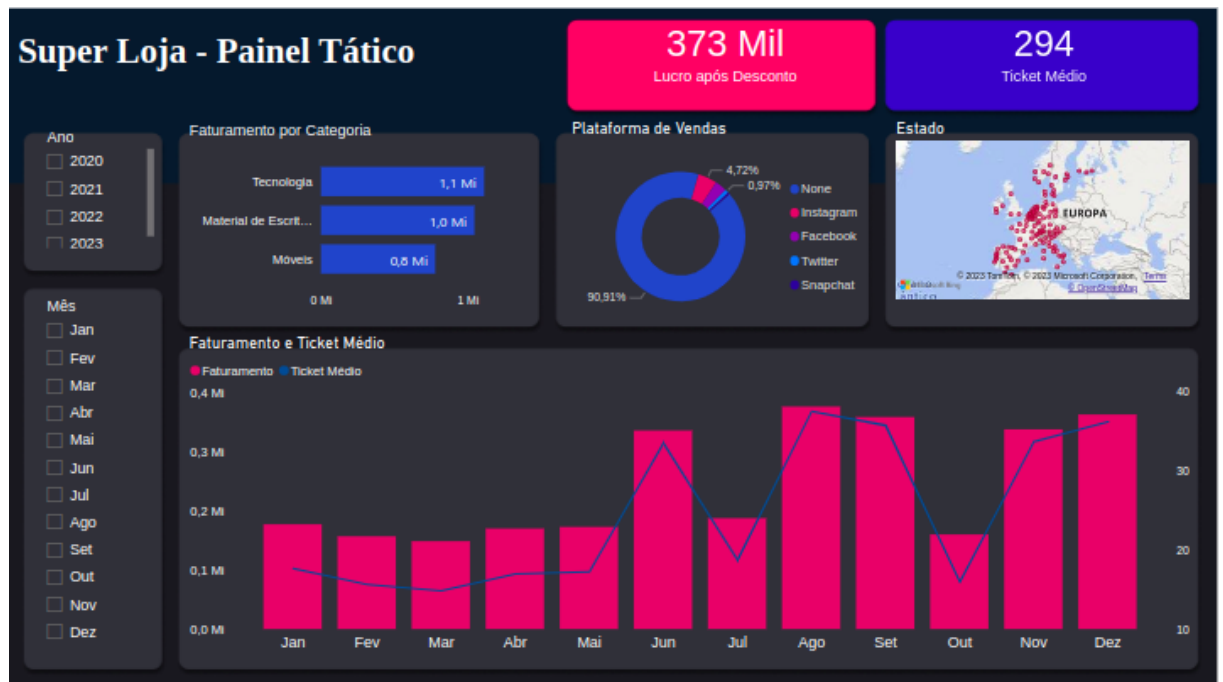
O painel tático apresenta na porção superior o total de lucro que obtivemos após o desconto e ao lado um card médio, assim temos uma visão de se o aumento ou a diminuição do lucro acompanha o aumento ou a diminuição do ticket médio.

Em seguida temos a filtragem por ano, ao lado é apresentada outra visão do faturamento, desta vez por categoria. Visualiza-se assim que, ao longo dos anos, algumas categorias aumentam seu faturamento.

Logo, temos e um gráfico de rosca apresenta o percentual de respostas de qual a plataforma mais utilizada para as compras online. O gráfico de mapa os estados que mais vendem.

Finalmente, observando-se o gráfico de colunas e linha que mostra o Faturamento e o ticket médio, nota-se que os meses de menor faturamento são março e outubro, em oposição, temos agosto e dezembro com meses de maior ticket médio.

Figura 58: Painel Tático



Fonte: Autor

5.3.3 Painel Operacional

O painel operacional mostra os dados focados no curto prazo. O ideal é que o responsável por enviar ou gerenciar quem despacha possa ter essas informações com maior facilidade.

Nesse painel é possível filtrar a visualização por anos considerando a data do pedido, onde o filtro localiza-se no canto superior direito.

Na posição inicial temos um gráfico de área empilhada, onde apresenta o faturamento por ano, região e categoria.

O gráfico de Gauge (velocímetro) apresenta o tempo médio para envio, em dias. Isto é, o tempo decorrido entre a compra e a entrega do pedido para a transportadora.

$$Tempoparaenvio = datadoenvio - adatadopedido$$

Ao lado, há um filtro com todas as regiões de vendas.

Na parte inferior, é apresentada outra visão do faturamento, desta vez por produtos. Visualiza-se assim que, ao longo dos anos, alguns produtos aumentam seu faturamento enquanto outras diminuíram. Top 10 mais vendido, dessa forma, temos os produtos mais vendidos em todos os anos registrados e que devem ter uma alta rotação de estoque.

E para finalizar mostramos também o Top 10 de produtos menos vendidos. Dessa maneira, podemos analisar os vinte produtos com vendas mais antigas, com isso entende-se que esses produtos que quase não foram vendidos, não precisam ser repostos. Caso ainda hajam itens no estoque é possível fazer uma promoção para tentar acabar com a mercadoria parada antes que fiquem obsoletos ou danificados.

Figura 59: Painel Operacional



Fonte: Autor

5.4 Modelo Canvas

| Problema | Resultados e Previsões | Aquisição de Dados |
|------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Analisar o Dataset de vendas, visando verificar o quão bem ou mal as vendas estão indo na Super Loja.</p> | <p>Gerar insights que podem influenciar em estratégias de vendas.</p> <p>Verificar quais são os produtos que mais vendem?</p> <p>Verificar quais os produtos que menos vendem?</p> <p>Verificar quais os maiores compradores?</p> <p>Analisar em quais regiões se concentram as maiores e menores taxas de vendas?</p> <p>Apresentar quais os melhores e piores meses de venda no ano.</p> | <p>Os dados de ambos os Datasets EU Superstore.xls e WhatsgoodlyData-6.csv foram coletados do site data.world</p> |
| Modelagem | Avaliação do Modelo | Preparação dos Dados |
| <p>Realizado análise nos datasets coletados, tanto de forma gráfica quanto análise descritiva dos dados utilizando a biblioteca do Pandas.</p> | <p>Para a avaliação dos resultados obtidos no modelo de vendas, foi avaliado o gráfico de vendas mensais, vendas anuais, produtos mais vendidos, ticket médio, conforme o Notebook em Python no diretório desse projeto.</p> | <p>Após a união dos datasets, os dados foram tratados, as colunas foram renomeadas, os dados duplicados foram removidos, e os dados desnecessários para análise também foram removidos.</p> |

6. Links

Todos os códigos desenvolvidos e a documentação utilizada são disponibilizados no repositório do Github.

<https://github.com/mariliafigueiredo/Trabalho-Puc---BI>

REFERÊNCIAS

CHIAVENATO, Idalberto. **Administração**: Teoria, processo e prática. 5. ed. – São Paulo: Manole, 2014.

NAKAMURA, André. **Comércio Eletrônico** riscos nas compras pela internet. 2011. 56 f. Dissertação (Tecnólogo em Processamento de Dados) Faculdade de Tecnologia São Paulo, São Paulo, 2011.

SILVA, Leandro. **Aumente suas Vendas com E-commerce**. Rio de Janeiro: Editora Ciência Moderna Ltda, 2009.

TRANSFORMAÇÃO DIGITAL. **Quais as vantagens do big data em vendas?** Disponível em <http://www.mma.gov.br/sitio/index.php?ido=conteudo.monta&idEstrutura=18>. Acesso em: 21/0/2023.