



**Universidade de Brasília
Faculdade de Tecnologia**

**MODELO PARA OTIMIZAR A ESTRATÉGIA DE
VENDAS DE COMPANHIA ONLINE: FRETE
GRÁTIS**

Natália Gleice Amorim Lemos

**TRABALHO DE GRADUAÇÃO
ENGENHARIA DE PRODUÇÃO**

Brasília
2023

**Universidade de Brasília
Faculdade de Tecnologia**

**MODELO PARA OTIMIZAR A ESTRATÉGIA DE
VENDAS DE COMPANHIA ONLINE: FRETE
GRÁTIS**

Natália Gleice Amorim Lemos

Trabalho de Graduação submetido como
requisito parcial para obtenção do grau de
Bacharel em Engenharia de Produção.

Orientador: Prof. Ph.D. Reinaldo Crispiniano Garcia

Brasília

2023

Lemos, Natália.

MODELO PARA OTIMIZAR A ESTRATÉGIA DE VENDAS DE
COMPANHIA ONLINE: FRETE GRÁTIS / Natália Gleice Amorim

Lemos; orientador Reinaldo Crispiniano Garcia. -- Brasília, 2023.

68 p.

Trabalho de Graduação em Engenharia de Produção --

Universidade de Brasília, 2023.

1.Otimização. 2. E-commerce. 3.Frete Grátis. 4.Data mining. I.
Garcia, Reinaldo Crispiniano II. ELABORAÇÃO DE MODELO PARA
OTIMIZAR A RETORNO DE VENDAS DE COMPANHIA ONLINE

**Universidade de Brasília
Faculdade de Tecnologia**

**MODELO PARA OTIMIZAR A ESTRATÉGIA DE
VENDAS DE COMPANHIA ONLINE: FRETE
GRÁTIS**

Natália Gleice Amorim Lemos

Trabalho de Graduação submetido como
requisito parcial para obtenção do grau de
Bacharel em Engenharia de Produção.

Brasília, 15 de Dezembro de 2023:

Prof. Ph.D. Reinaldo Crispiniano Garcia,
UnB/FT/EPR
Orientador

Prof. Ph.D. João Mello da Silva,
UnB/FT/EPR
Examinador interno

Brasília 2023

Agradecimentos

À minha mãe, Joice Amorim, a pessoa que mais se sacrificou pela minha educação, abrindo caminhos que ela mesma não teve a chance de trilhar. Mesmo distante fisicamente durante minha graduação, sua presença constante e apoio incondicional foram meu alicerce em todas as adversidades e alegrias desses 6 anos e meio.

Aos que foram essenciais para minha jornada até aqui, cuja presença agora vive somente em meu coração. Meu grande mentor e exemplo, Jeová Amorim, meu avô, sempre foi o primeiro a quem compartilhava as boas notícias, e sua constante motivação foi o impulso para concluir minha graduação e construir uma carreira sólida. Ao meu pai, Renato Lemos, cujo suporte e presença, foram fundamentais em minha trajetória, sem o que fez por mim eu jamais estaria onde estou hoje. Vô, Pai, eu consegui! Nós conseguimos!

À minha rede de apoio, que em tantos momentos foi meu suporte durante esta jornada, Jonatas Bueno e Desterro Marinho. Assim como meus tios, Ramon Lemos e Ronaldo Lemos, que sempre acreditaram em meu potencial e investiram em mim.

A meu companheiro, Guilherme Souza, presente em todos os momentos como meu porto seguro e maior admirador. Crescemos e prosperamos juntos ao longo desses anos, e guardarei para sempre na memória como a UnB nos uniu e nos fez evoluir.

À Universidade de Brasília, por proporcionar um ensino completo e diversificado. A UnB sempre foi meu objetivo, meu sonho como candanga. Sair daqui é motivo de orgulho, pois levo comigo bagagens tanto profissionais quanto pessoais. Sem essa experiência, eu não seria quem sou hoje.

E em especial, ao Reinaldo Crispiniano. Desde minha transição para Engenharia de Produção, você foi o professor que mais admirei. Suas palavras sempre nos incentivando a buscar o melhor, a nos destacar e a almejar mais na vida profissional. A faísca que acendeu com suas provocativas aulas foi crucial para meu desenvolvimento. Se alcancei este ponto, foi com seu apoio e estímulo.

A todos vocês, meu mais sincero e profundo, muito obrigada!

Resumo

A análise do impacto do frete grátis nas vendas do comércio eletrônico é de importância crucial no contexto atual do mercado. Este estudo investigou essa estratégia na plataforma Olist, focando na região Nordeste do Brasil, e buscou integrar teoria, metodologia e resultados para oferecer *insights* valiosos. A análise dos dados deste trabalho forneceu uma compreensão aprofundada do comportamento das vendas e identificou tendências sazonais e padrões de demanda, respaldando a eficácia da estratégia de frete grátis. Os resultados obtidos validaram o potencial do frete grátis como uma ferramenta estratégica para impulsionar as vendas no cenário do comércio eletrônico brasileiro. Essa análise abrangente e integrada contribui para o aprimoramento das estratégias de vendas e fornece às empresas ferramentas valiosas para maximizar seus retornos financeiros. Através da exploração de um referencial teórico que abordou a relação entre o frete grátis e estratégias de marketing e precificação, juntamente com a aplicação do modelo de previsão SARIMA, este trabalho atingiu seus objetivos com êxito.

Palavras-chave: Otimização. *E-commerce*. Frete Grátis. *Data mining*. SARIMA.

Abstract

The analysis of the impact of free shipping on e-commerce sales is of crucial importance in the current market context. This study investigated this strategy on the Olist platform, focusing on the Northeast region of Brazil, and sought to integrate theory, methodology, and results to provide valuable insights. The data analysis provided a deep understanding of sales behavior, identified seasonal trends and demand patterns, supporting the effectiveness of the free shipping strategy. The results obtained validated the potential of free shipping as a strategic tool to boost sales in the Brazilian e-commerce landscape. This comprehensive and integrated analysis contributes to the enhancement of sales strategies and provides valuable tools for companies to maximize their financial returns. Through the exploration of a solid theoretical framework that addressed the relationship between free shipping and marketing and pricing strategies, along with the application of the SARIMA forecasting model, this work successfully achieved its objectives.

Keywords: Optimization. E-commerce. Free Shipping. Data mining. SARIMA.

Lista de Figuras

Figura 1 – Número de empresas de varejo online.	12
Figura 2 – Receita Bruta de revenda do mercado online de varejo. . . .	13
Figura 3 – DISTRIBUIÇÃO GEOGRÁFICA DA INDÚSTRIA.	16
Figura 4 – Perfil do E-Commerce Brasileiro.	17
Figura 5 – Aplicação de PO e DM	26
Figura 6 – Distribuição de Publicações com <i>Data Mining e Optimization</i>	28
Figura 7 – Distribuição da Demanda no período, Regressão Linear.	32
Figura 8 – Distribuição da Demanda no período, Polinômio de Grau 3. . . .	33
Figura 9 – Distribuição da Demanda no período, Média Móvel.	34
Figura 10 – Pedidos por Região.	35
Figura 11 – Conjunto de Gráficos (a/h) Volume de Compras por Região e Categoria.	36
Figura 12 a) – Contraste da Distribuição da Demanda da categoria SE.	42
Figura 12 b) – Contraste da Distribuição da Demanda da categoria NE.	43
Figura 13 – Demanda em relação ao Frete - Região Nordeste e Sudeste	45
Figura 14 – Previsão de Frete no Nordeste.	47
Figura 15 – Comportamento do produto Beta no Sudeste.	48
Figura 16 – Comportamento do produto Beta no Nordeste.	49

Lista de tabelas

Tabela 1 – Quantidade de pedidos por status	31
Tabela 2 – Média de Frete por categoria e Região	45
Tabela 3 – SARIMA Frete do Nordeste para relógios_presentes	46
Tabela 4 – Teste de viabilidade de Frete grátis para o produto Beta	50

Lista de abreviaturas e siglas

DM – Data Mining

IBGE - Instituto Brasileiro de Geografia e Estatística

MEI - Microempreendedores Individuais

PAC - Pesquisa Anual de Comércio

PIB – Produto Interno Bruto

PO – Pesquisa Operacional

RMSE - Root Mean Square Error

SARIMA - Seasonal Autoregressive Integrated Moving Average

SEBRAE - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

Sumário

1. Introdução	12
1.1 Identificação da problemática	15
1.2 Objetivo geral.....	17
1.3 Objetivos específicos	18
1.4 Estrutura do projeto.....	19
2 Referencial Teórico.....	20
2.1 O despertar do <i>E-commerce</i> no Brasil.....	20
2.1.1 Olist: Uma Fusão Inovadora	20
2.2 Comportamento do consumidor de <i>E-commerce</i>	21
2.3 Taxas de Entrega e Custos de Envio na Maximização do Lucro	22
2.4 Pesquisa Operacional	23
2.4.1 <i>Data Mining</i>	24
2.4.2 Otimização	24
2.4.3 A Simbiose Produtiva entre DM e PO	26
3 Metodologia	29
3.1 Tratamento dos dados	29
3.2 Processo Exploratório: Compreendendo o Comportamento dos Dados	30
3.3 Compreendendo o Perfil dos Consumidores e a Distribuição Regional das Compras.....	35
3.4 Análise da Relação entre Frete e Demanda por Região e Categoria de Produtos	44
4 Resultados e Discussões.....	46
4.1 Modelo de Previsão de Frete Nordeste	46
4.2 Proposta teste de frete grátis para o Nordeste	48
5 Conclusão.....	52
REFERÊNCIAS	54
APÊNDICES.....	56

1. Introdução

Os *e-commerces*, ou comércio eletrônico, desempenharam um papel revolucionário nas últimas décadas, moldando significativamente o comportamento do consumidor. Estas mudanças são uma consequência direta da crescente acessibilidade à internet, da maior confiabilidade nas transações online e da expansão vertiginosa das opções de compras disponíveis. No mundo do comércio eletrônico, abarca-se tanto a comercialização de produtos físicos quanto de serviços intangíveis. Isto abrange todos os aspectos envolvidos no processo de aquisição, desde a obtenção de matérias-primas, passando pelas estratégias de marketing, procedimentos de pagamento e, por fim, o suporte legal antes, durante e após a compra (TASSABEHJI, 2013).

No Brasil, a internet tem se consolidado como um veículo essencial para as vendas. De acordo com dados da Pesquisa Anual de Comércio (PAC) do Instituto Brasileiro de Geografia e Estatística (IBGE), entre 2016 e 2020, testemunhamos um impressionante aumento de quase cinco vezes na quantidade de empresas de varejo online, como ilustrado na Figura 1. Além disso, a Receita Bruta proveniente da revenda no mercado online de varejo alcançou números ainda mais expressivos, conforme demonstrado na Figura 2.

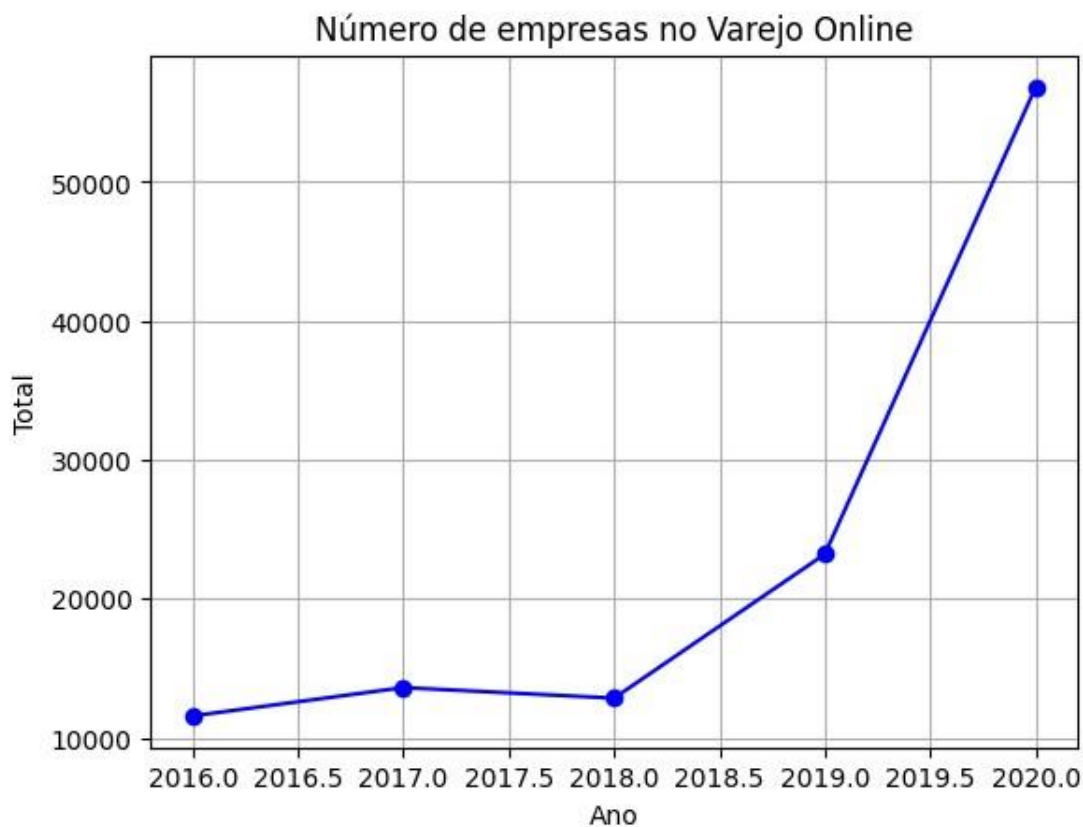


Figura 1 - Número de empresas de varejo online

Fonte: IBGE (2023).

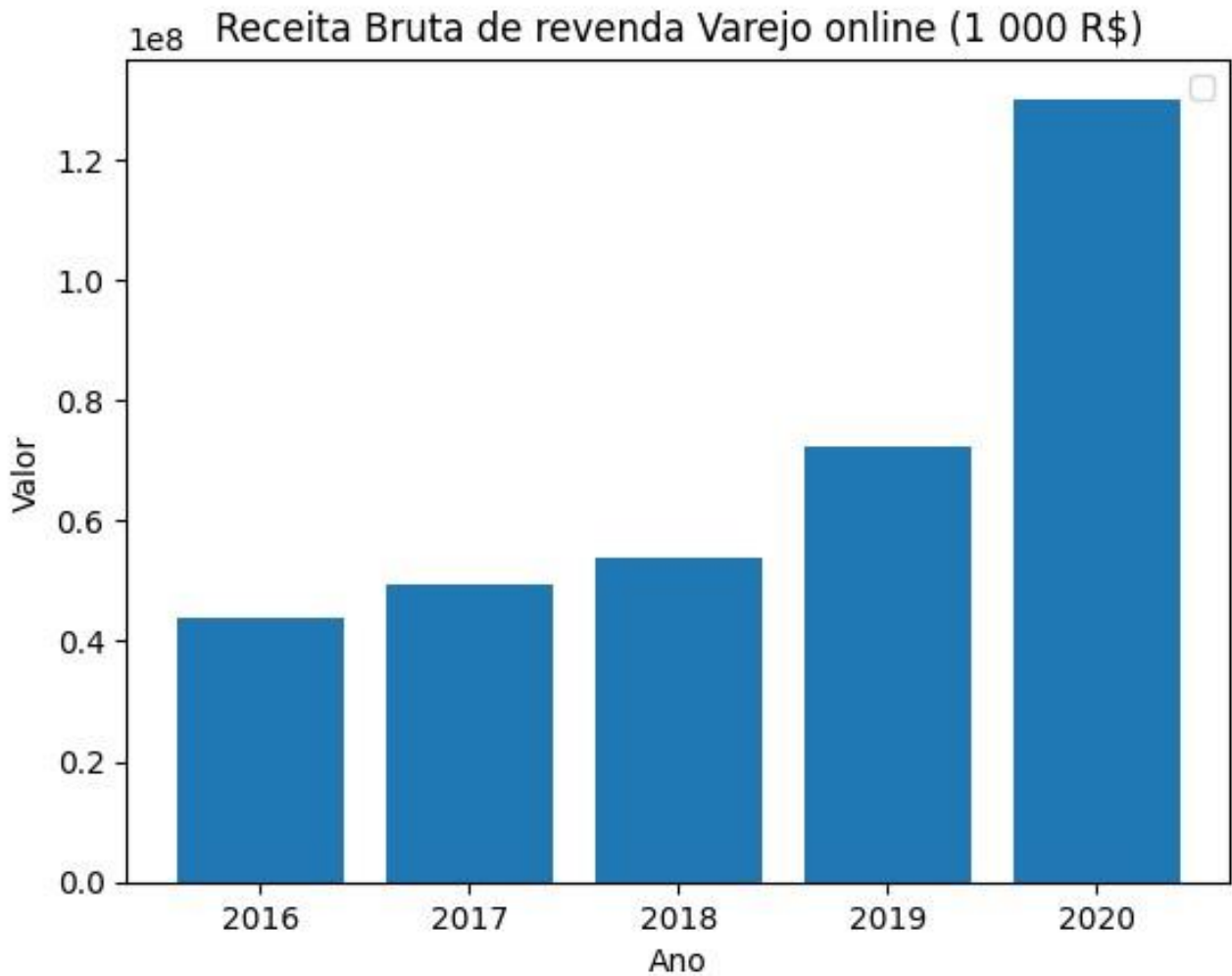


Figura 2 - Receita Bruta de revenda do mercado online de varejo

Fonte: IBGE (2023).

Com o contínuo crescimento desse tipo de comércio no Brasil, a otimização do retorno de vendas emerge como um aspecto crucial para as empresas atuantes no setor varejista, demandando uma análise meticulosa. Segundo Santos e Oliveira (2019), "a otimização do retorno de vendas pode ser definida como a maximização dos resultados financeiros por meio de estratégias de precificação e marketing".

De acordo com Zhao et al. (2019):

"Com o rápido desenvolvimento e progresso da sociedade, as compras online tornaram-se um componente integral na satisfação das necessidades diárias de consumo das pessoas. A tendência de utilizar o comportamento do cliente e a demanda para analisar os fatores que influenciam as vendas de produtos, com base em dados gerados pelos usuários, está ganhando destaque em todo o mundo."

Este artigo, visa comprovar que a estratégia de frete grátis é benéfica tanto para o incremento de vendas por parte do vendedor, quanto para aumentar o alcance a produtos por parte do consumidor na região do Nordeste brasileiro.

1.1 Identificação da problemática

No cenário do comércio eletrônico, as garantias oferecidas pelos vendedores desempenham um papel fundamental na redução dos riscos percebidos pelos consumidores, ao mesmo tempo que se apresentam como poderosas ferramentas de marketing para impulsionar as vendas (HART et al., 1992). O frete grátis é uma dessas garantias de entrega que os vendedores podem oferecer aos clientes em meio a um ambiente de comércio eletrônico. De fato, essa estratégia é uma das 14 variáveis analisadas no estudo de Zhao et al. (2019).

A pesquisa realizada por Zhao et al. (2019) comprovou que o impulsionamento do frete grátis é um fator determinante na disposição do consumidor para realizar compras online. No entanto, essa estratégia precisa ser viável financeiramente para que os empreendedores a adotem com segurança.

De acordo com o Serviço Brasileiro de Apoio às Micro e Pequenas Empresas (SEBRAE, 2023), com base em dados da Receita Federal Brasileira, algumas observações sobre as empresas fechadas em 2020 incluem:

- A taxa de mortalidade é maior no setor de comércio, com 30,2% das empresas fechando em 5 anos.
- Os Microempreendedores Individuais (MEIs) têm uma taxa de mortalidade de 29% em 5 anos.
- Muitas empresas fechadas carecem de preparo e capacitação.
- Muitos empreendedores falham em coletar informações relevantes ao criar seus negócios.

Diante desse cenário desafiador, a Olist se destaca como uma plataforma que apoia empreendedores ao longo de suas jornadas e os ajuda a atingir seus objetivos de maneira facilitada no ambiente online.

Nesse contexto, o foco deste trabalho se concentra na investigação do impacto do frete grátis nas vendas da plataforma Olist. A pesquisa visa determinar a partir de qual ponto de crescimento da demanda essa estratégia se torna vantajosa para a empresa em uma amostra. Compreender a influência do frete grátis no retorno de vendas é de suma importância, pois isso pode auxiliar na definição de políticas de precificação e estratégias de marketing que maximizem os resultados financeiros da empresa.

A empresa em análise, a Olist, iniciou suas operações em 2007 sob o nome de *Solidarium*, com o propósito inicial de gerar escalabilidade e visibilidade para artesãos no Paraná. No entanto, essa missão se expandiu ao empoderar o varejo digital e oferecer oportunidades de crescimento para pequenos, médios e grandes comércios (OLIST, 2021). Por meio dessa plataforma, empreendedores podem cadastrar-se e vender seus produtos em diversos marketplaces, acelerando seus resultados e estabelecendo sua presença online.

A problemática em questão é acentuada pela concentração da indústria brasileira na região Sudeste, um fenômeno que remonta à colonização do Brasil e à subsequente perpetuação dessa região como a mais desenvolvida do país. Conforme demonstrado na Figura 3, a participação no Produto Interno Bruto (PIB) da indústria atesta a atualidade desse fato.

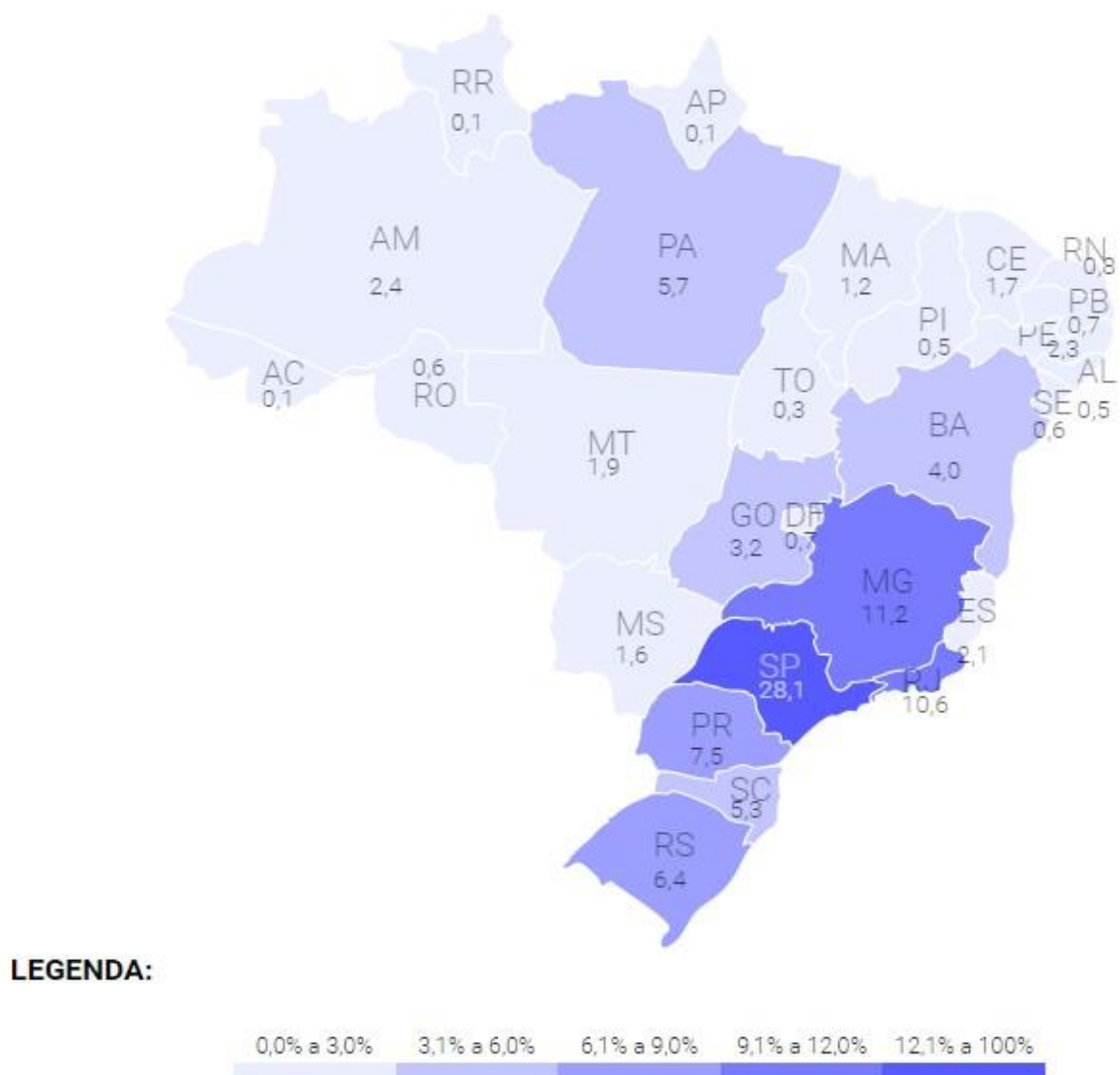


Figura 3 – DISTRIBUIÇÃO GEOGRÁFICA DA INDÚSTRIA

Fonte: Portal da indústria 2020

O perfil industrial brasileiro reflete o panorama do comércio eletrônico no país, que concentra a maior parte das empresas em São Paulo. Entre 2015 e 2018, as empresas sediadas em São Paulo representaram de 54% a 62% do total de comércios eletrônicos no Brasil. Esse período, especificamente o ano de 2018, marcou o auge das empresas de comércio eletrônico com foco em São Paulo, como evidenciado na Figura 4.

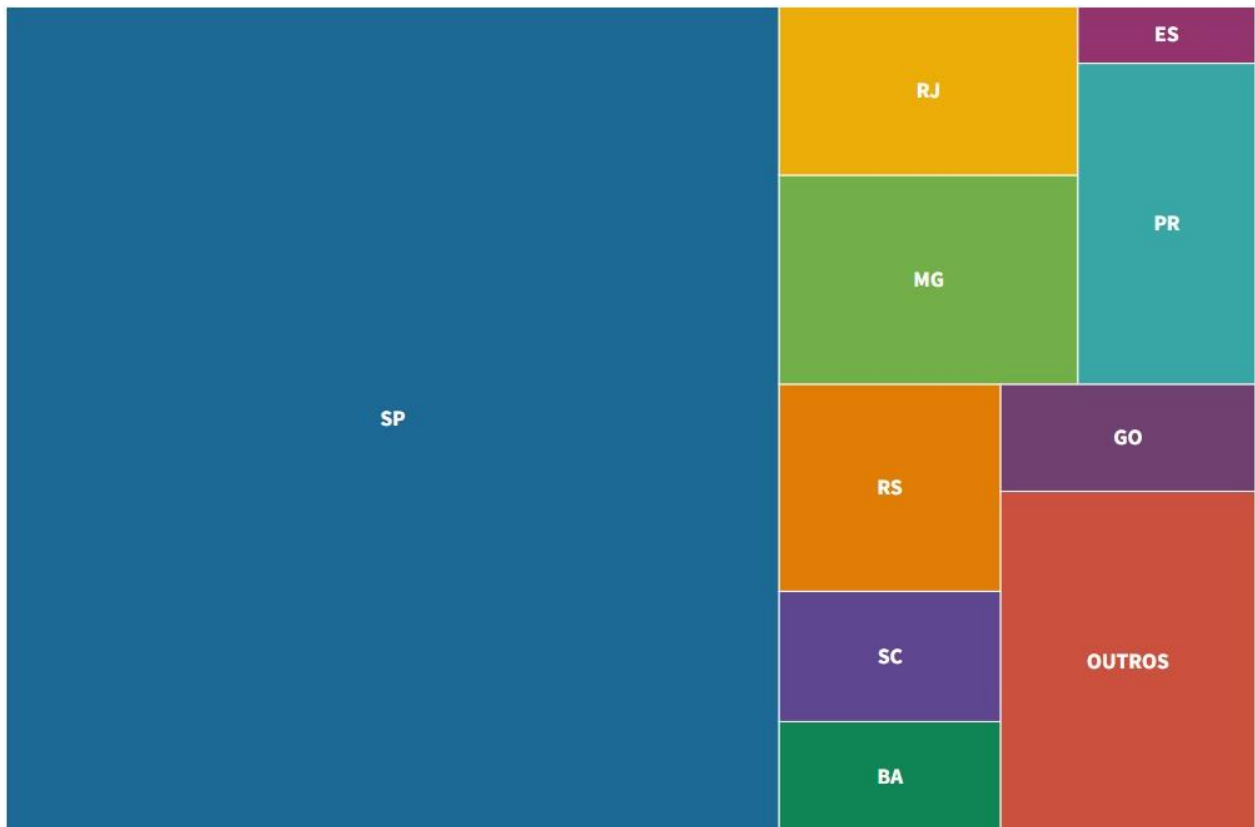


Figura 4 – Perfil do E-Commerce Brasileiro

Fonte: BigDataCorp

Essa concentração geográfica tem implicações diretas na aplicação e eficácia das estratégias de frete grátis, visto que atingem uma grande parcela da população.

1.2 Objetivo geral

O presente trabalho tem como objetivo analisar e propor um modelo para otimizar o retorno de vendas de uma companhia online, utilizando o estudo de caso do e-commerce Olist analisando a perspectiva do impacto do frete grátis sob a demanda.

1.3 Objetivos específicos

Para alcançar o objetivo proposto, serão utilizados dados de vendas do e-commerce Olist, compreendendo o período entre outubro de 2016 e outubro de 2018. Essa base de dados pública, intitulada "*Brazilian E-Commerce Public Dataset by Olist*", foi disponibilizada para uso acadêmico na plataforma *Kaggle*. O projeto de pesquisa se desdobrará em uma série de etapas, cujos objetivos específicos estão descritos a seguir:

Objetivo 1: Análise dos Dados de Vendas da Olist

A primeira etapa consiste na análise exploratória dos dados de vendas fornecidos pela Olist. Técnicas de visualização e estatística descritiva são empregadas para obter insights preliminares sobre o comportamento das vendas ao longo do período em questão.

Objetivo 2: Verificação de Padrões e Identificação da Metodologia

Nesta fase, serão identificados padrões sazonais, tendências e possíveis correlações nos dados de vendas. Além disso, será selecionada a metodologia mais adequada para a análise dos dados, considerando a natureza da série temporal e os objetivos do estudo.

Objetivo 3: Recorte e Análise de Amostras Média e Pequena

A fim de aprofundar as investigações, será realizado um recorte específico da amostra, focalizando em produtos de média demanda. Essa análise permitirá compreender o comportamento das vendas desses produtos e avaliar o impacto do frete grátis em sua comercialização. Em complemento, uma amostra de produtos de baixa demanda será selecionada para análise detalhada. Dessa forma, será possível investigar como o frete grátis influencia a venda de itens com menor procura e sua viabilidade econômica.

Objetivo 4: Desenvolvimento do Modelo

Com base nas análises realizadas, será proposto um modelo que leve em consideração diversos fatores, incluindo a implementação de frete grátis em determinados produtos. O modelo buscará otimizar o retorno de vendas da companhia online satisfazendo os clientes e consequentemente gerando lucro. A proposta é maximizar a Receita da empresa por meio do aumento das vendas, diminuindo os preços a serem pagos pelos clientes.

Objetivo 5: Previsão de Demanda SARIMA

Uma análise de previsão de demanda será realizada utilizando o modelo SARIMA (Seasonal Autoregressive Integrated Moving Average). Essa previsão auxiliará na tomada de decisões estratégicas, considerando a sazonalidade e os padrões de venda identificados.

Através da consecução desses objetivos específicos, o projeto almeja contribuir para o aprimoramento das estratégias de vendas da empresa, promovendo a eficiência operacional e a maximização dos retornos financeiros.

1.4 Estrutura do projeto

O primeiro capítulo terá a introdução e contextualização do tema, abordando a importância da otimização do retorno de vendas em uma companhia online. O segundo capítulo apresentará o embasamento teórico, abrangendo temas como Pesquisa Operacional, otimização e análise de dados aplicados ao contexto do comércio eletrônico. O terceiro capítulo detalhará a metodologia adotada para analisar os dados de vendas do e-commerce "Olist". O quarto capítulo exibirá os resultados obtidos e as discussões a respeito das análises realizadas. Por fim, o último capítulo que é a conclusão do projeto, apresenta as considerações finais e recomendações para o aprimoramento das estratégias de vendas da empresa.

2 Referencial Teórico

A seção de fundamentação teórica tem como propósito abordar os conceitos relevantes relacionados à problemática de pesquisa, a fim de fornecer uma base sólida para o trabalho em questão. Nesse sentido, serão explorados os tópicos que abrangem os recursos de otimização de vendas e o comportamento do cliente no ambiente online. Posteriormente, será realizado um aprofundamento no tema da pesquisa operacional, bem como no direcionamento da análise adotado.

2.1 O despertar do *E-commerce* no Brasil

A trajetória do comércio eletrônico no Brasil é marcada por marcos significativos que culminam com o ambiente empresarial atualmente existente no país. O pioneirismo desse setor teve início com a *Booknet* em 1995, uma loja de livros que viria a se transformar na renomada *Submarino* em 1999. Entretanto, foi na última década do século XX que o cenário do comércio online realmente consolidou sua presença, trazendo consigo uma onda de inovação e crescimento.

O ano de 1999 viu o surgimento da *Submarino*, que rapidamente se tornou uma referência em comércio eletrônico. No entanto, o grande ponto de inflexão aconteceu quando, em 2005, as Lojas Americanas expandiram suas operações para o ambiente online, abrindo espaço em seu portfólio de serviços. Essa decisão estratégica marcou um divisor de águas, trazendo uma gigante do varejo offline para o mundo online.

À medida que a Lojas Americanas estendeu seu alcance online, ela adquiriu ativos valiosos, como o *Shoptime* e o *Ingresso.com*. Além disso, foi incorporado o precursor do *Submarino*, consolidando sua posição como uma força dominante no mercado de comércio eletrônico brasileiro (GZH, 2023). Essa história ilustra o amadurecimento e a transformação do comércio eletrônico no Brasil, moldando o cenário atual de comércio online no país.

2.1.1 Olist: Uma Fusão Inovadora

O marketplace, um conceito intrinsecamente colaborativo, funciona como um espaço digital onde diversos vendedores e empresas terceirizadas se reúnem para listar e comercializar seus produtos ou serviços. Em um ambiente de marketplace, os consumidores têm à sua disposição uma ampla variedade de opções de compra, originárias de múltiplos vendedores. Tudo isso acontece dentro de uma única plataforma online, oferecendo uma experiência de compra

unificada. A característica central de um marketplace é a coexistência de diversos vendedores, cada um contribuindo com sua gama de produtos, enquanto a plataforma age como intermediária, facilitando transações seguras e eficientes.

O e-commerce tradicional, por outro lado, é um modelo no qual uma única empresa é responsável por operar sua própria loja online e vender produtos ou serviços diretamente aos consumidores finais. Os clientes interagem e fazem compras diretamente com a empresa proprietária da loja online. Essa abordagem proporciona à empresa um controle significativo sobre sua marca e a experiência do cliente, mas também implica desafios substanciais, como a construção de reconhecimento de marca e a atração de uma base de clientes.

A Olist, por sua vez, representa uma fusão criativa desses dois modelos distintos. A empresa brasileira oferece uma plataforma que capacita pequenos e médios vendedores a listar seus produtos em renomados *marketplaces*, como as Americanas, Magazine Luiza e Submarino. A Olist desempenha um papel fundamental como intermediária, oferecendo uma gama de serviços abrangentes, incluindo integração de catálogos, processamento de pedidos, atendimento ao cliente, gestão de pagamentos e até mesmo a logística de envio de produtos. Essa abordagem inovadora permite que os vendedores alcancem um público mais amplo e aproveitem uma infraestrutura de e-commerce robusta, sem a necessidade de criar e gerenciar suas próprias lojas online.

A distinção principal entre a Olist e um e-commerce tradicional reside na natureza de sua operação. Enquanto a Olist age como um conector que reúne vendedores e marketplaces líderes, um e-commerce tradicional é uma loja única operada por uma empresa que comercializa exclusivamente seus próprios produtos. A grande vantagem da Olist está na ampliação do alcance que proporciona aos vendedores, permitindo-lhes atingir um público mais vasto ao listar seus produtos em marketplaces renomados, como as Americanas. Essa flexibilidade é uma das muitas razões pelas quais a Olist se destaca no cenário do comércio eletrônico brasileiro.

2.2 Comportamento do consumidor de *E-commerce*

O estudo do comportamento do consumidor no contexto do comércio eletrônico, frequentemente referido como "*E-Customer Behavioural*" (Lantz e Hjort, 2013), representa uma área de pesquisa de fundamental importância. Isso se deve ao seu papel na compreensão dos padrões de comportamento que orientam as ações e decisões dos consumidores online. Este tópico transcende a mera análise de dados de compras e aborda uma variedade de aspectos que

vão desde as motivações de compra até a influência de estratégias específicas de marketing. Uma visão aprofundada desse campo revela nuances cruciais sobre como as empresas podem interagir eficazmente com seu público-alvo no ambiente digital.

Como exemplo, Lewis (2005) explora a retenção de clientes, identificando fatores críticos que influenciam esse processo. Suas descobertas ressaltam a influência significativa de políticas como o oferecimento de frete grátis. A política de frete grátis pode ser um fator determinante na decisão do cliente de continuar comprando de uma empresa específica. Esta observação destaca a importância da retenção de clientes como uma estratégia essencial para o sucesso contínuo de um negócio de comércio eletrônico.

O comportamento do consumidor também desempenha um papel crucial na aquisição de novos clientes. Lewis (2005) sugere que a política de aumento de cesta, que envolve incentivos para que os clientes comprem mais produtos ou serviços, é uma das estratégias mais influentes nesse contexto. Essa abordagem destaca a necessidade de compreender os gatilhos que motivam os consumidores a fazerem compras iniciais e como as empresas podem alavancar esses insights para atrair novos clientes.

O estudo do comportamento do consumidor de e-commerce é um campo multidimensional que exige uma análise aprofundada de diversas variáveis e métricas para otimizar as estratégias de retenção e aquisição de clientes em um ambiente online em constante evolução. Este campo é essencial para a eficácia das operações de comércio eletrônico e, conseqüentemente, para o sucesso a longo prazo das empresas nesse espaço altamente competitivo.

2.3 Taxas de Entrega e Custos de Envio na Maximização do Lucro

A relação intrincada entre as taxas de entrega e os custos de envio é um fator crítico que desempenha um papel decisivo na lucratividade das empresas (PYKE et al., 2001). Este ponto é enfatizado por Pyke et al. (2001), destacando que o equilíbrio adequado entre esses dois elementos é essencial para garantir o sucesso financeiro das operações de comércio eletrônico. As taxas de entrega, que representam os encargos repassados aos clientes para a entrega de suas compras, e os custos de envio, que refletem os gastos reais da empresa na efetivação dessas entregas, são dois componentes interligados que requerem análise e otimização cuidadosas.

A definição das taxas de entrega é uma tarefa crítica. Se estabelecidas muito altas,

elas podem sobrecarregar os clientes com custos adicionais, potencialmente desencorajando-os de realizar compras. Por outro lado, se estabelecidas muito baixas, correm o risco de não cobrir os custos reais de envio, o que, por sua vez, impactaria negativamente a lucratividade (RAI et al. 2008). Encontrar o equilíbrio ideal nesse processo é um desafio que as empresas enfrentam rotineiramente.

Nesse contexto, Pyke et al. (2001) enfatizam a importância da otimização dessa relação para construir uma estratégia de entrega verdadeiramente lucrativa. Ao examinar detalhadamente os custos de envio e considerar diferentes modelos de precificação, as empresas podem determinar a estrutura de taxas de entrega mais adequada, alinhada com sua estrutura de custos e as expectativas dos clientes. Essa estratégia de otimização não apenas visa assegurar que as taxas de entrega cubram adequadamente os custos associados, mas também busca torná-las atraentes para os clientes, promovendo a fidelização e a satisfação (LEWIS, 2005).

2.4 Pesquisa Operacional

A Pesquisa Operacional (PO) possui raízes profundas na história, com uma série de contribuições científicas que datam do século XVI. No entanto, foi durante a Segunda Guerra Mundial (1939 - 1945) que a PO atingiu seu auge, impulsionada pela necessidade global de alocar eficientemente recursos escassos em um cenário de guerra (HILLIER et al., 2013). Como uma disciplina interdisciplinar, a PO encontra sua base em áreas como matemática, lógica, estatística e ciência da computação, permitindo uma abordagem abrangente para lidar com desafios complexos enfrentados pelas organizações, frequentemente envolvendo uma miríade de restrições.

O objetivo essencial da Pesquisa Operacional é capacitar as organizações a tomar decisões fundamentadas, lógicas e vantajosas que melhorem o desempenho de suas operações (GUPTA, 1992). Ela serve como uma ferramenta valiosa para resolver problemas operacionais e estratégicos, com aplicações que variam desde o planejamento de rotas logísticas até a otimização de recursos financeiros em investimentos.

Em síntese um estudo de pesquisa operacional tem seis passos (HILLIER et al. 2013):

1. Definição do problema e coleta de dados;
2. Construção de um modelo matemático;
3. Derivação de soluções com base no modelo;
4. Teste das hipóteses levantadas; 5. Aplicação contínua do modelo;

6. Implementação do modelo.

2.4.1 Data Mining

O termo "Data Mining" (DM) emerge de um vasto mundo de dados, e sua essência é desvendar conhecimentos e padrões que, anteriormente, permaneciam ocultos, mas que detêm um potencial inestimável. Essa abordagem compreende uma série de etapas cruciais, desde a integração de dados provenientes de diversas fontes até o pré-processamento desses dados e a criação de um modelo por meio de algoritmos de aprendizado. Esse modelo, por sua vez, é aplicado para identificar insights valiosos e direcionar ações dentro de uma organização.

A Mineração de Dados transcende as fronteiras do tradicional, incorporando técnicas de diversas disciplinas, como estatística e aprendizado de máquina. Além disso, ela abrange uma ampla gama de aplicações, de análises de mercado a otimizações complexas, demonstrando sua versatilidade e relevância em várias áreas.

Embora a Mineração de Dados tradicionalmente se baseie em técnicas estatísticas e de aprendizado de máquina, vale ressaltar que muitos dos problemas nessa área também podem ser formulados como desafios de otimização (OLAFSSON et al., 2008). A otimização desempenha um papel essencial na busca por soluções ideais em meio a uma miríade de possíveis resultados, destacando ainda mais a amplitude e a complexidade desse campo.

2.4.2 Otimização

De acordo com Hillier (2013), a busca pela "otimalidade" representa o tema mais importante da Pesquisa Operacional (PO) e é uma faceta intrínseca à própria essência da disciplina. Ela implica em estabelecer objetivos claros, identificar restrições e modelar variáveis de decisão, com o propósito de atingir resultados ótimos. Esse processo é fundamentado em uma série de técnicas e abordagens, como a programação linear, algoritmos, e busca heurística, entre outras (BELFIORE, 2013).

Nesse contexto, este trabalho encontra inspiração em diversos modelos que visam a maximização de receita e a aplicação de princípios da programação dinâmica. Essa abordagem, amplamente explorada por autores como Boyd et al. (2003) e Previgliano et al. (2021), não se limita a essa, mas sim se beneficia da convergência de múltiplas perspectivas, resultando em uma abordagem abrangente e eficaz:

$$\begin{aligned} \max \quad & \sum_{i \in I} r_i x_i \\ \text{s.t.} \quad & \sum_{i \in I(l)} x_i \leq c_l, l \in L \\ & x_i \leq d_i, i \in I \\ & x_i \geq 0, i \in I \end{aligned}$$

Equação 1: maximização da receita

Onde, em paralelo:

- r_i : Receita;
- x_i : Demanda;
- l : Frete por distância;
- L : O conjunto de frete;
- i : Pedidos;
- c_l : Pedidos enviados;
- d_i : Quantidade de itens enviados;

Previgliano et al. (2021) destacam a aplicação da programação dinâmica como uma estratégia fundamental em questões relacionadas à otimização de receitas, particularmente quando pressupomos a capacidade do programa de alocar recursos de maneira eficiente. Neste trabalho, a ênfase recai sobre a proposta de otimização do retorno das vendas, adotando uma abordagem móvel e automatizada, ancorada na consideração do frete como fator determinante.

Nesse contexto, a incorporação do frete como um componente chave na otimização do retorno das vendas ressalta a importância de considerar as variáveis em constante mudança que afetam o desempenho financeiro de uma organização. Essa abordagem móvel e automatizada reflete a busca pela máxima eficiência em um ambiente de negócios cada vez mais competitivo e ágil.

2.4.3 A Simbiose Produtiva entre DM e PO

Embora os procedimentos de *Data Mining* (DM) e Pesquisa Operacional (PO) possam diferir em maturidade e abordagem, o propósito subjacente de sua aplicabilidade demonstra uma relação mutualística sólida. Isso pode ser claramente observado na Figura 5, conforme ilustrado por Meisel et al. (2010):

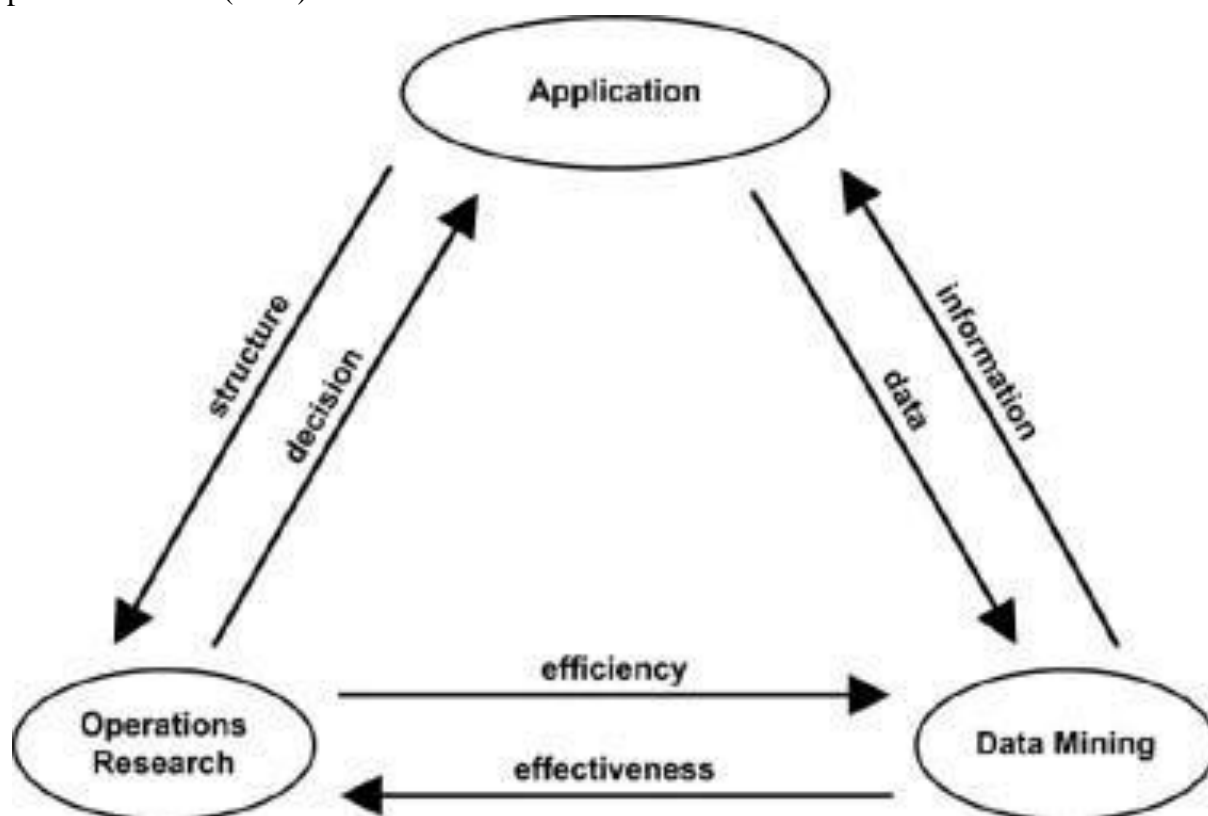


Figura 5 – Aplicação de PO e DM

Fonte: MEISEL, S. et al. 2010

A Figura 5 destaca a sinergia entre DM e PO, revelando como essas duas disciplinas, apesar de suas abordagens distintas, estão intrinsecamente ligadas. Ela sugere que, embora cada uma dessas áreas possa ser empregada independentemente, a sua verdadeira força emerge quando são combinadas e exploradas de forma conjunta.

Essa relação mutualística é evidenciada pelo fato de que a *Data Mining*, com sua capacidade de descobrir padrões e tendências em grandes conjuntos de dados, pode alimentar o processo de tomada de decisões na Pesquisa Operacional. A PO, por sua vez, fornece estruturas sólidas para otimizar operações e alocar recursos, potencializando o impacto das descobertas do DM.

Segundo Meisel (2010), três tipos de sinergias são distinguíveis:

1. A Pesquisa Operacional aumenta a eficiência da *Data Mining*: A aplicação de técnicas de Pesquisa Operacional na *Data Mining* pode melhorar a eficiência do

processo, otimizando etapas como pré-processamento de dados, seleção de variáveis relevantes e ajuste de parâmetros de algoritmos;

2. A Mineração de Dados aumenta a efetividade da Pesquisa Operacional por substituição: A *Data Mining* pode ser usada para substituir métodos tradicionais da Pesquisa Operacional, fornecendo insights mais precisos e atualizados sobre a estrutura e o comportamento do sistema de aplicação. Isso pode levar a decisões mais eficazes e otimizadas;
3. A *Data Mining* aumenta a efetividade da Pesquisa Operacional por refinamento: A Mineração de Dados pode complementar as técnicas da Pesquisa Operacional, fornecendo informações mais detalhadas e granulares sobre os dados e padrões do sistema de aplicação. Isso permite um refinamento nas formulações e modelos da Pesquisa Operacional, resultando em decisões mais precisas e melhoradas.

A figura 6 também reflete a crescente integração entre esses tópicos no âmbito bibliométrico. Como apontado por Meisel (2010), "recentemente, várias publicações de abordagens bem-sucedidas na interseção dos dois procedimentos têm surgido". Além disso, Olfasson (2008) destaca que "o campo da mineração de dados tem visto um grande interesse tanto da academia quanto da indústria".

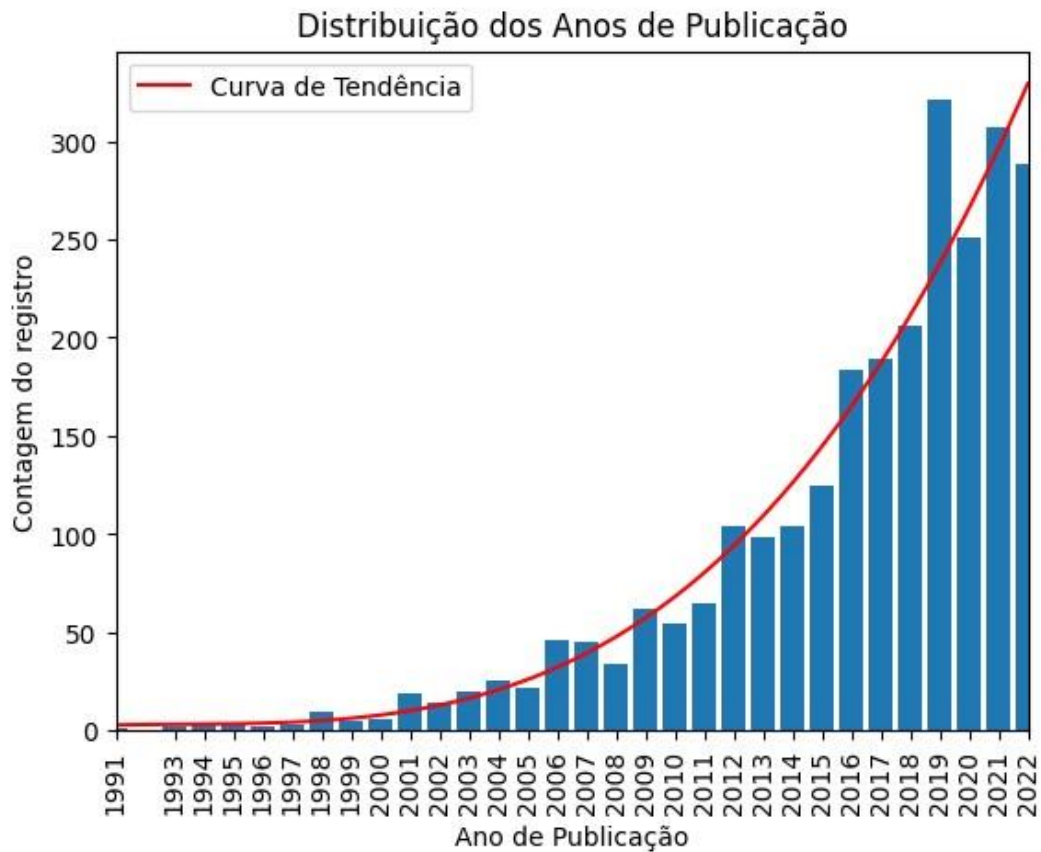


Figura 6 – Distribuição de Publicações com *Data Mining e Optimization*

Fonte: *Web of Science*

Essa integração atesta o potencial dessas disciplinas em se fortalecerem mutuamente, proporcionando soluções mais eficazes, decisões mais informadas e avanços notáveis em ambas as áreas. Ela ressalta a importância de uma abordagem holística na análise de dados e otimização de operações em ambientes dinâmicos e complexos.

3 Metodologia

O presente capítulo visa descrever todo o processo de desenvolvimento do estudo para a proposta de modelo de otimização de retorno de vendas de comércio eletrônico, baseado na aplicação de frete. Será dividido em etapas desde a explanação do tratamento dos dados, e como sucedeu as escolhas necessárias para que a proposta e modelo de predição pudesse a ser desenvolvido.

3.1 Tratamento dos dados

O conjunto de dados utilizado neste projeto foi obtido a partir da plataforma Kaggle e disponibilizado publicamente para fins educacionais (SIONEK, ANDRÉ, 2018). O período de vendas considerado abrange um extenso intervalo, desde outubro de 2016 até outubro de 2018, proporcionando um conjunto de dados de dois anos de informações ricas e variadas. Esses dados são compostos por várias tabelas, incluindo informações relacionadas a clientes, pagamentos, vendedores, geolocalização, avaliações dos clientes, detalhes dos itens dos pedidos, produtos e informações detalhadas sobre os próprios pedidos. A base de dados compreende mais de 100 mil linhas e é composta pelas seguintes informações:

- Tabela "Clientes": contendo informações como `customer_id`, `customer_unique_id`, `customer_zip_code_prefix`, `customer_city` e `customer_state`;
- Tabela "Pagamentos": incluindo dados como `order_id`, `payment_sequential`, `payment_type`, `payment_installments` e `payment_value`;
- Tabela "Vendedores": contendo informações sobre os vendedores, como `seller_id`, `seller_zip_code_prefix`, `seller_city` e `seller_state`;
- Tabela "Geolocalização": englobando os dados de geolocalização, tais como `geolocation_zip_code_prefix`, `geolocation_lat`, `geolocation_lng`, `geolocation_city` e `geolocation_state`;
- Tabela "Reviews": contendo informações relevantes sobre as avaliações dos clientes, incluindo `review_id`, `order_id`, `review_score`, `review_comment_title`, `review_comment_message` e `review_answer_timestamp`;
- Tabela "Itens": contendo dados relacionados aos itens dos pedidos, como `order_id`, `order_item_id`, `product_id`, `seller_id`, `shipping_limit_date`, `price` e `freight_value`;

- Tabela "Produtos": com informações sobre os produtos, como `product_id`, `product_category_name`, `product_name_length`, `product_description_length`, `product_photos_qty`, `product_weight_g`, `product_length_cm` e `product_height_cm`;
- Tabela "Pedidos": englobando dados sobre os pedidos realizados, tais como `order_id`, `customer_id`, `order_status`, `order_purchase_timestamp`, `order_approved_at`, `order_delivered_carrier_date`, `order_delivered_customer_date` e `order_estimated_delivery_date`.

Todos esses arquivos foram exportados em *Dataframes* dentro da plataforma *Visual Studio Code*, utilizando a linguagem de programação Python. O tratamento dos dados desempenhou um papel central nessa fase. Foi necessário remover valores nulos e eliminar possíveis duplicatas, procedimentos que fazem parte do estágio inicial de *Data Mining*. Foram realizadas a remoção de valores nulos e a eliminação de duplicatas, sendo estes procedimentos característicos de uma etapa inicial de *Data Mining*, que abrange a limpeza, integração, transformação e redução de dados (OLAFSSON et al., 2008), e foram essenciais para garantir que a base de dados estivesse em condições adequadas para análise. Detalhes específicos sobre as modificações feitas no *Dataframe* principal estão disponíveis no Apêndice.

3.2 Processo Exploratório: Compreendendo o Comportamento dos Dados

O processo exploratório é uma etapa essencial para compreender e orientar a análise diante de grandes volumes de dados, conforme destacado por Olafsson (2008). No contexto do presente projeto, utilizando a base de dados da empresa Olist, torna-se relevante examinar o padrão dos pedidos efetuados.

Conforme demonstrado na Tabela 1, a maioria dos pedidos da base de dados já se encontrava no estado de "entregue". Para a análise em questão, não foi necessária a realização de ajustes nos dados, visto que as informações referentes aos pedidos não entregues também continham valores de frete e outras informações de categorização.

Tabela 1 – Quantidade de pedidos por *status*

Status	Quantidade de pedidos
Entregue	114.859
Enviado	1.167
Cancelado	553
Faturado	370
Processando	370
Indisponível	7
Aprovado	3
Total	117.329

Fonte: Autoria própria.

A análise exploratória dos dados desempenha um papel fundamental na compreensão da dinâmica das vendas na plataforma Olist e na identificação de tendências e padrões significativos. Para explorar a sazonalidade da demanda ao longo do período analisado, foram utilizados gráficos de linhas e técnicas estatísticas relevantes.

Conforme evidenciado nas Figuras 7, 8 e 9, observou-se que o pico de vendas mais expressivo ocorreu nos meses de novembro e dezembro de 2017, impulsionado por eventos de varejo como a *Black Friday* e o Natal.

A Figura 7 ilustra a aplicação da regressão linear, uma técnica usada para modelar a relação entre duas variáveis contínuas

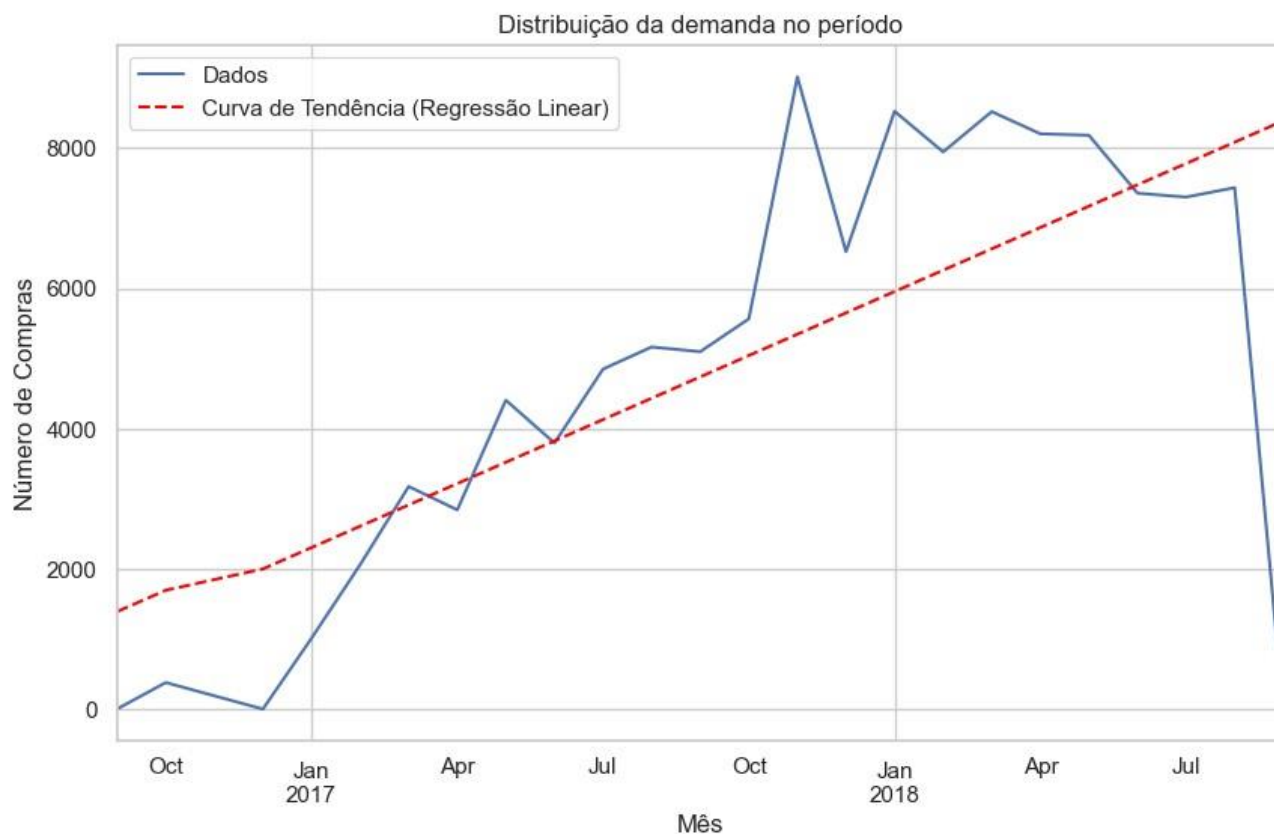


Figura 7 – Distribuição da Demanda no período, Regressão Linear

Fonte: Autoria Própria

Constatou-se que a demanda não apresenta uma tendência linear, mesmo com a linha reta ajustando aos pontos (x, y) , Mês e Número de Compras, pois não segue a regressão de forma contínua, sofrendo flutuações de picos e vales. Embora a demanda não siga uma tendência linear estrita, a análise revela uma tendência de crescimento geral ao longo do período.

$$y = mx + b$$

Equação 2 - Regressão linear Onde:

- y é a variável dependente (valor previsto ou a ser predito);
- x é a variável independente (valor conhecido);
- m é o coeficiente angular (inclinação da linha de regressão);
- b é o coeficiente linear (interceptação com o eixo y).

A fim de continuar a análise, a figura 8, conta com aplicação do Polinômio de terceiro grau a Distribuição da Demanda no período, em que é utilizada uma curva para de ajustar aos pontos do conjunto de dados.

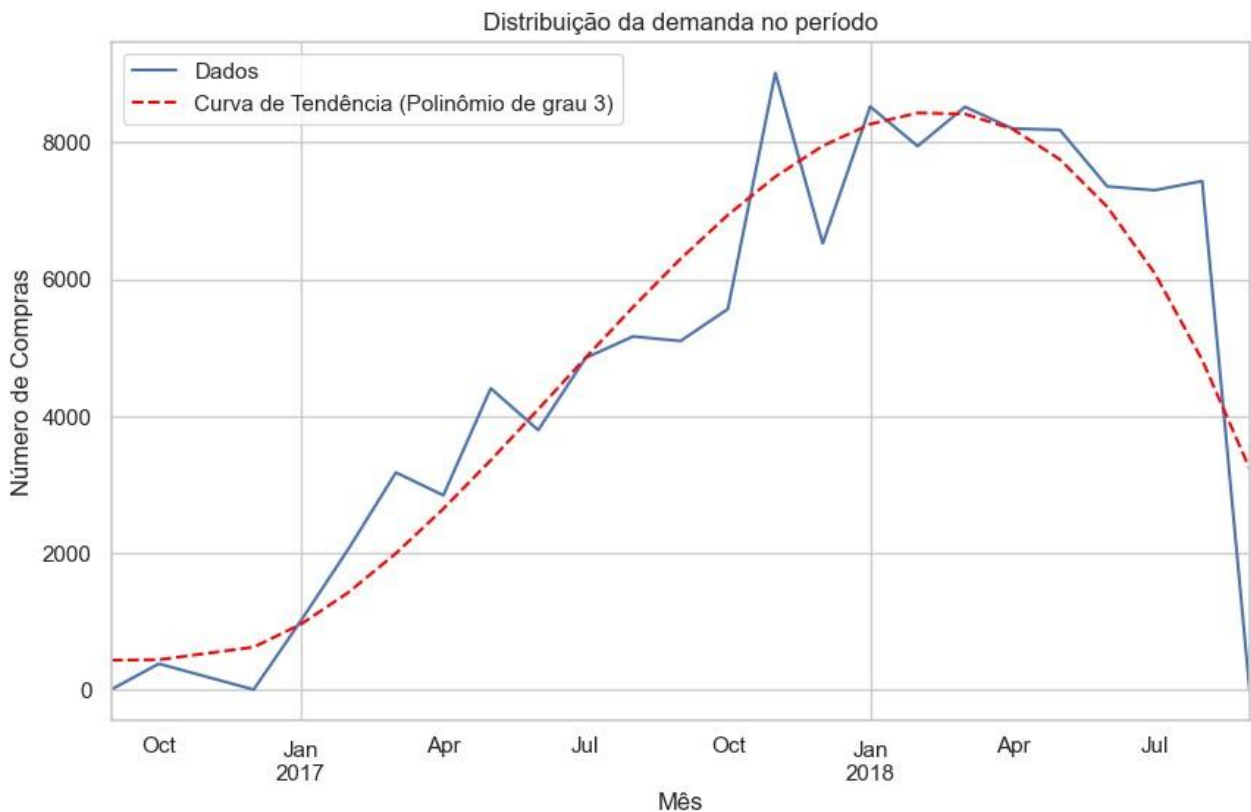


Figura 8 – Distribuição da Demanda no período, Polinômio de Grau 3

Fonte: Autoria Própria

A ideia por trás do polinômio de terceiro grau é encontrar a melhor curva cúbica que se encaixa nos pontos do conjunto de dados, minimizando a soma dos quadrados das diferenças entre os valores reais e os valores previstos pela curva.

O polinômio de terceiro grau é útil quando os dados mostram uma tendência curvilínea que não pode ser adequadamente descrita por uma linha reta. Ele pode ser aplicado em casos em que os dados têm padrões mais complexos ou formas não lineares, onde a relação entre as variáveis não pode ser representada efetivamente por uma regressão linear.

$$.y = ax^3 + bx^2 + cx + d$$

Equação 3 - Polinômio de Grau 3 Onde:

- y é a variável dependente (valor previsto ou a ser predito);
- x é a variável independente (valor conhecido);
- a, b, c e d são coeficientes que serão ajustados durante o processo de ajuste do polinômio.

O gráfico da figura 8, existe proximidade entre a curva de demanda polinomial cúbica

e dos dados reais o que demonstra que a proposta de modelos capazes de fazer previsões históricas com uso de dados de treinamento é viável e com tendência a acuracidade.

Entre as técnicas estatísticas mais importantes para a suavização de dados sazonais e a visualização de informações temporais, destaca-se a Média Móvel (Morettin e Bussab, 2017). Essa técnica é uma ferramenta valiosa para identificar padrões e tendências em séries temporais de dados, o que facilita a interpretação e análise dessas informações.

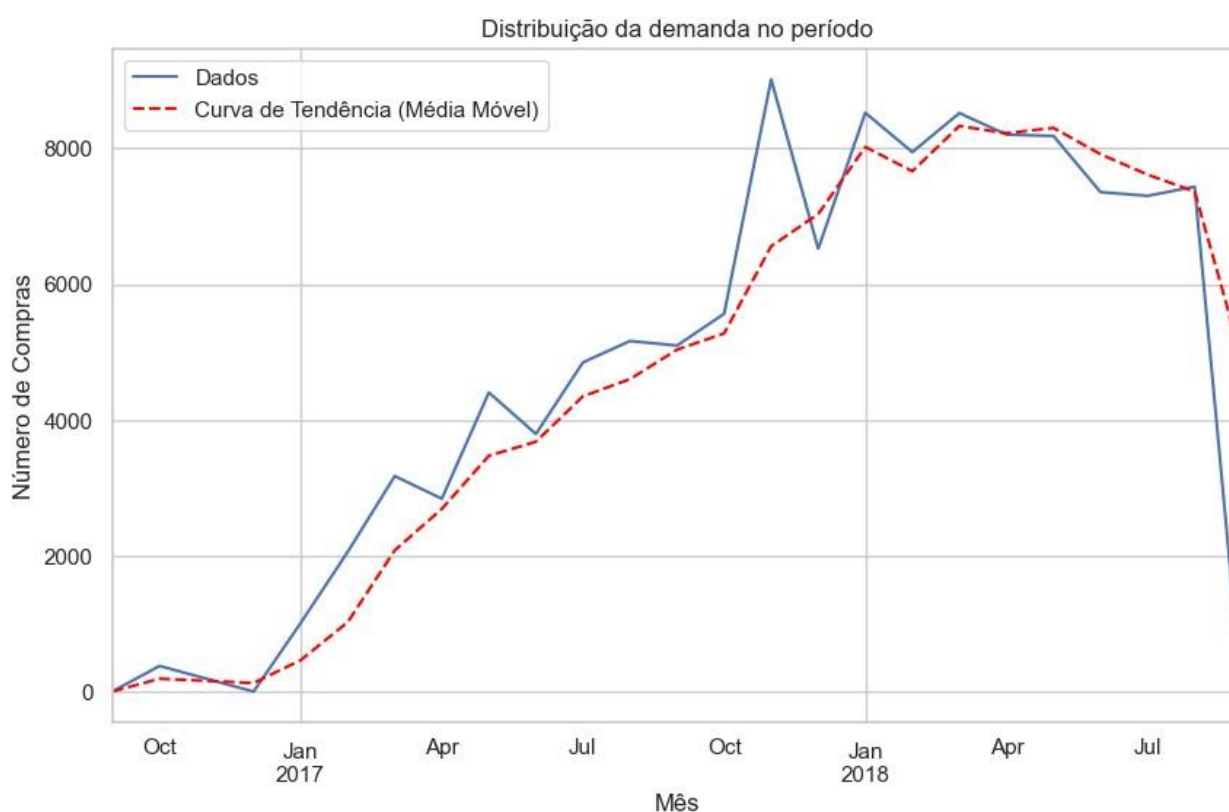


Figura 9 – Distribuição da Demanda no período, Média Móvel

Fonte: Autoria Própria

Neste estudo, foi utilizada a Média Móvel com uma janela de períodos de tamanho 3, caracterizada como pequena. Essa escolha proporciona uma média mais responsiva às mudanças recentes, mas também pode torná-la mais suscetível a ruídos. Apesar disso, o gráfico da Figura 9 demonstra uma proximidade satisfatória com os dados reais, validando a utilização dessa técnica para a suavização dos dados e a identificação de tendências.

$$SMA = \frac{\text{Soma dos últimos } n \text{ valores}}{n}$$

Equação 4 – Média Móvel Simples

3.3 Compreendendo o Perfil dos Consumidores e a Distribuição Regional das Compras

Após a análise da tendência da demanda, voltou-se a atenção para o perfil de compra dos consumidores na plataforma Olist. Descobriu-se que um total de 49.527 compras tiveram origem no estado de São Paulo, consolidando-o como o líder em vendas na plataforma. Essa expressiva contribuição do estado de São Paulo também é evidenciada no gráfico "Pedidos por Região" apresentado na Figura 10, onde a região Sudeste lidera os pedidos por estado, seguida pela região Sul em segundo lugar, e pela região Nordeste em terceiro.

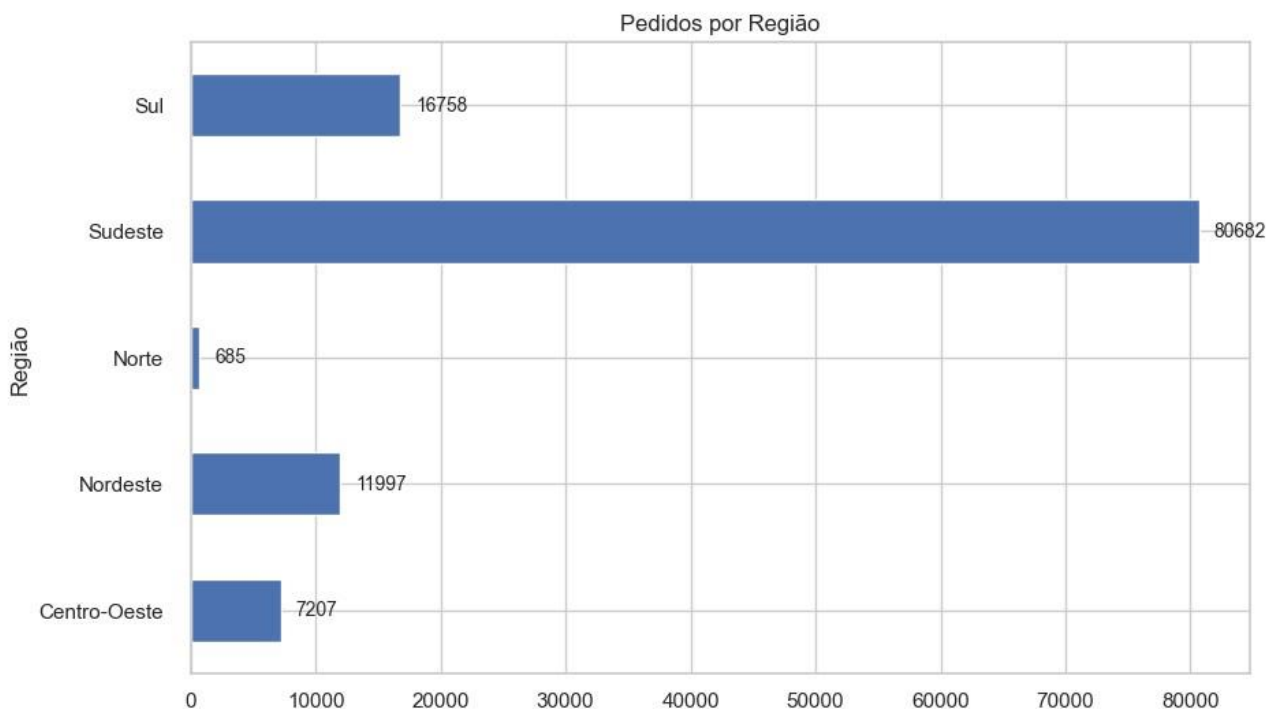


Figura 10 – Pedidos por Região

Fonte: Autoria Própria

Nesse cenário, verificou-se que o número total de categorias existentes para compra é de 74, abrangendo uma diversidade de segmentos, desde 'utilidades_domesticas', 'perfumaria', 'automotivo', 'pet_shop', 'papelaria', 'moveis_decoracao', até 'moveis_escritorio',

'ferramentas_jardim' e muitos outros. Dentre essas categorias, foram identificadas aquelas que representam os maiores volumes de compras em cada região. Os resultados foram destacados e apresentados na sequência de gráficos da Figura 11 - Conjunto de Gráficos Volume de Compras por Região e Categoria a) a h) – importante salientar que todos os gráficos possuem como eixo Y a Quantidade de pedidos.

Nos gráficos de 11 a) a 11 f), fica claro que a categorias são muito fortes no Sudeste, o que representa quase toda ou boa parte a demanda das categorias, o Sul ocupa o segundo lugar enquanto a demanda do Norte quase não existe, sendo de apenas 18 a 73 vendas no período de 2 anos.

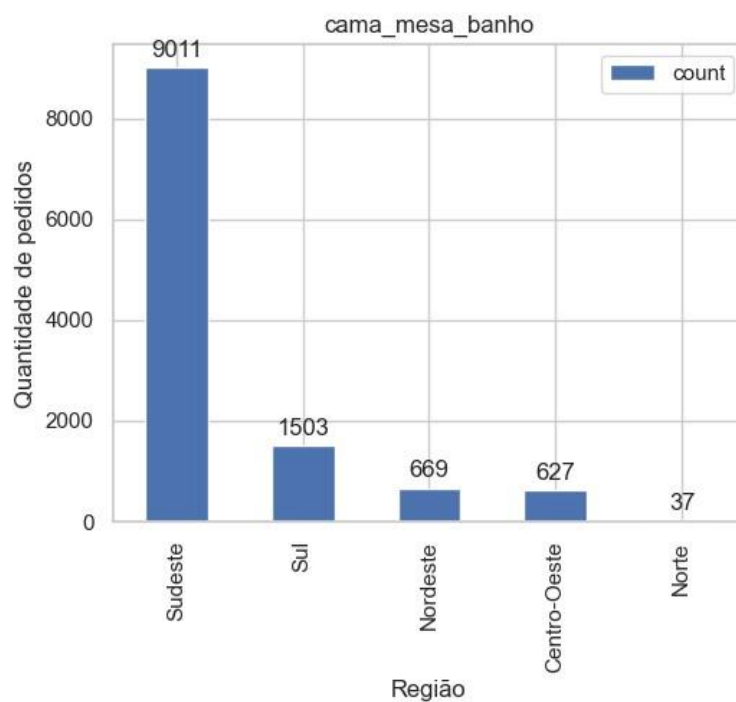


Figura 11 a) – Gráfico Volume de Compras por Região e Categoria de Cama, Mesa e Banho

Fonte: Autoria Própria

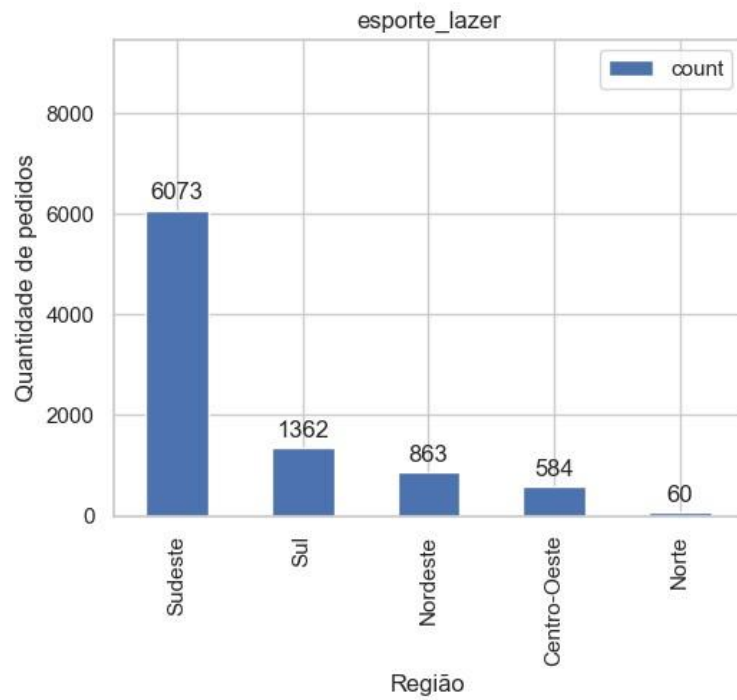


Figura 11 b) – Gráfico Volume de Compras por Região e Categoria de Esporte e Lazer

Fonte: Autoria Própria

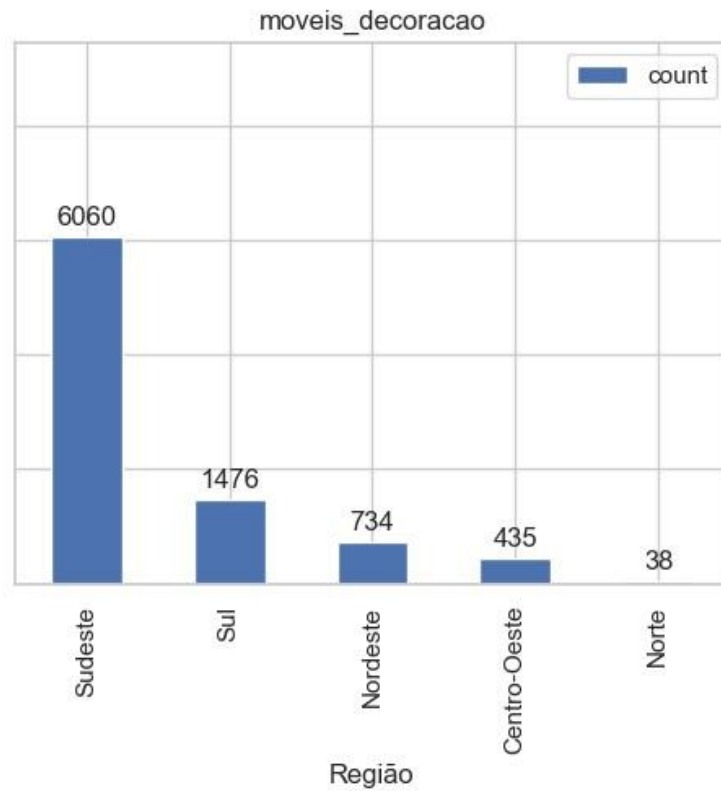


Figura 11 c) – Gráfico Volume de Compras por Região e Categoria de Moveis e Decoração

Fonte: Autoria Própria

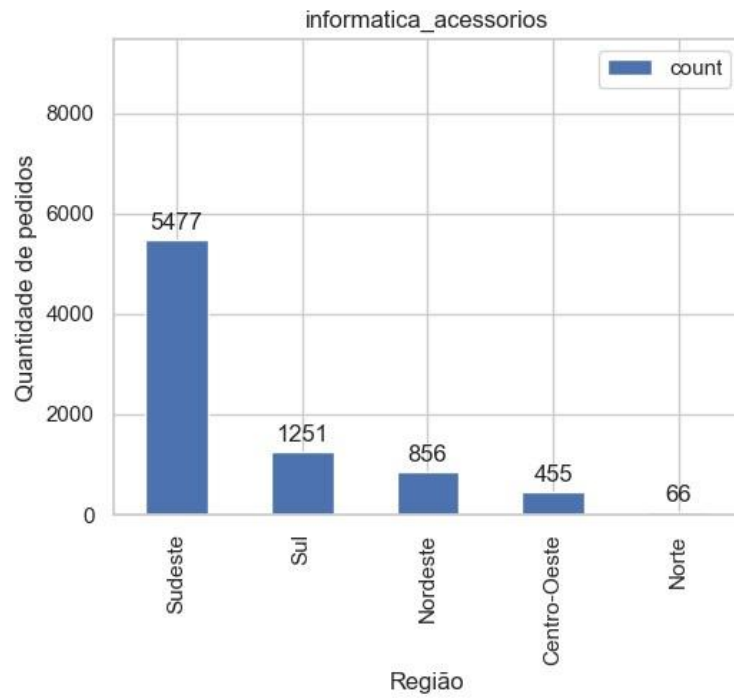


Figura 11 d) – Gráfico Volume de Compras por Região e Categoria de Informática e Acessórios

Fonte: Autoria Própria

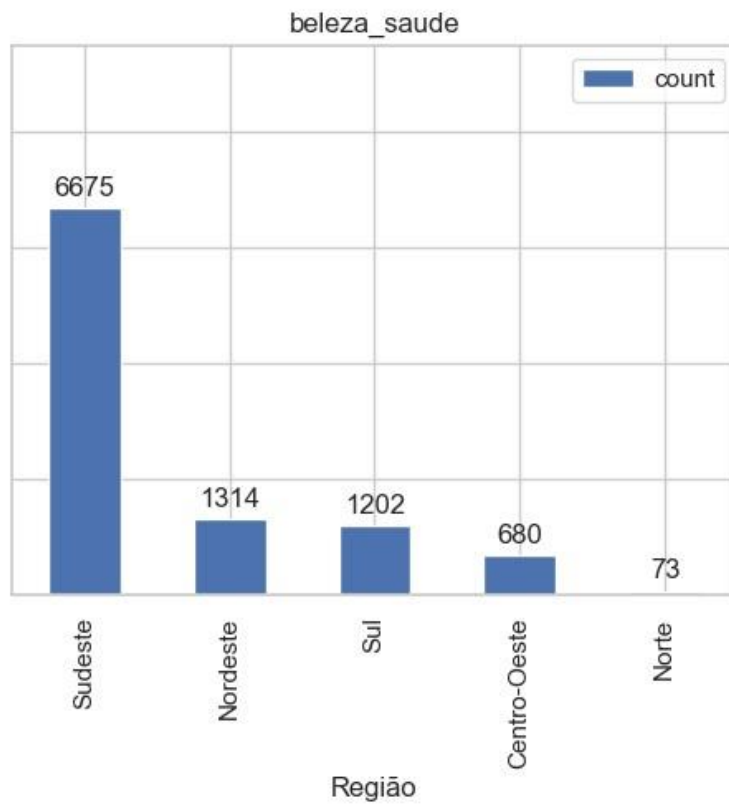


Figura 11 e) – Gráfico Volume de Compras por Região e Categoria de Beleza e Saúde

Fonte: Autoria Própria

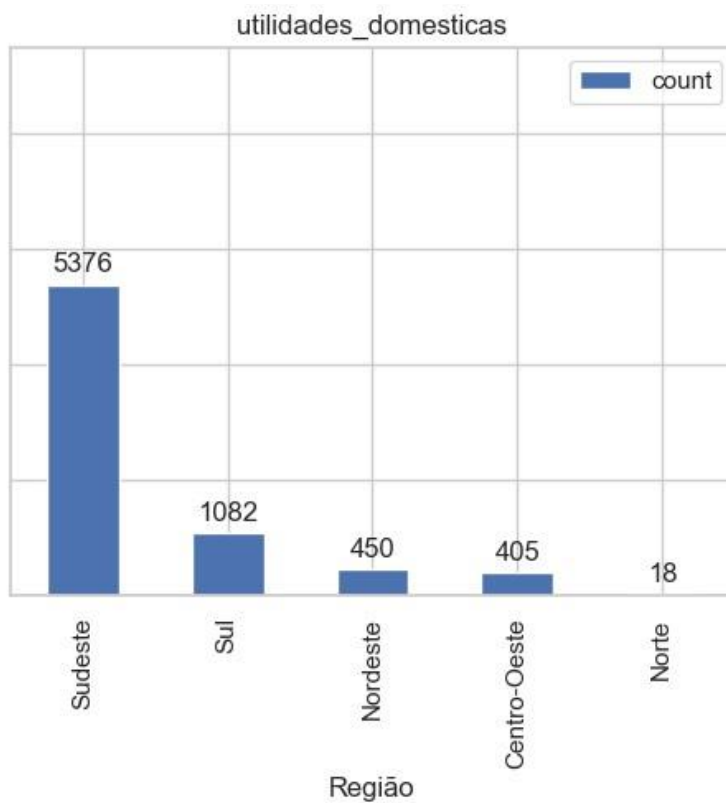


Figura 11 f) – Gráfico Volume de Compras por Região e Categoria de Utilidades Domésticas

Fonte: Autoria Própria

Nos gráficos de 11 g) e h), as categorias embora mantenham o Sudeste como região principal, já existe maior representatividade das outras regiões além Nordeste, que fica em segundo.

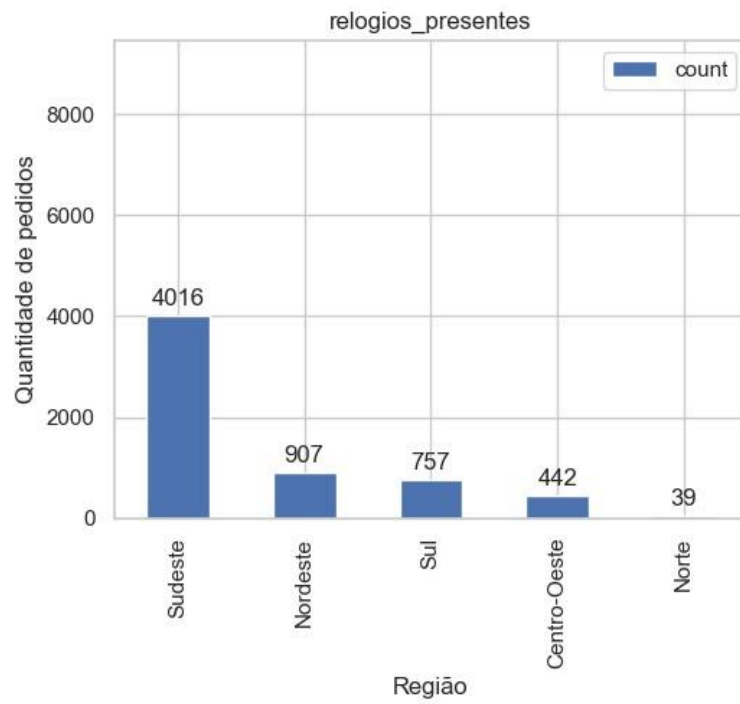


Figura 11 g) – Gráfico Volume de Compras por Região e Categoria de Relógios e Presentes

Fonte: Autoria Própria

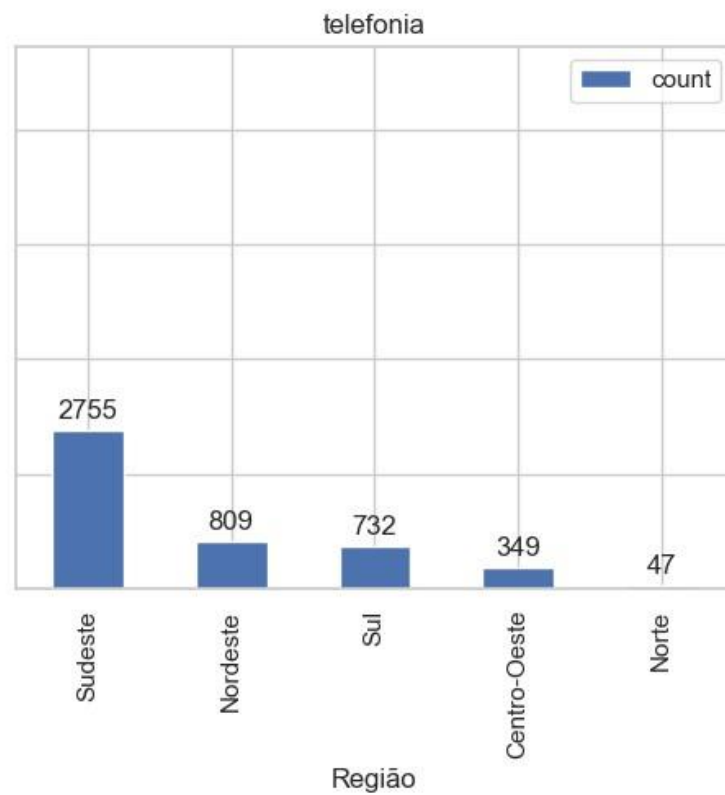


Figura 11 h) – Gráfico Volume de Compras por Telefonica

Fonte: Autoria Própria

A análise revela que as preferências dos consumidores variam significativamente de uma região para outra. Essa percepção é valiosa para aprimorar estratégias de marketing e promover produtos específicos com base nas preferências regionais. Além disso, a compreensão das categorias mais populares em cada região é fundamental para otimizar a oferta de produtos e garantir um estoque alinhado às demandas locais.

Diante do conjunto de dados, optou-se por realizar um estudo comparativo entre as regiões Sudeste e Nordeste, que representam, respectivamente, o primeiro e terceiro maior volume de vendas. Com esse objetivo em mente, o *dataframe* foi ajustado, e verificou-se que as categorias mais representativas para análise na região Nordeste seriam: 'beleza_saude', 'relogios_presentes', 'esporte_lazer' e 'informatica_acessorios'.

Com o intuito de compreender, detalhadamente, o comportamento da demanda nesse recorte, aplicou-se a Curva de Tendência de Média Móvel na Demanda, como pode ser observado na Figura 12 - Contraste da Distribuição da Demanda das categorias entre Sudeste e Nordeste.

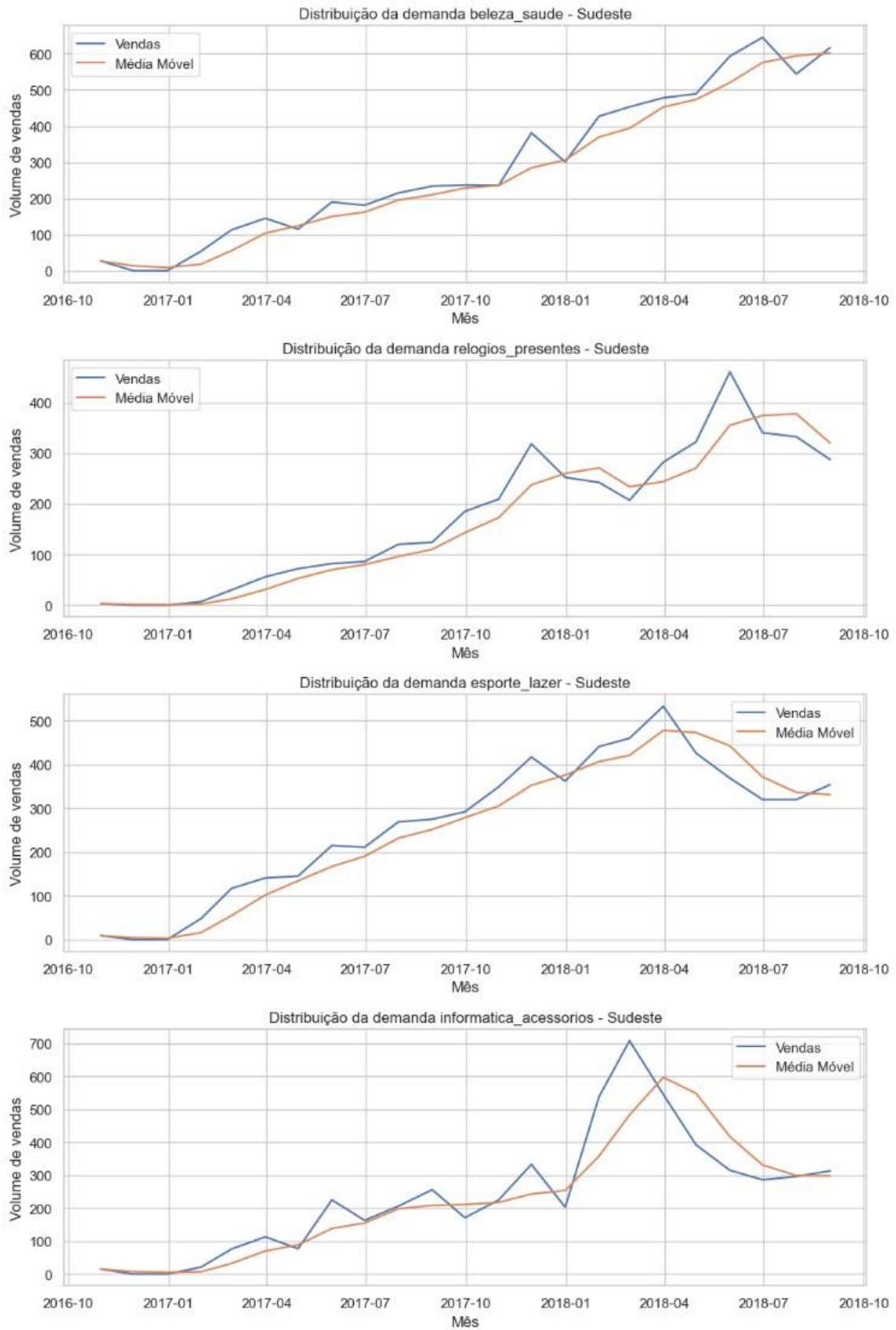


Figura 12 a) – Contraste da Distribuição da Demanda por categoria SE

Fonte: Autoria Própria

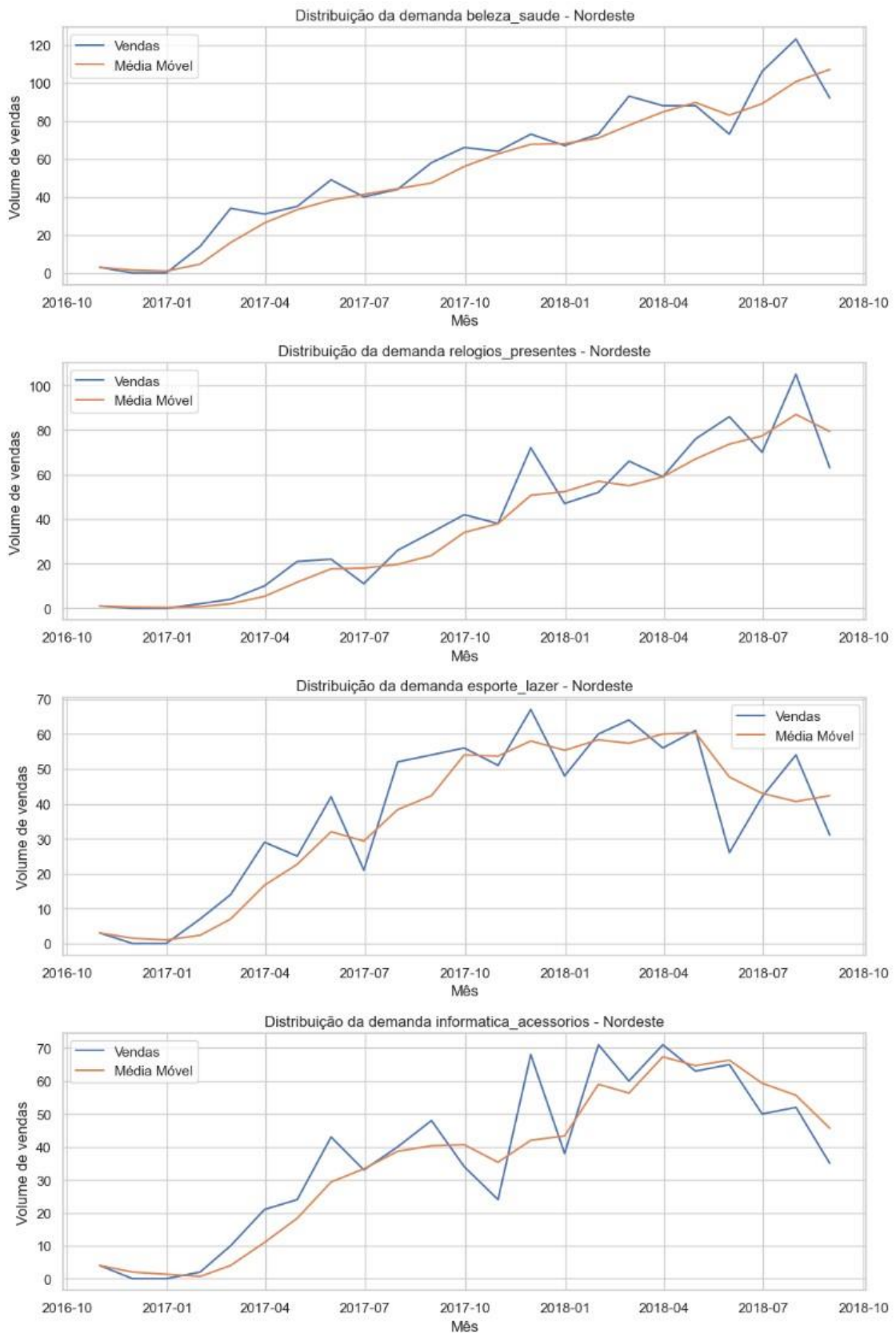


Figura 12 b) – Contraste da Distribuição da Demanda por categoria NE

Fonte: Autoria Própria

Ao analisar os gráficos, observa-se então que, de modo geral, todas as categorias apresentam vendas mais expressivas no Sudeste, o que era esperado. As categorias 'beleza_saude' e 'relogios_presentes' revelam uma tendência de crescimento em ambas as regiões. Por outro lado, a categoria 'esporte_lazer' tende à estabilidade desde março de 2017 no Nordeste, enquanto no Sudeste, evidencia-se um ponto de virada em março de 2018.

Por fim, a categoria 'informatica_acessorios' apresenta uma anomalia no primeiro trimestre de 2018 e, em seguida, retoma a tendência de estabilidade no mercado do Sudeste. Esta possível análise poderia ser estudada no futuro verificando se não houve, por exemplo, o lançamento de novo chip para computadores ou inovação disruptiva que replicasse o pico observado no Sudeste com o aumento de verbas. No entanto, no mercado nordestino, esse tipo de produto se caracteriza como muito mais cíclico, com muitos picos sazonais.

Essa análise aprofundada possibilita uma compreensão mais clara das nuances do comportamento do mercado em diferentes regiões, proporcionando insights valiosos para a tomada de decisões estratégicas.

Como explicitado, a compreensão de como as preferências dos consumidores variam regionalmente e como diferentes categorias de produtos se comportam em cada área é fundamental para a otimização de estoques, o planejamento de estratégias de marketing direcionadas e a alocação eficaz de recursos, buscando maximizar os retornos financeiros. Essa análise comparativa entre regiões distintas é um elemento chave para a proposição de no modelo de otimização de retorno de vendas baseado em frete.

3.4 Análise da Relação entre Frete e Demanda por Região e Categoria de Produtos

Para a etapa de cruzamento dos dados com informações de frete, foi compilada a Tabela 2 - Média de Frete por Categoria e Região, com o propósito de analisar a relação entre categorias de produtos, as regiões Nordeste e Sudeste, e as médias de frete aplicadas. Entretanto, as análises do próximo capítulo se concentrarão apenas no Nordeste para seguir-se a noção de afunilamento dos dados.

Tabela 2 – Média de Frete por categoria e Região

Região	Categoria	Média de Frete
Nordeste	beleza_saude	33.152004
	esporte_lazer	33.999781
	informatica_acessorios	27.386456
	relogios_presentes	26.436490
Sudeste	beleza_saude	15.806400
	esporte_lazer	18.303985
	informatica_acessorios	16.282827
	relogios_presentes	14.646492
Total		23.252304375

Fonte: Autoria própria.

Através da visualização do comportamento da demanda em relação ao frete, representado na Figura 13 - Demanda em relação ao Frete - Região Nordeste e Sudeste, fica evidente que o frete praticado para produtos na Região Nordeste é significativamente mais alto em comparação com o Sudeste, chegando a ultrapassar diversas vezes o patamar de R\$150,00 (pontos em vermelho). Além disso, o frete para a Região Nordeste apresenta uma distribuição em valores mais concentrados em torno de 50 reais.

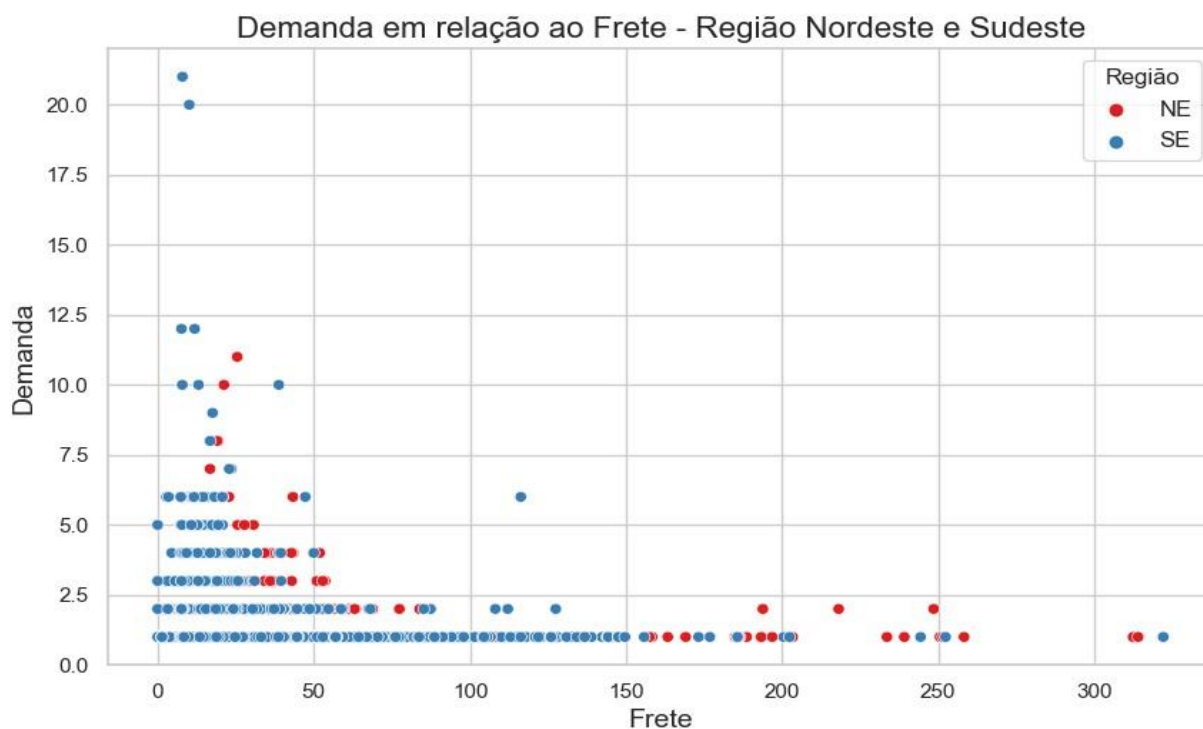


Figura 13 - Demanda em relação ao Frete - Região Nordeste e Sudeste

Fonte: Autoria Própria

4 Resultados e Discussões

A seção de Resultados e Discussões compreende os resultados obtidos com a implementação do modelo, seguindo as etapas delineadas na seção de metodologia. Isso envolve tanto o Modelo de Previsão de Frete para o Nordeste quanto a Proposta de Teste de Frete Grátis para a região Nordeste. O objetivo é apresentar conclusões relacionadas a cada um dos resultados cenários considerados.

4.1 Modelo de Previsão de Frete Nordeste

No processo de análise dos dados de frete para a região Nordeste, um modelo de previsão SARIMA foi desenvolvido com foco na categoria 'relógios_presentes'. Foram realizados testes com diversas combinações de ordens (p, d, q) e sazonalidades (P, D, Q, S), parâmetros estes descritos a seguir com o objetivo de encontrar a configuração que resultasse no menor erro de previsão, representado pelo *Root Mean Square Error* (RMSE). O modelo escolhido dentre os testes está apresentado na Tabela 3.

Tabela 3 – SARIMA Frete do Nordeste para relógios_presentes

order	seasonal_order	RMSE
(1, 1, 1)	(2, 1, 2, 60)	16.54643

Fonte: Autoria própria

O modelo SARIMA utiliza os parâmetros ‘order’ e ‘seasonal_order’ para modelar as características não sazonais e sazonais dos dados, respectivamente:

1. “order = (1, 1, 1)”:
 - ‘p (Ordem do componente AR) = 1’: Considera uma observação passada para prever a atual, usando termos de autorregressão;
 - ‘d (Ordem da diferenciação) = 1’: Realiza uma diferenciação nos dados para torná-los estacionários, evitando tendências;
 - ‘q (Ordem do componente MA) = 1’: Considera um erro passado para prever o erro atual, usando termos de média móvel;
2. ‘seasonal_order = (2, 1, 2, 60)’:
 - ‘P (Ordem do componente AR sazonal) = 2’: Considera duas observações passadas, separadas por 60 dias, para prever a atual, usando termos de autorregressão sazonal;

- ‘D (Ordem da diferenciação sazonal) = 1’: Realiza uma diferenciação sazonal nos dados para torná-los estacionários;
- ‘Q (Ordem do componente MA sazonal) = 2’: Considera dois erros passados, separados por 60 dias, para prever o erro atual, usando termos de média móvel sazonal;
- ‘S (Período sazonal) = 60’: Indica que os padrões sazonais se repetem a cada 60 dias.

A Figura 14 ilustra a previsão do frete para a categoria 'relógios_presentes' na região Nordeste, onde o modelo SARIMA captura as tendências e sazonalidades nos dados, com base nos parâmetros mencionados anteriormente. Observa-se que o modelo apresenta uma previsão precisa até março de 2018.

No entanto, após esse período, os dados começam a divergir do padrão devido ao último pico sazonal do Verão Nordestino, que não pôde ser totalmente previsto pelo modelo. Essa previsão desempenha um papel fundamental na tomada de decisões relacionadas ao gerenciamento de estoque, planejamento de recursos e otimização de custos de frete na região Nordeste, permitindo que as empresas se adaptem de forma eficaz às variações sazonais e à demanda flutuante. Essa análise enfatiza a importância de modelos de previsão avançados na aprimoração da logística e da gestão de recursos em empresas de comércio eletrônico.

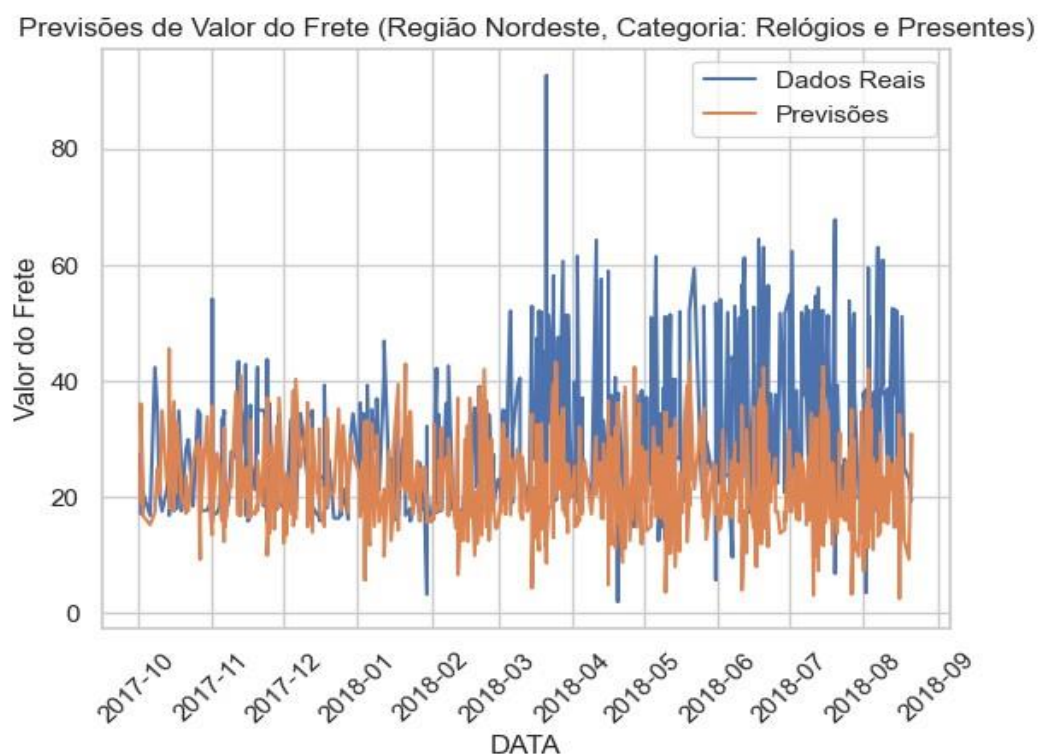


Figura 14 – Previsão de Frete no Nordeste

Fonte: Autoria Própria

4.2 Proposta teste de frete grátis para o Nordeste

Para a proposta de frete grátis, adotou-se uma estratégia de alta granularidade de dados, visando embasar a decisão com o máximo de informações disponíveis. Nesse contexto, o produto Beta foi escolhido devido ao seu alinhamento com as premissas do modelo hipotético de proposta. Especificamente, o produto Beta pertence à categoria de 'relógios_presentes' e já havia passado por um período promocional de frete grátis. Durante essa promoção, o produto obteve resultados satisfatórios tanto na região Sudeste (SE) quanto na região Nordeste (NE), conforme evidenciado nas Figuras 15 e 16, que destacam as vendas de relógios antes e depois da implementação do frete grátis nas respectivas regiões.

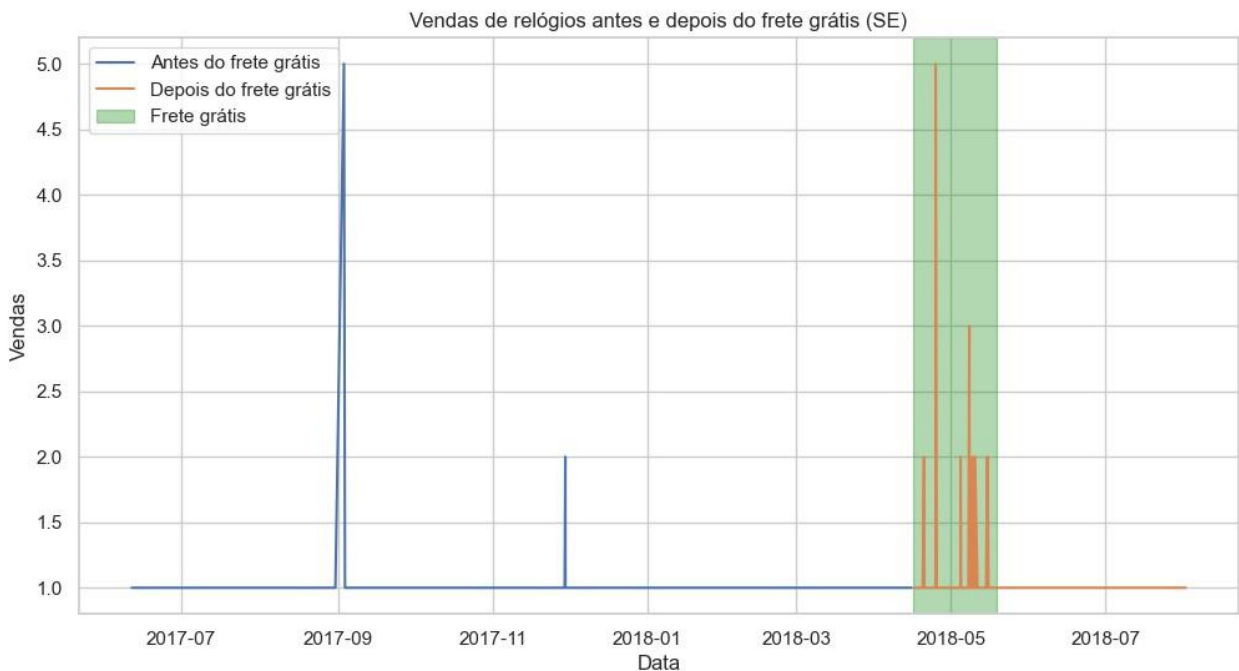


Figura 15 – Comportamento do produto Beta no Sudeste

Fonte: Autoria Própria



Figura 16 – Comportamento do produto Beta no Nordeste

Fonte: Autoria Própria

Essas figuras oferecem uma visão clara do impacto positivo do frete grátis nas vendas do produto Beta em ambas as regiões. A partir dessas informações, foi possível embasar a escolha do produto e da estratégia de frete grátis para o Nordeste. Essa análise aprofundada contribui para a formulação de estratégias de marketing e precificação direcionadas, visando impulsionar as vendas em uma região específica.

Após confirmar que o produto Beta atendia às premissas do modelo, a análise foi realizada para obter uma política que aumentasse o retorno financeiros da empresa. O preço fixo do produto escolhido foi de R\$ 99,9, correspondente ao preço praticado durante o período promocional de maio de 2018. Além disso, a demanda inicial da proposta foi fundamentada na demanda total do produto na região nordestina, totalizando 40 unidades. O valor médio de frete utilizado foi de R\$26,43, Tabela 2.

A proposta teste incorporou a variação do aumento da demanda, variando desde 1% até 100%. Esse espectro de variação da demanda visa avaliar a viabilidade a partir elasticidade da demanda. A Tabela 4 - Teste de viabilidade de Frete grátis para o produto Beta ilustra essa análise em detalhes, fornecendo informações cruciais para a formulação e avaliação da proposta de frete grátis.

A aplicação desse teste de viabilidade permite entender como a demanda reage a

diferentes níveis de aumento, possibilitando a identificação de cenários mais favoráveis e estratégias eficazes relacionadas à proposta de frete grátis para o Nordeste.

Tabela 4 – Teste de viabilidade de Frete grátis para o produto Beta

Aumento Demanda (%)	Receita (R\$)	Custo do Frete (R\$)	Lucro (R\$)	Vantajoso?
0	3996.0	1057.20	2938.80	Sim
5	4195.8	1110.06	3085.74	Sim
10	4395.6	1162.92	3232.68	Sim
15	4595.4	1215.78	3379.62	Sim
20	4795.2	1268.64	3526.56	Sim
25	4995.0	1321.50	3673.50	Sim
30	5194.8	1374.36	3820.44	Sim
35	5394.6	1427.22	3967.38	Sim
40	5594.4	1480.08	4114.32	Sim
45	5794.2	1532.94	4261.26	Sim
50	5994.0	1585.80	4408.20	Sim
55	6193.8	1638.66	4555.14	Sim
60	6393.6	1691.52	4702.08	Sim
65	6593.4	1744.38	4849.02	Sim
70	6793.2	1797.24	4995.96	Sim
75	6993.0	1850.10	5142.90	Sim
80	7192.8	1902.96	5289.84	Sim
85	7392.6	1955.82	5436.78	Sim
90	7592.4	2008.68	5583.72	Sim
95	7792.2	2061.54	5730.66	Sim
100	7992.0	2114.40	5877.60	Sim

Fonte: Autoria própria

Os resultados demonstram claramente a eficácia da estratégia de frete grátis para o produto Beta na região Nordeste. À medida que a demanda aumenta, a receita cresce consistentemente, com um custo de frete que se mantém estável. Isso indica que a empresa pode absorver o custo adicional do frete sem prejudicar substancialmente sua margem de lucro. Em todos os cenários analisados da elasticidade da demanda, o lucro líquido aumenta à medida que a demanda cresce, tornando a estratégia de frete grátis altamente vantajosa em comparação com a ausência de promoção. A análise respalda de forma sólida a viabilidade e o impacto positivo

dessa estratégia para o produto Beta na região Nordeste, reforçando a importância de considerar a adoção do frete grátis como uma ação promocional benéfica e lucrativa.

Sendo assim, considerando as premissas específicas do produto Beta, o teste revela uma resposta positiva à implementação do frete grátis na região Nordeste. Durante o teste, ficou claro que essa promoção resulta em economias significativas em comparação com a receita de referência, que é de R\$ 2.938,80, consolidando ainda mais a viabilidade e o impacto positivo dessa estratégia.

5 Conclusão

Este trabalho explorou a influência do frete grátis nas estratégias de marketing e precificação em ambientes de comércio eletrônico. A metodologia incorporou um modelo de previsão SARIMA para analisar sobre a dinâmica da demanda das vendas na plataforma Olist, com um enfoque específico na região Nordeste do Brasil.

Os objetivos estabelecidos foram cumpridos, contribuindo para o aprimoramento das estratégias de vendas da plataforma Olist, aprofundando então a compreensão do impacto do frete grátis. O primeiro objetivo direcionou a análise dos dados de vendas da Olist, proporcionando *insights* preliminares sobre o comportamento das vendas. O segundo objetivo buscou verificar padrões e identificar a metodologia mais apropriada, resultando na seleção da metodologia SARIMA para análise de séries temporais da demanda dos modelos. Este projeto permitiu assim uma análise detalhada do impacto do frete grátis em produtos de média e baixa demanda, fornecendo insights valiosos sobre a viabilidade econômica dessa estratégia. Um modelo foi proposto com a expectativa de avaliação do retorno de vendas com a proposta de frete grátis. Por fim, a previsão de demanda SARIMA foi realizada com a aplicação do modelo.

O trabalho desenvolvido está assim de acordo com a teoria de que a estratégia de promoção do frete grátis é uma tática eficaz para impulsionar as vendas e influenciar o comportamento do cliente no comércio eletrônico. Além disso, a metodologia empregada, incluindo o modelo SARIMA, permitiu uma análise das tendências de vendas, fornecendo dados que respaldam a eficácia dessa estratégia.

Este estudo não apenas contribui para a compreensão do impacto do frete grátis nas vendas da plataforma Olist, mas também reforça a importância de adotar abordagens de análise avançada, como a previsão SARIMA, na tomada de decisões em ambientes de comércio eletrônico. A possibilidade de capturar padrões sazonais e tendências em tempo real proporciona às empresas uma vantagem competitiva valiosa.

Portanto, os resultados deste estudo não apenas validam o potencial do frete grátis mas a combinação de teoria, metodologia e resultados contribui para uma visão holística que pode orientar as empresas na otimização de suas operações e na maximização de seus resultados financeiros. Contudo, é importante reconhecer as limitações deste estudo, os dados de domínio público e o foco restringiu-se ao produto Beta. Há oportunidade que pesquisas futuras ampliem a análise para uma gama mais diversificada de produtos e categorias, considerando colaborações

com empresas para acessar dados privados. Isso possibilitará a do frete grátis, avaliando seu impacto em diversas regiões do Brasil e em outras plataformas de e-commerce. Além disso, a inclusão de variáveis adicionais, como a qualidade do serviço ao cliente, pode enriquecer as análises futuras.

Em resumo, este estudo contribui para o entendimento das estratégias de frete grátis no contexto brasileiro, fornecendo informações valiosas para as empresas que desejam aprimorar suas operações e maximizar seus retornos financeiros. Ao integrar teoria, metodologia e resultados, abre-se um caminho promissor para pesquisas futuras e aprimoramentos nas estratégias de comércio eletrônico.

Referências

BELFIORE, P.; FÁVERO, L. P. **Pesquisa Operacional para cursos de Engenharia**. Elsevier Brasil, 2013. v. 1.

BOYD, E., BILEGAN, I. **Revenue Management and E-Commerce Management Science**, 2003

GZH. **Americanas foi fundada em 1929, no Rio; conheça a história**, Gauchazh, 2023
Disponível em: <https://encurtador.com.br/csN89>. Acesso em: Outubro de 2023

Hart CW, SCHLESINGER LA, MAHER D. **Guarantees come to professional service firms**. Sloan Manage Rev. 1992

Hillier, F. S.; Lieberman, G. J. **Introdução à pesquisa operacional**. McGraw Hill Brasil, 2013.

IBGE. **Instituto Brasileiro de Geografia e Estatística**. Disponível em:
<https://www.ibge.gov.br/estatisticas/economicas/comercio/9075-pesquisa-anual-decomercio.html?edicao=34674&t=o-que-e>

LANTZ B, HJOT K. **Real e-customer behavioural responses to free delivery and free returns**. Electron Commer Res 13, 183–198 (2013)

LEWIS MICHAEL. **The effect of shipping fees on customer acquisition, customer retention, and purchase quantities**. Journal of Retailing, 2006

LISKOV, B., SNYDER, A. **Abstraction and Specification in Program Development**. ACM Press, 1986.

MEISEL, S., MATTFELD, D. **Synergies of Operations Research and Data Mining**, European Journal of Operational Research, 2010

MORETTIN, P. BUSSAB, W., **ESTATÍSTICA BÁSICA**, 9 ed, Editora Saraiva. 2017

OLIST; SIONEK, ANDRÉ (2018). **Brazilian E-Commerce Public Dataset** by Olist Data set Kaggle <https://doi.org/10.34740/KAGGLE/DSV/195341/> Licença: [CC BY-NC-SA 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/) Acesso em: Novembro, 2022

OLIST. **Sobre o Olist**. Disponível em: <https://olist.com/pt-br/sobre-nos/>. Acesso em: 15/07/2023

OLAFSSON, S., XIAONAN, L., SHUNING W. **Operations research and data mining**
European Journal of Operational Research, 2008

PayPal Brasil e BigDataCorp. **Perfil do E-Commerce Brasileiro**. 2022
<https://public.flourish.studio/story/1683536/> Acesso em: Outubro de 2023

Portal da indústria **Perfil da indústria brasileira**. 2020 Disponível em:
<https://industriabrasileira.portaldaindustria.com.br/grafico/total/producao/#!/industria-total>
Acesso em: Outubro de 2023

PREVIGLIANO, F., VULCANO, G., **Managing Uncertain Capacities for Network Revenue Optimization** MANUFACTURING & SERVICE OPERATIONS MANAGEMENT 2021

PYKE S, JOHNSON ME, DESMOND P **E-Fulfillment, It's Harder Than It Looks** Supply Chain Management Review (January/February) (2001), pp. 26-32

RAI, H., VERLINDE, S. and MACHARIS, C. "**The “next day, free delivery” myth unravelled: Possibilities for sustainable last mile transport in an omnichannel environment**", International Journal of Retail & Distribution Management, (2019)

SANTOS, A. B.; OLIVEIRA, R. S. **Estratégias de otimização do retorno de vendas no comércio eletrônico**. Revista de Administração Online, v. 12, n. 3, p. 52-68, 2019.

SEBRAE. **Serviço Brasileiro de Apoio às Micro e Pequenas Empresas Taxa de Sobrevivência** Disponível em: <<https://sebrae.com.br/sites/PortalSebrae/artigos/a-taxa-desobrevivencia-das-empresas-no-brasil> >

TASSABEHJI, R. **Applying E-Commerce in Business**, SAGE Publications Ltd; First Edition, 2003

Z. ZHAO, J. WANG, H. SUN, Y. LIU, Z. FAN AND F. XUAN, "**What Factors Influence Online Product Sales? Online Reviews, Review System Curation, Online Promotional Marketing and Seller Guarantees Analysis**," IEEE Access, 2019, vol. 8, pp. 3920-3931

Apêndice

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import LinearRegression
#TRATANDO OS DADOS 1
# Carregando as tabelas CSV em dataframes customers_df
=
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB//PG/dados/olist_customers_dataset.csv')
customers_df.head()

payments_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_order_payments_dataset.csv')
payments_df.head()

sellers_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_sellers_dataset.csv')
sellers_df.head()

geolocation_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_geolocation_dataset.csv')
geolocation_df.head()

reviews_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_order_reviews_dataset.csv')
reviews_df.head()

items_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_order_items_dataset.csv')
items_df.head()
```



```
products_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_products_dataset.csv
')
products_df.head()

orders_df =
pd.read_csv('C:/Users/natal/OneDrive/UnB/UnB/PG/dados/olist_orders_dataset.csv')
orders_df.head()

# Juntando as tabelas com o ID
merged_df = pd.merge(orders_df, items_df, on='order_id')
merged_df = merged_df.merge(reviews_df, on='order_id') merged_df
= merged_df.merge(payments_df, on='order_id') merged_df =
merged_df.merge(products_df , on='product_id') merged_df =
merged_df.merge(customers_df , on='customer_id')
#merged_df.head()
print(merged_df.to_string())

##GRÁFICO/TABELA ORDER POR STATUS
order_status_counts = merged_df['order_status'].value_counts()
print(order_status_counts)
ax = order_status_counts.plot (kind='bar') plt.title('Orders
por status')
ax.bar_label(ax.containers[0], padding = 2) plt.show()

##GRÁFICOS DE CURVA DE TENDÊNCIA
# Convertendo a coluna de data para datetime
merged_df['order_purchase_timestamp'] =
pd.to_datetime(merged_df['order_purchase_timestamp'])

# Agrupando por mês e contando o número de compras em cada mês grouped
=
merged_df.groupby(merged_df['order_purchase_timestamp'].dt.to_period("M"))
order_purchase_counts = grouped['order_purchase_timestamp'].count()
# Criando o gráfico plt.figure(figsize=(10,
6))
```



```
order_purchase_counts.plot(kind='line') #
Adicionando título e labels dos eixos
plt.title('Distribuição da demanda no
período') plt.xlabel('Mês')
plt.ylabel('Número de Compras')
###GRÁFICO DE REGRESSÃO LINEAR
# Ajustando a regressão linear
X = np.arange(len(order_purchase_counts)).reshape(-1,
1) y = order_purchase_counts.values.reshape(-1, 1) reg
= LinearRegression().fit(X, y)
# Obtendo as previsões da curva de
tendência trend_line = reg.predict(X) #
Plotando a curva de tendência
plt.plot(order_purchase_counts.index, trend_line, color='red', linestyle='--
', label='Curva de Tendência (Regressão Linear)') plt.legend() plt.show()

###GRÁFICO DE POLÍNÔMIO
# Ajustando um polinômio de grau 2 aos dados para obter a curva de tendência
x = np.arange(len(order_purchase_counts))
coefficients = np.polyfit(x, order_purchase_counts, 3)
trend_line = np.polyval(coefficients, x)
# Plotando a curva de tendência
plt.plot(order_purchase_counts.index, trend_line, color='red', linestyle='--
', label='Curva de Tendência') plt.legend() plt.show()

##GRÁFICO DE MÉDIA MÓVEL
# Calculando a média móvel com janela de tamanho 3 (por exemplo)
rolling_average = order_purchase_counts.rolling(window=3, min_periods=1).mean()
# Plotando a média móvel como a curva de tendência
plt.plot(order_purchase_counts.index, rolling_average, color='red', linestyle='-
-', label='Curva de Tendência (Média
Móvel)') plt.legend() plt.show()
```



```
##GRÁFICO DE ESTADO
customer_state_counts = merged_df['customer_state'].value_counts()
# Criando a figura e os eixos fig, ax
= plt.subplots(figsize=(10, 6))
customer_state_counts.plot(kind='barh', ax=ax) # Plotando o gráfico de barras na
horizontal
plt.title('Orders por estado') # Adicionando título
for container in ax.containers: # Adicionando os valores nas barras
ax.bar_label(container, padding=8, fontsize=10)
ax.tick_params(axis='y', pad=8) # Ajustando espaçamento dos ticks do eixo y
plt.show()

#GRÁFICO POR REGIÃO # Agrupando por região merged_df['Região'] =
merged_df['customer_state'].map({'SP': 'Sudeste', 'RJ':
'Sudeste', 'MG': 'Sudeste', 'ES': 'Sudeste',
'RS': 'Sul', 'PR': 'Sul',
'SC': 'Sul',
'BA': 'Nordeste',
'PE': 'Nordeste', 'CE': 'Nordeste', 'PA': 'Nordeste',
'PB': 'Nordeste',
'RN': 'Nordeste', 'PI': 'Nordeste', 'MA': 'Nordeste',
'SE': 'Nordeste', 'AL':
'Nordeste',
'DF': 'Centro-
Oeste', 'GO': 'Centro-Oeste', 'MT': 'Centro-Oeste',
'MS': 'Centro-Oeste',
'TO': 'Centro-Oeste',
'RO': 'Norte', 'AM':
'Norte', 'AP': 'Norte', 'RR': 'Norte', 'AC': 'Norte'})

grouped = merged_df.groupby('Região')
customer_region_counts = grouped['Região'].size()
# Criando a figura e os eixos fig, ax
= plt.subplots(figsize=(10, 6))
customer_region_counts.plot(kind='barh', ax=ax)# Plotando o gráfico de barras na
horizontal
plt.title('Orders por Região')
```



```
for container in ax.containers: # Adicionando os valores nas barras
ax.bar_label(container, padding=8, fontsize=10)

ax.tick_params(axis='y', pad=8) # Ajustando espaçamento dos ticks do eixo y
plt.show()

#TRATANDO OS DADOS 2
merged_df['product_category_name'].fillna(value='sem cat', inplace=True)
#Renomeando celulas com NaN

unique_count = merged_df['product_category_name'].nunique() #Identificando
categorias disponíveis
unique_values = merged_df['product_category_name'].unique()
print(unique_values,unique_count)

counts = merged_df['product_category_name'].value_counts() #Quantidade
de ocorrencias de cada cat print(counts)

grouped = merged_df.groupby(['Região', 'product_category_name']) #Agupando dados
de categoria por região
counts2 = grouped['product_category_name'].count() print(counts2)

##GRÁFICOS TOP 8 CATEGORIAS POR REGIÃO - são escolhidas 8 das 74 para que haja
significancia de dados do Nordeste
grouped = merged_df.groupby(['product_category_name',
'Região']).size().reset_index(name='count') # Agrupar por região e categoria
# Obter as top 8 categorias mais frequentes
counts = merged_df['product_category_name'].value_counts() top_categories
= counts.nlargest(8).index
grouped = grouped[grouped['product_category_name'].isin(top_categories)]
grouped = grouped.sort_values(by='count', ascending=False) # Decrescente

fig, axes = plt.subplots(nrows=2, ncols=4, sharey=True, figsize=(20,
10))# Configurar o subplot axes = axes.flatten()
```

```

for i, category in enumerate(top_categories):# Plotar gráfico para cada
categoria
    region_group = grouped[grouped['product_category_name'] == category]
plot = region_group.plot(x='Região', y='count', kind='bar', ax=axes[i])
axes[i].set_title(category)
axes[i].set_ylabel('Quantidade de pedidos')

for p in plot.patches:# Adicionar a quantidade de pedidos acima de cada barra
    height = p.get_height()
    axes[i].annotate(f'{height}', xy=(p.get_x() + p.get_width() / 2,
height), xytext=(0, 3),
                    textcoords='offset points', ha='center', va='bottom')
# Ajustar layout e mostrar o
gráfico plt.tight_layout()
plt.show()

#TRATANDO OS DADOS 3
#Agrupando as colunas de interesse para afinar a análise
grouped_reg = merged_df.groupby(['Região', 'product_category_name',
pd.Grouper(key='order_purchase_timestamp', freq='M'), 'product_id',
        'freight_value']).size().reset_index(name='count')
print(grouped_reg)
# agrupando por Região, product_category_name, order_purchase_timestamp,
product_id e freight
grouped_df = merged_df.groupby(['Região', 'product_category_name',
'order_purchase_timestamp', 'product_id', 'freight_value', 'price'])
result_df = grouped_df.agg(demanda=('product_id', 'count')) # calculando a
contagem de ocorrências de cada product_id
result_df = result_df.reset_index()# resetando o índice do dataframe
print(result_df)# exibindo o resultado

```



```

##GRÁFICO DEMANDA PELO PERÍODO PARA CADA CATEGORIA - as 4 mais representativas
nordeste
#Filtrando categorias de interesse - as maiores de Nordeste e Sudeste
categorias_interesse = ['beleza_saude', 'relogios_presentes', 'esporte_lazer',
'informatica_acessorios']
df_filtrado =
result_df[result_df['product_category_name'].isin(categorias_interesse)]
print(df_filtrado)

# Lista de categorias de produto finais
categorias = ['beleza_saude', 'relogios_presentes', 'esporte_lazer',
'informatica_acessorios']
regioes = ['Sudeste', 'Nordeste'] # Lista de regiões (agora só com 'Sudeste' e
'Nordeste')
fig, axs = plt.subplots(nrows=len(categorias), ncols=len(regioes), figsize=(20,
15)) # Cria o subplot

for i, categoria in enumerate(categorias):# Para cada categoria e região,
cria um gráfico de volume de vendas por mês      for j, regioao in
enumerate(regioes):
    # Filtra as linhas da categoria e região correspondentes
    df_cat_regiao = df_filtrado[(df_filtrado['product_category_name'] ==
categoria) & (df_filtrado['Região'] == regioao)]

    # Agrupa os valores da coluna 'demanda' por mês e soma
vendas_por_mes =
df_cat_regiao.groupby(pd.Grouper(key='order_purchase_timestamp',
freq='M'))['demanda'].sum()

    # Plota o gráfico na posição (i, j)
axs[i, j].plot(vendas_por_mes)
axs[i, j].set_title(f'Distribuição da demanda {categoria} - {regiao}')
axs[i, j].set_xlabel('Mês')
axs[i, j].set_ylabel('Volume de
vendas') plt.tight_layout() plt.show()

#TABELA DE MÈDIA DE FRETE

```



```

# Agrupa por região, categoria e mês e calcula a média de frete e a soma da
demanda
df_filtrado_agrupado = df_filtrado.groupby(['Região', 'product_category_name',
pd.Grouper(key='order_purchase_timestamp', freq='M')]).agg({'freight_value':
'mean', 'demanda': 'sum'}).reset_index()

# Calcula a média de frete por região e categoria
media_frete = df_filtrado_agrupado.groupby(['Região',
'product_category_name'])['freight_value'].mean().reset_index()
# Exibe a tabela com a média de frete por região e categoria
print(media_frete)

##GRÁFICO DEMANDA EM RELAÇÃO AO FRETE - REGIÃO NORDESTE E SUDESTE
import seaborn as sns
# Definindo um dicionário para substituir o nome das regiões pelos seus
acrônimos regioes_dict = {'Sul': 'S', 'Sudeste': 'SE', 'Centro-Oeste': 'CO',
'Nordeste':
'NE', 'Norte': 'N'}
# Aplicando o dicionário ao dataframe
df_filtrado['Região'] = df_filtrado['Região'].replace(regioes_dict)
# Filtrando os dados para incluir apenas as regiões Nordeste (NE) e Sudeste
(SE) df_nordeste_sudeste = df_filtrado[df_filtrado['Região'].isin(['NE',
'SE'])] # Plot do gráfico de dispersão para a comparação entre Nordeste e
Sudeste sns.set(style="whitegrid") plt.figure(figsize=(10, 6))
g = sns.scatterplot(data=df_nordeste_sudeste, x='freight_value', y='demanda',
hue='Região', palette='Set1')
g.set_title("Demanda em relação ao Frete - Região Nordeste e Sudeste",
fontsize=16)
g.set_xlabel("Frete", fontsize=14)
g.set_ylabel("Demanda", fontsize=14)
g.legend(title='Região', fontsize=12)
plt.show()

##### ANÁLISES
#####

```



```
media_frete_por_preco = df_nordeste_sudeste.groupby(['Região',
'price'])['freight_value'].mean().reset_index()

# Indicação se o frete é gratuito ou não, considerando o valor máximo do frete
para cada região:
media_frete_por_preco['frete_gratis'] = False
media_frete_por_preco.loc[media_frete_por_preco['Região'] == 'NE',
'frete_gratis'] = media_frete_por_preco[media_frete_por_preco['Região'] ==
'NE']['freight_value'] <= media_frete_por_preco[media_frete_por_preco['Região']
== 'NE']['freight_value'].max()
media_frete_por_preco.loc[media_frete_por_preco['Região'] == 'SE',
'frete_gratis'] = media_frete_por_preco[media_frete_por_preco['Região'] ==
'SE']['freight_value'] == 0

#SARIMA
from statsmodels.tsa.statespace.sarimax import
SARIMAX from sklearn.metrics import
mean_squared_error import itertools
import matplotlib.pyplot as plt
import datetime as dt

df_nordeste_sudeste = df_filtrado[df_filtrado['Região'].isin(['NE', 'SE'])]
# Converter coluna 'order_purchase_timestamp' para formato de data
df_nordeste_sudeste['order_purchase_timestamp'] =
pd.to_datetime(df_nordeste_sudeste['order_purchase_timestamp'], format='%Y-%m')

# Criar um DataFrame apenas para a região Nordeste
df_nordeste = df_nordeste_sudeste[df_nordeste_sudeste['Região'] == 'NE']
# Filtrar a categoria 'relogios_presentes' e remover dados com frete igual a
zero
df_nordeste_relogios_presentes =
df_nordeste[(df_nordeste['product_category_name'] == 'relogios_presentes') &
(df_nordeste['freight_value'] != 0)]

# Dividir os dados em conjunto de treino e teste
```



```
mask = media_frete_por_preco_ARIMA['order_purchase_timestamp'] >= data_teste
treino =
media_frete_por_preco_ARIMA.loc[~mask].set_index('order_purchase_timestamp')
teste = media_frete_por_preco_ARIMA.loc[mask].reset_index(drop=True) #
Redefinir o índice para numérico

treino.index = treino.index.date
teste.index = teste.index.date

# Criar o modelo SARIMA e treinar com os dados de treino
modelo_sarima = SARIMAX(treino['freight_value'], order=(1, 1, 1),
seasonal_order=(2, 1, 2, 60))
modelo_sarima_treinado = modelo_sarima.fit()

# Fazer previsões para o conjunto de teste
previsoes = modelo_sarima_treinado.predict(start=len(treino), end=len(treino) +
len(teste) - 1)

# Calcular o erro RMSE
erro = mean_squared_error(teste['freight_value'], previsoes, squared=False)
print('RMSE:', erro)

# Plotar os dados reais
plt.plot(teste.index, teste['freight_value'], label='Dados Reais')
# Plotar as previsões
plt.plot(teste.index, previsoes, label='Previsões')

# Definir rótulos e título do
gráfico plt.xlabel('DATA')
plt.ylabel('Valor do Frete')
plt.title('Previsões de Valor do Frete (Região Nordeste, Categoria: Relógios e
Presentes)')

# Exibir a legenda
plt.legend()

# Exibir o gráfico
```



```

plt.show()

#### PROPOSTA A PARTIR DO PRODUTO
# Converter coluna 'order_purchase_timestamp' para formato de data
df_nordeste_norte_sudeste['order_purchase_timestamp'] =
pd.to_datetime(df_nordeste_norte_sudeste['order_purchase_timestamp'],
format='%Y-%m-%d')

# Filtrar os produtos específicos e as regiões Nordeste, Norte e
Sudeste P1 = ['53b36df67ebb7c41585e8d54d6772e08'] regioes = ['NE',
'SE'] df_filtrado =
df_nordeste_norte_sudeste[(df_nordeste_norte_sudeste['product_id'].isin(P1)) &
(df_nordeste_norte_sudeste['product_category_name'] == 'relogios_presentes') &
(df_nordeste_norte_sudeste['Região'].isin(regioes))]

# Agrupar por região, data e produto, calcular a soma das vendas
vendas_por_regiao_data_produto = df_filtrado.groupby(['Região',
'order_purchase_timestamp', 'product_id'])['demanda'].sum().reset_index()
# Dividir as vendas em antes e depois da aplicação do frete grátis
data_inicio_frete_gratis = '2018-04-16' data_fim_frete_gratis =
'2018-05-19' vendas_antes_frete_gratis =
vendas_por_regiao_data_produto[vendas_por_regiao_data_produto['order_purchase_timestamp'] < data_inicio_frete_gratis]
vendas_depois_frete_gratis =
vendas_por_regiao_data_produto[vendas_por_regiao_data_produto['order_purchase_timestamp'] >= data_inicio_frete_gratis]

# Plotar os gráficos das vendas por região
for regioao in regioes:
vendas_antes_regiao =
vendas_antes_frete_gratis[vendas_antes_frete_gratis['Região'] == regioao]
vendas_depois_regiao =
vendas_depois_frete_gratis[vendas_depois_frete_gratis['Região'] == regioao]

```

```
plt.figure(figsize=(12, 6))
plt.plot(vendas_antes_regiao['order_purchase_timestamp'],
vendas_antes_regiao['demanda'], label='Antes do frete grátis')
plt.plot(vendas_depois_regiao['order_purchase_timestamp'],
vendas_depois_regiao['demanda'], label='Depois do frete grátis')
plt.axvspan(data_inicio_frete_gratis, data_fim_frete_gratis,
color='green', alpha=0.3, label='Frete grátis')    plt.xlabel('Data')
plt.ylabel('Vendas')
plt.title(f'Vendas de relógios antes e depois do frete grátis
({regiao})')    plt.legend()    plt.show()

##COMPORTAMENTO DO PRODUTO
produto_id = '53b36df67ebb7c41585e8d54d6772e08'

# Filtrar as linhas correspondentes ao produto específico
df_produto = df_nordeste_norte_sudeste[df_nordeste_norte_sudeste['product_id']
== produto_id]

# Calcular a soma da demanda por região
soma_demanda_regiao = df_produto.groupby('Região')['demanda'].sum()

print('Soma da Demanda por Região:')
print(soma_demanda_regiao)

### TABELA DE PROPOSTA COM ANALISE DE ELASTICIDADE
# valores para aumento de demanda de 1 a 100
# Novos valores para aumento de demanda de 5 em 5% começando em 0 até
100% aumento_demanda = list(range(0, 101, 5)) price_fixo = 99.9
media_frete_ne = 26.43 demanda_inicial_ne = 40

# Função para calcular a receita adicional def
calcular_receita_adicional(aumento_demanda, price_fixo, demanda_inicial):
```

```
    demanda_total = demanda_inicial + (demanda_inicial * (aumento_demanda /
100))
    return demanda_total * price_fixo

# Função para calcular o custo do frete def
calcular_custo_frete(aumento_demanda, media_frete, demanda_inicial):
demanda_total = demanda_inicial + (demanda_inicial * (aumento_demanda /
100))
    return demanda_total * (media_frete)

# Função para calcular a diferença entre a receita adicional e o custo do frete
def calcular_diferenca(receita_adicional, custo_frete):
    return receita_adicional - custo_frete

# Função para determinar se a diferença vale a pena
def determinar_vale_a_pena(diferenca):
    return ['Sim' if diff > 0 else 'Não' for diff in diferenca]

# Tabela para a Região NE
tabela_ne = pd.DataFrame({'Aumento Demanda': aumento_demanda})
tabela_ne['Receita'] = calcular_receita_adicional(tabela_ne['Aumento Demanda'],
price_fixo, demanda_inicial_ne)
tabela_ne['Custo do Frete'] = calcular_custo_frete(tabela_ne['Aumento Demanda'],
media_frete_ne, demanda_inicial_ne)
tabela_ne['Lucro'] = calcular_diferenca(tabela_ne['Receita'], tabela_ne['Custo
do Frete'])
tabela_ne['Vantajoso?'] = determinar_vale_a_pena(tabela_ne['Lucro'])

# Exibir as tabelas
print()
print("Tabela para a Região NE:")
print(tabela_ne.to_string(index=False))
```