

**UNIVERSIDADE PAULISTA  
INSTITUTO DE CIÊNCIAS EXATAS E TECNOLOGIA**

André Luiz Oliveira Maldonado N7014J5  
Caio Teixeira de Lima Saraiva N7297F6  
Francieli dos Santos Muniz G2437E2  
Gabriel Aparecido Magalhães de Brito G191650  
João Vitor Soares de Paula G26BAH5  
Lais Falcochio de Camargo G2221J2

**SIMULAÇÃO DE E-COMMERCE COM PREVISÃO DE VENDAS:  
UM ESTUDO DE REGRESSÃO LINEAR E VISUALIZAÇÃO DE DADOS**

**SANTANA DE PARNAÍBA - SP**

**2024**

André Luiz Oliveira Maldonado N7014J5  
Caio Teixeira de Lima Saraiva N7297F6  
Francieli dos Santos Muniz G2437E2  
Gabriel Aparecido Magalhães de Brito G191650  
João Vitor Soares de Paula G26BAH5  
Lais Falcochio de Camargo G2221J2

**SIMULAÇÃO DE E-COMMERCE COM PREVISÃO DE VENDAS:  
UM ESTUDO DE REGRESSÃO LINEAR E VISUALIZAÇÃO DE DADOS**

Dissertação de Conclusão de Curso  
apresentada à Universidade Paulista – UNIP  
para a obtenção do título de bacharel em  
Ciência da Computação.

Orientador Prof. Me. Edy Carlos Hidemi  
Hayashida

**SANTANA DE PARNAÍBA – SP**

**2024**

## Ficha Catalográfica

### CIP - Catalogação na Publicação

SIMULAÇÃO DE E-COMMERCE COM PREVISÃO DE VENDAS: UM ESTUDO DE REGRESSÃO LINEAR E VISUALIZAÇÃO DE DADOS / Francieli Muniz; André Maldonado; Caio Saraiva; Gabriel Brito; João Vitor Paula; Lais Falchochio...[et al.]. - 2024.  
0090 f.

Trabalho de Conclusão de Curso (Graduação) apresentado ao Instituto de Ciência Exatas e Tecnologia da Universidade Paulista, Santana de Parnaíba, 2024.

Área de Concentração: Tecnologia.

Orientador: Prof. Me. Edy Carlos Hidemi Hayashida.

1. Tecnologia. 2. Pandemia. 3. Estatística. 4. Inteligência Artificial. 5. Ciência de Dados. I. Gabriel Brito; João Vitor Paula; Lais Falchochio, Francieli Muniz; André Maldonado; Caio Saraiva; . II. Hayashida, Edy Carlos Hidemi (orientador).

André Luiz Oliveira Maldonado N7014J5  
Caio Teixeira de Lima Saraiva N7297F6  
Francieli dos Santos Muniz G2437E2  
Gabriel Aparecido Magalhães de Brito G191650  
João Vitor Soares de Paula G26BAH5  
Lais Falcochio de Camargo G2221J2

## **SIMULAÇÃO DE E-COMMERCE COM PREVISÃO DE VENDAS: UM ESTUDO DE REGRESSÃO LINEAR E VISUALIZAÇÃO DE DADOS**

Dissertação de Conclusão de Curso  
apresentada à Universidade Paulista – UNIP  
para a obtenção do título de bacharel em  
Ciência da Computação.

Orientador Prof. Me. Edy Carlos Hidemi  
Hayashida

Aprovado em:

BANCA EXAMINADORA

\_\_\_\_\_ 28/11/2024

Prof. Me. Edy Carlos Hidemi Hayashida

Universidade Paulista – UNIP

\_\_\_\_\_ 28/11/2024

Prof. Me. Elias C de Oliveira

Universidade Paulista – UNIP

\_\_\_\_\_ 28/11/2024

Prof. Me. Henry Costa Ungaro

Universidade Paulista – UNIP

## DEDICATÓRIA

Eu, Francieli, dedico este trabalho em primeiro lugar à Deus e minha mãe, Luciana, que mesmo durante as dificuldades, nunca deixou que me faltasse motivação para estudar. À minha irmã, Amanda, por sempre me apoiar em minhas decisões e por aplaudir todas minhas conquistas. Ao meu marido, André Luiz, por estar ao meu lado durante toda minha caminhada.

Eu, André, dedico este trabalho em primeiro lugar à Deus, por me guiar e abençoar em cada passo desta jornada. À minha mãe, que sempre me apoiou e nunca mediu esforços para me ajudar, sou imensamente grato por seu amor e dedicação incansáveis. Ao meu pai e à minha esposa, cujo apoio constante foi fundamental para a realização deste trabalho, deixo meu sincero agradecimento. Sem vocês, este sonho não teria se tornado realidade.

Eu, João, dedico este trabalho primeiramente a minha mãe, Vera, que sempre me ensinou a fazer o certo com todo o amor e gratidão, obrigado pelos ensinamentos e carinho. Para meus queridos avós, cuja força, e alegria que foram proporcionados moldaram a pessoa que sou hoje.

Eu, Caio, dedico este trabalho aos meus pais, pela dedicação, pelo exemplo de força e perseverança e por todo o apoio que me proporcionaram ao longo desta jornada, sem vocês, esta conquista não teria sido possível. Aos meus irmãos, cujo suporte constante e encorajamento foram fundamentais em cada etapa deste processo. À minha avó, por suas palavras sábias e pelo carinho que sempre me incentivaram a buscar o melhor de mim. À minha namorada, Marcela, pelo companheirismo, paciência e apoio incondicional. Sua presença foi essencial para que eu mantivesse o foco e a determinação até a conclusão deste trabalho.

Eu, Lais, dedico este trabalho a minha mãe Adriana Falcochio por sempre me auxiliar quando tive dificuldades, ao meu pai Marcelo Camargo por sempre estar me motivando a continuar, ao meu irmão Lucas Falcochio por nunca me deixar desistir e ao meu namorado Bruno Rodrigues por sempre me apoiar em todas as minhas decisões.

Eu, Gabriel dedico este trabalho aos pilares da minha vida. Ao meu pai Clayton, pelo exemplo de força, coragem e determinação que sempre me inspirou a seguir em frente, mesmo nos momentos mais difíceis. À minha mãe Denisi, por seu amor incondicional e apoio inabalável, que me deram a segurança necessária para nunca

desistir dos meus sonhos. À minha noiva, Livia, cuja presença, compreensão e carinho foram fundamentais para que eu chegasse até aqui. Sua paciência e companheirismo me motivaram a alcançar este objetivo.

## **AGRADECIMENTOS**

Às nossas famílias, que nos apoiaram e nos incentivaram ao longo desta jornada desafiadora.

Aos nossos professores, por compartilharem seus conhecimentos e nos ampararem sempre que tínhamos dúvidas, em especial ao Professor Edy Hayashida, por nos orientar durante todo o projeto, sua orientação foi fundamental.

Ao nosso grupo, pela dedicação e comprometimento excepcionais durante toda a elaboração e desenvolvimento deste projeto. Sem todos vocês, nada disso seria possível. Nossos mais sinceros agradecimentos.

*“Eu acredito que, às vezes são as pessoas que ninguém espera nada, que fazem as coisas que ninguém consegue imaginar”.*

**Alan Turing**



## RESUMO

O presente Trabalho de Conclusão de Curso, intitulado "Simulação de *E-commerce* com Previsão de Vendas: Um Estudo de Regressão Linear e Visualização de Dados", tem como objetivo apresentar o impacto causado no comércio varejista após a pandemia de COVID-19. Ao longo desse estudo, destacamos a compreensão do comércio varejista no Brasil e como o crescimento desse setor evoluiu para o comércio eletrônico, após longos períodos de distanciamento social, devido a uma pandemia global.

Foi explorado o uso de algoritmos preditivos para influenciar a tomada de decisão da empresa fictícia CALFJ & Co. Através de dados históricos, do período de 2019 a 2023, foi desenvolvido um modelo utilizando Regressão Linear Simples, que identifica correlações entre duas variáveis (ano e vendas), permitindo a previsão comportamental das vendas para 2024.

O estudo apresenta a importância do Aprendizado de Máquina, Inteligência Artificial e Ciência de Dados aplicados na construção de técnicas preditivas. O modelo foi implementado utilizando a linguagem de programação Python, com a modelagem do banco de dados e integração dos dados utilizando o sistema gerenciador de banco de dados PostgreSQL. A página do *e-commerce* da CALFJ & Co foi desenvolvida utilizando a plataforma Unity e as previsões são apresentadas através de um *dashboard*, desenvolvido no PowerBI.

Este estudo busca contribuir para o debate sobre o comércio varejista e o impacto sofrido pós-pandemia, enfatizando o papel fundamental do uso de análise de dados e modelos preditivos como aliados para as empresas, a fim de otimizarem suas vendas baseadas em dados confiáveis.

**Palavras-Chaves:** Modelos Preditivos, Regressão Linear, Varejo, Pandemia, PowerBI, PostgreSQL, Python, Unity.

## **ABSTRACT**

The present thesis, entitled "Simulation of e-commerce with Sales Forecasting: A Study of Linear Regression and Data Visualization" aims to present the impact on retail trade after the COVID-19 pandemic. Throughout this study, we highlight the understanding of retail commerce in Brazil and how the growth of this sector evolved into e-commerce following long periods of social distancing due to a global pandemic.

The use of predictive algorithms to influence decision-making for the fictitious company CALFJ & Co was explored. Using historical data from 2019 to 2023, a model was developed using Simple Linear Regression, which identifies correlations between two variables (year and sales), allowing for behavioral sales forecasting for 2024.

The study emphasizes the importance of Machine Learning, Artificial Intelligence, and Data Science applied in the construction of predictive techniques. The model was implemented using the Python programming language, with database modeling and data integration using the PostgreSQL database management system. The CALFJ & Co e-commerce website was developed using the Unity platform, and the predictions are presented through a dashboard, developed in PowerBI.

This study seeks to contribute to the discussion on retail and the post-pandemic impact, emphasizing the fundamental role of data analysis and predictive models as key allies for companies to optimize their sales based on reliable data.

**Keywords:** Predictive Models, Linear Regression, Retail, Pandemic, PowerBI, PostgreSQL, Python, Unity.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Motivos de compra pelo aplicativo.....	25
Figura 2 - Indicadores do e-commerce .....	27
Figura 3 - Correlação Linear Positiva.....	30
Figura 4 - Correlação Linear Negativa .....	31
Figura 5 - Entidades.....	35
Figura 6 - Relacionamento e Cardinalidade.....	35
Figura 7 - Atributos .....	36
Figura 8 - Modelo Conceitual CALFJ&Co .....	37
Figura 9 - Entidade e Atributos .....	38
Figura 10 - Modelo Lógico CALFJ&Co .....	38
Figura 11 - Tabela do modelo físico CALFJ&Co.....	39
Figura 12 - Modelo Físico CALFJ&Co .....	40
Figura 13 - Console RDS CALFJ&Co.....	41
Figura 14 - Tabela Cliente CALFJ&Co.....	42
Figura 15 - Tabela Pessoa Física e Pessoa Jurídica CALFJ&Co.....	42
Figura 16 - Tabela Pedido CALFJ&Co.....	43
Figura 17 - Tabela Item pedido e Produto CALFJ&Co .....	43
Figura 18 - Tabela Fornecedor CALFJ&Co .....	43
Figura 19 - Tabela Estoque e Pagamento CALFJ&Co .....	43
Figura 20 - Gráfico de Dispersão Janeiro .....	52
Figura 21 - Gráfico de Dispersão Fevereiro.....	52
Figura 22 - Gráfico de Dispersão Março .....	53
Figura 23 - Gráfico de Dispersão Abril.....	53
Figura 24 - Gráfico de Dispersão Maio .....	54
Figura 25 - Gráfico de Dispersão Junho .....	54
Figura 26 - Gráfico de Dispersão Julho .....	55
Figura 27 - Gráfico de Dispersão Agosto.....	55
Figura 28 - Gráfico de Dispersão Setembro .....	56
Figura 29 - Gráfico de Dispersão Outubro .....	56
Figura 30 - Gráfico de Dispersão Novembro .....	57
Figura 31 - Gráfico de Dispersão Dezembro .....	57
Figura 32 - Gráfico de linhas com pontos de dispersão.....	58
Figura 33 - Tabela de Dados Reais vs Predição .....	58

Figura 34 - Resultados das métricas de validação .....	61
Figura 35 - Gráfico de Resíduos vs Predição .....	63
Figura 36 - Gráfico Q-Q dos Resíduos .....	63
Figura 37 - Paleta de Cores.....	64
Figura 38 - Logos.....	65
Figura 39 - Tela <i>home</i> .....	65
Figura 40 - Tela de Seleção.....	66
Figura 41 - Tela Carrinho.....	66
Figura 42 - Seleção de Endereço .....	67
Figura 43 - Seleção de Método de Pagamento .....	67
Figura 44 - Revisão de Pedido .....	68
Figura 45 - Finalização do Pedido .....	68
Figura 46 - Layout dashboard.....	69
Figura 47 - Página inicial dashboard.....	70
Figura 48 - Página de análise micro .....	70
Figura 49 - Página de análise de vendas.....	71
Figura 50 - Schema no PowerBI.....	72
Figura 51 - Modelo de Relação PowerBI .....	73

## LISTA DE ABREVIATURAS E SIGLAS

- ABNT – Associação Brasileira de Normas Técnicas
- ACID - Atomicidade, Consistência, Isolamento e Durabilidade
- ANVISA - Agência Nacional de Vigilância Sanitária
- API - Application Programming Interface (Interface de Programação de Aplicação)
- AWS - Amazon Web Services (Serviços Web Amazon)
- B2B - Business-to-Business (Empresa para Empresa)
- B2C - Business-to-Customer (Empresa para Cliente)
- CC - Cartão de Crédito
- CEO – Chief Executive Officer (Diretor Executivo)
- C2C – Customer-to-Customer (Cliente para Cliente)
- CNC - Confederação Nacional do Comércio de Bens, Serviços e Turismo
- COVID-19 - Coronavirus Disease 2019
- CSV - Comma-Separated Values (Valores Separados por Vírgulas)
- DDL - Data Definition Language (Linguagem de Definição de Dados)
- DER - Diagrama Entidade-Relacionamento
- DEX - Data Analysis Expressions
- DF - DataFrame
- DQL - Data Query Language (Linguagem de Consulta de Dados)
- DML - Data Manipulation Language (Linguagem de Manipulação de Dados)
- EDI - Eletronic Data Interchange (Intercâmbio Eletrônico de Dados)
- ETL - Extract, Transform, Load (Extrair, Transformar, Carregar)
- FK – Foreign Key (Chave Estrangeira)
- IA – Inteligência Artificial
- IaaS - Infrastructure as a Service (Infraestrutura como Serviço)
- IBGE – Instituto Brasileiro de Geografia e Estatística
- IBM – International Business Machines

MAE - Mean Absolute Error (Erro Médio Absoluto)

MER - Modelo Entidade-Relacionamento

MSE - Mean Squared Error (Erro Quadrático Médio)

ODBC - Open Database Connectivity (Conectividade de banco de dados aberto)

OMS - Organização Mundial de Saúde

PaaS - Platform as a Service (Plataforma como Serviço)

PK – Primary Key (Chave Primária)

RDS - Relational Database Service (Serviço de Banco de Dados Relacional)

RMSE - Root Mean Squared Error (Raiz do Erro Quadrático Médio)

SaaS - Software as a Service (Software como Serviço)

SARS-CoV-2 - Severe Acute Respiratory Syndrome Coronavirus 2 (Síndrome Respiratória Aguda Grave Coronavírus 2)

SEBRAE - Serviço Brasileiro de Apoio às Micro e Pequenas Empresas

SGBD - Sistema Gerenciador de Banco de Dados

SQL – Structured Query Language (Linguagem de Consulta Estruturada)

SSD - Solid State Drive (unidade de estado sólido)

SSL - Secure Sockets Layer (Camada de Soquetes Seguros)

UFSC - Universidade Federal de Santa Catarina

UI – User Interface (Interface do Usuário)

UNIVAG - Centro Universitário de Várzea Grande

UX – User Experience (Experiência do Usuário)

## Sumário

<b>1 INTRODUÇÃO</b> .....	<b>18</b>
1.1 <b>Motivação</b> .....	<b>19</b>
1.2 <b>Objetivos Gerais</b> .....	<b>19</b>
1.3 <b>Objetivos Específicos</b> .....	<b>20</b>
1.4 <b>Metodologia</b> .....	<b>21</b>
1.5 <b>Organização do Texto</b> .....	<b>22</b>
<b>2 REVISÃO DA LITERATURA</b> .....	<b>22</b>
2.1 <b>Contextualização da Pandemia COVID-19</b> .....	<b>23</b>
2.2 <b>Compreendendo o Comércio Varejista no Brasil e sua atuação durante a Pandemia</b> .....	<b>24</b>
2.3 <b>Crescimento do setor de e-commerce</b> .....	<b>25</b>
2.4 <b>A influência da Inteligência Artificial no contexto de Modelos Preditivos</b> .....	<b>27</b>
2.5 <b>Fatores externos que influenciam em modelos preditivos</b> .....	<b>28</b>
2.6 <b>Modelos Preditivos</b> .....	<b>28</b>
2.6.1 <b>Aplicações</b> .....	<b>29</b>
2.7 <b>Regressão Linear</b> .....	<b>29</b>
2.7.1 <b>Regressão Linear Simples</b> .....	<b>30</b>
2.7.2 <b>Regressão Linear Múltipla</b> .....	<b>31</b>
2.7.3 <b>Regressão Linear Logística</b> .....	<b>32</b>
2.7.4 <b>Resíduos</b> .....	<b>33</b>
<b>3 MATERIAIS E MÉTODOS</b> .....	<b>33</b>
3.1 <b>BR Modelo</b> .....	<b>34</b>
3.1.1 <b>Modelo Entidade-Relacionamento</b> .....	<b>34</b>
3.1.2 <b>Modelo Conceitual</b> .....	<b>35</b>
3.1.3 <b>Modelo Lógico</b> .....	<b>37</b>
3.2 <b>SQL Power Architect</b> .....	<b>38</b>
3.2.1 <b>Modelo Físico</b> .....	<b>39</b>
3.3 <b>Amazon Web Services</b> .....	<b>40</b>
3.3.1 <b>Relational Database Service</b> .....	<b>40</b>
3.4 <b>Azure Data Studio</b> .....	<b>41</b>
3.4.1 <b>PostgreSQL</b> .....	<b>41</b>
3.4.2 <b>Ingestão dos dados no PostgreSQL</b> .....	<b>44</b>

<b>3.5 Visual Studio Code</b> .....	<b>45</b>
3.5.1 Conexão com o RDS e criação do <i>DataFrame</i> .....	45
3.5.2 Algoritmo de Predição .....	46
3.5.3 Validação do Algoritmo de Predição .....	58
3.5.4 Análise e conclusão da confiabilidade do modelo .....	61
<b>3.6 Figma</b> .....	<b>64</b>
3.6.1 <i>Brand</i> e Paleta de Cores .....	64
3.6.2 <i>Mockup site</i> .....	65
3.6.3 <i>Mockup dashboard</i> .....	69
<b>3.7 PowerBI</b> .....	<b>69</b>
3.7.1 Conexão do Banco de Dados com o PowerBI .....	71
3.7.2 Modelo de Relação das tabelas no PowerBI .....	72
3.7.3 Linguagem DAX .....	73
3.7.4 Processo ETL .....	74
3.7.5 <i>Power Query</i> .....	74
3.7.6 <i>Link Dashboard</i> .....	74
<b>3.8 Unity</b> .....	<b>75</b>
3.8.1 Introdução ao Relatório de Desenvolvimento .....	75
3.8.2 Processo de desenvolvimento .....	76
3.8.3 Implementação de Funcionalidades .....	76
3.8.4 Soluções de Problemas Técnicos .....	77
3.8.5 Conclusão do Relatório de Desenvolvimento .....	78
3.8.6 Etapas futuras .....	78
3.8.7 <i>Link Site</i> .....	79
<b>3.9 Git e Github</b> .....	<b>79</b>
3.9.1 Repositórios .....	79
<b>3.10 Métodos</b> .....	<b>80</b>
<b>4 RESULTADOS</b> .....	<b>81</b>
4.1 Desempenho do Modelo de Regressão Linear Simples .....	81
4.1.1 Métricas de Avaliação .....	81
4.1.2 Análise dos Resíduos .....	82
4.1.3 Conclusão do Desempenho .....	82
4.2 Simulação do <i>site</i> .....	82
4.3 Visualização dos dados no PowerBI .....	82



4.4 Conclusão dos Resultados .....	82
<b>5 CONCLUSÃO.....</b>	<b>83</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>85</b>

## 1 INTRODUÇÃO

Os dados apresentam informações valiosas para a tomada de decisão, junto da análise de dados temos a análise estatística, que impactaram no surgimento de uma área chamada Ciência de Dados, utilizando Inteligência Artificial (IA) e *Machine Learning* (Aprendizado de Máquina), cria-se algoritmos que podem ser utilizados para a tomada de decisões.

A solução abordada neste projeto será através de um algoritmo de Regressão Linear, onde será identificadas correlações entre as variáveis que interferem positivamente ou negativamente no resultado esperado. "A regressão linear é uma ferramenta estatística fundamental para entender a relação entre variáveis em um conjunto de dados." (JAMES, 2013).

De acordo com Siegel (2013), algoritmos de IA são capazes de prever eventos futuros a partir de uma base de dados já existente, treinando o algoritmo ele indica uma probabilidade de ocorrência, trazendo assim uma tomada de decisão mais assertiva e eficiente, podendo antecipar tendências e se preparando para a sazonalidade do mercado, se posicionando melhor e mais estruturado para atender as demandas. Entretanto, de acordo com Grego (2017), existem fatores externos que podem influenciar de forma negativa os modelos preditivos, novos cenários econômicos ou sociais, como por exemplo crises financeiras que podem diminuir significativamente a procura por produtos, acarretando a ineficiência de modelos desenvolvidos em um cenário fora do esperado.

Será utilizado o modelo de regressão linear para desenvolver o algoritmo e assim conseguir realizar a predição de vendas no ano de 2024, da empresa fictícia de varejo de roupas chamada CALFJ & Co. Será apresentado um algoritmo desenvolvido em Python, modelagem e criação do banco de dados, será utilizado o Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL, também será realizada a integração via API (*Application Programming Interface* ou Interface de Programação de Aplicação) com o *front-end* do site que será criado utilizando *Unity* para ilustrar como esses dados serão armazenados, e por fim, ilustrar a correlação dos dados finais que influenciam as vendas futuras, todos exibidos através da ferramenta de visualização de dados PowerBI, trazendo uma visão "executiva" dos dados, com uma visualização simples, porém completa. O modelo do *dashboard* será desenvolvido

pelo grupo, utilizando o Figma, um editor gráfico de vetores e prototipagem de projetos de design, onde será criado toda a parte visual da empresa e dashboard.

## 1.1 Motivação

Devido a Pandemia do COVID-19, iniciada no final de 2019, foram adotadas medidas de distanciamento social em todo o país, muitas lojas físicas precisaram ser fechadas a fim da tentativa de reduzir o contágio do vírus. Em 2020 as vendas *online*, tiveram um grande aumento, contendo mais de 4 milhões de novos clientes só no começo do ano. De acordo com a pesquisa da Ebit-Nielsen (2020), o Brasil registrou um crescimento acima da média em relação a novos consumidores do *e-commerce*, conforme a pesquisa, durante o primeiro semestre de 2020, as lojas *online* obtiveram o crescimento de faturamento de 47%.

O estudo da tecnologia é muito abrangente e com diversas oportunidades de inovação, na escolha desse projeto, buscamos um tema onde pudéssemos abordar a vasta área de *e-commerce* e o aumento significativo de compras *online*, trazer uma solução tecnológica e eficaz para o mercado de varejo brasileiro através de análises preditivas.

Serão aplicados os conhecimentos adquiridos durante o curso de Ciência da Computação, expandindo o projeto para além do desenvolvimento, mas também em análises estatísticas, matemáticas e de dados.

## 1.2 Objetivos Gerais

O objetivo deste trabalho é apresentar, compreender e analisar a importância da implementação de modelos preditivos no mercado de varejo brasileiro.

Através de dados estatísticos e modelos matemáticos será possível prever aumento ou diminuição de vendas, utilizando como variável as vendas históricas da empresa e previsão para o ano de 2024. Além disso, visa também explorar o uso de Inteligência Artificial durante a elaboração do algoritmo, treinamento do modelo de regressão linear e utilização de métricas para avaliar a funcionalidade do algoritmo.

O modelo será desenvolvido utilizando Regressão Linear, uma equação matemática é utilizada para estimar a condição de uma variável  $Y$ , em vista dos valores de outras variáveis  $X$ .

Será elaborado o modelo como um todo, desde análise das equações matemáticas, criação do algoritmo utilizando a linguagem de programação Python, modelagem e criação do banco de dados utilizando a linguagem de programação SQL e o Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL, criação da página de simulação de um *e-commerce* e integração do *front-end* ao banco de dados via API, e como finalização, apresentar os dados estatísticos obtidos através do algoritmo em um *dashboard* desenvolvido utilizando PowerBI.

### 1.3 Objetivos Específicos

Com a motivação baseada na predição de vendas do varejo brasileiro através de cálculos matemáticos, para alcançar o objetivo geral, serão executados os seguintes objetivos específicos:

1. Modelar e criar o banco de dados da empresa fictícia CALFJ & Co, utilizando a linguagem de programação SQL e o SGBD PostgreSQL para armazenar os dados necessários para a predição;
2. Identificar as variáveis que podem influenciar as vendas e determinar sua relevância dentro da predição;
3. Desenvolver o algoritmo de regressão linear utilizando a linguagem de programação Python;
4. Validar o desempenho do algoritmo, a fim de garantir que ele seja de fato eficaz;
5. Simular uma página de *e-commerce*, integrar essa página via API para coletar dados e armazenar no banco de dados;
6. Implementar um *dashboard* interativo, utilizando PowerBI para visualização dos resultados do modelo preditivo e métricas, destacando ideias relevantes para tomada de decisões;
7. Realizar ajustes necessários no modelo, considerando os dados obtidos durante o desenvolvimento;
8. Documentar todo o desenvolvimento, incluindo métodos e dificuldades;
9. Apresentar os resultados alcançados.

## 1.4 Metodologia

O objetivo desse projeto é demonstrar como o uso da inteligência artificial pode melhorar a eficiência e ser mais assertivo na tomada de decisões utilizando informações extraídas de um banco de dados relacional. Através de um desenvolvimento simultâneo com a análise de dados e análises estatísticas, o trabalho visa proporcionar mais valor para este sistema. Os procedimentos metodológicos que serão abordados adotam um método qualitativo - portanto uma pesquisa descritiva aplicada - cujo foco é voltado na subjetividade presente na investigação de dados somado à IA.

O Trabalho de Conclusão de Curso será dividido em duas etapas primordiais: a primeira consiste em uma extensa revisão teórica e matemática voltada para os conteúdos de estatística probabilística e modelos de regressão linear. Para estabelecer uma base teórica concreta sobre os temas, o grupo estudará materiais presentes em livros, artigos científicos e trabalhos acadêmicos relevantes. Já para informações complementares, será realizada a consulta de repositórios técnicos e artigos digitais presentes em sites conceituados.

A segunda fase do projeto refere-se ao desenvolvimento de um modelo preditivo, juntamente com a modelagem e criação do Banco de Dados e visualização da saída através de um *dashboard*. Tendo em vista o viés acadêmico proposto, o grupo seguirá a formatação da Associação Brasileira de Normas Técnicas (ABNT) bem como a constante revisão bibliográfica dos conceitos base. Desse modo, a pesquisa busca ter uma concepção mais profunda de como a IA, integrada a análise de dados, pode desencadear na melhoria do consumo, eficiência e tomada de decisão das interfaces e *dashboards* na experiência e usabilidade do usuário (UI/UX).

Em suma, este trabalho representa um esforço coletivo para ampliar o nosso entendimento sobre o uso de modelos preditivos no mercado brasileiro de varejo. Sempre tendo em vista a inovação no mercado de trabalho bem como o impacto positivo nas atividades exercidas no cotidiano dos analistas.

## 1.5 Organização do Texto

No capítulo 1, é apresentado o conteúdo introdutório de toda a proposta do trabalho, enfatizando motivação, objetivos, métodos e organização do texto.

No capítulo 2 será apresentado a revisão de literatura, onde levantaremos os temas abordados de forma geral, manifestaremos estudos sobre área de dados e estatística, empresas varejistas, e como um sistema preditivo é vantajoso para as empresas brasileiras no mercado atual.

O capítulo 3 visa demonstrar todos os instrumentos principais e métodos utilizados para criação do modelo matemático que prevê as vendas da empresa fictícia CALFJ & Co, e conterá todo o processo de desenvolvimento. Também terá como objetivo demonstrar todo o resultado ao longo da criação do trabalho acadêmico, mostraremos como será feita a criação do algoritmo de regressão linear, o banco de dados, o *site* da empresa integrado via API e a demonstração dos resultados em um *dashboard*.

O capítulo 4 tem como conteúdo a finalização do trabalho acadêmico e as expectativas alcançadas que nos levarão a novos horizontes na busca de novas tecnologias. Este capítulo será também reservado às considerações finais, conclusões e referências de onde provém a investigação de artigos científicos, pesquisas acadêmicas, livros e base de dados empregados para o auxílio da pesquisa.

## 2 REVISÃO DA LITERATURA

A fim de prover uma noção geral sobre como o tópico da aplicação de um modelo preditivo através de regressão linear, fazendo uso de inteligência artificial pode influenciar em análises do mercado varejista no Brasil, em um período pós-pandêmico, onde de acordo com a Organização Pan-Americana de Saúde (2022), o termo "pandemia" refere-se à distribuição geográfica de determinada doença, e não está diretamente ligada a gravidade da doença. Ao referir-se em distribuição geográfica, claramente engloba o impacto econômico no país, uma vez que afeta todo o território.

Segundo o Instituto Brasileiro de Geografia e Estatística (2023), mesmo durante a pandemia, em 2021 o Brasil apresentou um saldo positivo de empresas que se

mantiveram no mercado, onde 926 mil empresas entraram e 605 mil saíram, o saldo final foi de 321 mil empresas, um número crescente, quando comparado a 2019 e 2020. Através dos dados apresentados pelo IBGE é claramente exposto a influência e necessidade de soluções tecnológicas, tendo em vista que o varejo no Brasil passou por um baque e precisou passar por adaptações abruptas durante a pandemia, a fim de manter o comércio fluindo.

Verifica-se que há uma necessidade de aprofundamento no tema de estudo para permitir a existência de tamanha flexibilidade em função dos cálculos matemáticos envolvidos, sempre focados no resultado almejado através da sua utilização.

## **2.1 Contextualização da Pandemia COVID-19**

No final de 2019, o governo chinês relatou o primeiro caso do novo coronavírus. A população da cidade de Wuhan, estava presenciando o início de uma pandemia global, segundo Machado (2020), os residentes da cidade chinesa começaram a apresentar sintomas de uma síndrome respiratória aguda grave e seu crescimento exponencial afetou milhares de pessoas em um curto período. Em 11 de março de 2020, a Organização Mundial da Saúde (OMS) declarou a doença COVID-19 (*Coronavirus Disease 2019*) como uma pandemia global causada por um novo coronavírus que foi denominado SARS-CoV-2.

Em outubro de 2023, o Brasil contava com aproximadamente 37,7 milhões de casos, onde cerca de 686,9 mil resultaram no óbito das pessoas infectadas, conforme os dados do *Coronavirus Resource Center* (Centro de Recursos sobre Coronavírus, em tradução livre), mantido pela *Johns Hopkins University*. Foram desenvolvidas ao menos sete vacinas em todo o mundo, sendo quatro, autorizadas pela Agência Nacional de Vigilância Sanitária – ANVISA para uso no Brasil (ANVISA, 2021). A CoronaVac, criada pelo Instituto Butantan juntamente com biofarmacêutica chinesa Sinovac, e as outras criadas fora do país, vacinas das farmacêuticas Pfizer, Janssen e a AstraZeneca, sendo a vacina da Pfizer o primeiro imunizante aprovado pelo Organização Mundial de Saúde – OMS (BEE, 2022).

## 2.2 Compreendendo o Comércio Varejista no Brasil e sua atuação durante a Pandemia

Pretendendo apresentar uma visão geral do tema, regressão linear e previsão de vendas, observa-se a necessidade da utilização de modelos preditivos no varejo brasileiro após um longo período de pandemia, e o sucesso das empresas varejistas no Brasil possuem como base a capacidade de observação e a adaptabilidade às tendências de consumo (GIULIANI, 2014). A adaptação dos varejistas também é uma característica fundamental para o funcionamento do setor.

Com a expansão do comércio varejista brasileiro, manter-se atualizado sobre modelos de gestão baseada em novas tecnologias é essencial para alcançar estratégias e vantagens competitivas.

Diante do cenário pandêmico de COVID-19 que se iniciou em 2020, inúmeras esferas da sociedade e economia sofreram alterações, com base em informações do SEBRAE (2020) as empresas precisaram se adaptar e inovar rapidamente para atender as novas necessidades de uma sociedade que por questões sanitárias precisava manter-se em casa. “33,5% das empresas em funcionamento reportaram que a pandemia teve efeitos negativos sobre a empresa”. (IBGE, 2020)

Uma das principais estratégias adotadas pelos varejistas para minimizar os danos aos negócios diante de um período de instabilidade foi investir em vendas *online* através de *e-commerces*, redes sociais e aplicativos. O comércio eletrônico é o processo de compra e venda de produtos através de meios eletrônicos, utilizando dispositivos como celulares e computadores. A palavra *e-commerce* é uma abreviatura de *eletronic commerce*, comércio eletrônico em tradução direta.

De acordo com CNC (2021), um dos setores do varejo brasileiro mais atingido pela pandemia de COVID 19 foi o setor de vestuários, assim como livrarias e artigos automotivos.

Apesar de toda incerteza que assombrava os comerciantes com estabelecimentos comerciais físicos durante a pandemia, muitas empresas rapidamente se adaptaram e inauguraram lojas *online*, os *e-commerces*. Buscando amenizar grandes perdas financeiras, um dos setores que mais se destacou na premissa de estabelecimentos *online* foi o comércio varejistas de vestuário e acessórios. De acordo com a Ebit Nielsen (2020), em 2020 houve um aumento de

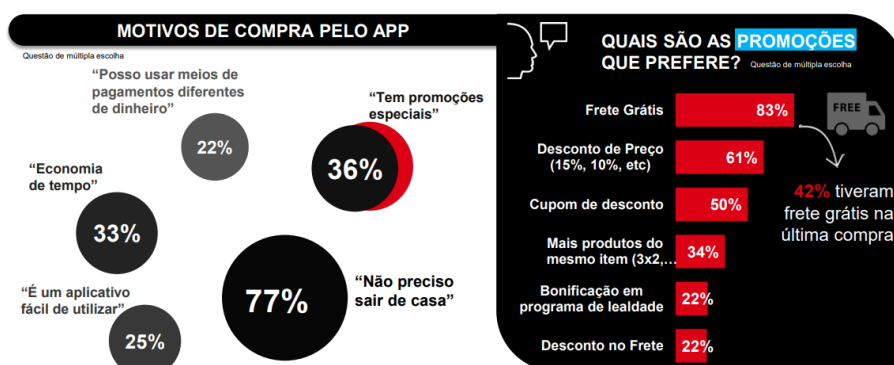


22% na quantidade de pedidos *online* no segmento vestuário quando comparada a 2019.

De acordo com Kotler e Armstrong (2015), as empresas têm buscado utilizar inúmeras abordagens digitais e sociais, focando no objetivo de alcançar clientes e estabelecer relacionamento com eles a partir de qualquer lugar.

De acordo com figura 1, a motivação dos compradores através de meios digitais, 77% informaram que efetuam compras *online* pelo motivo “Não preciso sair de casa”, seguido por motivos econômicos, 36% compraram por “Ter promoções especiais” e 33% por “Economia de tempo”. Um fator que mais causa impacto nos clientes é o fato das promoções escolhidas possuírem frete grátis.

Figura 1 - Motivos de compra pelo aplicativo



Fonte: Webshoppers (2020). 42ª Edição

### 2.3 Crescimento do setor de e-commerce

O comércio eletrônico originou-se nos Estados Unidos, na década de 1980 onde grandes empresas varejistas iniciaram o processo de modernização na solicitação de pedidos aos fornecedores, mas foi em 1993 que surgiram *softwares*, tornando o uso de computadores mais comuns. O conceito de navegar (*browser*) entre páginas eletrônicas começou a ser disseminado em meados de 1995, onde de acordo com Santos (2018), surgiram empresas como Amazon e eBay que se tornaram líderes no setor.

Conforme descrito por Novaes (2007), os principais tipos de *e-commerce* durante a primeira década do século XXI, eram:

- *Electronic Data Interchange* (EDI): Que consistia em transferência eletrônica automática de dados entre computadores de empresas;
- Comércio Eletrônico *Business-to-Business* (B2B): O conceito de páginas eletrônicas na Internet começou a ser apresentado por fornecedores, onde empresas clientes podiam obter e trocar informações com fornecedores.
- Comércio Eletrônico *Business-to-Customer* (B2C): Nesse comércio, o comprador pessoa física, através de seu computador ou dispositivo pessoal, realiza buscas e compras através da Internet.

Além dos tipos citados, Bagatini e Laimer (2019) apresentam também o *Customer-to-Customer* (C2C), modelo de *e-commerce* utilizado pelo aplicativo OLX, onde consumidores transacionam produtos para outros consumidores.

De acordo com artigo publicado pelo escritório JGM Advogados (2024), a implementação do comércio eletrônico inicialmente não foi tão bem aceita, pois, ao transacionar através de Internet, as operações ficavam propensas a fraude. Com a evolução desse nicho comercial, foram desenvolvidas abordagens para mitigar fraudadores e transações fraudulentas, acarretando assim no aumento significativo de transações via *e-commerces*.

Observando do ponto de vista geográfico, o comércio eletrônico possibilitou a liberdade dos consumidores, que através de dispositivos móveis como *smartphones*, analisarem preços e variedades de produtos, no âmbito nacional e global.

Um exemplo de *e-commerce* de sucesso é a Shopee, empresa de origem asiática, fundada em 2015 inicialmente em sete países, sendo eles Singapura, Malásia, Indonésia, Tailândia, Taiwan, Vietnã e Filipinas.

Conforme evidenciado no site Oficial da Shopee (2024), em 2019 a gigante laranja, como é conhecida, iniciou suas operações fora do sudeste asiático e escolheu o Brasil como sede de suas operações fora da Ásia. Inicialmente, manteve o comércio de forma branda, mas com os eventos que viriam a ocorrer em 2020 com o início da pandemia de COVID-19, atingiu números grandiosos de vendas através da plataforma, uma vez que a população se viu obrigada a manter-se em casa devido o isolamento social.

De acordo com o Jornal Estadão (2024), hoje a Shopee conta com mais de 3 milhões de empreendedores locais (Brasil), 43,6 milhões de usuários ativos e lidera o ranking de *downloads* do app ao redor do mundo, com mais de 200 milhões de

*downloads*, tendo como principais concorrentes no Brasil a chinesa Shein e o Mercado Livre, de origem argentina.

Através da figura 2, podemos analisar os números do *e-commerce* em 2023, com uma ênfase no faturamento de 185,7 bilhões de reais e 395,1 milhões de pedidos.

Figura 2 - Indicadores do e-commerce



Fonte: Abcomm

## 2.4 A influência da Inteligência Artificial no contexto de Modelos Preditivos

A primeira ação da Inteligência Artificial (IA) que pode ser considerada foi realizada por Warrem Macculloh e Walter Pitts em 1943, o estudo é conhecido como “*A logical calculus of the ideas immanent in nervous activity*”, o estudo tem como premissa que o neurônio artificial recebe um sinal como entrada, é multiplicado por um peso e é comparado com um fator discriminante, quando a saída é maior, é igual a um, caso o contrário, é zero.

A Inteligência Artificial está ligada diretamente a Ciência da Computação, com o intuito de se concentrar no desenvolvimento de algoritmos que são capazes de realizar ações que necessariamente apenas o homem conseguiria fazer. Hoje a IA é aplicada em diversas situações, como por exemplo: Assistentes Virtuais, Análise de Dados, Finanças, na Medicina, entre outros muitos trabalhos.

Conforme artigo publicado pela empresa Alura (2024), dentre esses algoritmos, existem os preditivos. Um algoritmo de previsão é utilizado em Análises de Dados e Inteligência Artificial para estimar resultados futuros com base em dados anteriores. Estes algoritmos são criados para detectar padrões nos dados e utilizar esses mesmos padrões para efetuar projeções ou previsões.

De acordo com Santos (2018), os algoritmos preditivos têm a finalidade de aprender com os dados já existentes para ser capaz de prever eventos futuros. Registros de dados são comumente aplicados em diversas áreas, indo desde previsões climáticas até análises do mercado financeiro, diagnósticos médicos, recomendações de produtos no *e-commerce*, projeção da demanda e assim por diante. Normalmente utilizando algoritmos como Regressão Linear, Árvores de

Decisão, Redes Neurais e/ou *Random Forest* (muitas vezes não se limitando a apenas um tipo de algoritmo).

## **2.5 Fatores externos que influenciam em modelos preditivos**

Segundo Greco (2017), os fatores externos que influenciam em modelos preditivos incluem mudanças no ambiente econômico, situações imprevistas, como desastres naturais. Essas variáveis podem ter um impacto profundo no cenário em que os modelos foram desenvolvidos, tornando as previsões imprecisas ao não considerar novos cenários econômicos ou sociais. Por exemplo, crises financeiras podem diminuir a procura por produtos, enquanto mudanças rápidas nos comportamentos de consumo podem invalidar previsões baseadas em dados históricos.

Além disso, inovações tecnológicas e mudanças nas regulamentações são elementos externos de importante relevância. Novas tecnologias têm potencial de transformar o completamente o mercado, com isso, fazendo que modelos antigos percam sua validade. Mudanças nas legislações ou políticas ambientais, também podem impactar diretamente os dados utilizados nos modelos, resultando em desvios nas previsões.

A qualidade dos dados e fatores em um período específico e condição climática podem ter um impacto significativo. Dados de baixa qualidade ou desatualizados, além da ausência de adaptações a variação sazonais ou ambientais, podem resultar em distorções e prejudicar a eficácia dos modelos. Portanto, é essencial incorporar mecanismos de atualização para intensificar esses impactos externos e aprimorar a precisão das previsões.

## **2.6 Modelos Preditivos**

De acordo com Monteiro (2008), modelos preditivos correspondem a uma regra (equação), que fazendo a utilização de informações históricas e/ou atuais, prevê um evento ou tendências futuras. Existem diversos modelos preditivos já inseridos no mercado em diversas áreas de negócio. As empresas estão cada vez mais em busca de soluções, a fim de mitigar os riscos, expansões, prevenção de fraude, entre outros objetivos de acordo com as necessidades da empresa.

### 2.6.1 Aplicações

Segundo Assunção (2012), podemos apresentar os seguintes modelos, acompanhados de uma breve descrição de suas funções.

Na área de negócio de Crédito, temos o *Credit Score* que é utilizado para classificar solicitações de crédito de acordo com o risco de inadimplência envolvido no cliente. Já no segmento Cobrança, podemos apresentar o *Collection Score*, utilizado para classificar pessoas inadimplentes, levando em conta a probabilidade de quitação de seus débitos.

No setor de Recursos Humanos, de acordo com Cantarella (2022), existe o Pymetrics que se destaca no setor de recrutamento por usar inteligência artificial para facilitar a seleção de candidatos. A Pymetrics analisa perfis de indivíduos usando métodos estatísticos para determinar aqueles com maior probabilidade de ter sucesso na jornada de trabalho, ajudando empresas a contratar quem mais combina com as devidas necessidades por meio de testes e jogos interativos. Segundo Frida Polli, CEO da Pymetrics, 50 (cinquenta) empresas já utilizam seus serviços, entre elas grandes companhias como Accenture, Unilever, McDonald's e Banco JP.

No segmento Saúde, conforme descrito por Santos (2019), já existem áreas que utilizam a inteligência preditiva para identificar de maneira antecipada óbitos em idosos. Esta tecnologia está sendo usada para auxiliar os profissionais da saúde a tomarem decisões mais inteligentes e precisas, devido às informações que são fornecidas através da tecnologia. Com isso, é certo dizer que atualmente eles contam com uma ferramenta de trabalho a mais para cuidar e acompanhar seus pacientes.

## 2.7 Regressão Linear

A origem do termo "regressão" remonta a Francis Galton, que, em 1886, publicou um artigo com o intuito de elucidar o motivo dos pais de estatura elevada tendem a ter filhos com estatura média inferior à deles, enquanto pais de estatura reduzida tendem a ter filhos com estatura média superior. (GALTON, 1886)

De acordo com a *Amazon Web Services*, a Regressão Linear é um método estatístico de análise de dados que torna possível mensurar a influência de uma ou diversas variáveis independentes sobre uma dependente. Sendo modelada a partir de

uma equação linear contendo as variáveis independentes e dependentes. (AWS 1, 2024)

A partir desses modelos de regressão linear – sejam eles simples ou múltiplos - usando essas variáveis (independentes e dependentes) brutas, podem ser convertidas em dados mais refinados que serão usados para traçar previsões estatísticas.

### 2.7.1 Regressão Linear Simples

De acordo com Martins (2019), a regressão linear simples é representada pela equação  $Y = \alpha + \beta X$ , onde:

$Y$  = Variável Dependente (sujeita a mudanças ou variações);

$X$  = Variável Independente (contém fatores que influenciam a variável dependente);

$\alpha$  = Interceptação em  $y$  quando  $x = 0$ ;

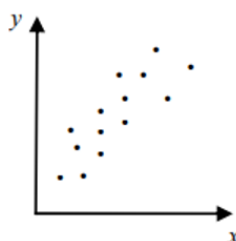
$\beta$  = Coeficiente angular da linha de regressão (determina se a relação entre as variáveis é grande ou pequena).

Utiliza-se regressão linear simples, quando apenas uma variável independente interfere na variável dependente, porém, para que seja efetiva, deve-se considerar a correlação entre as variáveis e analisar como elas se relacionam.

Segundo Yamauti (2013), podemos utilizar o diagrama de dispersão, uma forma de representar graficamente os pares  $(x, y)$ , onde ao visualizar um padrão em relação aos pares ordenados, é possível supor que há correlação entre as variáveis.

Conforme a figura 3, a correlação linear positiva ocorre quando as variáveis são ascendentes, ou seja, em quanto uma aumenta a outra também aumenta.

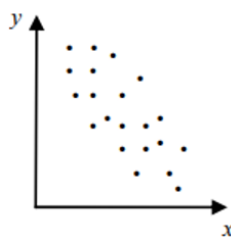
Figura 3 - Correlação Linear Positiva



Fonte: Repositório PUC

Na figura 4 é apresentada a correlação linear negativa onde ocorre o oposto da positiva, as variáveis são descendentes.

Figura 4 - Correlação Linear Negativa



Fonte: Repositório PUC

Yamauti (2013) apresenta um exemplo de regressão linear, analisando o preço de aluguel de automóveis de uma agência, onde a taxa de aluguel ( $Y$ ) é determinada pela equação  $Y = 8 + 0,15x$ , em que  $Y$  representa a taxa de aluguel em reais (R\$) e  $x$  a distância percorrida em quilômetros. Conforme exemplificado, ao percorrer 100 km, a taxa de aluguel seria de R\$ 23,00. O autor ressalta, no entanto, que esta equação possui uma margem de erro devido a diversos fatores não controlados que podem influenciar no preço final do aluguel, como por exemplo, condições dos veículos, sazonalidade e demanda. Para conseguirmos analisar a margem de erro, utilizamos a seguinte equação:  $Y = aX + \beta + \varepsilon$ , onde  $\varepsilon$  refere-se a margem de erro.

Na prática não existe uma reta perfeita, parte da dispersão irá para o erro aleatório, porém, permite representar a dependência de uma variável quantitativa em relação a outra por meio da equação linear e prever valores da variável  $Y$  a partir do valor da variável  $X$ .

“A reta de regressão é utilizada em previsões, isto é, para prever o valor de  $Y$ , para um dado valor de  $X$ . No entanto estas previsões não devem contemplar valores de  $X$  fora do intervalo do  $XiS$ , uma vez que o facto de a reta se ajustar bem aos pontos dados não significa que sirva para fazer *extrapolações*”. (MARTINS, 2019)

Uma forma de avaliar a correlação entre as variáveis é o Método de Pearson onde  $R^2 = r^2$ . O coeficiente de correlação de Pearson é muito utilizado em análises de correlação simples, enquanto o  $R^2$  é mais associado à análise de regressão, embora ambos sejam considerados indicadores da força da relação entre variáveis. A análise de regressão pode usar o coeficiente de correlação de Pearson para analisar a correlação entre uma variável dependente e uma ou mais variáveis independentes. O  $R^2$  é uma medida comum de ajuste do modelo nesse contexto.

### 2.7.2 Regressão Linear Múltipla

De acordo com Coelho (2014), neste modelo a análise é feita a partir da modelagem da relação linear entre uma variável de desfecho contínua e múltiplas variáveis utilizadas como preditoras – contínuas ou categóricas. Sendo assim a grande diferença entre a regressão linear simples e a múltipla é a quantidade de variáveis preditoras (regressão linear simples faz uso de apenas uma).

O uso da regressão linear múltipla constantemente se dá quando há a necessidade de investigar a relação que variáveis tem com a variável de desfecho. Sendo assim, possibilitando a visualização do efeito que uma variável preditora tem sobre a outra.

A regressão linear múltipla pode ser descrita pela equação  $Y = a + \beta_1X_1 + \beta_2X_2$ .

$Y$  = Variável Dependente (sujeita a mudanças ou variações);

$X$  = Variável Independente (contém fatores que influenciam a variável dependente);

$a$  = Interceptação em  $Y$  quando  $X = 0$ ;

$\beta$  = Coeficiente angular da linha de regressão (determina se a relação entre as variáveis é grande ou pequena).

Sendo a diferença entre a regressão linear simples e a regressão linear múltipla, apenas a adição de mais uma variável preditora ( $\beta_2X_2$ , por exemplo) que pode ser adicionada sem limites ( $\beta_nX_n$ ). O cálculo de margem de erro e de correlação entre as variáveis se mantem da mesma forma que é realizada na regressão linear simples.

É válido ressaltar que uma parte importante da regressão linear múltipla (assim como na simples) é a qualidade da análise dos dados que serão utilizados como variáveis, já que é importante analisar se suas variáveis correspondem aos dados a serem tratados, criando mais transparência do modelo e mais poder de previsão.

### 2.7.3 Regressão Linear Logística

Conforme apresentado por Figueira (2006), a regressão logística é um exemplo distinto já que funciona de forma diferente. Ela é usada para prever a probabilidade de um evento ocorrer, sendo sua variação de 0 a 1. Onde 0 está relacionado a eventos que são improváveis de acontecer e 1 a eventos que muito provavelmente podem



acontecer. É inteiramente baseada em funções logarítmicas para realizar o cálculo da linha de regressão.

Sendo assim, não basta ter um modelo bem desenvolvido, existe ainda a necessidade de análise desses dados e como o seu consumo para o modelo de regressão linear escolhido (simples, múltiplo ou até logístico) é realizado – se está correto e eficiente, além de gerar mais transparência nos resultados.

A regressão linear por conta de sua versatilidade e variedade de modelos se tornou muito popular em diversos campos de pesquisa. Dado essas vantagens, a regressão linear – seja ela qual for, simples ou múltipla – não serve apenas para fins de previsão uma vez que deve ser provado também eficiente para descrever sistemas.

#### 2.7.4 Resíduos

Os Resíduos de um Modelo Preditivo, segundo Rodrigues (2012), trata-se da diferença entre a previsão e os valores reais, em caso de resíduos pequenos, significa que o modelo fez uma boa previsão e caso o resíduo seja muito grande, o modelo errou bastante. Essa métrica ajuda a entender se o modelo está acertando ou errando e contribui para melhorar as previsões, quanto mais baixo for o resíduo, melhor é o modelo.

Se o resíduo for positivo ou negativo, ele demonstra o tamanho e a direção do erro, tornando possível ajustes para tornar o modelo cada vez mais preciso.

Equação para cálculo dos resíduos:

$$\text{Resíduos} = y - \hat{y}$$

Onde  $y$  é o valor real e  $\hat{y}$  os valores previstos.

### 3 MATERIAIS E MÉTODOS

O objetivo deste capítulo é apresentar as ferramentas utilizadas durante o desenvolvimento de todo o projeto, desde linguagens de programação e versionamento de códigos, plataformas de desenvolvimento e ferramentas que auxiliaram na criação do modelo e implementação do banco de dados, desenvolvimento do algoritmo, criação do *front end* e visualização dos dados adquiridos através do modelo preditivo.

### 3.1 BR Modelo

O software BR Modelo foi a escolha para a criação do Diagrama Entidade-Relacionamento (DER), através dos dados obtidos do Modelo Entidade-Relacionamento (MER), da loja de roupas CALFJ & Co.

De acordo com Candido (2005), esse *software* foi desenvolvido em 2005, como uma ferramenta de código aberto e gratuita, com foco na modelagem de banco de dados relacionais, foi criada pelo autor, como trabalho de conclusão de curso de especialização em banco de dados pelas Universidade Federal de Santa Catarina (UFSC) e Centro Universitário de Várzea Grande (UNIVAG), após o autor constatar a falta de uma ferramenta nacional para tal finalidade.

#### 3.1.1 Modelo Entidade-Relacionamento

Foi desenvolvido o seguinte Modelo Entidade-Relacionamento:

1. Queremos coletar os dados pessoais de nossos clientes, pessoa física ou jurídica. No caso de Pessoa Física o seu CPF e no caso de Pessoa Jurídica o CNPJ. Além disso, também coletar e armazenar o seu nome, endereço, telefone e *e-mail*.
2. O produto principal do *e-commerce* são roupas e estas roupas têm informações associadas a elas como o nome, categoria, gênero, tamanho, cor e o valor.
3. As roupas são fornecidas por fornecedores e precisamos ter guardados o nome do fornecedor, o nome de contato, o *e-mail* e no máximo 2 telefones.
4. Sabemos que não podemos ter a mesma roupa vindo de vários fornecedores pois a roupa é exclusiva de um fornecedor.
5. Nosso cliente pode comprar uma ou mais roupas através de um pedido de compra.
6. Os clientes podem efetuar o pagamento através de apenas três formas de pagamento (Pix, Cartão de Crédito ou Boleto) e o pagamento poderá ser associado a três status (Sucesso, Falha e Cancelado).

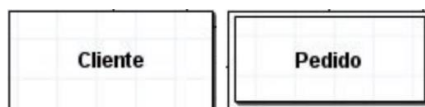
Através do modelo apresentado, observou-se as seguintes entidades e atributos:

- Entidades fortes: clientes e fornecedores;
- Entidades fracas: roupas e pedidos de compra;
- Atributos: nome, endereço, telefone, *e-mail*, categoria, gênero, tamanho, cor, valor, disponibilidade, forma de pagamento e status;
- Atributos de especialização: CPF e CNPJ.

### 3.1.2 Modelo Conceitual

Através das informações obtidas no MER, foi desenvolvido o seguinte Diagrama Entidade-Relacionamento, de forma técnica denominado Modelo Conceitual (Alto nível), utilizando o *software* BR Modelo, a fim de representar as entidades, onde a representação ocorre através de um retângulo para entidades fortes, e um retângulo duplo para entidades fracas, conforme apresentado na figura 5.

Figura 5 - Entidades



Fonte: Autoria própria

Pode-se observar na figura 6 que o relacionamento entre as entidades é representado por um losango, e existem relacionamentos binários, ternários ou n-ários. É durante o relacionamento que é definida a cardinalidade – ou conectividade – entre as entidades.

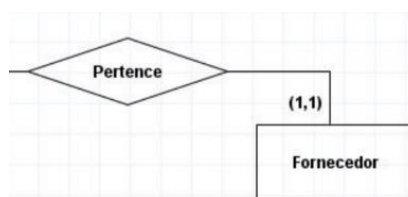
As cardinalidades podem ser:

1:1 – Lê-se um para um;

1:n – Lê-se um para muitos;

n:n - Lê-se muitos para muitos.

Figura 6 - Relacionamento e Cardinalidade

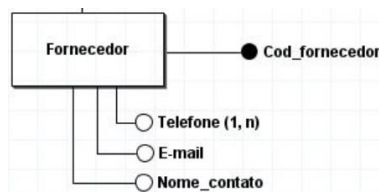


Fonte: Autoria própria

Os atributos são as características ligadas as entidades, no DER são representados por círculos, conforme representado na figura 7. Atributos podem ser classificados como:

- Atômico: Possuem valores únicos que não podem ser divididos em partes menores para formar novos atributos, como exemplo temos o CPF;
- Composto: Seus valores podem ser divididos em partes menores, como exemplo, endereços, que podem ser divididos em rua, estado e cidade;
- Multivalorado: Podem possuir vários valores associados a ele, como telefones, onde uma pessoa pode ter mais de um;
- Atributo Chave: Utilizado para identificar de forma única uma entidade, como exemplo, Código do Produto, ele é único pois cada produto recebe apenas um código distinto;
- Atributo Chave Primária: Atributos que nunca se repetem, não podem possuir valor *null* e somente entidades fortes possuem chave primária.

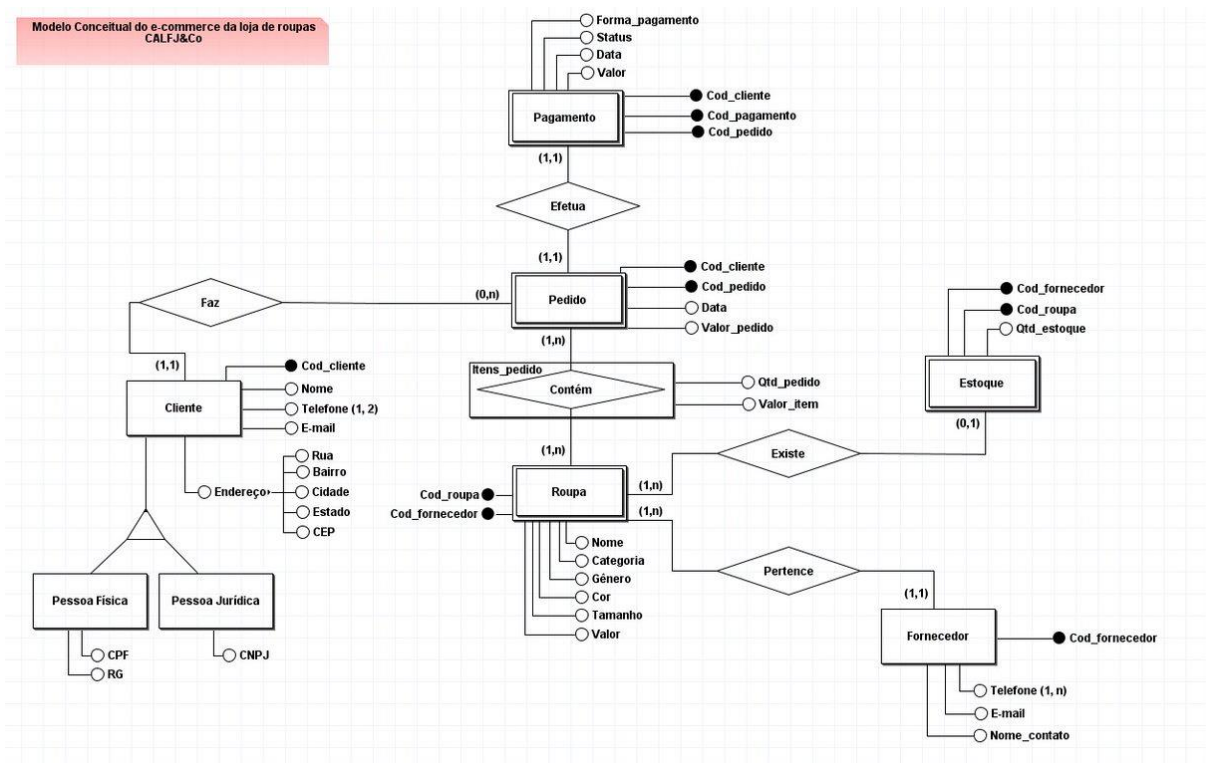
Figura 7 - Atributos



Fonte: Autoria própria

Na figura 8, é possível observar o Modelo Conceitual completo, desenvolvido de acordo com o Modelo Entidade-Relacionamento da CALFJ & Co.

Figura 8 - Modelo Conceitual CALFJ&amp;Co



Fonte: Autoria própria

### 3.1.3 Modelo Lógico

Segundo a *Amazon Web Services*, o modelo lógico de dados é criado a partir das descrições dos dados representadas no Modelo Conceitual. Ele descreve como os dados serão armazenados no banco de dados, identificando as entidades, os atributos, as chaves primárias e estrangeiras e os seus relacionamentos. (AWS 2, 2024)

Cada tabela no Modelo Lógico descreve todas as entidades, em retângulos, com seus cabeçalhos, muito semelhantes à estrutura de tabelas, e seus respectivos atributos, dispostos em forma de campos, dentro dos retângulos, preenchidos com o nome do atributo, evidenciando as *Primary Keys* (PK) e as *Foreign Keys* (FK), respectivamente Chaves Primárias e Chaves Estrangeiras, em tradução direta. Um exemplo de tabela no Modelo Lógico pode ser visualizado na figura 9.

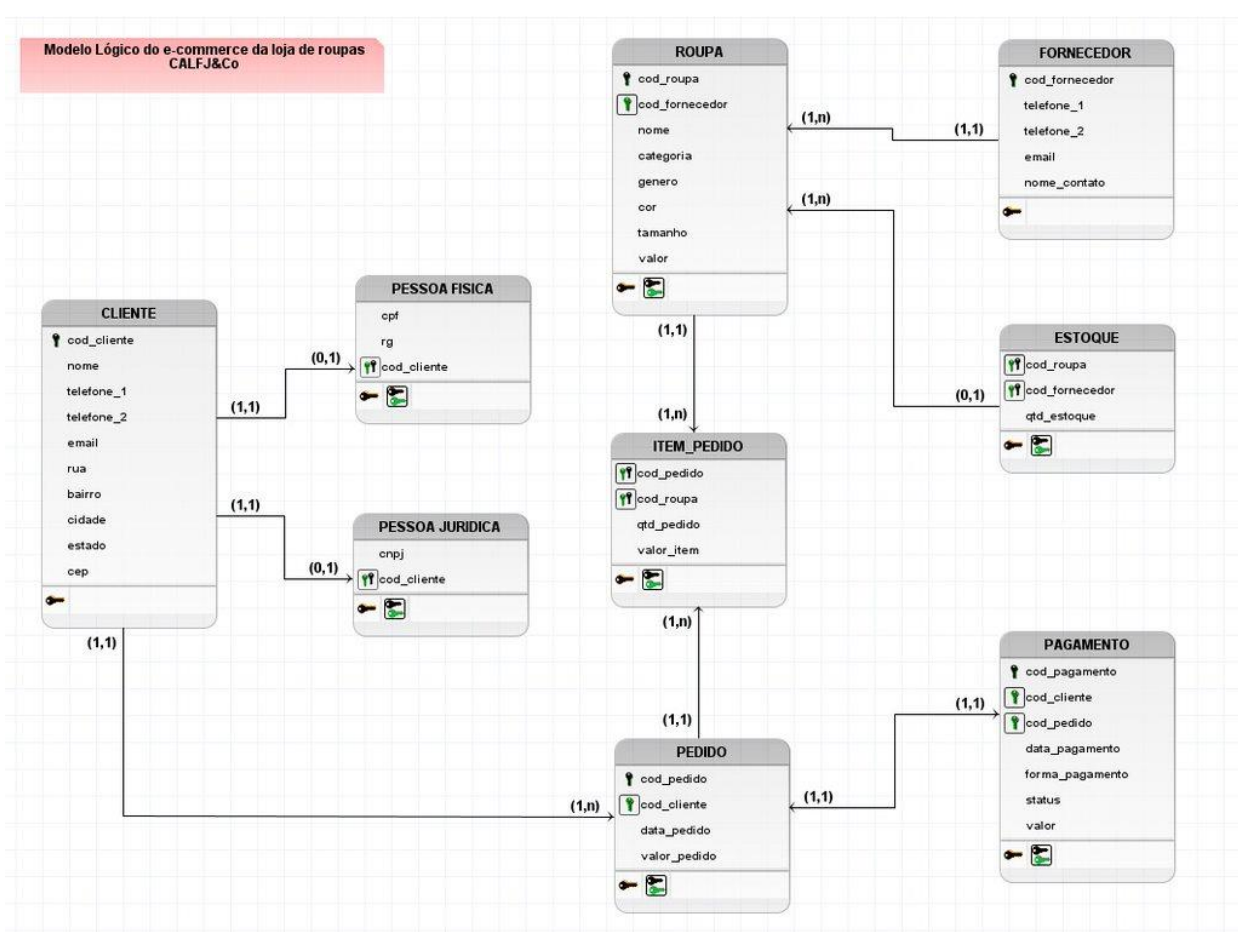
Figura 9 - Entidade e Atributos



Fonte: Autoria própria

Na figura 10, é possível observar o Modelo Lógico completo, desenvolvido de acordo com o Modelo Conceitual da CALFJ & Co.

Figura 10 - Modelo Lógico CALFJ&amp;Co



Fonte: Autoria própria

### 3.2 SQL Power Architect

De acordo com a Best Of BI, empresa desenvolvedora, o SQL Power Architect é uma ferramenta de modelagem de dados, amplamente utilizada por arquitetos de dados, DBAs (*Database Administrator*), Analistas e *Designers*, que apresenta funcionalidades e recursos amplamente técnicos para que os usuários façam modelos

de banco de dados, engenharia reversa de banco de dados e geração de metadados ETL.

O software foi escolhido para suprir a necessidade da criação do Modelo Físico do projeto, uma vez que o BR Modelo não permite tal criação, apenas Modelos Conceituais e Lógicos.

### 3.2.1 Modelo Físico

O Modelo Físico fornece detalhes que auxiliam os administradores de banco de dados e desenvolvedores a implementarem o banco de dados propriamente dito. Nesse modelo, as entidades chamam-se “Tabelas”, e os atributos, “Colunas”.

Dentro desse modelo é necessário descrever quais são os dados que serão aceitos nas colunas das tabelas, os tipos de dados normalmente são definidos em categorias: tipos numéricos, de data e de *string*, são as mais conhecidas. A figura 11 demonstra a estrutura de uma tabela no Modelo Físico.

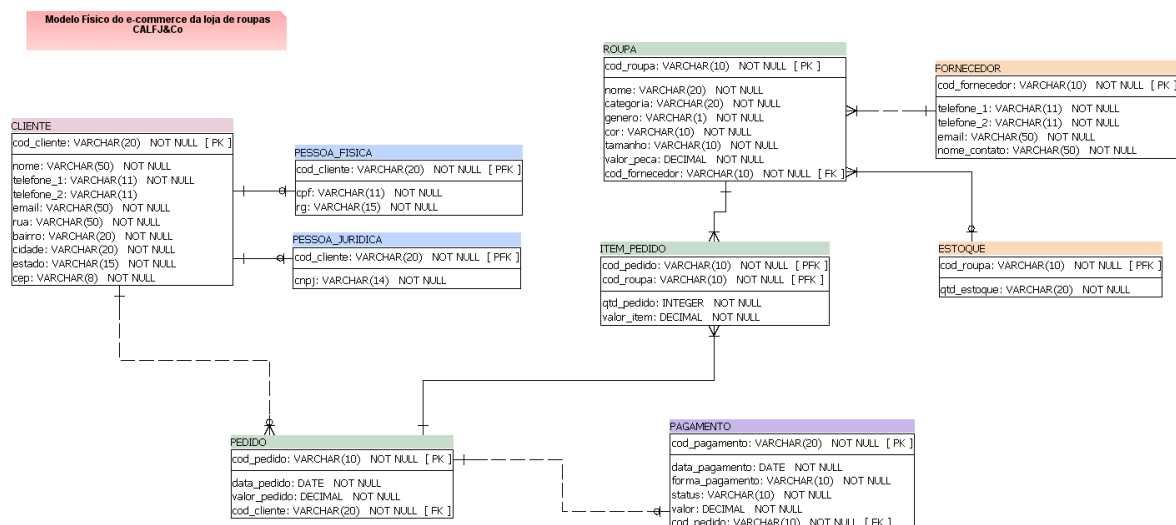
Figura 11 - Tabela do modelo físico CALFJ&Co

PESSOA_JURIDICA
cod_cliente: VARCHAR(20) NOT NULL [ PFK ]
cnpj: VARCHAR(14) NOT NULL

Fonte: Autoria própria

Na figura 12, é possível observar o Modelo Físico completo, desenvolvido de acordo com o Modelo Lógico da CALFJ & Co.

Figura 12 - Modelo Físico CALFJ&amp;Co



Fonte: Autoria própria

### 3.3 Amazon Web Services

O *Amazon Web Services* (AWS), de acordo com Kewate (2022), é o fornecedor mais confiável de computação em nuvem, que oferece ótimos serviços sob demanda, onde é utilizado dimensionamentos de recursos em nuvem, privacidade, integridade e disponibilidade dos dados. Os serviços da AWS estão disponíveis a preços acessíveis, sem necessidade de nenhum investimento prévio, o cliente apenas paga pelos serviços utilizados regularmente. De forma generalista, a AWS é um *IaaS* (*Infrastructure as a Service*), mas também oferta recursos como *PaaS* (*Plataform as a Service*) e *SaaS* (*Software as a Service*), *Infraestrutura como Serviço*, *Plataforma como Serviço* e *Software como Serviço*, em tradução direta, respectivamente.

#### 3.3.1 Relational Database Service

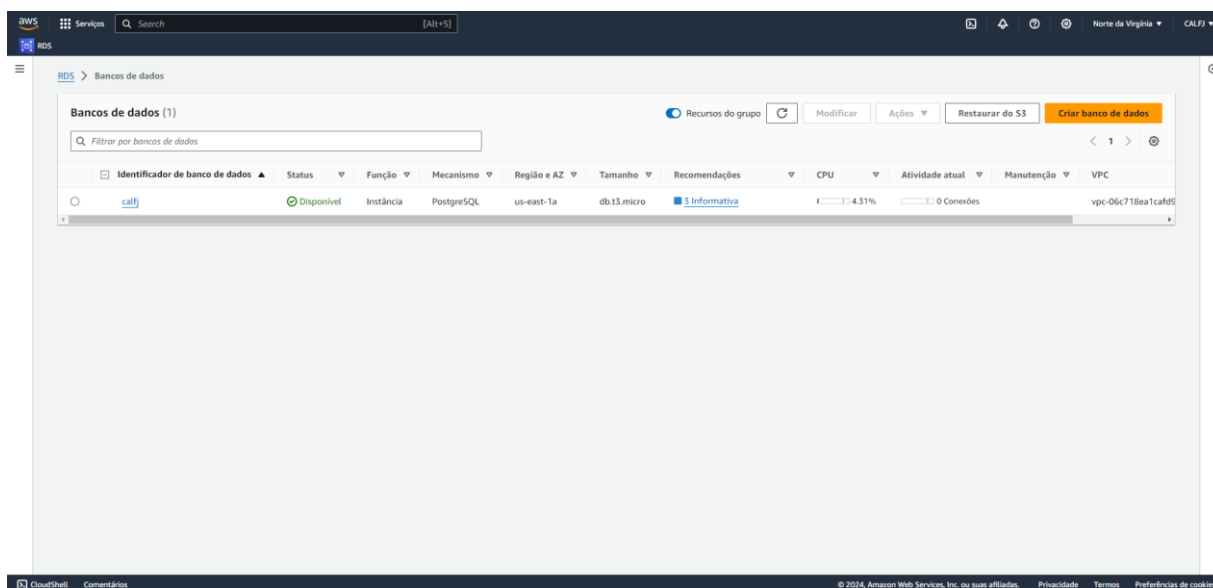
O serviço escolhido foi o *Relational Database Service* (RDS), conforme site oficial da AWS, o RDS é um *PaaS* (*Plataform as a Service*), uma coleção de serviços gerenciados que facilita a configuração, operação e escalabilidade de dados em nuvem. Possui suporte para Amazon Aurora, PostgreSQL, MySQL, MariaDB e SQL Server. Além disso é um produto com 12 meses para teste gratuito, com limite de uso de 750 horas de instância por mês, 20 GB de armazenamento de banco de dados de



uso geral (SSD) e 20 GB de armazenamento para *backups* de banco de dados e DB *Snapshots*. (AWS 3, 2024)

Têm como benefícios a facilidade de uso, alta disponibilidade, escalabilidade, segurança e economia, uma vez que você apenas paga pelos recursos que usa. A figura 13 demonstra o console no RDS.

Figura 13 - Console RDS CALFJ&Co



Fonte: Autoria própria

## 3.4 Azure Data Studio

De acordo com a *Microsoft*, o Azure Data Studio é uma ferramenta de análise e gerenciamento de dados moderna, multiplataforma e de código aberto, com conectividade com bancos de dados locais e em nuvem.

Possui capacidade de conexão imediata com SQL da Azure, e amplas extensões de suporte, incluindo MySQL e PostgreSQL. Oferece uma interface de editor de código moderna, terminal integrado e integração de controle de versionamento de código-fonte (Git).

### 3.4.1 PostgreSQL

O SGBD escolhido foi o PostgreSQL, que de acordo com o *Google Cloud*, é um sistema de código aberto, lançado sob a licença PostgreSQL que é compatível com banco de dados relacional (SQL) e não relacional (JSON) e funções SQL avançadas.

Conta com diversas extensões robustas, que incluem recuperação pontual, controle de acesso, simultaneidade de várias versões e espaços de tabela. Por ser código aberto, é totalmente gratuito e pode ser executado em quase todos os principais sistemas operacionais, como Linux, Microsoft e Mac OS X. Ainda conforme o Google *Cloud*, "O PostgreSQL também oferece suporte a propriedades ACID (atomicidade, consistência, isolamento e durabilidade)".

Durante a criação do banco de dados, foram abordadas as seguintes linguagens do SQL e seus respectivos comandos:

**DDL (*Data Definition Language*):** Abrange os comandos de interação com os objetos do banco, através dos comandos *CREATE*, *ALTER* e *DROP*.

**DML (*Data Manipulation Language*):** Abrange os comandos de interação com os dados dentro das tabelas, através dos comandos *INSERT*, *DELETE* e *UPDATE*.

**DQL (*Data Query Language*):** São comandos de consulta, muito utilizados após a criação e inserção de dados dentro do banco, sendo o principal comando, o *SELECT*.

Da figura 14 a 19, é possível visualizar a criação de todas as tabelas da CALFJ & Co quem compõem o Banco de Dados.

Figura 14 - Tabela Cliente CALFJ&Co

```

1 CREATE TABLE cliente (
2     cod_cliente SERIAL,
3     nome VARCHAR(50) NOT NULL,
4     telefone_1 VARCHAR(11) NOT NULL,
5     telefone_2 VARCHAR(11),
6     email VARCHAR(50) NOT NULL,
7     rua VARCHAR(50) NOT NULL,
8     bairro VARCHAR(20) NOT NULL,
9     cidade VARCHAR(20) NOT NULL,
10    estado VARCHAR(15) NOT NULL,
11    cep VARCHAR(8) NOT NULL,
12    PRIMARY KEY(cod_cliente)
13 );

```

Fonte: Autoria própria

Figura 15 - Tabela Pessoa Física e Pessoa Jurídica CALFJ&Co

```

1 CREATE TABLE pessoa_fisica (
2     cod_cliente SERIAL NOT NULL,
3     cpf VARCHAR(11) NOT NULL,
4     rg VARCHAR(15) NOT NULL,
5     FOREIGN KEY (cod_cliente) REFERENCES cliente (cod_cliente)
6 );
7
1 CREATE TABLE pessoa_juridica (
2     cod_cliente SERIAL,
3     cnpj VARCHAR(14) NOT NULL,
4     FOREIGN KEY (cod_cliente) REFERENCES cliente (cod_cliente)
5 );

```

Fonte: Autoria própria

Figura 16 - Tabela Pedido CALFJ&amp;Co

```

1 CREATE TABLE pedido (
2     cod_pedido SERIAL,
3     cod_cliente SERIAL,
4     data_pedido DATE NOT NULL,
5     valor_pedido DECIMAL NOT NULL,
6     PRIMARY KEY (cod_pedido),
7     FOREIGN KEY (cod_cliente) REFERENCES cliente (cod_cliente)
8 );

```

Fonte: Autoria própria

Figura 17 - Tabela Item pedido e Produto CALFJ&amp;Co

```

1 CREATE TABLE item_pedido (
2     cod_pedido SERIAL,
3     cod_produto SERIAL,
4     qtd_pedido INTEGER NOT NULL,
5     valor_item DECIMAL NOT NULL,
6     FOREIGN KEY(cod_pedido) REFERENCES pedido (cod_pedido),
7     FOREIGN KEY(cod_produto) REFERENCES produto (cod_produto)
8 );

1 CREATE TABLE produto (
2     cod_produto SERIAL,
3     cod_fornecedor SERIAL,
4     nome VARCHAR(20) NOT NULL,
5     categoria VARCHAR(20) NOT NULL,
6     genero VARCHAR(1) NOT NULL,
7     cor VARCHAR(10) NOT NULL,
8     tamanho VARCHAR (10) NOT NULL,
9     valor_unidade DECIMAL NOT NULL,
10    PRIMARY KEY (cod_produto),
11    FOREIGN KEY(cod_fornecedor) REFERENCES fornecedor (cod_fornecedor)
12 );

```

Fonte: Autoria própria

Figura 18 - Tabela Fornecedor CALFJ&amp;Co

```

1 CREATE TABLE fornecedor (
2     cod_fornecedor SERIAL,
3     telefone_1 VARCHAR(11) NOT NULL,
4     telefone_2 DECIMAL(11),
5     email VARCHAR(50) NOT NULL,
6     nome_contato VARCHAR(50) NOT NULL,
7     PRIMARY KEY (cod_fornecedor)
8 );

```

Fonte: Autoria própria

Figura 19 - Tabela Estoque e Pagamento CALFJ&amp;Co

```

1 CREATE TABLE estoque (
2     cod_produto SERIAL,
3     qtd_estoque VARCHAR(20),
4     FOREIGN KEY(cod_produto) REFERENCES produto (cod_produto)
5 );

1 CREATE TABLE pagamento (
2     cod_pagamento SERIAL,
3     cod_pedido SERIAL,
4     data_pagamento DATE NOT NULL,
5     forma_pagamento VARCHAR(10) NOT NULL,
6     status VARCHAR(10) NOT NULL,
7     PRIMARY KEY (cod_pagamento),
8     FOREIGN KEY(cod_pedido) REFERENCES pedido (cod_pedido)
9 );

```

Fonte: Autoria própria

### 3.4.2 Ingestão dos dados no PostgreSQL

Foi desenvolvido um algoritmo em Python, com o intuito de facilitar a ingestão dos dados em massa para popular o banco de dados. Através de arquivos csv, foi possível inserir aproximadamente 4,5 milhões de registros instantaneamente no banco. O algoritmo utilizado pode ser visualizado no código a seguir:

```

""" Projeto: Trabalho de Conclusão de Curso UNIP
Descrição: Carregamento dos dados
Autores: André Maldonado, Caio Teixeira, Francieli Muniz, Gabriel Aparecido,
João Soares e Lais Falcochio
(c) """
import pandas as pd
from sqlalchemy import create_engine
from dotenv import load_dotenv
import os

# Carrega as variáveis de ambiente
load_dotenv()

# Configuração do Banco de Dados
# os.getenv pega os valores das variáveis de ambiente
DB_HOST = os.getenv('DB_HOST')
DB_NAME = os.getenv('DB_NAME')
DB_USER = os.getenv('DB_USER')
DB_PASSWORD = os.getenv('DB_PASSWORD')

# Cria a conexão com o Banco de Dados
# Cria o engine de conexão para interagir com o banco e fazer consultas.
DATABASE_URL = f'postgresql://{DB_USER}:{DB_PASSWORD}@{DB_HOST}/{DB_NAME}'
engine_conexao = create_engine(DATABASE_URL)

# Path para os arquivos excel
path = r'C:\Users\Francieli Muniz\OneDrive\Área de Trabalho\Área de
Trabalho\Francieli dos Santos Muniz\Docs\Faculdade\8 - TCC I e II\Ingestion'

# Lê os arquivos excel
pedido = pd.read_excel(f'{path}\pedido.xlsx')
item_pedido = pd.read_excel(f'{path}\item_pedido.xlsx')
pagamento = pd.read_excel(f'{path}\pagamento.xlsx')
cliente = pd.read_excel(f'{path}\cliente.xlsx')
fornecedor = pd.read_excel(f'{path}\fornecedor.xlsx')
pessoa_fisica = pd.read_excel(f'{path}\pessoa_fisica.xlsx')
produto = pd.read_excel(f'{path}\produto.xlsx')

# Insere os dados nas tabelas
def insert_data(table_name, dataframe):

```

```

dataframe.to_sql(table_name, engine_conexao, if_exists='append',
index=False)

# Insere os dados nas tabelas, para não ter problemas com as PK e FK
insert_data('cliente', cliente)
insert_data('fornecedor', fornecedor)
insert_data('pessoa_fisica', pessoa_fisica)
insert_data('produto', produto)
insert_data('pedido', pedido)
insert_data('item_pedido', item_pedido)
insert_data('pagamento', pagamento)

print("Dados carregados com sucesso!")

```

### 3.5 Visual Studio Code

De acordo com a *Microsoft*, criado em 2015, o *Visual Studio Code* apresenta ambientes simples e práticos, é uma ferramenta de código aberto e gratuita projetada para edição de código e desenvolvimento *web*, a ferramenta é aceita nos principais sistemas operacionais, como Linux, *Microsoft Windows* e Mac OS X.

Por se tratar de código aberto, a *Microsoft* incentiva a comunidade a desenvolver extensões para o *Visual Studio Code*, com o objetivo de ser cada vez mais maleável com as tecnologias presentes no mercado.

#### 3.5.1 Conexão com o RDS e criação do *DataFrame*

Iniciou-se o desenvolvimento através da criação de um algoritmo em Python, que pudesse acessar e manipular a base de dados armazenado no RDS da AWS. As credenciais do banco de dados foram salvas em um arquivo “.env”, com o intuito de atribuir mais segurança e integridade ao banco.

Para manipulação e análise dos dados, foi utilizada a biblioteca Pandas, juntamente com a biblioteca SQLAlchemy para conectar-se ao banco de dados. Os dados foram salvos em um arquivo pickle (.pkl), o tipo de arquivo foi escolhido devido sua eficiência em termos de desempenho e espaço, quando comparado com outros formatos como json ou csv. Outra grande vantagem que influenciou na escolha do formato do arquivo, foi o fato do .pkl ser muito bom para uso em *Machine Learning*, pois o modelo uma vez treinado, é possível carregá-lo novamente sem a necessidade de retreiná-lo. O código desenvolvido pode ser visualizado abaixo:

```

""" Projeto: Trabalho de Conclusão de Curso UNIP
Descrição: RDS conexão

```

```

Autores: André Maldonado, Caio Teixeira, Francieli Muniz, Gabriel Aparecido,
João Soares e Lais Falcochio
(c) ""
from dotenv import load_dotenv
import os
import pandas as pd
from sqlalchemy import create_engine, text

# Carrega as variáveis de ambiente
load_dotenv()

# Cria a URL de conexão com sqlalchemy
DATABASE_URL =
f"postgresql://{os.getenv('DB_USER')}:{os.getenv('DB_PASSWORD')}@{os.getenv('D
B_HOST')}/{os.getenv('DB_NAME')}"

# Cria o engine de conexão para interagir com o banco e fazer consultas.
engine_conexao = create_engine(DATABASE_URL)

# Função para extrair os dados e armazenar em um DataFrame
def get_data(query, engine):

    # Usa a conexão com o engine
    with engine.connect() as conexao:
        df = pd.read_sql(text(query), conexao)
        return df

# Conecta com o banco, extrai os dados e apresenta uma amostra
try:
    consulta = 'SELECT * FROM pagamento'
    data = get_data(consulta, engine_conexao)
    print(data.head())

    # Salva o DataFrame em um arquivo .pkl
    data.to_pickle("dados_vendas.pkl")
    print("Arquivo dados_vendas.pkl criado com sucesso")
except Exception as excecao:
    print("Erro ao conectar ao banco de dados:", excecao)

```

### 3.5.2 Algoritmo de Predição

Com o objetivo de: "Criar um modelo preditivo de vendas realizadas com sucesso, mês a mês para 2024, considerando as vendas da CALFJ & Co entre 2019 e 2023."

E com base no questionamento hipotético da área de negócios da empresa: "Queremos investir mais dinheiro em campanhas de marketing no ano de 2024. Sem um investimento inicial, qual é a previsão de vendas bem-sucedidas, baseadas nos anos anteriores, com índice de erro abaixo de 10%, para o ano de 2024?"

Para a criação do modelo, foram utilizadas as bibliotecas Pandas e Numpy para manipulação dos dados, Scikit-Learn, por ser uma das mais populares bibliotecas de *Machine Learning*. Dentro da biblioteca foi utilizada a classe *LinearRegression*, para que fosse possível aplicar o modelo de regressão linear e Matplotlib.pyplot para a criação dos gráficos.

```
""" Projeto: Trabalho de Conclusão de Curso UNIP
Descrição: Algoritmo de Predição
Autores: André Maldonado, Caio Teixeira, Francieli Muniz, Gabriel Aparecido,
João Soares e Lais Falcochio
(c) """
import pandas as pd
import numpy as np
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt
```

O algoritmo carregou o *DataFrame* criado anteriormente na conexão com o banco de dados, fez a conversão e filtros, conforme a solicitação da área de negócios.

```
# Carrega o DataFrame criado no connection_rds.py
data = pd.read_pickle("dados_vendas.pkl")

# Converte a coluna 'data_pagamento' em formato datetime
data['data_pagamento'] = pd.to_datetime(data['data_pagamento'],
errors='coerce')

# Filtra pagamentos com status 'Sucesso'
data_sucesso = data[data['status'] == 'Sucesso']
```

O próximo passo consistiu na inicialização e iteração das listas para armazenamento dos dados, de acordo com os meses de cada ano (2019 a 2023), filtrando os dados de cada mês, calculando a soma das vendas de cada mês, armazenando as vendas e criando um *DataFrame* para a análise do modelo.

```
# Inicializa as listas para armazenagem dos dados
meses = list(range(1, 13))
vendas_2019 = []
vendas_2020 = []
vendas_2021 = []
vendas_2022 = []
```

```

vendas_2023 = []
previsoes_2024 = []

# Itera em cada mês
for mes in meses:
    # Filtra o respectivo mês de 2019 a 2023
    meses_19 = data_sucesso[(data_sucesso['data_pagamento'].dt.year == 2019) &
(data_sucesso['data_pagamento'].dt.month == mes)]
    meses_20 = data_sucesso[(data_sucesso['data_pagamento'].dt.year == 2020) &
(data_sucesso['data_pagamento'].dt.month == mes)]
    meses_21 = data_sucesso[(data_sucesso['data_pagamento'].dt.year == 2021) &
(data_sucesso['data_pagamento'].dt.month == mes)]
    meses_22 = data_sucesso[(data_sucesso['data_pagamento'].dt.year == 2022) &
(data_sucesso['data_pagamento'].dt.month == mes)]
    meses_23 = data_sucesso[(data_sucesso['data_pagamento'].dt.year == 2023) &
(data_sucesso['data_pagamento'].dt.month == mes)]

    # Calcula a soma de vendas para cada mês, de cada ano
    meses_19_vendas = meses_19['valor_pedido'].sum()
    meses_20_vendas = meses_20['valor_pedido'].sum()
    meses_21_vendas = meses_21['valor_pedido'].sum()
    meses_22_vendas = meses_22['valor_pedido'].sum()
    meses_23_vendas = meses_23['valor_pedido'].sum()

    # Armazena os dados de 2019 a 2023
    vendas_2019.append(meses_19_vendas)
    vendas_2020.append(meses_20_vendas)
    vendas_2021.append(meses_21_vendas)
    vendas_2022.append(meses_22_vendas)
    vendas_2023.append(meses_23_vendas)

# Cria um DataFrame com os dados de vendas de todos os meses, todos os
anos
dados_vendas = pd.DataFrame({
    'ano': [2019, 2020, 2021, 2022, 2023],
    'vendas': [meses_19_vendas, meses_20_vendas, meses_21_vendas,
meses_22_vendas, meses_23_vendas]
})

```

Em continuação, foi criado o modelo de Regressão Linear e o método *fit* foi utilizado para calcular o coeficiente angular e o intercepto que melhor se ajustam na relação entre  $X$  e  $Y$ . O algoritmo faz a predição utilizando os coeficientes calculados, para prever o valor da variável dependente (vendas), para a variável independente (ano) e armazenar os dados da predição.

```

# Modelo de Regressão Linear
X = dados_vendas[['ano']] # Variável Independente

```



```

y = dados_vendas['vendas'] # Variável Dependente

model = LinearRegression()
model.fit(X, y)

# Predição de vendas para o respectivo mês em 2024
meses_2024_vendas_predicao = model.predict(np.array([[2024]]))

# Armazena as previsões
previsoes_2024.append(meses_2024_vendas_predicao[0])

```

Foi plotado um gráfico de dispersão para cada mês, evidenciando os dados reais e a previsão para 2024.

```

# Cria o gráfico de dispersão para o respectivo mês
plt.figure(figsize=(10, 6))

# Plota os pontos de dados reais
plt.scatter(X, y, color='blue', label='Dados Reais')

# Plota a linha de regressão
plt.plot(X, model.predict(X), color='red', label='Linha de Regressão')

# Adiciona o ponto previsto para 2024
plt.scatter([2024], meses_2024_vendas_predicao, color='green',
label='Previsão 2024')

# Configurações do gráfico
plt.title(f'Regressão Linear - Mês: {mes:02d}')
plt.xlabel('Ano')
plt.ylabel('Vendas')
plt.legend()
plt.grid(True)

# Exibe o gráfico
plt.show()

```

Após a predição, os valores foram convertidos para apresentar um formato mais legível. Em seguida, plotou-se um Gráfico de Dispersão, juntamente com uma linha de tendência para as vendas de 2019 a 2023 e as previsões para 2024, com seus respectivos títulos, mês a mês.

```

# Formata os valores com duas casas decimais e separadores de milhar
vendas_2019 = [f'{valor:,.2f}' for valor in vendas_2019]
vendas_2020 = [f'{valor:,.2f}' for valor in vendas_2020]
vendas_2021 = [f'{valor:,.2f}' for valor in vendas_2021]
vendas_2022 = [f'{valor:,.2f}' for valor in vendas_2022]
vendas_2023 = [f'{valor:,.2f}' for valor in vendas_2023]

```

```

previsoes_2024 = [f'{prev:,.2f}' for prev in previsoes_2024]

# Converte os meses para strings
meses_str = [f'{mes:02d}' for mes in meses]

# Cria a figura do eixo do gráfico
fig, ax = plt.subplots(figsize=(14, 6))

# Plota no gráfico as vendas de 2019
ax.scatter(meses, [float(valor.replace(',','')) for valor in vendas_2019],
color='orange', Label='Vendas 2019')
ax.plot(meses, [float(valor.replace(',','')) for valor in vendas_2019],
color='orange')

# Plota no gráfico as vendas de 2020
ax.scatter(meses, [float(valor.replace(',','')) for valor in vendas_2020],
color='purple', Label='Vendas 2020')
ax.plot(meses, [float(valor.replace(',','')) for valor in vendas_2020],
color='purple')

# Plota no gráfico as vendas de 2021
ax.scatter(meses, [float(valor.replace(',','')) for valor in vendas_2021],
color='grey', Label='Vendas 2021')
ax.plot(meses, [float(valor.replace(',','')) for valor in vendas_2021],
color='grey')

# Plota no gráfico as vendas de 2022
ax.scatter(meses, [float(valor.replace(',','')) for valor in vendas_2022],
color='red', Label='Vendas 2022')
ax.plot(meses, [float(valor.replace(',','')) for valor in vendas_2022],
color='red')

# Plota no gráfico as vendas de 2023
ax.scatter(meses, [float(valor.replace(',','')) for valor in vendas_2023],
color='green', Label='Vendas 2023')
ax.plot(meses, [float(valor.replace(',','')) for valor in vendas_2023],
color='green')

# Plota as previsões para 2024
ax.scatter(meses, [float(prev.replace(',','')) for prev in previsoes_2024],
color='blue', Label='Previsões 2024')
ax.plot(meses, [float(prev.replace(',','')) for prev in previsoes_2024],
color='blue')

# Títulos
ax.set_title('Vendas de 2019 a 2023 x Previsões de Vendas para 2024')
ax.set_xlabel('Mês')
ax.set_ylabel('Vendas')
ax.set_xticks(meses)

```

```
ax.set_xticklabels(meses_str, rotation=45)
ax.grid(True)
ax.legend()
```

De maneira a tornar a visualização de todos os valores mais assertiva, também foi criada uma tabela, com o de-para entre as vendas com sucesso entre 2019 e 2023, e as previsões para 2024.

```
# Cria a tabela com os dados do gráfico
tabela_dados = pd.DataFrame({
    'Mês': meses_str,
    'Vendas 2019': vendas_2019,
    'Vendas 2020': vendas_2020,
    'Vendas 2021': vendas_2021,
    'Vendas 2022': vendas_2022,
    'Vendas 2023': vendas_2023,
    'Previsão 2024': previsoes_2024
})

# Cria a figura para a tabela
fig_table, ax_table = plt.subplots(figsize=(10, 6))

# Plota a tabela
ax_table.axis('off')
tabela = ax_table.table(cellText=tabela_dados.values,
                        colLabels=tabela_dados.columns,
                        cellLoc='center',
                        loc='center',
                        colColours=["lightpink"] * len(tabela_dados.columns))

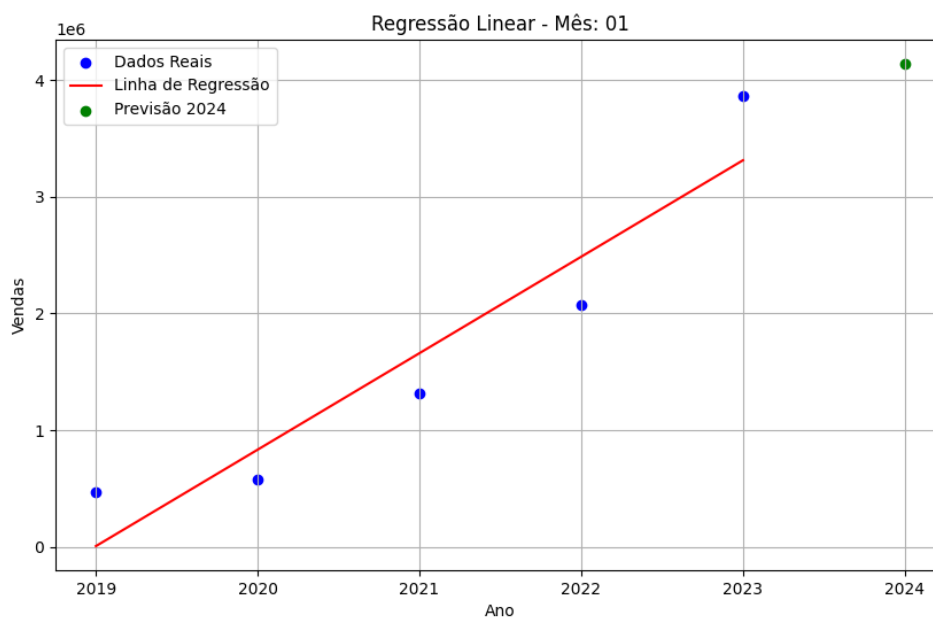
# Ajusta o layout para evitar sobreposição
plt.subplots_adjust(left=0.1, right=0.7, wspace=0.5)

# Exibe o gráfico e a tabela
plt.show()

# Salva as vendas em um arquivo .pkl para uso no código de validação
pd.to_pickle({'vendas_2019': vendas_2019, 'vendas_2020': vendas_2020,
             'vendas_2021': vendas_2021, 'vendas_2022': vendas_2022, 'vendas_2023':
             vendas_2023, 'previsoes_2024': previsoes_2024}, 'dados_para_validacao.pkl')
```

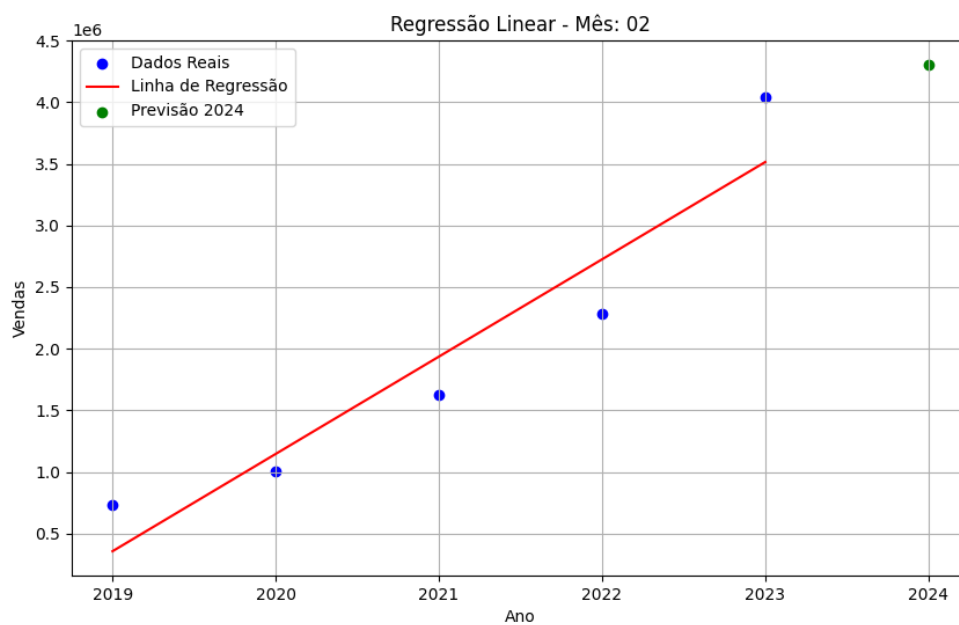
Entre as figuras 20 a 33, é possível analisar os gráficos obtidos através do algoritmo:

Figura 20 - Gráfico de Dispersão Janeiro



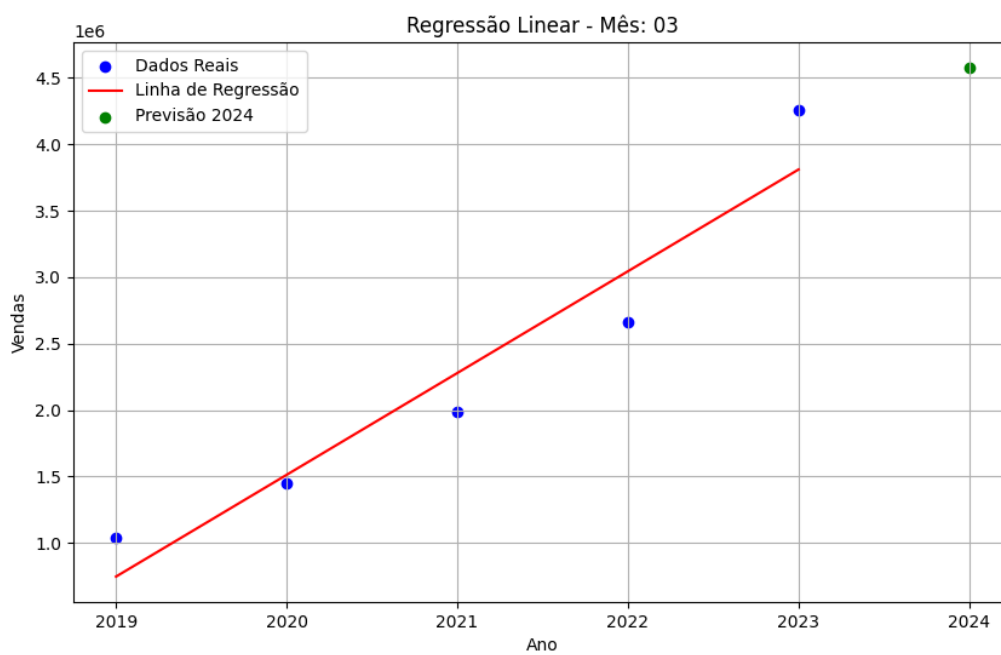
Fonte: Autoria própria

Figura 21 - Gráfico de Dispersão Fevereiro



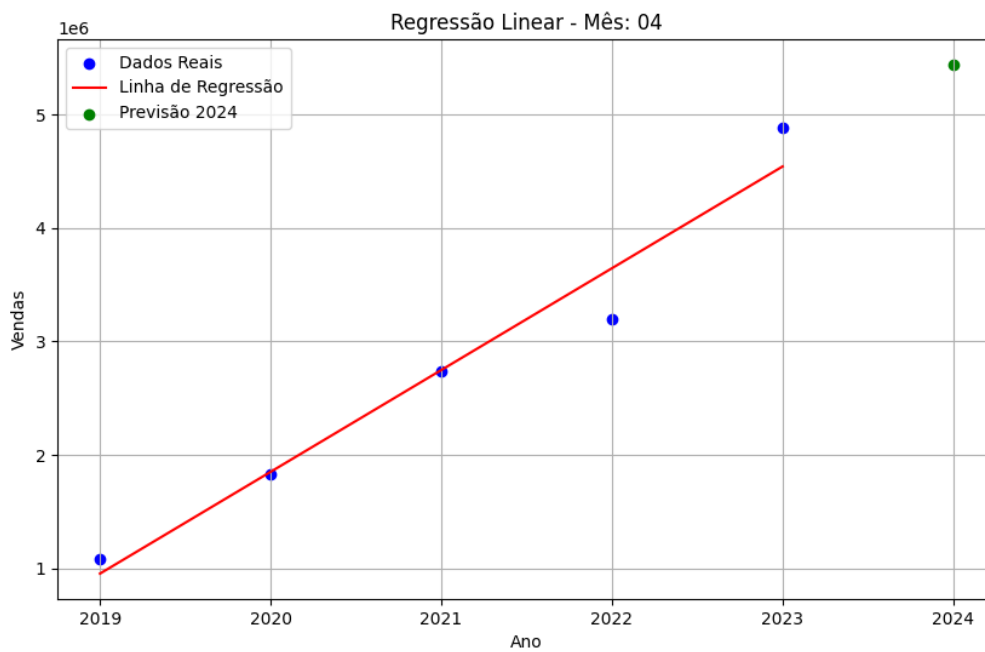
Fonte: Autoria própria

Figura 22 - Gráfico de Dispersão Março



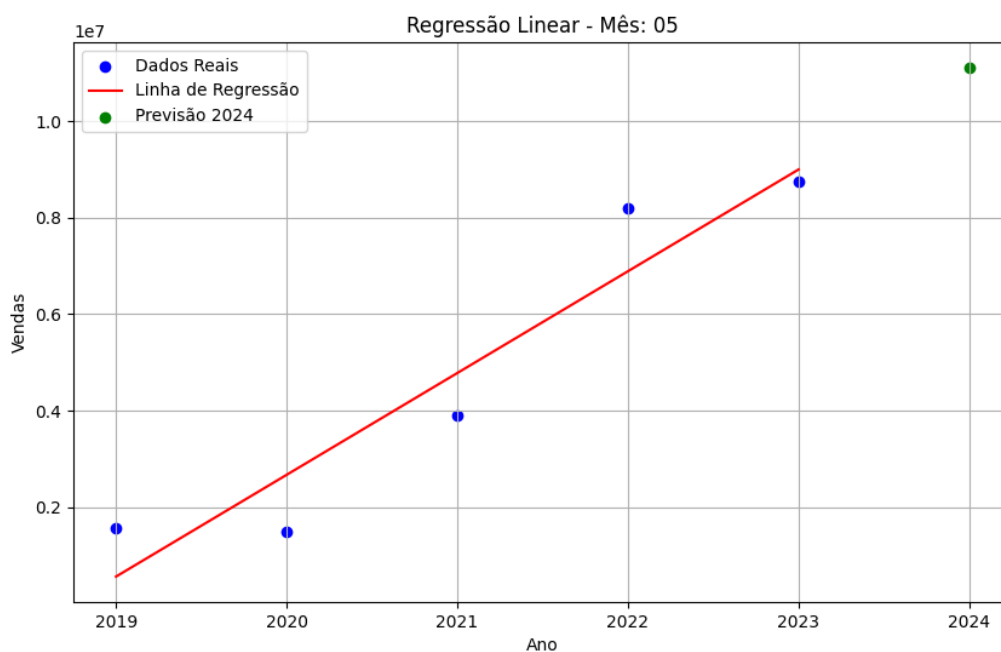
Fonte: Autoria própria

Figura 23 - Gráfico de Dispersão Abril



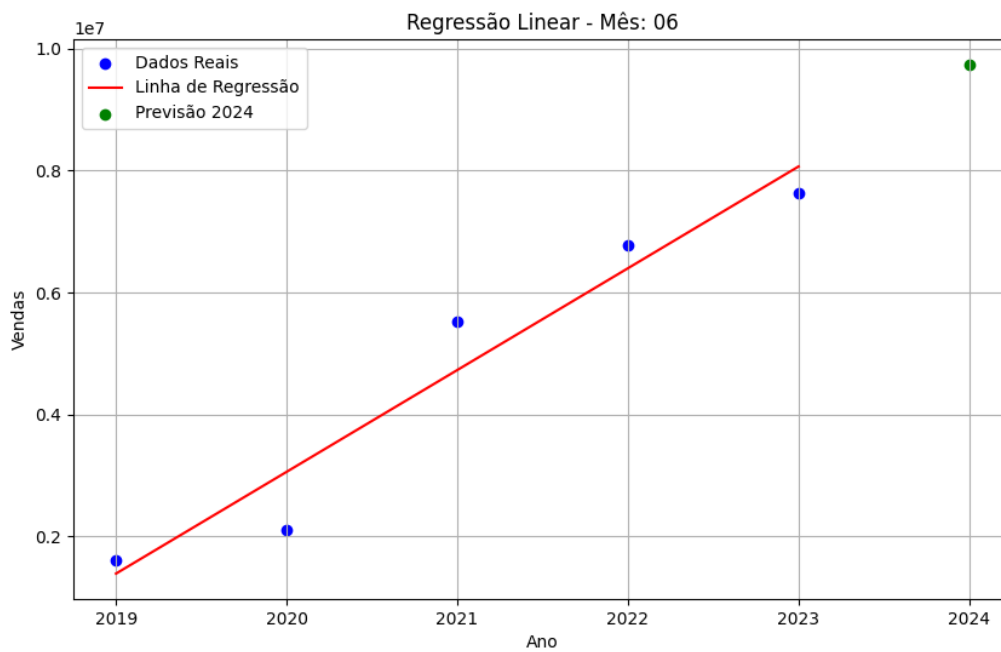
Fonte: Autoria própria

Figura 24 - Gráfico de Dispersão Maio



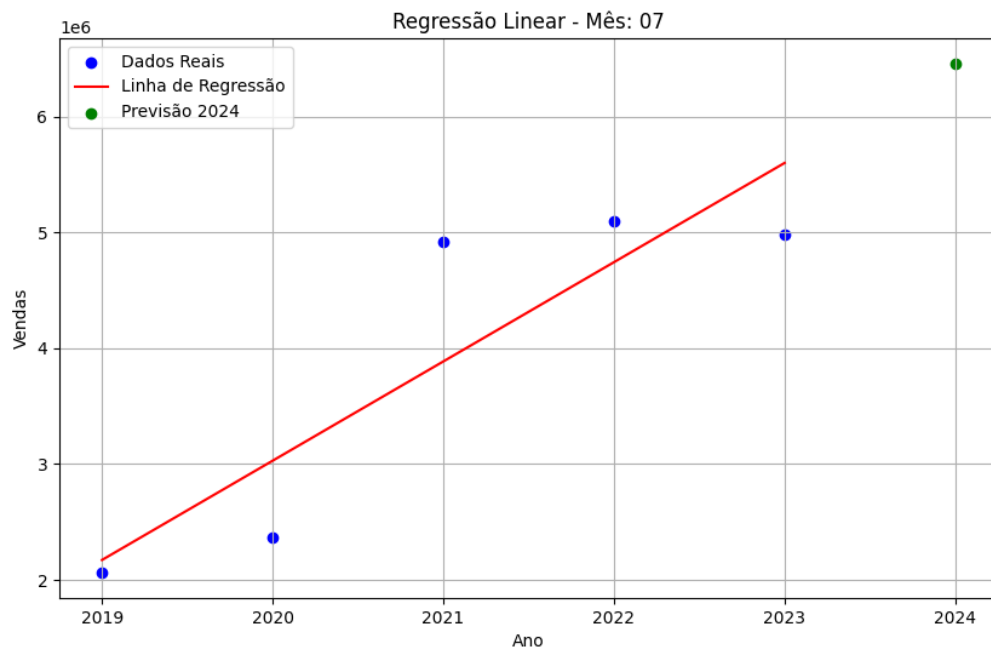
Fonte: Autoria própria

Figura 25 - Gráfico de Dispersão Junho



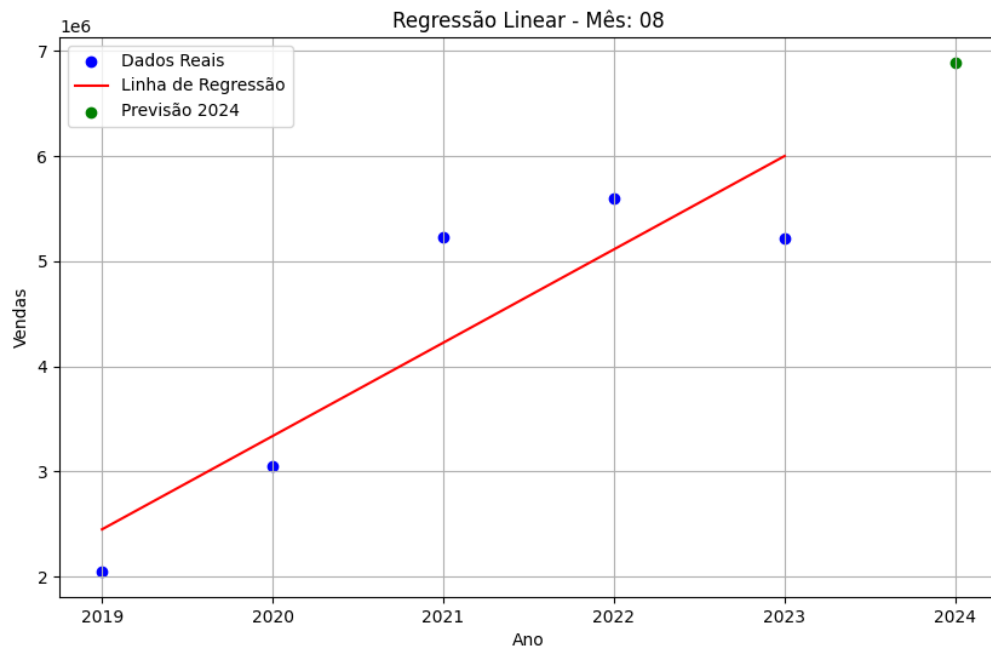
Fonte: Autoria própria

Figura 26 - Gráfico de Dispersão Julho



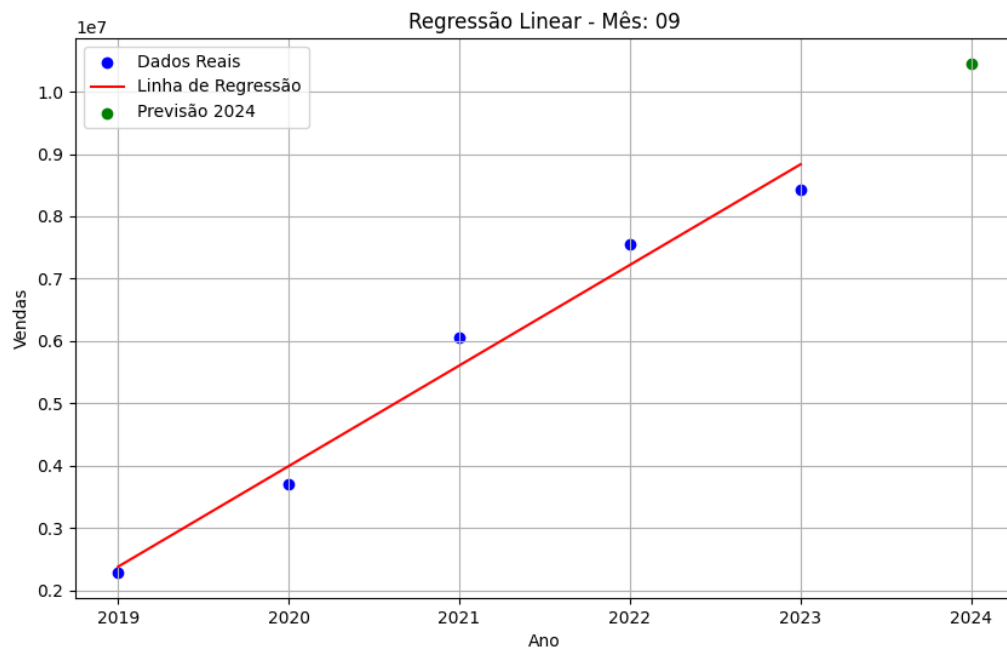
Fonte: Autoria própria

Figura 27 - Gráfico de Dispersão Agosto



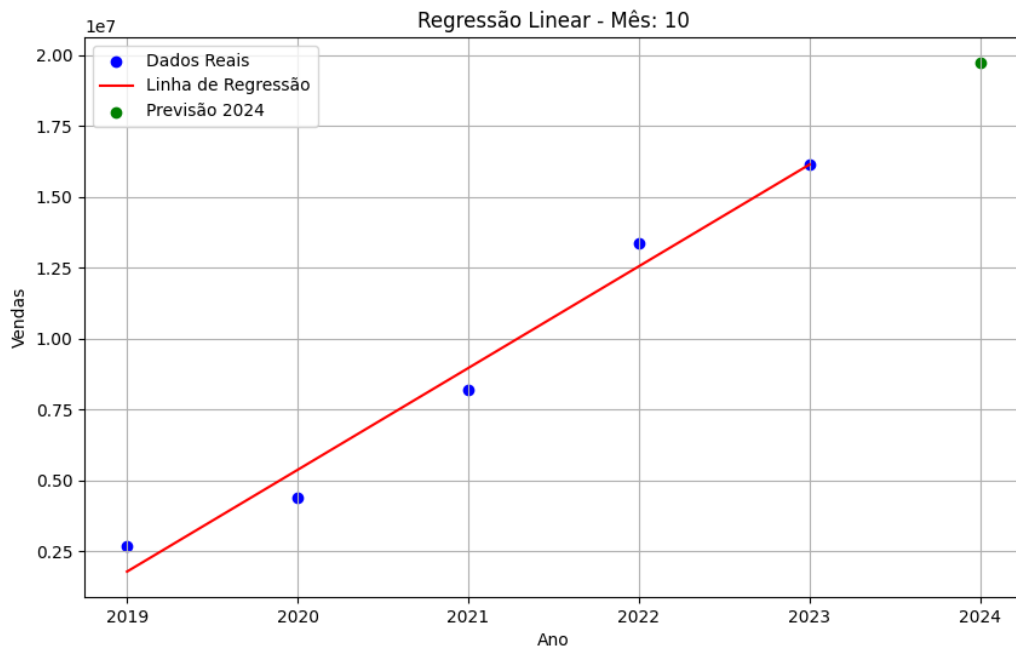
Fonte: Autoria própria

Figura 28 - Gráfico de Dispersão Setembro



Fonte: Autoria própria

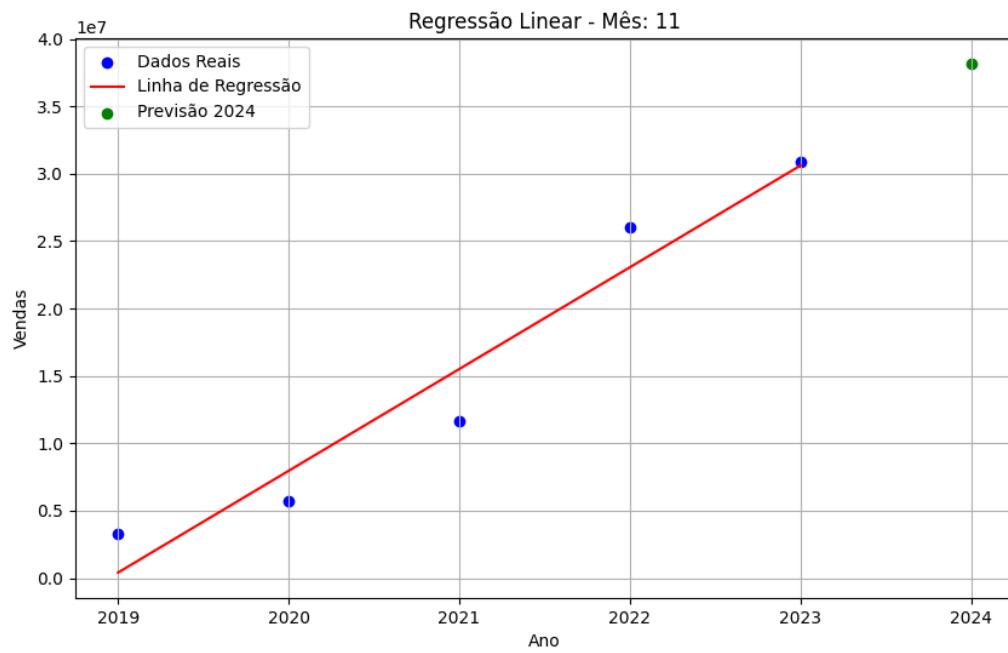
Figura 29 - Gráfico de Dispersão Outubro



Fonte: Autoria própria

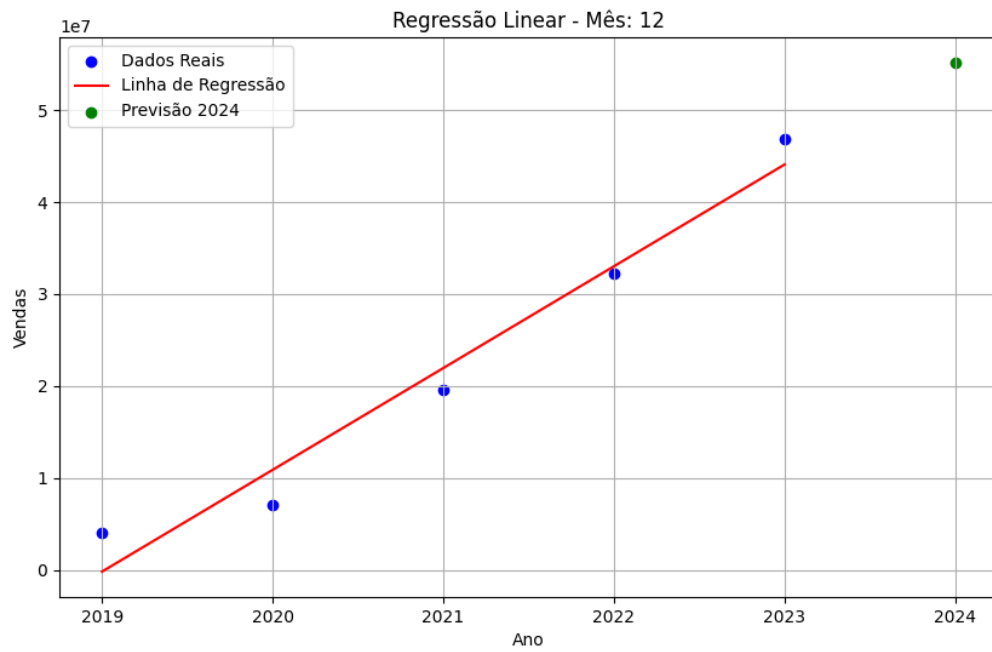


Figura 30 - Gráfico de Dispersão Novembro



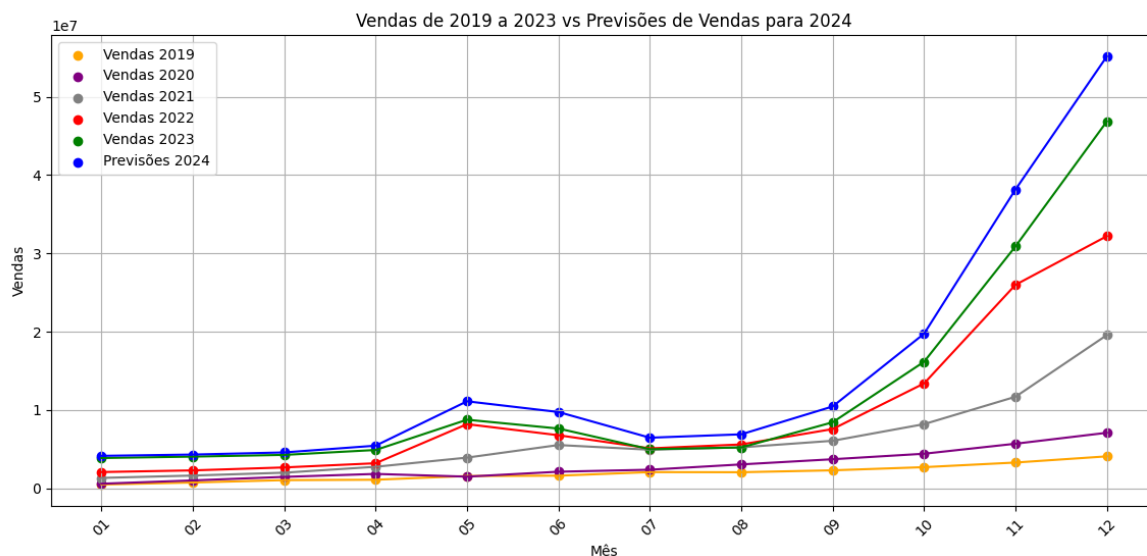
Fonte: Autoria própria

Figura 31 - Gráfico de Dispersão Dezembro



Fonte: Autoria própria

Figura 32 - Gráfico de linhas com pontos de dispersão



Fonte: Autoria própria

Figura 33 - Tabela de Dados Reais vs Predição

Mês	Vendas 2019	Vendas 2020	Vendas 2021	Vendas 2022	Vendas 2023	Predição 2024
01	473,001.80	578,014.00	1,319,391.10	2,069,283.00	3,859,934.80	4,139,465.44
02	733,564.10	1,000,227.30	1,620,850.60	2,280,705.90	4,040,882.80	4,303,780.94
03	1,036,055.50	1,451,719.90	1,986,987.70	2,659,387.60	4,259,954.90	4,575,461.07
04	1,087,234.20	1,830,223.80	2,742,271.50	3,194,212.40	4,888,012.60	5,438,054.52
05	1,563,365.20	1,482,527.90	3,913,060.10	8,186,692.70	8,753,746.00	11,105,356.30
06	1,604,633.90	2,111,648.00	5,521,191.20	6,772,324.70	7,626,940.20	9,738,934.39
07	2,063,102.00	2,367,774.90	4,920,112.90	5,094,382.10	4,984,177.10	6,456,537.02
08	2,046,877.90	3,045,596.10	5,223,964.30	5,595,842.20	5,212,324.80	6,889,263.03
09	2,290,608.00	3,706,333.20	6,062,996.10	7,561,457.30	8,425,524.40	10,446,870.87
10	2,694,050.70	4,408,030.60	8,188,730.60	13,389,236.20	16,156,635.40	19,739,249.20
11	3,285,089.80	5,688,592.10	11,686,223.50	25,987,380.10	30,891,779.20	38,161,462.98
12	4,073,999.80	7,085,047.50	19,567,055.90	32,204,282.90	46,864,165.90	55,168,780.68

Fonte: Autoria própria

### 3.5.3 Validação do Algoritmo de Predição

Com o intuito de verificar e validar os resultados obtidos através do modelo, foi desenvolvido um algoritmo de validação.

Seguindo as práticas implementadas nos algoritmos de ingestão e conexão com o banco de dados, o algoritmo de validação também se iniciou com a importação das bibliotecas pertinentes para a avaliação dos resultados.

Foram utilizadas as bibliotecas Pandas para manipulação de dados, Numpy para realização de operações numéricas, Matplotlib.pyplot e Seaborn para

visualização dos dados, Scipy.stats para funções estatísticas, Sklearn.metrics para métricas de avaliação e Sklearn.linear\_model para implementação de regressão linear.

```
""" Projeto: Trabalho de Conclusão de Curso UNIP
Descrição: Validação do Algoritmo de Predição
Autores: André Maldonado, Caio Teixeira, Francieli Muniz, Gabriel Aparecido,
João Soares e Lais Falcochio
(c) """
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import scipy.stats as stats
from sklearn.metrics import mean_squared_error, mean_absolute_error
from sklearn.linear_model import LinearRegression
```

Os próximos passos carregaram os dados gerados no algoritmo de predição, prepararam os dados convertendo-os em valores numéricos como preparação para o modelo de validação e criou o *DataFrame* para análise de resíduos.

```
# Carrega as vendas e previsões geradas no sales_algorithm.py
resultados = pd.read_pickle('dados_para_validacao.pkl')

# Converte as vendas e previsões para formato numérico
meses = list(range(1, 13))
vendas_2019 = [float(venda.replace(',', '')) for venda in resultados['vendas_2019']]
vendas_2020 = [float(venda.replace(',', '')) for venda in resultados['vendas_2020']]
vendas_2021 = [float(venda.replace(',', '')) for venda in resultados['vendas_2021']]
vendas_2022 = [float(venda.replace(',', '')) for venda in resultados['vendas_2022']]
vendas_2023 = [float(venda.replace(',', '')) for venda in resultados['vendas_2023']]
previsoes_2024 = [float(prev.replace(',', '')) for prev in resultados['previsoes_2024']]

# Prepara os dados para o modelo
dados_venda = pd.DataFrame({
    'ano': [2019, 2020, 2021, 2022, 2023],
    'vendas': [np.mean(vendas_2019), np.mean(vendas_2020),
np.mean(vendas_2021), np.mean(vendas_2022), np.mean(vendas_2023)]
})

# Modelo de regressão linear
```

```
X = dados_venda[['ano']]
y = dados_venda['vendas']

model = LinearRegression()
model.fit(X, y)

# Adiciona as previsões do DataFrame original para fazer a análise de resíduos
dados_venda['previsoes'] = model.predict(X)
dados_venda['residuos'] = dados_venda['vendas'] - dados_venda['previsoes']
```

Um Gráfico de Resíduos foi plotado para verificar se existem padrões não capturados pelo modelo, seguido por um Gráfico Q-Q (*Quantile-Quantile*) para verificar se os resíduos se ajustam em linha reta.

```
# Plota o gráfico de resíduos
plt.figure(figsize=(10, 6))
sns.scatterplot(x=dados_venda['previsoes'], y=dados_venda['residuos'])
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel('Previsões')
plt.ylabel('Resíduos')
plt.title('Resíduos vs Previsões')
plt.show()

# Normalidade dos resíduos
fig = plt.figure(figsize=(10, 6))
stats.probplot(dados_venda['residuos'], dist="norm", plot=plt)
plt.title('Gráfico Q-Q dos Resíduos')
plt.show()
```

Por fim, calculou-se os valores para as métricas MAE, MSE, RMSE e R<sup>2</sup>, juntamente com os percentuais de erro.

```
# Calcula as métricas de validação
mae = mean_absolute_error(dados_venda['vendas'], dados_venda['previsoes'])
mse = mean_squared_error(dados_venda['vendas'], dados_venda['previsoes'])
rmse = np.sqrt(mse)
r2 = model.score(X, y)

# Formatação dos valores das métricas para incluir separadores de milhar e
exibição dos valores
print(f'R²: {r2:.3f}')
print(f'MAE: {mae:,.2f}')
print(f'MSE: {mse:,.2f}')
print(f'RMSE: {rmse:,.2f}')

# Calcula a média dos valores reais
vendas_media = dados_venda['vendas'].mean()
```

```
# Percentual do MAE (Erro Médio Absoluto) e RMSE (Raiz do Erro Quadrático Médio)
percentual_mae = (mae / vendas_media) * 100
percentual_rmse = (rmse / vendas_media) * 100

print(f'Percentual de erro MAE: {percentual_mae:.2f}%')
print(f'Percentual de erro RMSE: {percentual_rmse:.2f}%')
```

### 3.5.4 Análise e conclusão da confiabilidade do modelo

O primeiro código (Algoritmo de Predição), teve como objetivo principal a previsão das vendas para o ano de 2024, com base em dados históricos das vendas, sendo utilizado na fase de desenvolvimento do modelo.

Já o segundo código (Validação do Algoritmo de Predição) teve como objetivo a validação das previsões feitas inicialmente, validando o desempenho e precisão do modelo de regressão linear simples, sendo utilizado na fase de validação do modelo.

Como métricas de avaliação, foram avaliados os seguintes pontos que podem ser observados na figura 34.

Figura 34 - Resultados das métricas de validação

```
R²: 0.975
MAE: 547,758.42
MSE: 375,604,994,563.86
RMSE: 612,866.21
Percentual de erro MAE: 8.40%
Percentual de erro RMSE: 9.39%
```

Fonte: Autoria própria

$R^2$  (Coeficiente de Determinação)

$$R^2 = \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

0.975

De acordo com Martins (2018), o valor de  $R^2 \approx 1$  demonstra que os pontos apresentados na dispersão estão próximos da reta de regressão, sendo assim, o modelo desenvolvido explica variabilidade dos dados em 97,5%, significando que o modelo é eficaz na captura de variações nas vendas.

MAE (Erro Médio Absoluto)

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

547.758,42 com percentual de erro 8,40% da média das vendas.

O modelo apresenta média de erros absolutos entre as previsões e os valores reais, significando que as previsões estão errando por 8,40% da média das vendas, sendo um valor relativamente baixo.

MSE (Erro Quadrático Médio)

$$\frac{1}{n} \sum (y - \hat{y})^2$$

375.604.994.563,86

Conforme descrito por Chicco (2021), o MSE é mais sensível a *outliers* e a média dos quadrados dos erros penaliza erros maiores quando comparado ao MAE pois seus erros são elevados ao quadrado. Ele é interpretado juntamente com o RMSE.

RMSE (Raiz do Erro Quadrático Médio)

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

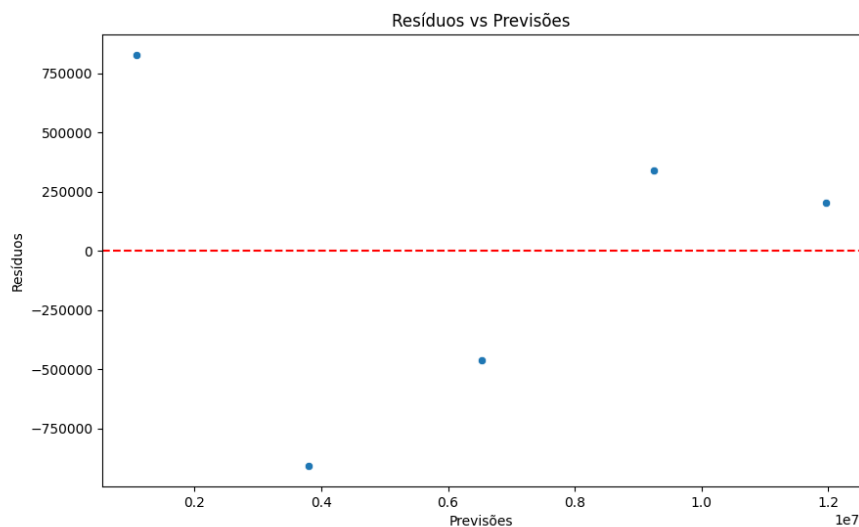
612.866,21 com percentual de erro 9,39% da média das vendas.

Seguindo a descrição de Chicco (2021), o RMSE é a raiz do MSE que fornece a medida de erros da variável dependente (vendas), o MSE e RMSE são relacionadas através da raiz quadrada. Através do modelo, é possível concluir que as previsões estão errando por 9,39% da média das vendas, sendo também um valor relativamente baixo.

Gráfico de Resíduos vs Previsões

De acordo com Rodrigues (2012), o gráfico deve apresentar os pontos aleatoriamente e sem nenhum padrão em torno da reta  $y = 0$ . Com essa análise é possível verificar na figura 35, que a variância dos resíduos é constante.

Figura 35 - Gráfico de Resíduos vs Predição

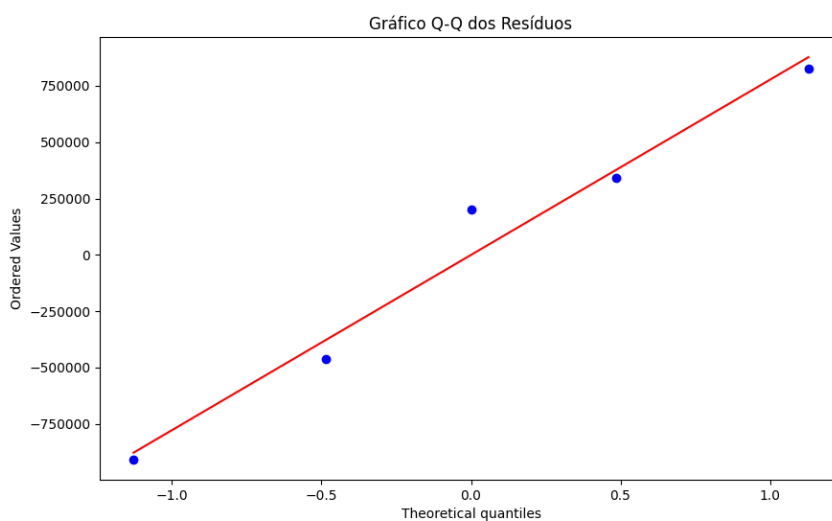


Fonte: Autoria própria

### Normalidade dos Resíduos

Segundo a descrição de Rodrigues (2012), a normalidade dos resíduos pode ser verificada quando os resíduos se aproximam da reta diagonal, sem um afastamento significativo, ao analisar a figura 36, é possível concluir que os resíduos atendem a normalidade.

Figura 36 - Gráfico Q-Q dos Resíduos



Fonte: Autoria própria

Concluiu-se que o modelo está fornecendo previsões de qualidade, com bom ajuste dos dados e os erros são relativamente baixos (dentro do esperado pela área de negócio da empresa CALFJ & Co, se mantendo abaixo de 10% tanto no MAE quanto no RMSE). O modelo é aceitável e eficaz para realizar previsões e entrega um resultado confiável com erros que compõem uma pequena parte das vendas médias, o MSE é muito grande, o que também é esperado devido a ampliação dos erros que são elevados ao quadrado.

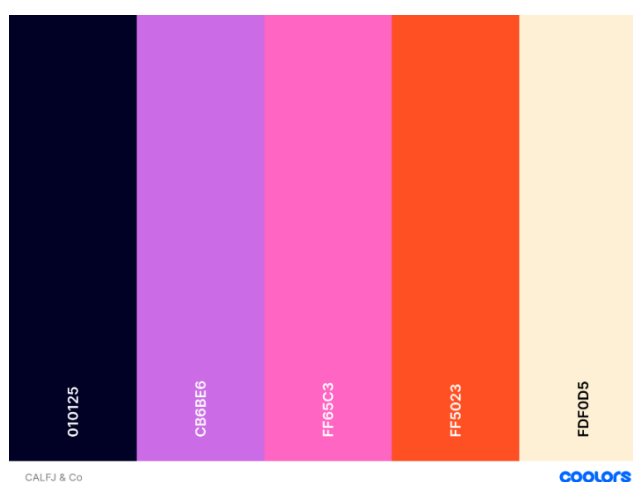
### 3.6 Figma

De acordo com Chung (2021), o Figma é um serviço *online* para desenvolvimento de interfaces e prototipagem. Foi desenvolvido por Evan Wallace e Dylan Field em 2015 e inicialmente seus serviços eram disponibilizados apenas para convidados. Atualmente a plataforma Figma é gratuita e oferece todas as ferramentas necessária para a fase de design de protótipos, sejam simples ou complexos.

#### 3.6.1 Brand e Paleta de Cores

A paleta de cores conta com os códigos hexadecimais que podem ser visualizados na figura 37.

Figura 37 - Paleta de Cores



Fonte: Autoria própria

Como forma de identidade visual, também foram desenvolvidos os logos, para identificação da marca, que podem ser visualizados na figura 38.



Figura 38 - Logos

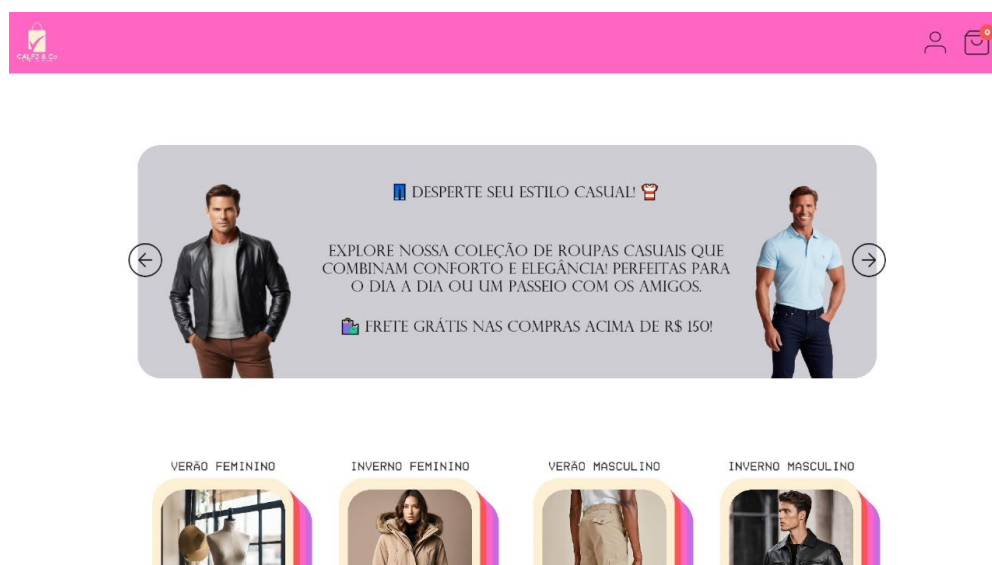


Fonte: Autoria própria

### 3.6.2 Mockup site

O site é composto por tela inicial (*home*), tela de seleção, tela de carrinho (*checkout*), tela de seleção de endereço, tela de seleção de pagamento, tela de revisão e tela de conclusão do pedido.

Na figura 39, é apresentada a tela *home* do site, que é composta por um cabeçalho com logo, botão de conta, botão de *checkout*, e carrossel com apresentação de anúncios, quatros botões de direcionamento.

Figura 39 - Tela *home*

Fonte: Autoria própria

A figura 40 demonstra a tela de seleção de roupas, sendo possível visualizar informações sobre a peça escolhida, o botão de selecionar cor, tamanho, adicionar ao carrinho e fechar a aba.

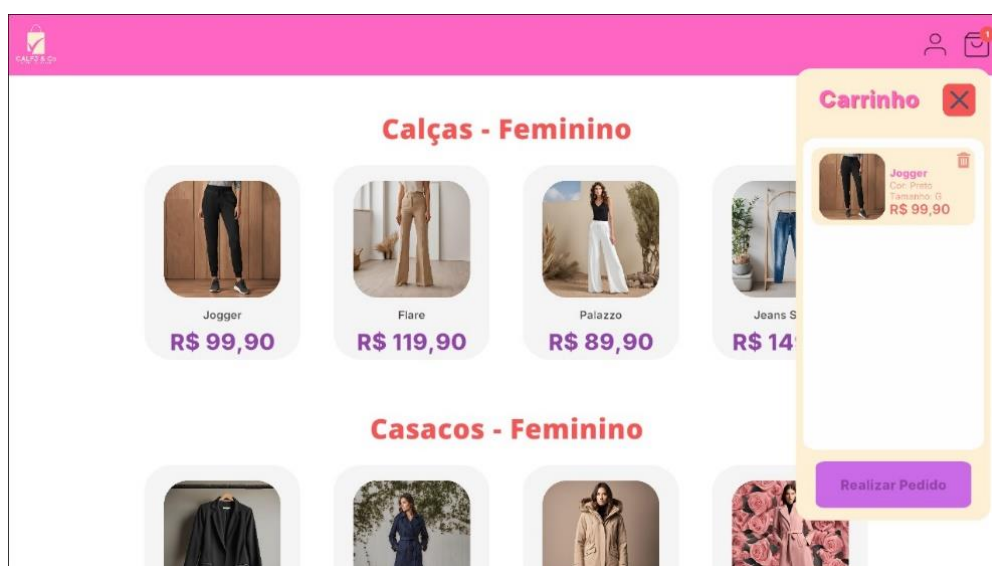
Figura 40 - Tela de Seleção



Fonte: Autoria própria

Na figura 41 é possível observar a tela do carrinho de compras, com os itens inseridos e botão para realizar o pedido.

Figura 41 - Tela Carrinho



Fonte: Autoria própria

A tela de seleção de endereço mostra as opções de endereços para entrega, cadastrados pelos usuários, conforme figura 42.

Figura 42 - Seleção de Endereço



Fonte: Autoria própria

A tela para seleção do método de pagamento pode ser visualizada na figura 43.

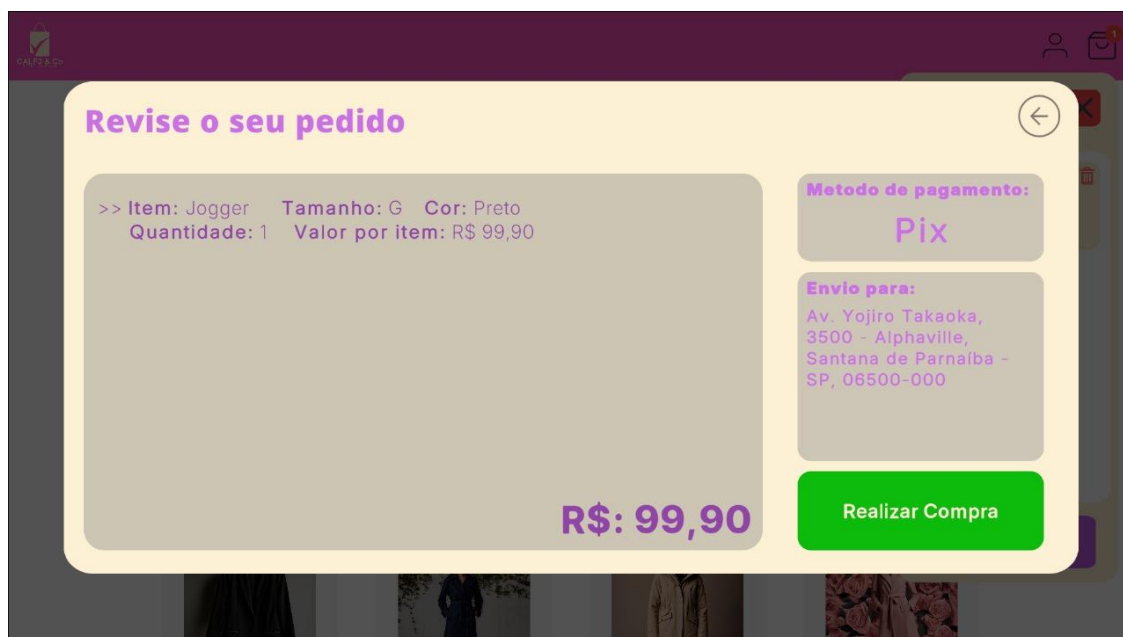
Figura 43 - Seleção de Método de Pagamento



Fonte: Autoria própria

Na figura 44, é possível visualizar a tela de revisão do pedido, onde é apresentado os dados de envio, pagamento e informações sobre os itens selecionados.

Figura 44 - Revisão de Pedido



Fonte: Autoria própria

Na figura 45, é possível visualizar a tela que é apresentada ao usuário após a realização de um pedido.

Figura 45 - Finalização do Pedido

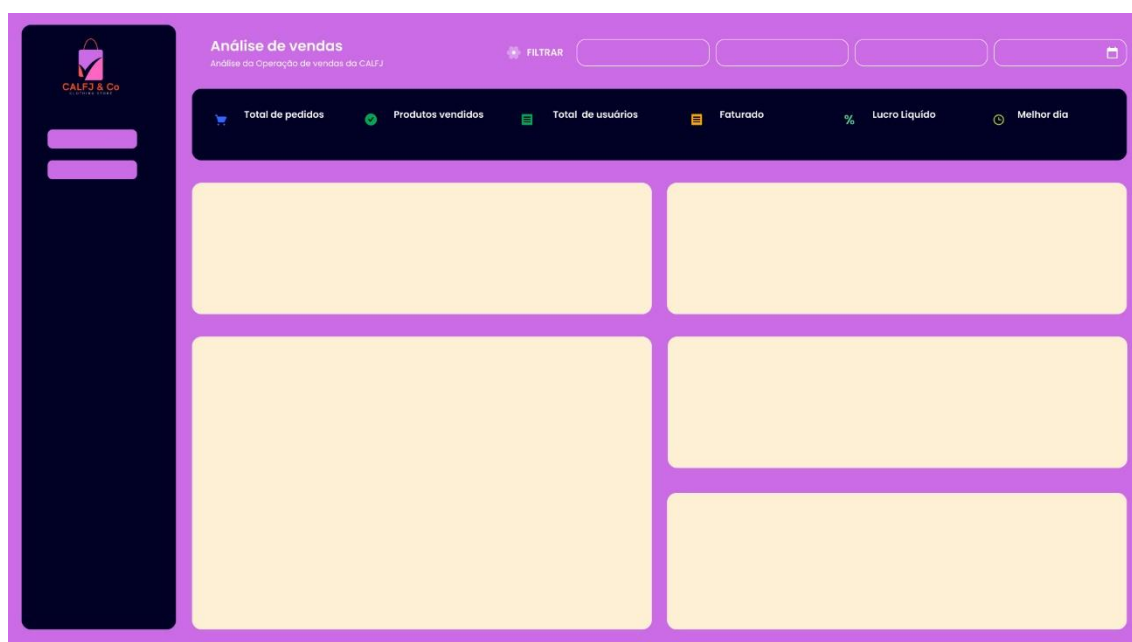


Fonte: Autoria própria

### 3.6.3 Mockup dashboard

Para o *layout* do *dashboard*, foi desenvolvido um modelo simples, com botões de navegação lateral e menu de resumo de indicadores na parte superior. Foi determinado campos de filtros a serem incluídos de acordo com a necessidade do negócio, e um filtro de data/período para personalizar a análise do usuário. A resolução do *layout* é *full hd* (1920x1080). Na figura 46 é possível visualizar o *layout* criado no Figma e a aplicação no PowerBI.

Figura 46 - Layout dashboard



Fonte: Autoria própria

## 3.7 PowerBI

De acordo com a *Microsoft*, o Power BI é um *software*, que transforma fontes de dados não relacionadas em *insights* visualmente atrativos e interativos.

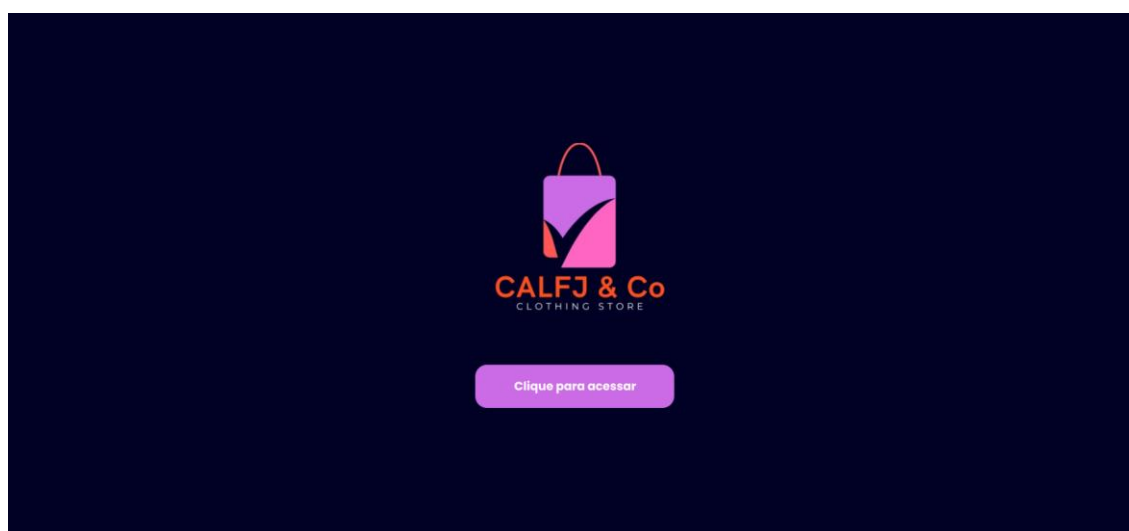
“Os dados podem estar em uma planilha do Excel ou em uma coleção de data *warehouses* híbridos locais ou baseados na nuvem. Com o Power BI, você pode se conectar facilmente a fontes de dados, visualizar e descobrir conteúdo importante e compartilhá-lo com todas as pessoas que quiser.”  
(MICROSOFT, 2024)

Nesta ferramenta é possível fazer todo o processo de ETL (*Extract Transform and Loading*), Extração, Transformação e Carregamento em tradução direta. Os dados podem estar em diversos tipos de fontes de dados, como tabelas e bancos de dados. Com essa ferramenta é possível criar visuais e trazer vida para os dados,

podendo criar *insights* e *dashboards* interativos, transformando informações complexas em indicadores de fácil compreensão. Além disso com o *PowerQuery* é possível fazer o processo de transformação, mudando o formato dos dados, fazendo filtros ou qualquer mudança com a linguagem DAX.

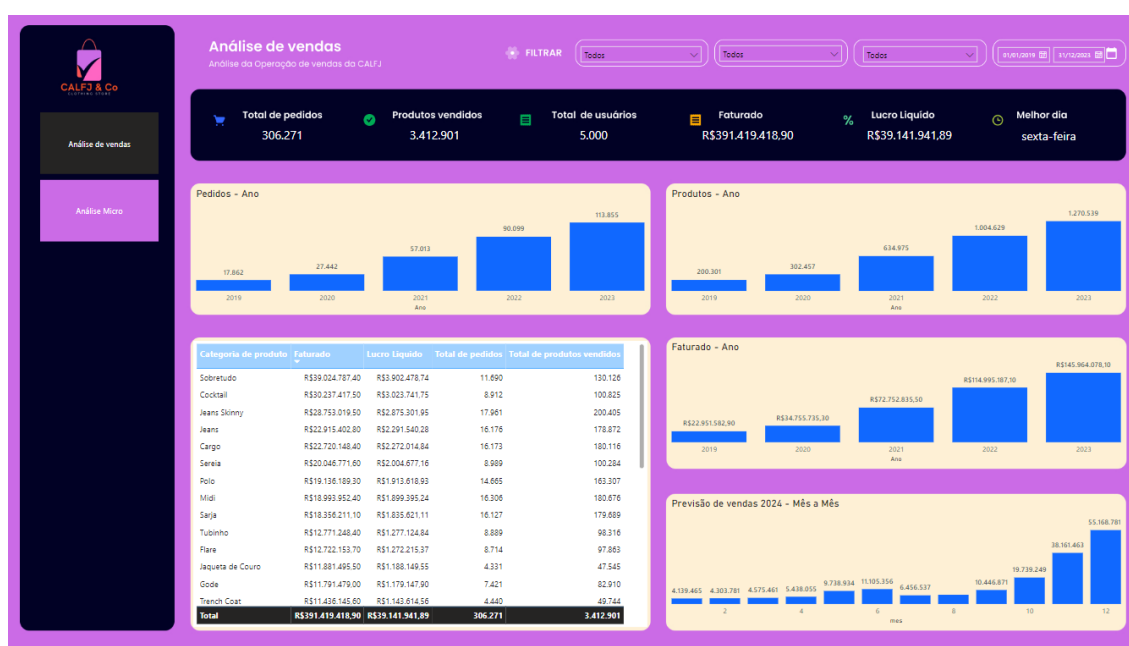
Nas figuras 47, 48 e 49 é possível observar a tela inicial do dashboard no PowerBI com descrição dos botões de acesso, página de análise de vendas e página de análise micro.

Figura 47 - Página inicial dashboard



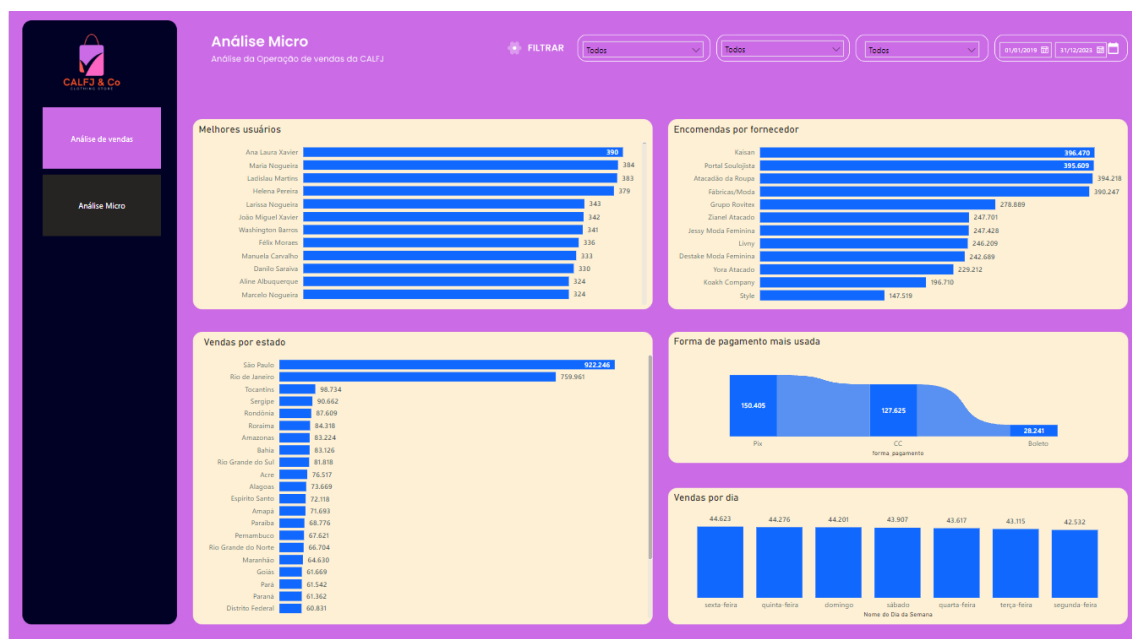
Fonte: Autoria própria

Figura 48 - Página de análise micro



Fonte: Autoria própria

Figura 49 - Página de análise de vendas



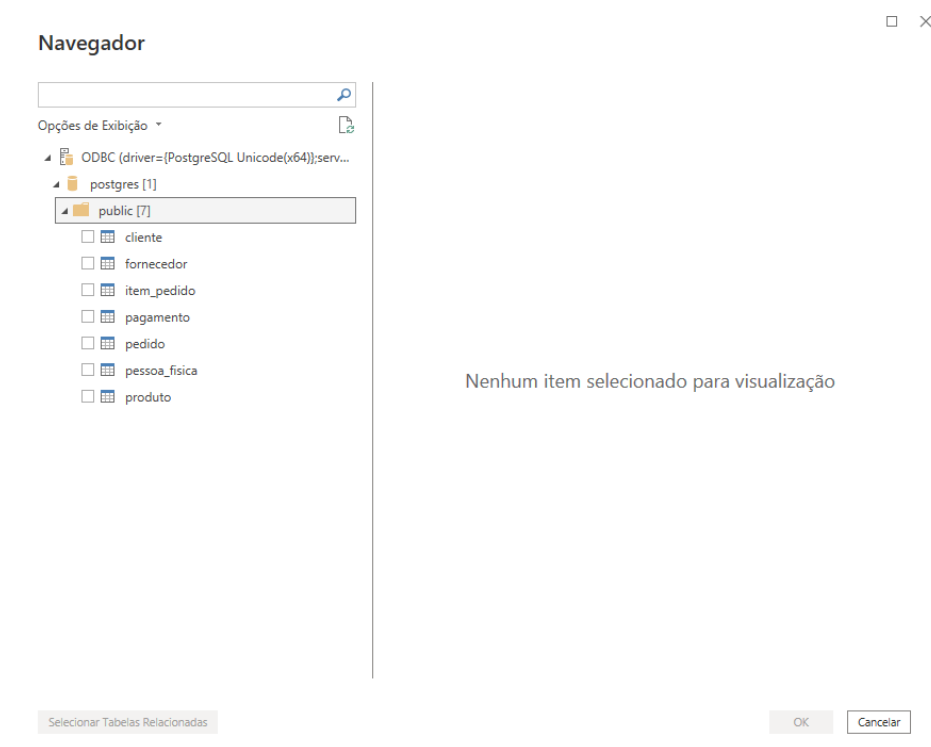
Fonte: Autoria própria

### 3.7.1 Conexão do Banco de Dados com o PowerBI

Inicialmente, seria utilizado o conector nativo do PowerBI para PostgreSQL, porém, houve dificuldades com essa conexão, e a fim de contornar o problema, foi realizada a conexão via ODBC (*Open Database Connectivity*), através da *string* de conexão, utilizando o certificado digital SSL (*Secure Sockets Layer*) para garantir que todos os dados transmitidos entre o cliente e servidor de dados serão criptografados, garantindo segurança e integridade das informações.

Após a conexão, foi possível acessar o *schema* e visualizar todas as tabelas que foram importadas para o PowerBI, conforme figura 50.

Figura 50 - Schema no PowerBI



Fonte: Autoria própria

Uma vez que a tabela é importada, pode-se manter o PowerBI atualizado de maneira automática, sem a necessidade de refazer o procedimento de conexão.

### 3.7.2 Modelo de Relação das tabelas no PowerBI

Na figura 51 é possível observar todo o relacionamento das tabelas no PowerBI, que foram resumidos em:

**Relacionamento:** Tabela fornecedor ⇔ Tabela Produto

- fornecedor.cod\_fornecedor (PK) ⇔ produto.cod\_fornecedor (FK)
- Tipo de relacionamento: Um para muitos (1:\*);
- Direção do Filtro cruzado: Único

**Relacionamento:** Tabela Produto ⇔ Tabela item\_pedido

- produto.cod\_produto (PK) ⇔ item\_pedido.cod\_produto (FK)
- Tipo de relacionamento: Muitos para um (\*:1);
- Direção do Filtro cruzado: Único

**Relacionamento:** Tabela pessoa\_fisica ⇔ Tabela cliente

- pessoa\_fisica.cod\_cliente (PK) ⇔ cliente.cod\_cliente (PK)



- Tipo de relacionamento: um para um (1:1);
- Direção do Filtro cruzado: Ambos

**Relacionamento:** Tabela cliente ↔ Tabela pedido

- cliente.cod\_cliente (PK) ↔ pedido.cod\_cliente (FK)
- Tipo de relacionamento: um para um (1:1);
- Direção do Filtro cruzado: Único

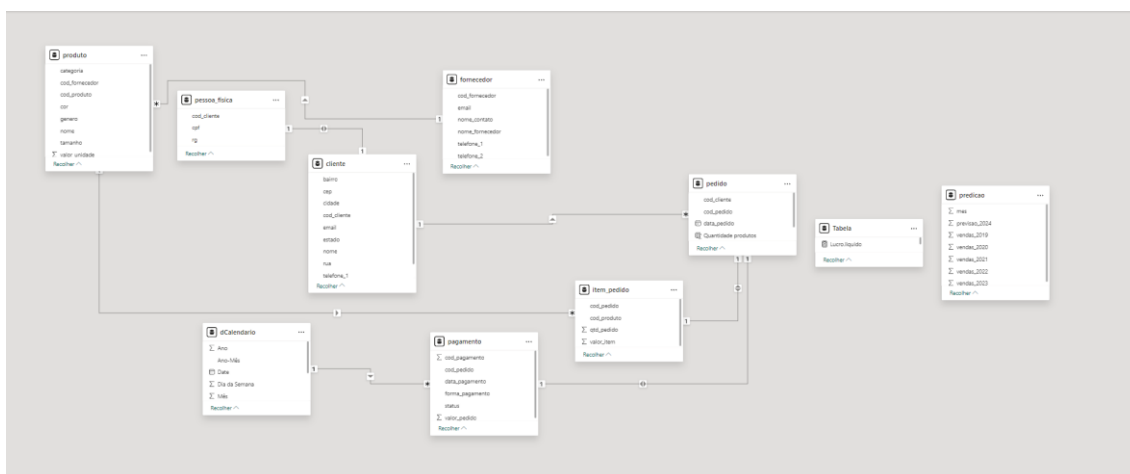
**Relacionamento:** Tabela pedido ↔ Tabela item\_pedido

- pedido.cod\_pedido (PK) ↔ item.cod\_pedido (FK)
- Tipo de relacionamento: um para um (1:1);
- Direção do Filtro cruzado: Ambos

**Relacionamento:** Tabela pagamento ↔ Tabela pedido

- pedido.cod\_pedido (PK) ↔ pagamento.cod\_pedido (FK)
- Tipo de relacionamento: um para um (1:1);
- Direção do Filtro cruzado: Ambos

Figura 51 - Modelo de Relação PowerBI



Fonte: Autoria própria

### 3.7.3 Linguagem DAX

De acordo com a *Microsoft (2024)*, a *Data Analysis Expressions (DAX)*, é uma linguagem utilizada em análises no PowerBI, as fórmulas incluem operadores, funções e valores para realizar consultas em dados e cálculos avançados.

A DAX auxilia na criação de novas informações a partir de dados que já existem na base de dados, contribuindo para trazer *insights* de uma forma mais assertiva.

Quando a base de dados é importada da fonte pelo processo de *Extract*, o PowerBI já permite a criação de análises a partir das informações extraídas, com a DAX também é possível fazer cálculos dentro da base, comparar vendas e crescimento entre períodos, possibilitando tirarmos o máximo proveito dos dados.

#### 3.7.4 Processo ETL

De acordo com a *International Business Machines* (IBM, 2024), o método de *Extract, Transform e Load* (ETL), é um processo de integração de dados de longa data usado para combinar dados de várias fontes em um conjunto de dados único. O ETL limpa e organiza os elementos de forma a atender necessidades específicas de *business intelligence*, como relatórios mensais, mas também pode lidar com análises mais avançadas, que podem melhorar processos de *back-end* ou experiências de usuário final.

#### 3.7.5 Power Query

Segundo Seminger (2024) da equipe *Microsoft Learn*, Power Query é uma ferramenta de desenvolvimento e preparação de dados. Atua com interface gráfica para obter dados de fontes e um editor do *Power Query* para aplicar as transformações.

Utilizando *Power Query* é possível processar os dados através do conceito extrair, transformar e carregar, conhecido como “ETL”. O objetivo desta interface é auxiliar as transformações de dados com seu editor simples e amigável de se interagir, desta maneira facilitando para o usuário.

#### 3.7.6 Link Dashboard

<https://app.powerbi.com/view?r=eyJrIjoiZTNhYTI1NGMtYTA0Ni00NGZkLTlkMjU0N2I3ZGIwYjUyMDNmliwidCI6IjM2MjdmZTMwLWFjODItNGRkYi1iMGFjLTUxMDcxNjdjYzc1NSJ9&pageName=4778ca44cd1512cb609e>

### 3.8 Unity

Conforme descrito pela empresa *Unity Technologies*, *Unity* é um *game engine* 2D e 3D, que permite aos desenvolvedores a criação de conteúdos interativos em diversas plataformas, como computadores, consoles, dispositivos móveis e realidade virtual e aumentada.

Foi criado em 2004 por David Helgason, Joachin Ante e Nicholas Francis, sendo criado inicialmente com foco em Mac OS X, mas com o passar do tempo tornou-se um dos mais populares e poderosos motores de criação da indústria, possui como base a linguagem de programação C#. Oferece uma licença gratuita, permitindo o desenvolvedor utilizar a maior parte de duas funcionalidades básicas, para projetos mais complexos ou para fins comerciais, também oferece assinaturas e recursos diferentes.

#### 3.8.1 Introdução ao Relatório de Desenvolvimento

Este relatório tem como objetivo detalhar as atividades realizadas durante esta etapa do desenvolvimento do projeto de software, focando nos aspectos técnicos e nas decisões de design adotadas. A etapa atual visa a criação de um sistema dinâmico para gerenciar um catálogo de itens com funcionalidades interativas, como a manipulação de eventos de interface gráfica, a integração com um banco de dados remoto e a interação do usuário com os itens visualizados. Além disso, o sistema foi projetado para oferecer uma experiência intuitiva e eficiente, adaptando-se às necessidades do usuário final.

O desenvolvimento foi realizado utilizando a Unity, uma das mais robustas plataformas de criação de jogos e interfaces interativas, empregando a linguagem C#. As comunicações constantes via chat desempenharam um papel fundamental na formulação de soluções, permitindo iterações rápidas e resoluções eficientes de problemas. A metodologia de desenvolvimento adotada seguiu uma abordagem ágil, permitindo adaptações durante o processo e garantindo que as entregas fossem realizadas de maneira contínua e incremental, como veremos nas seções a seguir.

### 3.8.2 Processo de desenvolvimento

O processo de desenvolvimento foi cuidadosamente estruturado para seguir um ciclo iterativo e incremental. Esta abordagem, baseada nos princípios da metodologia ágil, permitiu que o projeto avançasse em pequenos incrementos, com cada etapa sendo testada e validada antes de se prosseguir para a próxima. Isso proporcionou flexibilidade para realizar ajustes e melhorias à medida que problemas técnicos ou de usabilidade surgiam. Cada iteração focou em um conjunto específico de funcionalidades, o que garantiu um processo contínuo de refinamento e otimização.

No início da etapa, foi realizada uma análise das necessidades do sistema, identificando os requisitos principais, como a integração com um catálogo de produtos, o uso de JSON para o carregamento dinâmico de dados e a implementação de uma interface gráfica que permitisse uma interação fluida com o usuário. Durante o desenvolvimento, a equipe enfrentou desafios relacionados à exibição dinâmica de itens e à performance do sistema, que foram superados por meio de discussões colaborativas e ajustes no código.

Essa fase do projeto também incluiu a criação de protótipos para testar as funcionalidades principais, como a manipulação de *scrolls* e a instância dinâmica de *prefabs*. Tais protótipos serviram como base para ajustes posteriores, permitindo refinar aspectos relacionados à experiência do usuário, como tempos de resposta, suavidade nas animações e facilidade de navegação pelo catálogo.

### 3.8.3 Implementação de Funcionalidades

Nesta etapa do desenvolvimento, uma série de funcionalidades foi implementada, cada uma voltada para tornar o sistema mais eficiente e intuitivo. As principais funcionalidades desenvolvidas incluem:

- Instanciação Dinâmica de *Prefabs*

Um dos pilares desta fase foi a criação de um sistema que permite a instanciação dinâmica de itens do catálogo a partir de um arquivo JSON. O JSON contém informações detalhadas sobre cada item, como nome, cor, tamanho e valor, que são mapeadas para os elementos visuais da interface gráfica em Unity. Cada item do catálogo é representado por um *prefab* que, quando instanciado, carrega as informações correspondentes e se adapta à interface do usuário.

Esse sistema foi projetado para ser flexível, permitindo futuras expansões. Por exemplo, novos atributos podem ser facilmente adicionados ao JSON e integrados à interface gráfica, sem a necessidade de grandes modificações no código base. A abordagem adotada permite a reutilização de componentes e melhora a eficiência no gerenciamento de dados.

- Manipulação de *ScrollView*

Outro aspecto crítico foi a implementação de um *ScrollView* que permite ao usuário navegar por um catálogo potencialmente vasto de itens, sem comprometer a usabilidade do sistema. Um dos desafios enfrentados foi garantir que o *ScrollView* funcionasse de maneira fluida, mesmo com múltiplos itens instanciados dinamicamente.

A implementação incluiu otimizações no sistema de rolagem, garantindo que o desempenho não fosse afetado por grandes volumes de itens. Para isso, foram utilizados eventos de atualização parcial da interface (*Lazy Loading*), carregando novos itens apenas quando necessário, à medida que o usuário percorre a lista.

- Interações Dinâmicas

Um componente importante da experiência do usuário foi a adição de interações dinâmicas, como o evento de *hover*. Ao passar o mouse sobre os itens do catálogo, mudanças visuais são acionadas, como a alteração na cor do texto e o destaque dos produtos. Além disso, os itens respondem a cliques, permitindo que o usuário visualize detalhes adicionais ou adicione o item diretamente ao carrinho.

Essas interações dinâmicas foram implementadas com *EventTriggers*, que permitem a definição de comportamentos específicos para cada evento de interação do usuário, como '*PointerEnter*' (ao passar o mouse) e '*PointerClick*' (ao clicar no item). Isso resultou em uma interface mais interativa e intuitiva, onde o usuário pode explorar o catálogo de maneira mais eficiente e envolvente.

#### 3.8.4 Soluções de Problemas Técnicos

Durante o desenvolvimento, surgiram diversos desafios técnicos, os quais foram resolvidos com base em discussões colaborativas e testes contínuos. A seguir estão os principais problemas enfrentados e suas respectivas soluções:

- Problemas com *ScrollView*

Inicialmente, houve dificuldades em fazer com que o *ScrollView* funcionasse corretamente enquanto o *mouse* estivesse sobre os *prefabs* instanciados. O comportamento padrão da Unity não direcionava os eventos de rolagem corretamente quando os itens dinâmicos estavam ativos, o que causava um comportamento inesperado na rolagem da interface. A solução foi encontrada por meio da redirecionamento manual dos eventos de *scroll* para o componente *ScrollRect* do *ScrollView* pai, garantindo que os eventos fossem processados corretamente mesmo quando os itens dinâmicos estivessem em foco.

- Hover sobre Prefabs

Outro desafio foi a implementação do efeito de *hover* sobre os itens instanciados. Era necessário garantir que o sistema fosse capaz de detectar quando o *mouse* passava sobre os *prefabs* e, ao mesmo tempo, preservar o desempenho do sistema. A solução foi o uso de *EventTriggers*, permitindo uma resposta imediata e eficiente às interações do usuário. Esse recurso possibilitou a criação de uma experiência mais rica e fluida, sem sacrificar a performance.

### 3.8.5 Conclusão do Relatório de Desenvolvimento

A conclusão desta etapa do projeto resultou na entrega de um sistema robusto e dinâmico para a instanciação e gerenciamento de itens de catálogo. As funcionalidades desenvolvidas atendem plenamente às necessidades identificadas, proporcionando uma interface intuitiva e interativa. A combinação de uma estrutura de dados flexível, baseada em JSON, com a interface gráfica de *Unity* e a manipulação eficiente de eventos resultou em um sistema coeso e pronto para futuras expansões.

Os desafios enfrentados ao longo do desenvolvimento, como a manipulação de eventos no *ScrollView* e a implementação de interações dinâmicas com *hover* e clique, foram superados com sucesso. O desenvolvimento iterativo e o foco na solução de problemas técnicos garantiram que o sistema fosse entregue dentro do prazo e com alta qualidade.

### 3.8.6 Etapas futuras

Com base no sucesso desta etapa, o projeto apresenta diversas oportunidades para evolução. As futuras etapas de desenvolvimento podem incluir:

- Aprimoramento da Responsividade do Design

Com a crescente utilização de dispositivos móveis, um próximo passo natural seria o aprimoramento da interface gráfica para garantir que o sistema seja completamente responsivo. Isso envolverá a adaptação da interface a diferentes tamanhos de tela e resoluções, garantindo uma experiência de usuário consistente tanto em *desktops* quanto em dispositivos móveis.

### 3.8.7 Link Site

<https://calfjco.web.app/>

## 3.9 Git e Github

Conforme documentação oficial, o Git é um sistema de controle de versionamento de código distribuído, é gratuito e de código aberto. Foi desenvolvido em 2005 por Linus Torvalds, projetado para atender demandas desde projetos pequenos a muito grandes, sempre prezando por velocidade e eficiência. Possui vantagens como fácil aprendizagem do sistema, tem um desempenho rápido e ocupa pouco espaço. (GIT, 2024)

De acordo com Longen (2023), o GitHub é um serviço que hospeda o sistema de controle de versão Git, é baseado em nuvem e permite colaboração de desenvolvedores em projetos compartilhados e mantém os registros detalhados dessas mudanças. É uma ferramenta essencial para desenvolvedores, permitindo acompanhar as mudanças feitas no código base, sempre registrando quem executou a alteração, também permite restauração de código modificado ou removido.

Foi devido essas vantagens que escolhemos esse sistema e ferramenta para hospedar nossos repositórios do projeto, permitindo que todos os integrantes do grupo pudessem colaborar com o código-fonte.

### 3.9.1 Repositórios

[https://github.com/Caio-Saraiva/Project\\_TC](https://github.com/Caio-Saraiva/Project_TC)

### 3.10 Métodos

A CALFJ & Co é uma loja online especializada em varejo de roupas, que oferece uma experiência de compra totalmente digital. Como *e-commerce* (comércio eletrônico), a empresa tem como objetivo proporcionar comodidade e praticidade aos clientes, permitindo que eles comprem seus produtos sem a necessidade de se deslocarem fisicamente até uma loja. A CALFJ & Co busca inovação, garantindo que o processo de compra seja simples, rápido e seguro.

A empresa utiliza um conjunto de estratégias no intuito de melhorar a experiência de compra de clientes no ambiente online. Além disso, a loja oferece uma interface intuitiva e acessível para todos os públicos. O site foi projetado para ser funcional em diferentes dispositivos, como *smartphones*, *tablets* e *desktops*, garantindo que a navegação e a compra sejam fluidas e sem complicações. A empresa também foca na diversificação dos meios de pagamento, facilitando a jornada de compras.

A seguir, os principais métodos adotados pela loja para a eficiência e segurança do consumidor. Essas abordagens visam tornar o processo de compra mais acessível, além de soluções personalizadas que atendem diversos tipos de público.

Foi desenvolvido um sistema de previsão e validação que permite a tomada de decisão de forma estratégica, baseada em históricos. Esse sistema é apoiado por um dashboard interativo, que centraliza e visualiza métricas de desempenho e resultados preditivos de forma prática e acessível para times de negócio.

#### Predição:

Foi desenvolvido um algoritmo de predição a fim de analisar resultados futuros com base em dados passados.

#### Validação:

Assim como a predição, foi desenvolvido um algoritmo de validação para que seja realizada a verificação e validação das informações geradas.

#### Dashboard:

Também foi desenvolvido um *dashboard* interativo para facilitar a visualização dos resultados do modelo e dados da empresa, consumidos diretamente do banco de dados. Esse painel foi desenvolvido com o intuito de ser uma ferramenta prática e visual, permitindo o acompanhamento dos indicadores mais relevantes.



## 4 RESULTADOS

Os resultados apresentados nos tópicos seguintes, demonstram a aplicação de um modelo de regressão linear simples, que tem o intuito de prever as vendas futuras com base em dados históricos, de uma empresa de varejo de roupas. Além disso, um site foi desenvolvido como forma de simulação de um *e-commerce* e para facilitar a visualização dos dados, um *dashboard* no PowerBI foi criado.

### 4.1 Desempenho do Modelo de Regressão Linear Simples

O desempenho do modelo foi avaliado com o objetivo de prever as vendas de 2024, através dos dados históricos da empresa fictícia CALFJ & Co. Durante o processo de desenvolvimento foram utilizados códigos focados na predição e códigos para avaliar o desempenho, que são apresentados a seguir.

#### 4.1.1 Métricas de Avaliação

A fim de medir a qualidade do modelo, foram calculadas as seguintes métricas:

- Coeficiente de Determinação ( $R^2$ ): O valor de  $R^2$  foi de 0,975, indicando 97,5% de variação nas vendas. É um valor elevado que mostra que o modelo se ajusta muito bem aos dados.
- Erro Médio Absoluto (MAE): O MAE foi de 547.758,42, que representa o percentual de erro de 8,40% da média das vendas. Significando que o modelo errou por aproximadamente 547 mil reais, sendo considerado aceitável dentro da margem de erro estipulada.
- Erro Quadrático Médio (MSE): O MSE foi de 375.604.994.563,86, sendo um valor elevado devido sua característica de amplificar os erros elevando-os ao quadrado.
- Raiz do Erro Quadrático Médio (RMSE): O RMSE foi de 612.866,21, que representa o percentual de erro de 9,39% da média das vendas, sendo considerado aceitável dentro da margem de erro estipulada.

As análises demonstram que o modelo gera uma variação com boa precisão, uma vez que o MAE e RMSE se mantêm abaixo de 10%.

#### 4.1.2 Análise dos Resíduos

Gráfico de Resíduos vs Previsões: O gráfico gerado atende aos padrões conhecidos, onde os resíduos devem estar distribuídos de forma aleatória em torno da linha  $y = 0$  e com variância constante.

Normalidade dos Resíduos: Os resíduos se alinham sem grandes desvios com a reta diagonal, demonstrando que os erros seguem uma distribuição normal.

#### 4.1.3 Conclusão do Desempenho

Baseado nas análises das métricas e gráficos, é possível afirmar que o modelo se ajusta de boa maneira aos dados históricos e os erros observados se mantêm dentro do esperado (<10%), isso mostra que o modelo é confiável para realizar previsões e é uma boa ferramenta para auxiliar em tomada de decisões estratégicas.

#### 4.2 Simulação do *site*

O *site* desenvolvido utilizando *Unity*, simula um *e-commerce* onde os usuários podem visualizar os produtos disponíveis para compras, sem dificuldade, trafegar por todo site analisando preços e interagindo com a 'sacola' de vendas, até a finalização do pedido, que é armazenado diretamente no banco de dados da empresa.

#### 4.3 Visualização dos dados no PowerBI

O *dashboard* permite a visualização dos dados que estão presentes no banco de dados e projeções que foram geradas pelo modelo preditivo. Os gráficos são interativos e permitem a análise dos dados de maneira mais simples. Filtros foram implementados para permitir uma melhor visualização do usuário.

#### 4.4 Conclusão dos Resultados

Os resultados obtidos através do modelo desenvolvido, mostram um resultado eficaz da utilização de regressão linear simples para o cenário apresentado, validando de forma positiva seu uso. Além disso, a integração do modelo com os dados obtidos

através de interações entre os usuários do site e o *dashboard*, permitiram uma visualização clara dos dados.

## 5 CONCLUSÃO

O presente trabalho de conclusão de curso, intitulado "Simulação de *E-commerce* com Previsão de Vendas: Um Estudo de Regressão Linear e Visualização de Dados", buscou aprofundar a compreensão sobre modelos preditivos, com ênfase no mercado de varejo brasileiro e a adaptabilidade que o mercado sofreu durante a pandemia de COVID-19, que privou a sociedade de algumas atividades, como por exemplo a compra presencial. Durante esse período houve um grande aumento nas vendas através do comércio eletrônico, desde compra de alimentos à compra de itens permanentes. No decorrer do desenvolvimento desse estudo, ficou evidente que esse período pandêmico, impulsionou de uma forma significativa a modalidade de compras online, e através de novos estudos e tecnologias, as empresas se tornaram cada vez mais adeptas ao meio eletrônico.

A principal abordagem proposta a este trabalho, foi o estudo e análises de modelos preditivos, com a finalidade de aplicar a uma simulação de caso específica, enfatizando o uso de regressão linear simples, que melhor se adaptava ao caso que foi determinado, de avaliar a previsão de vendas de 2024 da empresa fictícia CALFJ & Co, baseada nos dados históricos disponíveis. Modelos preditivos apresentam uma abordagem muito robusta para diversos setores da sociedade, que podem ser desde previsões envolvendo dados de saúde, até análises climáticas e financeiras. As análises preditivas representam uma maneira de apoiar empresas na tomada de decisão estratégica, pois a partir de dados estruturados é possível realizar modelos de previsão que atendem diversas áreas internas, como pagamentos, *marketing*, logística etc.

Com a ideia de realizar uma estrutura muito próxima a realidade empresarial, foi desenvolvido um projeto completo, abrangendo *front-end*, *back-end*, análise e engenharia de dados, e métodos de visualização. Foi desempenhado um papel crucial durante a elaboração deste projeto, a separação da equipe para cada finalidade citada e a realização das atividades em *sprints* de duas semanas. A partir do início do desenvolvimento, surgiram desafios, oportunidades e dificuldades em relação a escopo e tecnologias utilizadas.

De maneira macro, para o desenvolvimento do *front-end* foram utilizadas a *Unity* e *Figma*, já na estruturação dos dados, análise, algoritmo de predição e validações, foram utilizados *softwares* como o *BR Modelo* e *Power Architect* e os serviços da *Amazon Web Services*, *Azure Data Studio*, *PostgreSQL* e *Python*, para o desenvolvimento final de visualização dos dados, foram utilizados *PowerBI*, *Power Query* e *DAX*. Todo o desenvolvimento culminou em um gigante projeto de aprendizagem de serviços, linguagens de programação, *softwares* e métricas.

Por fim, este projeto contribuiu de maneira significativa para o desenvolvimento do grupo, tivemos a oportunidade de aplicar todo nosso conhecimento e estudos obtidos no decorrer do curso e ampliou a visão de que a tecnologia utiliza de maneira correta, é uma grande aliada em diversas esferas da sociedade.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABComm - **Associação Brasileira de Comércio Eletrônico. Pesquisas.** Disponível em: <https://abcomm.org/pesquisas/>. Acesso em: 24 abr 2024.

ALURA. **Inteligência artificial: o que é e para que serve?** Disponível em: <<https://www.alura.com.br/artigos/inteligencia-artificial-ia>>. Acesso em: 28 abril 2024.

ARAÚJO, F. H. A.; FERNANDES, L. H. S. **Lighting the populational impact of COVID-19 vaccines in Brazil.** 2022.

ARAÚJO, M.; CALDWELL, L.; SANTOS, M.; SOUZA, I; ROSA, P.; SANTOS, A.; BATISTA, L. **Morbimortalidade pela COVID-19 segundo raça/cor/etnia: a experiência do Brasil e dos Estados Unidos.** *Saúde em Debate*, 2020. Disponível em: <https://doi.org/10.1590/0103-11042020E412>.

ASSUNÇÃO, F. **Estratégias para tratamento de variáveis com dados faltantes durante o desenvolvimento de modelos preditivos.** Dissertação (Mestrado em Ciências) - Instituto de Matemática e Estatística da Universidade de São Paulo, São Paulo, 2012.

Amazon Web Services – AWS 3. **Amazon RDS (Serviço de Banco de Dados Relacional).** Disponível em: <https://aws.amazon.com/pt/rds/> Acesso em: 28 fev 2024

Amazon Web Services – AWS 2. **The Difference Between Logical and Physical Data Model.** Disponível em: <https://aws.amazon.com/pt/compare/the-difference-between-logical-and-physical-data-model/> Acesso em: 22 fev 2024

Amazon Web Services – AWS 1. **O que é regressão linear?** Disponível em: <https://aws.amazon.com/pt/what-is/linear-regression/> Acesso em: 30 mar 2024

BAGATINI, F; LAIMER, C. (2019). **O Contexto do e-commerce no Brasil: Análise do Desempenho do Varejo Online no Período de 2003 a 2018.** In: **Congresso Latino-Americano de Varejo e Consumo: "Building Purpose Through Stakeholders in Retailing"**, São Paulo, 2019.

Best of BI. **SQL Power Architect - Data Modeling.** Disponível em: <https://bestofbi.com/products/sql-power-architect-data-modeling/> Acesso em: 22 mar 2024

Brasil. Ministério da Saúde. **Governo Federal investiu mais de R\$ 540 bilhões para o enfrentamento da pandemia no Brasil.** Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2022/dezembro/governo-federal-investiu-mais-de-r-540-bilhoes-para-o-enfrentamento-da-pandemia-no-brasil>. Acesso em: 20 mar 2024.

Brasil. **Ministério da Saúde. Governo federal investiu R\$ 38 bilhões em vacinas COVID-19 e criou logística inédita permitindo alta cobertura vacinal dos brasileiros.** Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2022/dezembro/governo-federal-investiu-r-38-bilhoes-em-vacinas-covid-19-e-criou-logistica-inedita-permitindo-alta-cobertura-vacinal-dos>



EBIT-NIELSEN (2020). **E-Commerce no Brasil cresce 47% no primeiro semestre, maior alta em 20 anos.** Disponível em: <https://static.poder360.com.br/2020/08/EBIT-ecommerce-Brasil-1semestre2020.pdf>

Estadão. **Nos bastidores da Shopee: saiba mais sobre a ascensão da plataforma no Brasil.** 04 mar 2024. Disponível em: <https://www.estadao.com.br/economia/nos-bastidores-da-shopee-saiba-mais-sobre-a-ascensao-da-plataforma-no-brasil/> Acesso em: 24 abr. 2024.

ESTEVIÃO, A. **COVID-19. Artigo de Opinião.** Acta Radiológica Portuguesa, volume 32, número 1, páginas 5-6, 2020.

FIGUEIRA, C. **Modelos de regressão logística.** 2006. Universidade Federal do Rio Grande do Sul. Instituto de Matemática. Programa de Pós-Graduação em Matemática. Disponível em: <https://lume.ufrgs.br/bitstream/handle/10183/8192/000569815.pdf?sequence=1&isAllowed=y>.

Figma. Disponível em: <https://www.figma.com/> Acesso em 01 abr 2024.

GALTON, F. (1886). **Regression Towards Mediocrity in Hereditary Stature.** Journal of the Anthropological Institute of Great Britain and Ireland, 15, 246-263. <https://www.jstor.org/stable/2841583>

GIT. **Documentação do Git.** Disponível em: <<https://git-scm.com/doc>>. Acesso em: 03 mai 2024.

GIULIANI, A. **Administração de varejo para pequenas e médias empresas** 1. ed. Jundiaí – SP. Paco Editorial, 2014.

Google Cloud. **PostgreSQL vs. SQL Server: comparando recursos e funcionalidades.** Disponível em: [https://cloud.google.com/learn/postgresql-vs-sql?hl=ptbr#:~:text=O%20PostgreSQL%20tamb%C3%A9m%20%C3%A9%20altamente,MVCC\)%20e%20espa%C3%A7os%20de%20tabela](https://cloud.google.com/learn/postgresql-vs-sql?hl=ptbr#:~:text=O%20PostgreSQL%20tamb%C3%A9m%20%C3%A9%20altamente,MVCC)%20e%20espa%C3%A7os%20de%20tabela). Acesso em 08 mar 2024.

GRECO, D. **Modelos preditivos: para que servem e quais são os diferentes tipos?** Minas Gerais, 10 mai 2017. Disponível em: <https://marketingpordados.com/analise-de-dados/modelos-preditivos-para-que-servem-e-quais-sao-os-diferentes-tipos/>. Acesso em: 04 set de 2024.

IBGE. **Em 2021, saldo de empresas que entraram e saíram do mercado cresceu pelo terceiro ano seguido.** 26 out. 2023. Disponível em <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/38171> Acesso em: 01 abr 2024.

IBGE. **Pesquisa Pulso Empresa: Impacto da COVID-19 nas Empresas.** 15 jun. 2020. Disponível em <https://covid19.ibge.gov.br/pulso-empresa/> Acesso em: 08 mar 2024.

IBM. **O que é ETL (extrair, transformar, carregar)?** Disponível em: <<https://www.ibm.com/br-pt/topics/etl>>. Acesso em: 09 mai 2024.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning with Applications in R** 1. ed. Califórnia. Editora Springer, 2013.

JGM Advogados. **Compra e venda na internet: quais os riscos e como se proteger?** Disponível em: <https://jgm.adv.br/publicacao/compra-e-venda-na-internet-quais-os-riscos-e-como-se-proteger/>. Acesso em: 09 mai 2024.

JHU. **Johns Hopkins University. Coronavirus Resource Center**. 2022. Disponível em: <https://coronavirus.jhu.edu/region/brazil>

KEWATE, N.; RAUT, A.; DUBEKAR, M.; RAUT, Y.; PATIL, A. (2022). **A Review on AWS - Cloud Computing Technology. International Journal for Research in Applied Science & Engineering Technology (IJRASET)**, Disponível em: <https://www.ijraset.com/best-journal/a-review-on-aws-cloud-computing-technology>

KOTLER, P.; ARMSTRONG, G. **Princípios de Marketing**. 15. ed., Pearson Prentice Hall, 2015.

LONGEN, A. **"O que é GitHub?"** 21 jul. 2023. Disponível em: <<https://www.hostinger.com.br/tutoriais/o-que-github>>. Acesso em: 03 mai 2024.

MACHADO, A. B.; RICHTER, M. F. **Sustainability in times of pandemic (Covid-19)**. RECIMA21, [S.l.], v. 1, n. 2, p. 270, 2020. Disponível em: <https://doi.org/10.47820/recima21.v1i2.25>.

MARTINS, M. (2018). **Coeficiente de determinação**, Rev. Ciência Elem., V6(01):024. Disponível em: [doi.org/10.24927/rce2018.024](https://doi.org/10.24927/rce2018.024)

MARTINS, M. (2019). **Regressão linear simples**. Revista Ciência Elementar, Disponível em: <http://doi.org/10.24927/rce2019.045>

MEDEIROS, E. **A luta dos profissionais de saúde no enfrentamento da COVID-19**. Acta Paul Enferm. 2020;33:e-EDT20200003

Mercado e Consumo. **Compras online crescem 26% em 1 ano no Brasil, mostra pesquisa**. Disponível em: <https://mercadoeconsumo.com.br/03/03/2023/ecommerce/compras-online-crescem-26-em-1-ano-no-brasil-mostra-pesquisa/>. Acesso em: 25 fev 2024.

MICROSOFT. **Aprenda noções básicas sobre o DAX no Power BI Desktop**. Disponível em: <<https://learn.microsoft.com/pt-br/power-bi/transform-model/desktop-quickstart-learn-dax-basics>>. Acesso em: 10 mai 2024.

MICROSOFT. **O que é o Azure Data Studio?** Disponível em: <https://learn.microsoft.com/pt-br/azure-data-studio/what-is-azure-data-studio> Acesso em: 19 fev 2024.



MICROSOFT. **Visão geral do Power BI.** Disponível em: <<https://learn.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>>. Acesso em: 28 abril 2024.

MONTEIRO, L. (2008). **Modelos preditivos de conforto térmico: quantificação de relações entre variáveis microclimáticas e de sensação térmica para avaliação e projeto de espaços abertos.** Tese de doutorado, Faculdade de Arquitetura e Urbanismo, Universidade de São Paulo. Disponível em: [https://www.teses.usp.br/teses/disponiveis/16/16132/tde-25032010-142206/publico/Monteiro\\_2008\\_Tese.pdf](https://www.teses.usp.br/teses/disponiveis/16/16132/tde-25032010-142206/publico/Monteiro_2008_Tese.pdf)

NOVAIS, M. (2007). **E-commerce: na Internet com sucesso.** São Paulo: Érica.

Organização Pan-Americana de Saúde. (2022). **Emergência de Saúde Pública de Importância Internacional.** Disponível em: <https://www.paho.org/pt/covid19/historico-da-pandemia-covid-19#:~:text=Em%2011%20de%20mar%C3%A7o%20de,pa%C3%ADses%20e%20regi%C3%B5es%20do%20mundo>. Acesso em: 27 mar 2024.

PALLARÉS, V.; GÓRRIZ-ZAMBRANO, C.; MORILLAS, C.; Llisterri, J.L.; & GORRIZ, J.L. **COVID-19 y enfermedad cardiovascular y renal: ¿Dónde estamos? ¿Hacia dónde vamos?** Acta Bioquímica Clínica Latinoamericana, volume ePii: S1138359320301441.

PAULA, J; CAMILO, L; SIQUEIRA, E. **A corrida pela vacinação contra a Covid-19 no Brasil: como as fake news e o descaso do governo podem influenciar na imunização?** <https://cjt.ufmg.br/2021/02/03/acorrida-pela-vacinacao-contr-a-covid-19-no-brasil-como-as-fake-news-e-o-descaso-dogoverno-podem-influenciar-na-imunizacao/> Acesso em: 17 mar 2024.

PEARSON, K. (1900). **On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.** Philosophical Magazine Series 5, 50(302), 157-175.

Psicometria Online. **O que devo saber sobre o poder preditivo da regressão linear?** Disponível em: <https://psicometriaonline.com.br/o-que-devo-saber-sobre-o-poder-preditivo-da-regressao-linear/>. Acesso em: 24 fev 2024.

QUEIROZ, M.; SOUSA, F.; PAULA, G.; **Educação e Pandemia: impactos na aprendizagem de alunos em alfabetização.** Disponível em: <https://revistas.uece.br/index.php/ensinoemperspectivas/article/view/6057/5383>

RODRIGUES, S. **Modelo de Regressão Linear e suas Aplicações.** Covilhã, 2012, Disponível em: <https://ubibliorum.ubi.pt/bitstream/10400.6/1869/1/Tese%20Sandra%20Rodrigues.pdf> f. Acesso em: 18 ago. 2024.

SANTOS, A; COSTA, C. **Características gerais do varejo no Brasil**. BNDES, biblioteca digital, 1997. Disponível em: [https://web.bndes.gov.br/bib/jspui/bitstream/1408/7125/2/BS%2005%20Caracteristicas%20gerais%20do%20varejo%20no%20Brasil\\_P.pdf](https://web.bndes.gov.br/bib/jspui/bitstream/1408/7125/2/BS%2005%20Caracteristicas%20gerais%20do%20varejo%20no%20Brasil_P.pdf). Acesso em: 07 mar 2024.

SANTOS, D; BASTOS, B; GABRIEL, J. (2018). **Vendas no varejo eletrônico (via internet) no Brasil antes e depois da popularização dos smartphones**. Disponível em: <https://ojs.brazilianjournals.com.br/ojs/index.php/BASR/article/view/543/467> Acesso em: 25 abr 2024

SANTOS, H. “**Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de São Paulo, Brasil**” São Paulo, 2019. Disponível em: <https://www.scielo.br/j/csp/a/jyhKL6G4dZhcbchMD6bcS8s/?format=pdf&lang=pt>. Acesso em: 10 set de 2024.

SANTOS, H. **Dinâmica Ressonante: Estratégia para Análise de Séries Temporais**. São Paulo: Universidade de São Paulo, 2018. Disponível em: <[https://www.teses.usp.br/teses/disponiveis/6/6141/tde-09102018-132826/publico/HellenGeremiasdosSantos\\_DR\\_ORIGINAL.pdf](https://www.teses.usp.br/teses/disponiveis/6/6141/tde-09102018-132826/publico/HellenGeremiasdosSantos_DR_ORIGINAL.pdf)>. Acesso em: 28 abril 2024.

SEBRAE. **Resumo executivo. Impactos da covid-19 nos pequenos negócios**. 16 jul. 2020. Disponível em: [https://bibliotecas.sebrae.com.br/chronus/ARQUIVOS\\_CHRONUS/bds/bds.nsf/502a5cbd31ce3ffb3d5472dd5751b84f/\\$File/19629.pdf](https://bibliotecas.sebrae.com.br/chronus/ARQUIVOS_CHRONUS/bds/bds.nsf/502a5cbd31ce3ffb3d5472dd5751b84f/$File/19629.pdf). Acesso em: 05 mar 2024.

SEMINGER, D; “**O que é Power Query?**”. Disponível em: <https://learn.microsoft.com/pt-br/power-query/power-query-what-is-power-query>. Acesso em: 09 mai 2024.

SIEGEL, E. (2013). **Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die**. 1. ed. Editora Wiley Publishing, 2013.

SORDI, L; MAGALHÃES, I.; CASSELHAS, D.; ANDRADE, M. (2020). **O Papel da Imunidade Inata na COVID-19**. REVISTA CIÊNCIAS EM SAÚDE. Disponível em: <https://doi.org/10.21876/rcshci.v10i3.997>. Acesso em: 16 mar 2024.

Shopee Brasil. **Sobre a Shopee**. Disponível em: <https://careers.shopee.com.br/about>. Acesso em: 25 abr 2024

TURING TALKS. **Turing Talks #10: Introdução à Predição**. Disponível em: <<https://medium.com/turing-talks/turing-talks-10-introdu%C3%A7%C3%A3o-%C3%A0-predi%C3%A7%C3%A3o-a75cd61c268d>>. Acesso em: 28 abril 2024.

Unity. Disponível em: <https://unity.com/pt/> Acesso em 03 abr 2024.

Visual Studio Code. Disponível em: <https://code.visualstudio.com/> Acesso em 28 mar 2024.

Webshoppers (2020). 42 ed. Edição Free. Ebit/Nielsen

YAMAUTI, M. (2013). **Regressão Linear Simples nos livros de Estatística para cursos de Administração: um estudo didático**. Pontifícia Universidade Católica de São Paulo. Disponível em: <https://repositorio.pucsp.br/bitstream/handle/10958/1/Marcelo%20Massahiti%20Yamauti.pdf>