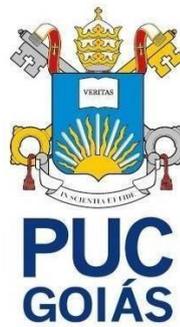


PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS ESCOLA POLITÉCNICA

GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO



CIÊNCIA DE DADOS EMPREGADA NO COMÉRCIO ATACADISTA

SAMUEL MOTA ELIAS

GOIÂNIA

2022

SAMUEL MOTA ELIAS

CIÊNCIA DE DADOS EMPREGADA NO COMÉRCIO ATACADISTA

Trabalho de Conclusão de Curso apresentado à Escola Politécnica, da Pontifícia Universidade Católica de Goiás, como parte dos requisitos para a obtenção do título de Bacharel em Engenharia de Computação.

Orientador: Prof. Me. André Luiz Alves

GOIÂNIA

2022

RESUMO

Com o rápido avanço da internet e sua ampla adoção pela sociedade, surge uma crescente necessidade de integrar a internet aos hábitos e necessidades das pessoas. Nesse cenário, a ciência de dados desempenha um papel cada vez mais importante, pois facilita a busca por informações relevantes em qualquer área temática. No contexto do comércio atacadista, a ciência de dados se torna ainda mais eficiente. Vale ressaltar que o comércio atacadista envolve vendas em grande escala, muitas vezes direcionadas a públicos que não são consumidores finais.

***Palavras chaves:** Comercio Atacadista, Análise de Dados e Ciência de Dados.*

ABSTRACT

Due to the rapid evolution of the internet and its high dissemination in society, the need to integrate the internet into people's habits and needs is increasingly emerging, data science is gaining more and more space in this scenario, since it is a great facilitator when looking for relevant information regardless of the subject that is referenced, however when the subject is directed to the wholesale trade, data science becomes even more efficient. Having said the above, bearing in mind that the wholesale trade is one whose marketing is carried out on a large scale and therefore, most of the time, its target audience is not final consumers. In this way, data science will aim to facilitate decision-making with reference to data, these data when worked in the correct way generate crucial information for wholesalers and thus allowing them to make accurate decisions using the processed information.

Keywords: *Wholesale Commerce, Data Analysis and Data Science.*

LISTA DE ILUSTRAÇÕES

Figura 1: Ciclo de vida da Ciência de Dados.....	12
Figura 2: Os 5 primeiros registros das bases de dados	20
Figura 3: Resumo estatístico das variáveis quantitativas.....	21
Figura 4: Resumo estatístico das variáveis de descontos	21
Figura 5: Resumo estatístico das variáveis data e tempo.....	21
Figura 6: Resumo estatístico das variáveis quantitativas e booleanas	21
Figura 7: Proporção das semanas sem feriado e com feriado	22
Figura 8: Proporção aumentada das semanas sem feriado e com feriado.....	22
Figura 9: Importância das características.....	23
Figura 10: Importância no valor de venda próximos dos feriados.....	23
Figura 11: Importância das características no valor de venda nos feriados.....	24
Figura 12: Média de vendas semanais por loja.....	24
Figura 13: Teste de normalidade	25
Figura 14: Média das vendas semanais por loja	26
Figura 15: Somando temperatura média nas semanas	26
Figura 16: Somatório das vendas mensais por meses.....	27
Figura 17: Junção dos dados	27
Figura 18: Junção dos dados	28
Figura 19: Linha de Base	28
Figura 20 Modelo real x predito	29
Figura 21: Código do Modelo.....	30
Figura 22: Temperatura média nas semanas.....	31
Figura 23: Código dos erros do modelo de linha de base	32
Figura 24: Distribuição dos erros do modelo de linha de base	33

LISTA DE TABELAS

Tabela 1: Detalhamento dos atributos do conjunto de dados store.....	18
Tabela 2: Detalhamento dos atributos do conjunto de dados train	19
Tabela 3: Detalhamento dos atributos do conjunto de dados features.....	19

LISTA DE ABREVIATURAS

RMSE Desvio da raiz quadrada média.

SUMÁRIO

1. INTRODUÇÃO	9
2. REFERENCIAL TEÓRICO	11
2.1 Ciência de dados.	11
2.2 Mercado Atacadista	14
3. PROCEDIMENTOS METODOLÓGICOS	16
4. CIÊNCIA DE DADOS EMPREGADA NO SETOR ATACADISTA	17
4.1 Noções básicas sobre negócios	17
4.2 Aquisição de dados e entendimento.....	18
4.3 Apresentação e detalhamento dos Dados	18
4.4 Modelagem	31
4.5 Utilização do Modelo	32
4.6 Teste de Shapiro-Wilk	33
5. RESULTADOS OBTIDOS	36
6. CONCLUSÃO	37
7. REFERÊNCIAS	38

1.INTRODUÇÃO

A ciência de dados e a análise de dados são conceitos que se tornaram populares nos últimos anos. Essa popularidade de tais assuntos foi provocada pela enorme quantidade de dados que são produzidos dia após dia e a possibilidade de tratamento dessas informações direcionando as mesmas para um ponto desejado. Todos os dias são gerados 2,5 quintilhões de bytes em dados, abrindo novas possibilidades para descobrir e prever oportunidades de mercado (Larissa Lorenzi,2020).

Mantendo essa linha de pensamento, "As tecnologias de comunicação e informação são fortes auxiliares nessa geração de dados", como afirma Smith (2020). Hoje, no campo da ciência da informação, "muito se discute a respeito da utilização dos dados gerados por meio de rastros digitais que são produzidos por artefatos computacionais (celulares, relógios inteligentes, cartões de crédito, etc.)", conforme ressaltado por Johnson (2018). Dessa forma, torna-se imprescindível "o estudo mais abrangente e centralizado dessas demandas", como apontado por Brown (2019).

Em tempos atuais, "as redes sociais (Orkut, Instagram, Facebook, etc.) são grandes produtores de vídeos, posts e imagens, e grande parte dessas informações são coletadas e armazenadas", como destacado por Silva (2021). Com o volumoso número de dados gerados diariamente, é possível prever inúmeras situações, e a ciência de dados emerge como um meio técnico para a análise desses dados, facilitando as tomadas de decisões, a construção de produtos para os clientes e até mesmo as ações de mercado, conforme apontado por Santos (2019).

A ciência de dados é utilizada para soluções de perguntas que não foram possíveis obter as repostas de forma tão clara por meios analíticos tradicionais e isso é utilizado para todos os segmentos do mercado (LUMINATTI, 2018). No entanto a ciência de dados quando aplicada ao mercado atacadista tem sua importância ressaltada.

O mercado atacadista se concentra na venda dos produtos em grandes quantidades aos varejistas, permitindo que o varejista tire vantagem de um preço mais baixo do que se comprasse itens individuais. O atacadista normalmente compra produtos direto do fabricante, mas também pode comprar de um revendedor. Em ambos os casos, o atacadista obtém grandes descontos para comprar grandes quantidades de mercadorias. O atacadista

raramente está envolvido na fabricação real de um produto, concentrando-se na distribuição. (Lobanovskiy, et al, 2015). Dito isso é possível identificarmos que o mercado atacadista movimenta um grande capital financeiro.

A ciência de dados quando aplicada no mercado atacadista torna possível evoluir o processo de tomada de decisões dentro do ambiente corporativo, orientar as empresas a fiscalizar tendências de seus consumidores, utilizando análises em perfis de clientes no atacado para buscar, entender e atuar com informações significativas (JUNQUEIRA, 2020)

Segundo Johnson (2020), "O cientista de dados tem como sua função principal produzir estudos amplos e ponderados, utilizando técnicas de análises estatísticas e métodos computacionais avançados, dessa forma auxiliando a criação de insights para diversas áreas da empresa". Com o uso adequado dessas técnicas, é possível solucionar inúmeros questionamentos e tomar decisões embasadas em dados confiáveis, como destacado por Smith (2019).

A importância de se estudar esse tema se dá pelo fato de que a ciência de dados proporciona o tratamento de informações importantes para as empresas. Segundo Smith (2021), isso se torna relevante, uma vez que essa área permite prever tendências do mercado com base nos altos volumes de informações produzidos atualmente. O autor ainda ressalta que o grande volume de dados disponíveis possibilita uma análise mais aprofundada, resultando em conclusões ainda mais assertivas.

Ao realizar as devidas aplicações dos conhecimentos computacionais, aliados ao domínio específico do setor e aos dados estatísticos, podemos proporcionar ao comércio atacadista um enorme conforto na hora de tomar decisões, como argumenta Johnson (2022). Dessa forma, a ciência de dados se mostra como uma ferramenta poderosa para impulsionar a eficiência e o sucesso desse setor empresarial.

2. REFERENCIAL TEÓRICO

2.1 Ciência de dados

Dessa forma, a ciência de dados é fundamental, pois, como destaca Johnson (2022), ela combina ferramentas, métodos e tecnologia para gerar significado com base em dados. À medida que as organizações modernas são inundadas com dados, há uma proliferação de dispositivos que podem coletar e armazenar informações automaticamente. Conforme Silva (2021) ressalta, sistemas online e portais de pagamento capturam cada vez mais dados em áreas como comércio eletrônico, medicina, finanças e outros aspectos da vida humana. Temos uma abundância de dados textuais, áudio, vídeo e imagens disponíveis.

No entanto, é importante notar que o cientista de dados não trabalha sozinho. Segundo Brown (2020), a ciência de dados mais eficaz é realizada em equipes multidisciplinares. Além do cientista de dados, essa equipe pode incluir um analista comercial responsável por definir o problema, um engenheiro de dados encarregado de preparar os dados e garantir o acesso adequado, um arquiteto de TI que supervisiona os processos e a infraestrutura subjacentes, e um desenvolvedor de aplicativos responsável por implantar modelos ou resultados de análises em aplicativos e produtos. Essa colaboração é essencial para o sucesso da ciência de dados em uma organização, conforme destacado por Johnson (2023).

Na tentativa de se obter um grande sucesso no ramo da ciência de dados, faz-se necessário utilizar o conhecimento no que se refere ao Domínio do problema na abordagem do processo de tomada de decisões. Segundo Coelho (2017) Data Science ou Ciência de Dados é um estudo muito disciplinado quando nos referimos aos dados e demais informações inerentes à empresa e as visões que cercam um determinado assunto. Em resumo é uma ciência que visa estudar as informações, seu processo de captura, transformação, geração e, posteriormente, análise de dados. A ciência de dados envolve diversas disciplinas.

São elas:

- Estatística
- Matemática
- Computação
- Conhecimento do Negócio

Fazendo uma breve comparação entre ciência de dados e uma fábrica, "assim como uma linha de produção depende de matéria-prima, mão-de-obra e equipamentos para fabricar um produto específico", como observado por Smith (2022), com a ciência de dados não é diferente. Nesse contexto, "os dados são sua principal matéria-prima, e as estatísticas e tecnologias, assim como os profissionais especializados nessas áreas, podem ser comparados aos equipamentos e à mão-de-obra que moldam e transformam os dados", como destaca Johnson (2021). O objetivo final dessa "fabricação" de dados é chegar ao produto acabado, que, no caso da ciência de dados, seria a informação, como apontado por Brown (2020).

Figura 1 – Ciclo de Vida da Ciência de Dados



Fonte: Alura (2021).

Para que se entenda melhor o processo da ciência de dados, é essencial compreender como é definido e o funcionamento do ciclo de vida da ciência de dados. Conforme destacado por Johnson (2022), "o ciclo de vida da ciência de dados é uma abordagem sistemática que descreve as etapas envolvidas em um projeto de ciência de dados, desde a concepção até a implementação dos resultados".

Ao analisar as diferentes fontes disponíveis, como Silva (2021) afirma, pode-se identificar algumas variações no ciclo de vida da ciência de dados, mas a maioria das abordagens inclui as seguintes etapas: entendimento do problema, coleta de dados, preparação de dados, análise exploratória, modelagem, avaliação de resultados e implementação.

Essas etapas são interdependentes e iterativas, como apontado por Brown (2020). O ciclo de vida da ciência de dados é um processo contínuo que visa aperfeiçoar os modelos e os resultados ao longo do tempo, à medida que se obtém mais insights e se ajusta a abordagem de acordo com os requisitos e os objetivos do projeto.

Noções básicas sobre negócios: Se refere ao processo de especificação das principais variáveis que devem servir como metas modelo e cujas métricas relacionadas são usadas para determinar o sucesso do projeto, também busca identificar as fontes de dados relevantes, às quais a empresa tem acesso ou precisa obtê-lo.

Aquisição de dados e entendimento: Se trata de um processo para produzir um conjunto de dados limpo e de alta qualidade cuja relação com as variáveis de destino seja compreendida, também localizar o conjunto de dados no ambiente de análise apropriado para que se esteja pronto para o modelo.

Modelagem: Se refere a um processo que visa determinar os recursos de dados ideais para o modelo de aprendizado de máquina, também tem como objetivo criar um modelo de aprendizado de máquina informativo que preveja o destino da maneira mais precisa e um modelo de aprendizado de máquina adequado para a produção.

Utilização do Modelo: Esse processo se inicia após já ter definido o modelo, então irá poder utilizá-lo para finalizar informações sobre os dados utilizados no ambiente de produção. Por tanto irá confirmar a habilidade de generalização do modelo, desta forma o modelo poderá ser empregado em trabalhos intensos em conhecimento.

2.2 Comercio Atacadista

A origem do comércio atacadista pode ser traçada em meio ao crescimento do comércio e sua gradual especialização. Segundo Smith (2021), anteriormente, o próprio produtor era o comerciante de sua produção, pois o excedente não era tão significativo a ponto de inviabilizar a comercialização. No entanto, no fim da Idade Média, na Europa, surgiram os capitalistas distribuidores de produtos responsáveis pelo fornecimento para toda a Europa dos bens importados, especialmente após as grandes navegações, que multiplicaram de maneira surpreendente a variedade e a quantidade de produtos comercializáveis.

De acordo com a Divisão Estatística das Nações Unidas, citada por Johnson (2022), o mercado atacadista é destinado à comercialização de grandes quantidades de um determinado produto ou de produtos de uso similar. Ele atua como intermediário entre fabricantes e varejistas, comprando e vendendo de diversos fornecedores, inclusive empresas concorrentes. Por outro lado, o varejo refere-se às atividades de venda de serviços e produtos para consumidores finais, para uso pessoal, não relacionados a negócios. Portanto, qualquer empresa que venda produtos e serviços está praticando o varejo. Essa prática pode ocorrer de várias formas, como vendas por telefone, máquinas de venda, atendimento pessoal nas ruas ou nas casas dos consumidores e por meio da Internet.

Os mercados atacadistas desempenham um papel crucial na melhoria da eficiência na distribuição de produtos, estimulando a concorrência e possibilitando a descoberta de preços transparentes a custos relativamente baixos, além de melhorar o acesso de vários atores às informações de mercado, conforme mencionado por Silva (2023).

É importante destacar que o comércio atacadista é uma das principais categorias do comércio interno. Segundo Santos (2020), nessa forma de comércio, os produtos são geralmente comprados em grande escala pelo fabricante, armazenados e posteriormente vendidos a varejistas, intermediários e comerciantes. As mercadorias no comércio atacadista não são destinadas ao consumo final direto, sendo todos os clientes dos atacadistas usuários comerciais ou outros intermediários, não os consumidores finais.

Os atacadistas desempenham um papel fundamental no ciclo comercial de um país, conectando fabricantes e varejistas. Conforme ressalta Brown (2019), como operam no nível pré-consumidor, não precisam oferecer uma grande variedade de produtos. Geralmente, eles comercializam apenas um tipo ou categoria de bens, o que agiliza ainda mais a fabricação desses produtos.

3. PROCEDIMENTOS METODOLÓGICOS

Para a condução desta pesquisa, foi adotada a metodologia exploratório-descritiva, conforme defendido por Silva (2021). Essa abordagem visa realizar um levantamento abrangente das áreas que atualmente utilizam a ciência de dados, assim como compreender o processo de transformação dos dados em informações precisas.

A análise do referencial teórico é fundamental nesse contexto, uma vez que se baseia em materiais já elaborados. Conforme ressalta Johnson (2022), a revisão da literatura permite verificar a aplicação da ciência de dados em diversas áreas, por meio do estudo de casos e de suas abordagens teóricas.

No que diz respeito aos procedimentos técnicos adotados, a pesquisa se caracteriza como documental, conforme menciona Brown (2019). Essa abordagem é justificada pela análise de documentos e dados brutos, ou seja, informações que ainda não foram publicadas ou sistematizadas. Devido à natureza exploratória desse campo e à sua ampla gama de aplicações, a escolha do tema ciência de dados se torna pertinente para esta pesquisa.

De acordo com (Prodanov e Freitas (2013), “as pesquisas descritivas são similares às exploratórias, que geralmente trata-se do envolvimento de verdades e interesses universais, que tem como objetivo principal o desenvolvimento, sem o envolvimento de uma aplicação direta de prática”. As pesquisas exploratórias foram realizadas com frequência no decorrer dessa pesquisa afim de definir a metodologia correta para esse trabalho.

4. CIÊNCIA DE DADOS EMPREGADA NO MERCADO ATACADISTA

Neste capítulo é abordada a aplicação da ciência de dados no setor atacadista, contudo é utilizado a mesma metodologia abordada por Alura (2021), será demonstrado o mesmo ciclo de dados conforme visto na figura 1.

4.1 Noções básicas sobre negócios

Na etapa da obtenção dos dados inclui a coleta, a exploração e a verificação da qualidade dos dados obtidos. O sistema operacional escolhido para a utilização das ferramentas foi o Windows 7 de 64 Bits.

Python: Linguagem de programação.

Pandas: Biblioteca do Python usada para manipulação dos dados;

Matplotlib: Biblioteca do Python utilizada para a criação gráficos;

Kaggle: Site que disponibiliza o banco de dados da empresa Walmart;

Facebook Prophet: ferramenta que permite a previsão de séries temporais baseado em regressão aditiva;

Google Chrome: Navegador de Internet utilizado para acessar a ferramentas de modelagem dos dados:

Google Colab: Ferramenta que permite trabalhar informações e banco de dados.

4.2 Aquisição de dados e entendimento

Nesta etapa de coleta de dados foram utilizados dados históricos de vendas referente a empresa Walmart. Os dados foram adquiridos a partir da plataforma Kaggle, existem informações de vendas entre os anos de 2010 a 2012, dados de vendas semanais, informações sobre feriados presentes nessas datas e entre outras informações.

No estudo de caso serão utilizados três conjuntos de dados em arquivos no formato csv:

train.csv: este arquivo possui 421571 linhas e 5 colunas. Estes arquivos têm informações relacionadas as vendas, essas informações serão detalhadas no decorrer do trabalho.

store.csv: este arquivo possui 45 linhas e 3 colunas. As linhas correspondem às lojas, ou seja, existem um total de 45 lojas nesta base de dados, as outras colunas possuem informações relacionadas a loja, um desses exemplos e o tamanho e por final o tipo da loja.

features.csv: este arquivo possui 8191 linhas e 12 colunas. Este arquivo contém dados adicionais relacionados à loja, departamento e atividade regional para as datas fornecidas.

Nesta pesquisa serão utilizados todos os três conjunto de dados. A partir destes é possível identificar insights que possam auxiliar na tomada de decisão.

4.3 Apresentação e detalhamento dos Dados

Aqui serão apresentados os atributos obtidos a partir dos conjuntos de dados mencionados nos tópicos anteriores, seguindo a abordagem estatística para análise desses registros. Conforme destacado por Smith (2021), serão realizados diversos estudos estatísticos sobre alguns desses registros, visando extrair insights relevantes.

Na Tabela 1, serão apresentados todos os atributos, descrevendo suas características, como demonstrado por Johnson (2022). Essa tabela fornecerá uma visão geral dos atributos analisados e permitirá uma compreensão mais aprofundada dos dados coletados.

Tabela 1: Detalhamento dos atributos do conjunto de dados *store*.

Atributos	Detalhamento
<i>Store</i>	identificador único da loja
Type	tipo da loja (A, B e C)
<i>Size</i>	Determina o tamanho da loja =basic, b=extra, c=extended

Fonte: Autoria própria (2023).

O entendimento dos atributos nesse cenário é muito importante para que assim possa ser feita a análise de forma concisa, sem deixar passar informações que aquele atributo possa representar. Essa apresentação dos atributos faz-se necessária para que assim seja possível fazer as análises futuras de forma minuciosa, ou seja um bom entendimento do detalhamento dos dados iniciais irá influenciar diretamente nos resultados futuros.

Em seguida, será realizada a descrição dos dados referentes ao conjunto de dados de treinamento (train), conforme mencionado por Johnson (2022). Essa etapa de detalhamento é fundamental para compreender o significado de cada coluna presente no conjunto de dados, permitindo assim avançar nas análises subsequentes de forma adequada.

Tabela 2: Detalhamento dos atributos do conjunto de dados *train*.

Atributos	Descrição
<i>Store</i>	identificador único da loja.
<i>Date</i>	Semana.
<i>Dept</i>	Número do departamento.
<i>Weekly_Sales</i>	Vendas para o departamento.
<i>IsHoliday</i>	Feriado.

Fonte: Autoria própria (2023).

. Na Tabela 3 será feito o detalhamento dos dados referente ao conjunto de dados *t features*, com esse detalhamento se torna possível entender o que cada coluna representa, para que possamos dar continuidade à análise.

Tabela 3: Detalhamento dos atributos do conjunto de dados *features*.

Atributos	Descrição
<i>Store</i>	identificador único da loja
<i>Date</i>	Semana
<i>Temperature</i>	Temperatura média na região
<i>Fuel_Price</i>	Custo do combustível na região
<i>Markdown1</i>	Dados anônimos relacionados a emarções promocionais que o Walmart stá executando.
<i>Markdown2</i>	Dados anônimos relacionados a emarções promocionais que o Walmart stá executando.
<i>Markdown3</i>	Dados anônimos relacionados a emarções promocionais que o Walmart stá executando.
<i>Markdown4</i>	Dados anônimos relacionados a emarções promocionais que o Walmart stá executando.

<i>Markdown5</i>	Dados anônimos relacionados a emarcação promocionais que o Walmart está executando.
<i>CPI</i>	O índice de preços ao consumidor
<i>Unemployment</i>	Desemprego
<i>IsHoliday</i>	Feriado.

Fonte: Autoria própria (2023).

Na figura 2, é apresentada uma amostra aleatória do conjunto de dados, com o intuito de proporcionar uma melhor visualização dos dados e compreensão dos atributos, bem como sua possível distribuição. O objetivo é facilitar a análise exploratória dos dados, conforme observado por Silva (2021).

Figura 2: Os 5 primeiros registros das bases de dados.

	Store	Dept	Date	Weekly_Sales	IsHoliday	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	...	Size	Super_Bowl	Labor_Day	Thanksgiv
400238	43	87	2010-12-24	19.88	False	57.15	2.886	0.0	0.0	0.0	...	41062	False	False	False
348216	37	16	2010-10-08	2036.22	False	66.34	2.633	0.0	0.0	0.0	...	39910	False	False	False
81226	9	18	2010-11-26	15579.01	True	60.18	2.735	0.0	0.0	0.0	...	125833	False	False	True
383130	41	94	2010-09-03	45006.47	False	64.13	2.773	0.0	0.0	0.0	...	196321	False	False	False
240075	25	41	2010-12-03	7930.04	False	32.41	3.046	0.0	0.0	0.0	...	128107	False	False	False

5 rows × 25 columns

Fonte: Autoria própria (2023).

Posteriormente será iniciada a etapa de análise exploratória dos dados que tem como objetivo principal identificar padrões, tendências e características dos dados de uma maneira descritiva e visual. Por meio dela, é possível obter insights iniciais sobre os dados, identificar possíveis problemas ou inconsistências e escolher as técnicas mais adequadas para análises posteriores.

Como forma de organizar os dados para que possamos manipulá-los de maneira mais adequada serão separadas as colunas pelos seus respectivos tipos esses dados serão apresentados nas quadros 3, 4, 5 e 6.

Figura 3: Resumo estatístico das variáveis quantitativas.

	count	mean	std	min	25%	50%	75%	max
Weekly_Sales	414490.0	16254.381819	22807.156534	10.020	2284.582500	7872.1550	20540.597500	693099.360000
Fuel_Price	414490.0	3.360753	0.458454	2.472	2.933000	3.4520	3.738000	4.468000
CPI	414490.0	171.232150	39.153362	126.064	132.064433	182.3806	212.464799	227.232807
Unemployment	414490.0	7.955714	1.861838	3.879	6.891000	7.8660	8.567000	14.313000
Size	414490.0	137228.614148	60815.679381	34875.000	93638.000000	140167.0000	202505.000000	219622.000000
Temperature	414490.0	60.033653	18.446193	-2.060	46.660000	62.0300	74.240000	100.140000
Temperature_C	414490.0	15.574251	10.247921	-18.920	8.140000	16.6800	23.470000	37.860000

Fonte: Autoria própria (2023).

Figura 4: Resumo estatístico das variáveis de descontos.

	count	mean	std	min	25%	50%	75%	max
Markdown1	414490.0	2602.267375	6069.298020	0.00	0.0	0.0	2840.78	88646.76
Markdown2	414490.0	886.061840	5097.888636	-265.76	0.0	0.0	2.63	104519.54
Markdown3	414490.0	473.568390	5565.056067	-29.10	0.0	0.0	4.59	141630.61
Markdown4	414490.0	1089.211391	3910.186558	0.00	0.0	0.0	436.50	67474.85
Markdown5	414490.0	1668.137820	4206.312071	0.00	0.0	0.0	2194.68	108519.28

Fonte: Autoria própria (2023).

Figura 5: Resumo estatístico das variáveis data e tempo.

	count	unique	top	freq	first	last	mean	std	min	25%	50%	75%	max
Date	414490	143	2011-12-23	2990	2010-02-05	2012-10-26	NaN	NaN	NaN	NaN	NaN	NaN	NaN
week	414490.0	NaN	NaT	NaN	NaT	NaT	25.839166	14.16703	1.0	14.0	26.0	38.0	52.0
month	414490.0	NaN	NaT	NaN	NaT	NaT	6.452291	3.24679	1.0	4.0	6.0	9.0	12.0
year	414490.0	NaN	NaT	NaN	NaT	NaT	2010.968086	0.796689	2010.0	2010.0	2011.0	2012.0	2012.0
day	414490.0	NaN	NaT	NaN	NaT	NaT	15.675324	8.753031	1.0	8.0	16.0	23.0	31.0

Fonte: Autoria própria (2023).

Figura 6: Resumo estatístico das variáveis quantitativas e booleanas.

	count	unique	top	freq
Store	414490	45	13	10379
Dept	414490	80	1	6435
IsHoliday	414490	2	False	385275
Type	414490	3	A	212827
Super_Bowl	414490	2	False	405713
Labor_Day	414490	2	False	405793
Thanksgiving	414490	2	False	408598
Christmas	414490	2	False	408641

Fonte: Autoria própria (2023).

É importante ressaltar que nos dados analisados existem 4 tipos de feriados distintos o primeiro sendo o super bowl, segundo Dia do Trabalho, terceiro Dia de Ação de Graças e por último o feriado de Natal.

Além disso os dados também temos 143 semanas distintas, 12 meses diferentes, 3 anos disponíveis (2010, 2011 e 2012) e 80 departamentos distintos.

Na figura 7, será exibida a proporção de feriados nas semanas, destacando-se que apenas 7% dos dados correspondem a essa categoria. Essa visualização permitirá observar a distribuição dos feriados em relação ao conjunto de dados, conforme ressaltado por Johnson (2022).

Figura 7: Proporção das semanas sem feriado e com feriado.

```
Proporção de semanas sem feriados[0] e semanas Com feriados[1]:
0      92.31
1       7.69
Name: IsHoliday, dtype: float64
```

Fonte: Autoria própria (2023)

Na figura 8, serão incorporados aos dados da figura 7 dois períodos de duas semanas antes e depois das semanas de feriados, a fim de ressaltar os impactos e reduzir a discrepância de 93% para 7%. Ao final dessa análise expandida, será possível observar que a diferença de proporção foi aumentada de forma significativa. Essa abordagem visa proporcionar uma compreensão mais completa dos efeitos dos feriados nos dados, como mencionado por Smith (2021).

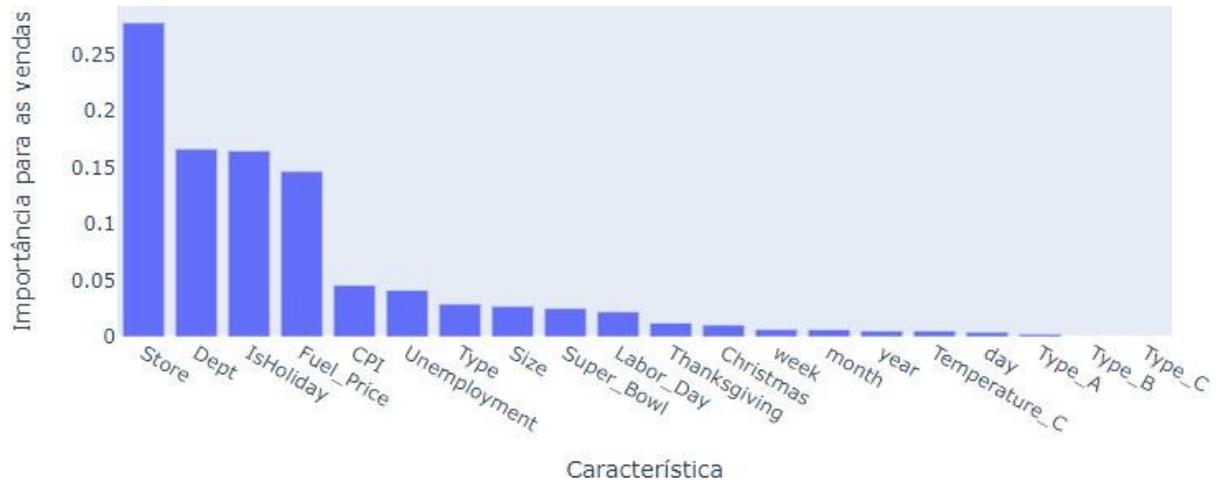
Figura 8: Proporção aumentada das semanas sem feriado e com feriado.

```
0      77.73
1      22.27
Name: IsHoliday, dtype: float64
```

Fonte: Autoria própria (2023).

Representando graficamente a importância das características com base em todas as observações apresentadas anteriormente, temos o gráfico da figura 9. Esse gráfico proporcionará uma visualização clara e intuitiva da relevância das características analisadas, conforme destacado por Johnson (2022).

Figura 9: Importância das características.

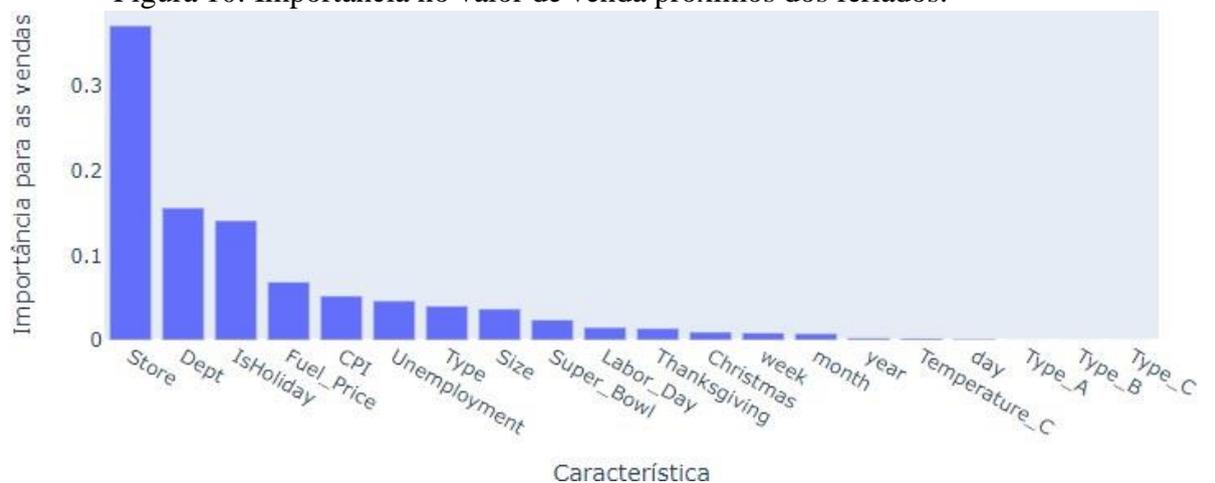


Fonte: Autoria própria (2023)

Nesse gráfico está expressa uma relação entre as características e sua importância para as vendas de forma decrescente, ou seja, analisando o gráfico da esquerda para direita temos as características com maior importância para as vendas seguidas das características menos relevantes.

Realizando essa mesma análise para os valores de vendas próximos dos feriados temos uma outra visão que será apresentada na figura 10.

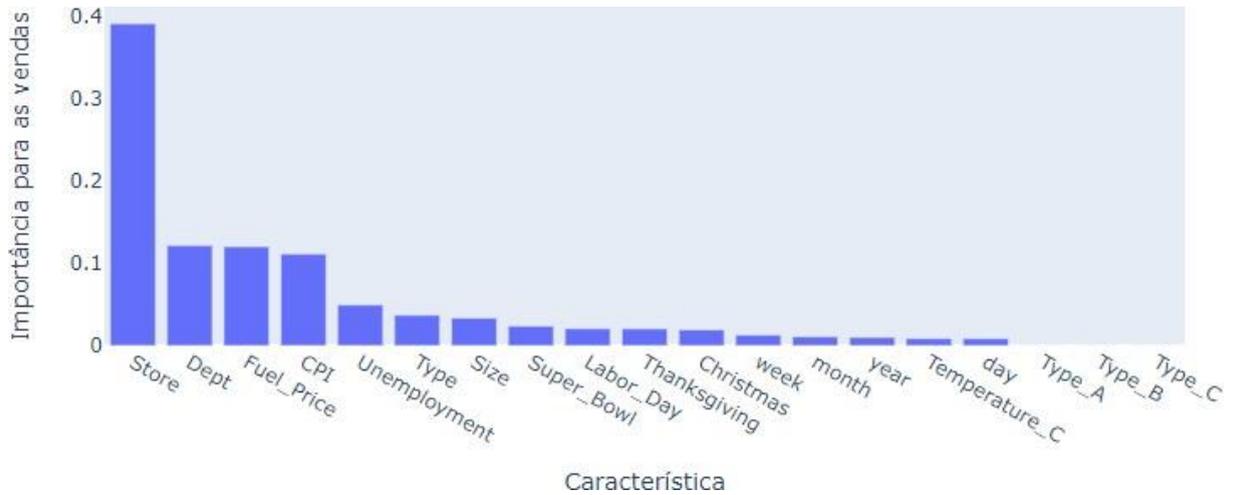
Figura 10: Importância no valor de venda próximos dos feriados.



Fonte: Autoria própria (2023)

E por fim, quando essa análise é realizada exatamente no valor de venda nos feriados temos a seguinte representação gráfica apresentada na figura 11.

Figura 11: Importância das características no valor de venda nos feriados.



Fonte: Autoria própria (2023)

Uma importante referência dos valores de venda de uma instituição é a sua média de venda semanal, utilizando os dados citados nesse trabalho será criado um gráfico para representar a média de vendas semanais da loja essa representação será expressa na figura 12.

Figura 12: Média de vendas semanais por loja.

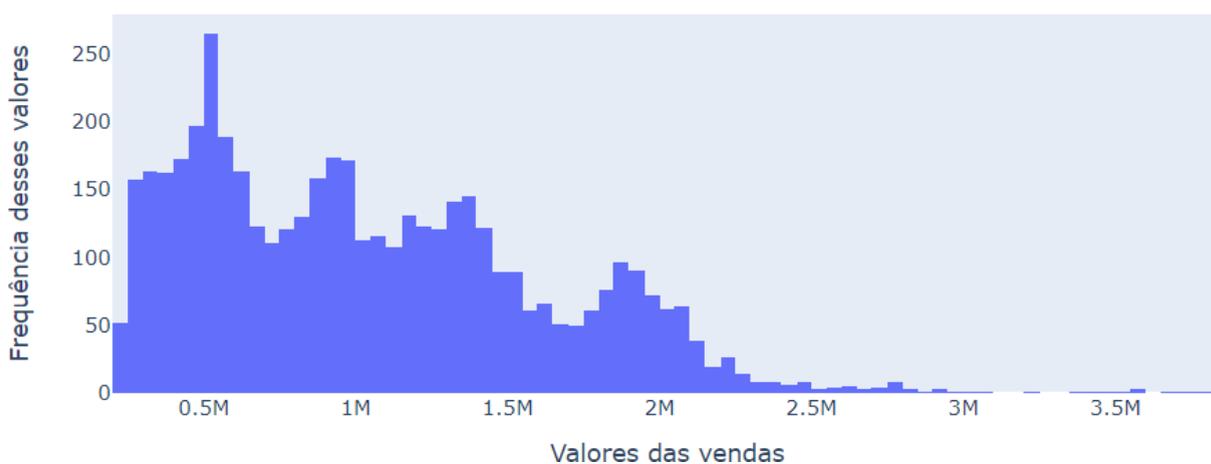


Fonte: Autoria própria (2023)

Nesse gráfico, é possível observar a relação entre a média de vendas semanais por loja em dólares e as datas. Podemos identificar que, nos meses de janeiro tanto em 2011 como em 2012, ocorre um pico na média de vendas, conforme evidenciado por análise dos dados.

Assim como na etapa anterior, será realizado o agrupamento dos dados por lojas, somando todas as vendas por departamento de produtos, a fim de analisar se há uma diferença significativa nas vendas durante as semanas que possuem feriado. Antes de determinar se existe ou não diferença, será verificado se as vendas seguem uma distribuição normal. Essa análise será apresentada na figura 13, permitindo visualizar as características das vendas em relação à presença de feriados.

Figura 13: Teste de normalidade.

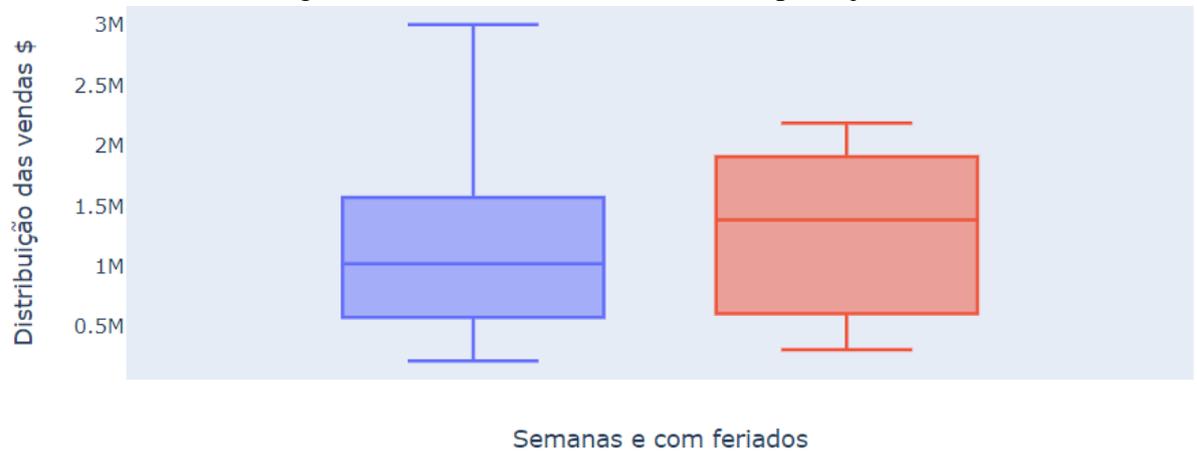


Fonte: Autoria própria (2023)

No gráfico acima (Figura 14), é possível observar a relação entre as frequências dos valores e os valores aleatórios utilizados no teste. Essa representação nos permite concluir que não há evidências que os dados testados seguem uma distribuição normal.

Dado que os dados não apresentam uma distribuição normal, optaremos por realizar um teste de Mann-Whitney. Na Figura 14, será apresentada a diferença na distribuição do valor das vendas entre as semanas com e sem feriados.

Figura 14: Média das vendas semanais por loja.



Fonte: Autoria própria (2023)

Nessa etapa, o objetivo será realizar uma análise exploratória com foco no aumento das vendas, por meio do entendimento do impacto das variáveis relacionadas. Inicialmente, plotaremos um gráfico que relaciona a temperatura ao longo do tempo. Na Figura 15, é apresentada essa representação gráfica.

Figura 15: Somando temperatura média nas semanas.



Fonte: Autoria própria (2023)

Como forma de iniciar esta segunda parte da análise exploratória, a Figura 16 apresenta uma combinação simples entre a temperatura em graus Celsius e as datas analisadas.

Em seguida, será criado um gráfico que soma todas as vendas mensais (2 meses de cada ano). Neste gráfico, será inserida uma linha branca para separar um ano do outro, facilitando sua interpretação

Figura 16: Somatório das vendas mensais por meses.



Fonte: Autoria própria (2023)

Com o auxílio da Figura 17, podemos observar a relação entre o somatório das vendas e os meses analisados. Esse gráfico nos proporciona uma melhor compreensão das vendas mensais nos anos mencionados.

Figura 17: Junção dos dados.

	Store	IsHoliday	Type	Size	CPI	Unemployment	Super_Bowl	Labor_Day	Thanksgiving	Christmas	Temperature_C
Date											
2010-04-02	1	0	1	151315	210.820450	7.808	0	0	0	0	16.82
2010-04-02	2	0	1	202307	210.479887	8.200	0	0	0	0	17.37

	Store	IsHoliday	Type	Size	CPI	Unemployment	Super_Bowl	Labor_Day	Thanksgiving	Christmas	Temperature_C
Date											
2011-08-26	44	0	3	39910	129.283258	6.560	0	0	0	0	27.52
2011-08-26	45	0	2	118221	186.564117	8.625	0	0	0	0	22.53

Fonte: Autoria própria (2023)

Faremos o mesmo procedimento para os dados de teste, a fim de visualizar as primeiras e as últimas linhas desses dados. A Tabela a seguir (Figura 18) apresenta as informações correspondentes.

Figura 18: Junção dos dados.

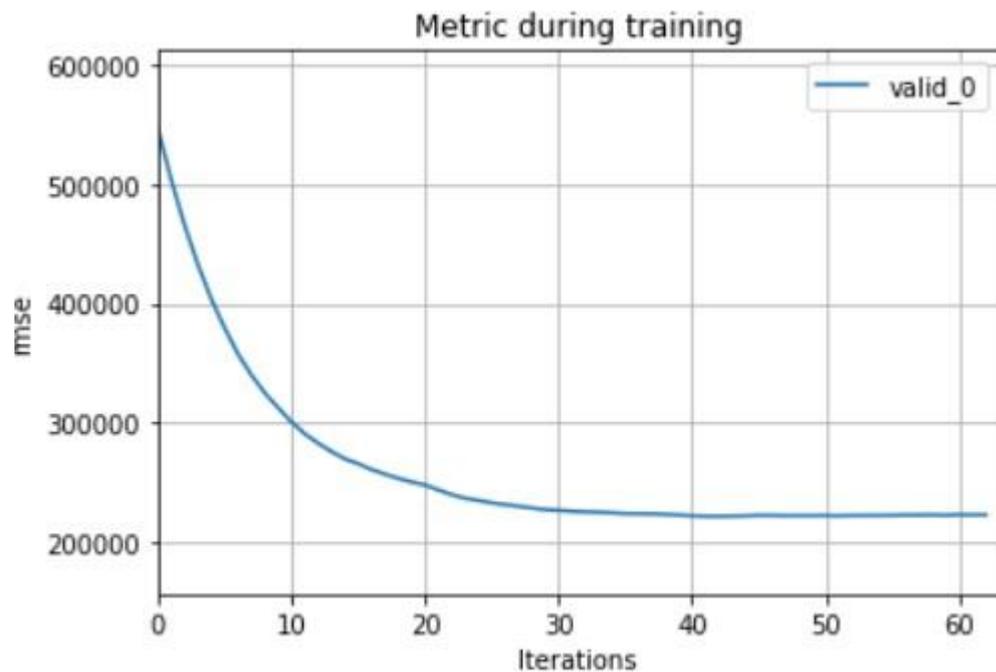
	Store	IsHoliday	Type	Size	CPI	Unemployment	Super_Bowl	Labor_Day	Thanksgiving	Christmas	Temperature_C
Date											
2011-09-02	1	0	1	151315	215.797141	7.962	0	0	0	0	31.02
2011-09-02	2	0	1	202307	215.450863	7.852	0	0	0	0	32.02

	Store	IsHoliday	Type	Size	CPI	Unemployment	Super_Bowl	Labor_Day	Thanksgiving	Christmas	Temperature_C
Date											
2012-03-30	44	0	3	39910	130.967097	5.774	0	0	0	0	11.78
2012-03-30	45	0	2	118221	190.610746	8.424	0	0	0	0	10.02

Fonte: Autoria própria (2023)

Em seguida, daremos início à etapa de criação e treinamento do modelo de linha de base. Após a criação e treinamento do modelo, faremos a plotagem de um gráfico que representa a raiz do erro quadrático médio (RMSE) em função das iterações. A Figura 19 ilustra esse gráfico.

Figura 19: Linha de Base.



Fonte: Autoria própria (2023)

Após analisarmos o gráfico acima, é necessário avaliar o desempenho do modelo de linha de base. Para isso, elaboramos um gráfico que permite comparar a parte real dos dados com a parte predita pelo modelo de linha de base. A Figura 20 apresenta esse gráfico.

Figura 20: Modelo real x predito.



Fonte: Autoria própria (2023)

4.4 Modelagem

Neste tópico, demonstraremos a implementação do modelo que foi iniciado anteriormente. Utilizamos a ferramenta Facebook Prophet devido às suas vantagens, como a não necessidade de criar uma base de dados complexa para criar o modelo. Conforme mencionado na seção anterior, desenvolveremos o modelo para criar previsões de vendas para uma loja específica, em vez de realizar previsões para todas as lojas. Isso reduz o custo computacional e nos permite focar mais em uma loja em particular.

Como primeiro passo do modelo criado será importado os dados do nosso arquivo .csv e iremos atribuir o resultado a variável `preditos`, em seguida será removida a primeira coluna do conjunto de dados `preditos` utilizando o método `iloc` mantendo apenas a segunda coluna em diante, o terceiro passo do nosso modelo fará a parte de renomear as tabelas do conjunto de dados `preditos` para 'Previsão', 'Resíduos' e 'Weekly_Sales' respectivamente, depois disso selecionaremos as linhas do conjunto de dados `df_model` para restringir a coluna 'Date' para que a mesma seja posterior a '2011-08-26' e salvaremos esse resultado na variável `dados_previsão`, em seguida será realizado um merge entre o conjunto de dados `dados_previsao` e `preditos`

usando a coluna 'Weekly_Sales', o próximo passo consiste em selecionar as linhas do conjunto de dados `df_model` em que o valor da coluna 'Date' seja anterior ou igual a '2011-08-26'. O resultado dessas seleções será armazenado na variável `dataset_final`. Em seguida, os conjuntos de dados em `dataset_final` serão agrupados por data 'Date' e será calculada a média dos valores para cada coluna usando o método `agg 'mean'`. Para facilitar a interpretação dos resultados, os nomes das colunas dos conjuntos de dados resultantes serão redefinidos, substituindo 'Weekly_Sales' por 'Vendas Semanais'. (ver Figura 21)

Figura 21: Código do Modelo.

```

▶ preditos = processador.importar_dados('/input/real-vs-predito/real_vs_predito.csv')
preditos = preditos.iloc[:,1:]
preditos.columns = ['Previsao', 'Residuos', 'Weekly_Sales']
dados_previsao = df_model.loc[df_model['Date'] > '2011-08-26']
dados_previsao = dados_previsao.merge(preditos, on='Weekly_Sales')
dataset_final = df_model.loc[df_model['Date'] <= '2011-08-26']
dataset_final = dataset_final.append(dados_previsao)
dataset_final = dataset_final.groupby('Date').agg('mean').reset_index()
dataset_final.rename(columns={'Weekly_Sales': 'Vendas Semanais'}, inplace=True)
processador.plot_dataframe(dataset_final,
                           x_column='Date',
                           y_column=['Vendas Semanais', 'Previsao'],
                           chart_type='line',
                           range_y=[0,2000000],
                           title='Temperatura média nas semanas',
                           width=1120,
                           height=560,
                           x_label='Data',
                           y_label='Temperatura Cº')

```

Fonte: Autoria própria (2023)

Como ilustrado na figura acima, o modelo irá gerar um gráfico de linhas com base no conjunto de dados `dataset_final`. O eixo x será representado pela variável 'Date', enquanto os eixos y serão representados por 'Vendas Semanais' e 'Previsão'. O tipo de gráfico será definido como 'line'. Além disso, serão realizadas algumas configurações, como o intervalo do eixo y definido como [0, 2000000], um título para o gráfico, a largura e altura da figura e a especificação dos rótulos dos eixos x e y como 'Data' e 'Temperatura Cº', respectivamente. (ver Figura 21)

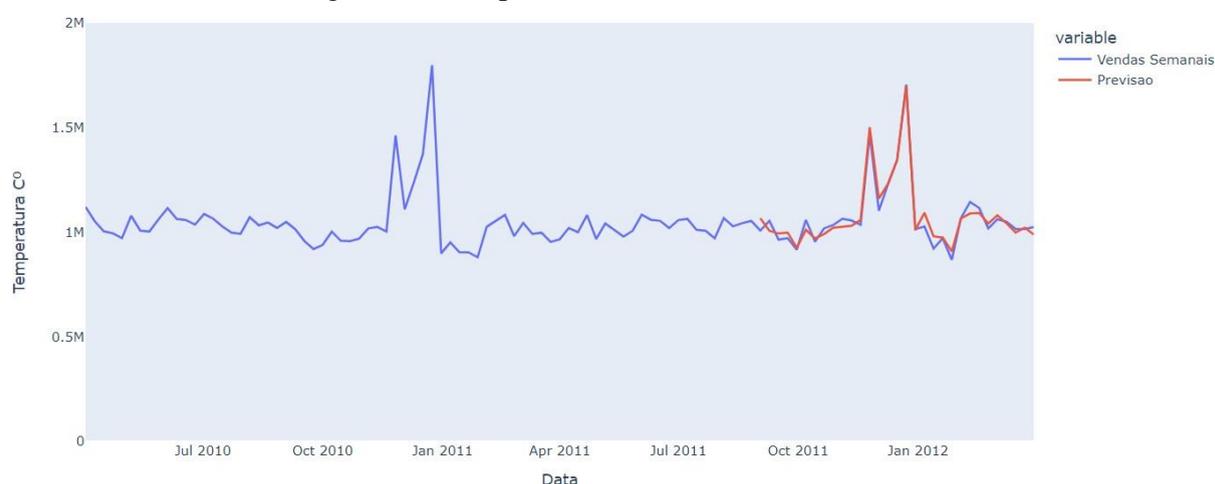
4.5 Utilização do modelo

Após definirmos o modelo em questão, torna-se possível a utilização dele para concluir algumas informações a respeito dos dados aplicado no ambiente produtivo. Como informado

acima os experimentos de criação e o melhoramento do modelo foram realizados no Facebook Prophet e a tabela de predição do teste exportado para a utilização dele no desenvolvimento desse trabalho.

A figura a seguir apresenta um gráfico que ilustra o modelo desenvolvido com base nas características e definições especificadas na implementação. O objetivo desta figura é proporcionar uma melhor visualização da previsão gerada pelo modelo em relação aos dados utilizados. O gráfico destaca a comparação entre os valores previstos e os dados reais.

Figura 22: Temperatura média nas semanas.



Fonte: Autoria própria (2023)

No gráfico acima, apresenta-se uma previsão das vendas semanais destacada em azul. Através do gráfico, é possível observar a relação entre as vendas semanais, a temperatura e a data das vendas. A linha vermelha representa a previsão das vendas semanais. Ao comparar a parte vermelha do gráfico, que representa a previsão, com o valor em azul, que representa o valor real das vendas semanais, é evidente que a previsão se aproxima significativamente dos valores reais. Essa proximidade sugere que o modelo criado está realizando previsões consistentes.

4.6 Teste de Shapiro-Wilk

Neste tópico, realizaremos o teste de Shapiro-Wilk como forma de validar o modelo, com base na figura 23. Primeiramente, chamamos a função `plot_dataframe` para criar um histograma dos valores 'Resíduos' do conjunto de dados preditos. Em seguida, calculamos o teste de normalidade Shapiro-Wilk nos dados da coluna 'Resíduos' dos conjuntos de dados preditos utilizando a função `shapito`. Os resultados são atribuídos às variáveis `stats_teste` e

p_valor_teste. Em seguida, definimos o nível de significância como 0.05 e verificamos se o valor de p_valor_teste é menor que o nível de significância definido. Caso seja, exibimos a mensagem "Rejeitar H0!!! Não há evidências de que os dados testados sejam normalmente distribuídos", juntamente com o valor de p_valor_teste arredondado para 6 casas decimais. Caso contrário, exibimos a mensagem "H0 não pode ser rejeitada!!! Há evidências de que os dados testados sejam normalmente distribuídos", juntamente com o valor de p_valor_teste arredondado para 6 casas decimais.

Figura 23: Código dos erros do modelo de linha de base.

```
processador.plot_dataframe(preditos['Residuos'],
                           chart_type='histogram',
                           title='Teste de normalidade nos dados do modelo final',
                           x_label='Distribuição dos erros do modelo final',
                           y_label='Frequência dos valores',
                           showlegend=False)

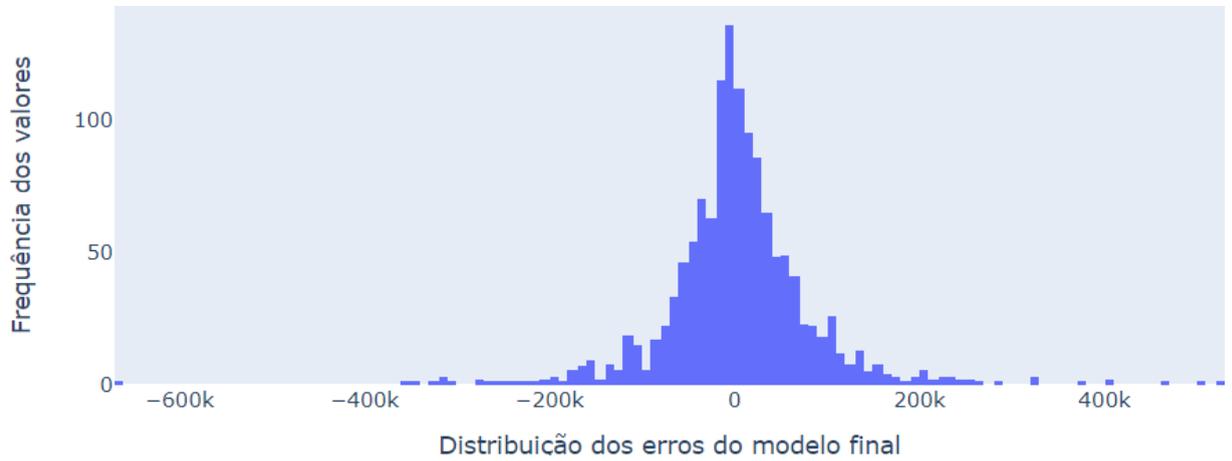
stats_teste, p_valor_teste = stats.shapiro(preditos['Residuos'])
signific = 0.05

if p_valor_teste < signific:
    print(f'p valor: {round(p_valor_teste,6)} Rejeitar H0!!!')
else:
    print(f'p valor: {round(p_valor_teste,6)} H0 não pode ser rejeitada!!!')
```

Fonte: Autoria própria (2023)

A figura acima retorna um gráfico de histograma que representa a normalidade do modelo criado em produção. Executando esse código será gerado o histograma da figura abaixo.

Figura 24: Distribuição dos erros do modelo de linha de base.



Fonte: Autoria própria (2023)

Analisando o histograma acima podemos chegar à conclusão de que com um p valor igual a 0.0 Rejeita H_0 , sendo assim não há evidências de que os dados testados são normalmente distribuídos. Portanto, a hipótese nula de que os dados seguem uma distribuição normal foi rejeitada com muita confiança estatística. Isso sugere que os dados possam seguir uma distribuição diferente da normal ou que possuam algum tipo de desvio significativo em relação à normalidade.

5. RESULTADOS OBTIDOS

Considerando que o objetivo principal do trabalho desenvolvido não é a tomada imediata de decisões, mas sim a lapidação dos dados usando conceitos e recursos básicos da ciência de dados para verificar as vendas no comércio atacadista, é viável adotar exemplos hipotéticos para auxiliar na tomada de decisões, conforme realizado na apresentação e detalhamento dos dados, assim como no modelo criado.

Levando em consideração o exposto acima, os gráficos criados podem auxiliar nas tomadas de decisões. Por exemplo, a imagem 17 do trabalho representa uma previsão das vendas com base na análise da temperatura média das semanas em que ocorreram as vendas. Isso permite que a empresa contrate uma equipe maior nos dias de maior volume de vendas e providencie um estoque maior para atender à demanda nesses dias. Além disso, possibilita que a empresa invista mais capital nos dias de menor volume de vendas para estimular o consumo nesse período. O trabalho também identifica a influência dos feriados nas vendas, possibilitando analisar e melhorar as vendas nas semanas em que esses feriados ocorrem. Dessa forma, o modelo desenvolvido se torna uma ferramenta valiosa para empresas que desejam ter uma visão abrangente da análise de seus dados de vendas, possibilitando previsões com base nessas informações e auxiliando na tomada de decisões fundamentais para o crescimento dessas instituições.

6. CONCLUSÃO

Tendo em vista o modelo de ciclo de vida dos dados na ciência de dados, é possível afirmar que as informações desempenham um papel extremamente importante no processo de tomada de decisão das empresas. Nesse contexto, a ciência de dados assume um papel fundamental, tanto para as empresas em geral quanto para o setor de comércio atacadista. Ao analisarmos a necessidade de extrair informações de dados brutos, percebemos que esse processo pode ser extremamente desafiador. No entanto, com o auxílio da ciência de dados, essa tarefa se torna mais simples e acessível.

Nesse trabalho, foi desenvolvido um modelo fundamentado na ciência de dados aplicada ao setor atacadista, utilizando os dados da empresa Walmart. Através desse modelo, foram criados relatórios e gráficos de alta relevância para auxiliar a empresa Walmart em suas tomadas de decisões. Um aspecto importante desse modelo foi a capacidade de realizar previsões com base nas vendas relacionadas à temperatura média das semanas em que ocorreram as vendas. Para viabilizar isso, foram utilizadas ferramentas essenciais, como o Google Colab e o Facebook Prophet, que facilitam o tratamento dos dados, permitem a análise de dados históricos de forma simplificada e possibilitam a construção de previsões com base em tendências, sazonalidades e feriados.

Diante do exposto, pode-se concluir que a ciência de dados é um instrumento de extrema importância, que muitas vezes se torna essencial. A partir de dados brutos, é possível realizar análises e identificar informações relevantes para as empresas. Além disso, a ciência de dados possibilita a realização de previsões, fornecendo uma base científica para a tomada de decisões cruciais que auxiliam no crescimento dessas organizações. É importante ressaltar que esse trabalho também contribuirá para proporcionar aos novos estudantes uma maior exposição e contato com a ciência de dados, ampliando seu conhecimento nessa área em constante expansão.

REFERÊNCIAS

AGRAWAL, Himanshu; CHAFLE, Girish; GOYAL, Sunil; MITTAL, Sumit; MUKHERJEA, Sougata. An Enhanced Extract-Transform-Load System for Migrating Data in Telecom Billing. 2008 IEEE 24th International Conference on Data Engineering, [s. l.], 2008.

ALVES, Yago Modesto; SILVA, Edna de Mello. O jornalismo selfie como prática jornalística nas mídias sociais: casos e exemplos. In: II Jornada Interdisciplinar do Programa de Pós-graduação em Comunicação e Sociedade, 2017, Palmas. Anais da II Jornada Interdisciplinar do Programa de Pós - Graduação em Comunicação e Sociedade: Comunicação, Inovação e Tecnologias, 2017. p. 114-120.

AMARAL, Fernando. Introdução à ciência de dados: mineração de dados e big. Rio de Janeiro: ALTA Books, 2016.

CORNELIUS JUNIOR, Romeu. USO DA MINERAÇÃO DE DADOS NA IDENTIFICAÇÃO DE ALUNOS COM PERFIL DE EVASÃO DO ENSINO SUPERIOR. 2015. 138 f. TCC (Graduação) - Curso de Ciência da Computação, Universidade de Santa Cruz do Sul, Santa Cruz do Sul, 2015. Cap. 3. Disponível em: <https://repositorio.unisc.br/jspui/bitstream/11624/535/1/Romeu_Cornelius_Junior_-_TCC_-_Final.pdf>. Acesso em: 20 maio 2023.

DAVENPORT, Thomas. Big Data no trabalho: Derrubando Mitos e descobrindo oportunidades. Rio de Janeiro: Alta Books, 2017.

HAGEN, L. et al. Processes, potential benefits, and limitations of big data analytics: A case analysis of 311 data from city of miami. In: Proceedings of the 20th Annual International Conference on Digital Government Research. New York, NY, USA: Association for Computing Machinery, 2019. (dg.o 2019), p. 1–10. ISBN 9781450372046. Disponível em: <<https://doi.org>>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Séries históricas e Estatísticas – Comercio – Índices mensais do comercio varejista. Disponível em: <https://serieestatisticas.ibge.gov.br/series.aspx?no=2&op=0&vcodigo=MC69&t=receita-nominal-varejo-indice-acumulado-12>. Acesso em: 05/05/2023.

JUNIOR, Jose Carlos Da Silva Freitas; MAÇADA, Antonio Carlos Gastaud; OLIVEIRA, Mirian; BRINKHUES, Rafael Alfonso. BIG DATA E GESTÃO DO CONHECIMENTO: DEFINIÇÕES E DIRECIONAMENTOS DE PESQUISA. REVISTA ALCANCE, [S. l.], p. 04-22, 15 maio 2023.

KAGGLE. Walmart Store walmart_analise_e_previsao_de_vendas, promotion, and competitor data. In: Featured Prediction Competition. [S. l.], 5 ago. 2018. Disponível em: <https://www.kaggle.com/code/uclitonviana/walmart-analise-e-previsao-de-vendas>. Acesso em: 19 maio. 2023

MATA, KESLEY BRENNER DA COSTA. E-COMMERCE: ANÁLISE DE DADOS SOBRE O COMÉRCIO ELETRÔNICO NO BRASIL. Escola de Ciências Exatas e da Computação, da Pontifícia Universidade Católica de Goiás, [s. l.], 2021.

NETTO, Antonio Valerio. CIÊNCIA DE DADOS EM SAÚDE: CONTRIBUIÇÕES E TENDÊNCIAS PARA APLICAÇÕES. Revista Saúde.Com, [s. l.], 2023.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

WITTEN, Ian H.; FRANK, Eibe; HALL, Mark A.. Data Mining Practical Machine Learning Tools and Techniques. 3. ed. [s.l.]: Morgan Kaufmann, 2011. 664 p.

APÊNDICE



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS

PUCGABINETE DO REITOR

Av. Universitária, 1069 Setor Universitário

Caixa Postal 86 • CEP 74605-010

GOIÁS Goiânia • Goiás • Brasil

Fone: (62) 3946.1000 www.pucgoias.edu.br reitoria@pucgoias.edu.br

RESOLUÇÃO n.º 038/2020- CEPE
ANEXO 1 APÊNDICE ao TCC

Termo de autorização de publicação de produção acadêmica

O(A) estudante Samuel Mota Elias do Curso de Engenharia da Computação, matrícula 20171003301570 telefone: (62)98307-6189 e-mail samuel.mota19@hotmail.com na qualidade de titular dos direitos autorais, em consonância com a Lei nº 9.610/98 (Lei dos Direitos do autor), autoriza a Pontifícia Universidade Católica de Goiás (PUC Goiás) a disponibilizar o Trabalho de Conclusão de Curso intitulado Ciência de dados aplicada ao comercio atacadista gratuitamente, sem ressarcimento dos direitos autorais, por 5 (cinco) anos, conforme permissões do documento, em meio eletrônico, na rede mundial de computadores, no formato especificado (Texto (PDF); Imagem (GIF ou JPEG); Som (WAVE, MPEG, AIFF, SND)•, Vídeo (MPEG, MWV, AVI, QT)•, outros, específicos da área; para fins de leitura elou impressão pela internet, a título de divulgação da produção científica gerada nos cursos de graduação da PUC Goiás.

Goiânia, 14 de abril de 2023

Assinatura do(s) autor(es): *Samuel Mota Elias*

Nome completo do autor: Samuel Mota Elias

Assinatura do professor-orientador: _____

Nome completo do professor-orientador: