

## **ANÁLISE E PREVISÃO DO VOLUME DE VENDAS DO SETOR VAREJISTA DE SUPERMERCADOS E HIPERMERCADOS BRASILEIROS**

**JOÃO PEDRO VIEIRA PORTARO**

UNIVERSIDADE PRESBITERIANO MACKENZIE

**RAFAELA COSTA MARTINS DE MELLO DOURADO**

FACULDADE DE ECONOMIA, ADMINISTRAÇÃO, CONTABILIDADE E ATUÁRIA DA UNIVERSIDADE DE SÃO PAULO - FEA

**ORLANDO YESID ESPARZA ALBARRACIN**

UNIVERSIDADE PRESBITERIANA MACKENZIE (MACKENZIE)

Agradecimento à orgão de fomento:

Não se aplica.

# ANÁLISE E PREVISÃO DO VOLUME DE VENDAS DO SETOR VAREJISTA DE SUPERMERCADOS E HIPERMERCADOS BRASILEIROS

## 1 INTRODUÇÃO

O setor de varejo tem se revelado um pilar de extrema importância para a economia brasileira ao longo da última década. Sua relevância advém de sua posição estratégica na cadeia de distribuição de produtos e serviços, e da proximidade direta com os consumidores finais (SEBRAE, 2023). Estabelecimentos deste setor são responsáveis por fornecer uma ampla variedade de produtos alimentícios, bens de consumo e serviços diretamente aos consumidores finais por todo o país. Em termos de geração de serviços e empregos, o setor varejista é um importante empregador, fornecendo uma ampla gama de oportunidades de trabalho para diferentes níveis de habilidade e educação (Novo varejo, 2023), desde caixas, reposição de estoque, atendimento ao cliente até funções gerenciais. Esses empregos contribuem para o crescimento econômico, proporcionando renda e estabilidade para muitas famílias brasileiras.

No ano de 2021, o varejo movimentou mais de R\$1,99 trilhão, o que equivaleu para aquele ano, 22,9% do PIB brasileiro (SEBRAE, 2023). Em termos quantitativos de serviços e geração de empregos, tal setor empregou 25,8% dos trabalhadores de carteira assinada do Brasil em 2021, isto é, um a cada quatro trabalhadores de carteira assinada são empregados pelo varejo.

Com o constante desenvolvimento e evolução tecnológica, o varejo tem encontrado desafios na última década, devido as mudanças de relações e interações entre consumidores e vendedores/empresas. Exemplificando tais desafios, é plausível citar a pandemia decorrida por conta da COVID-19. Durante um período, os consumidores ficaram confinados em suas residências, o que modificou definitivamente a estrutura do setor varejista e a maneira de como os mesmos se relacionariam, a partir daquele ocorrido, com os clientes. Neste contexto, a pandemia resultou em uma grande ascensão de plataformas digitais que passaram a ter o comércio eletrônico como principal fonte de renda dentro do setor varejista.

A ciência e a análise de dados, termos amplamente disseminados a partir dos anos 2000, empregam técnicas estatísticas juntamente com recursos computacionais, para possibilitar trabalhar, estudar, entender e analisar grandes conjuntos de dados. Com os avanços tecnológicos, foi se tornando cada vez mais possível ter computadores em ambientes corporativos, já que os custos para isso reduziram gradativamente, juntamente com as suas dimensões e tamanhos, que deixou ainda mais viável obter computadores em empresas familiares ou de pequeno e médio porte. Atualmente, a ciência de dados é indispensável em inúmeras áreas, como por exemplo: Vendas, Marketing, Financeiro, Operações.

Nos setores de Marketing e Vendas, é possível utilizar *Data Science* para entender o comportamento do cliente, quais são seus padrões e em quais meses do ano eles estão mais dispostos a gastar dinheiro. Criar estratégias e planos para minimizar custos e maximizar lucros com base nos padrões encontrados através da análise do perfil de consumo dos clientes faz-se necessário dentro das empresas do século XXI. Em Finanças, na área de contabilidade, a ciência de dados atua para prever valores da empresa e o fluxo de caixa, analisar DREs (Demonstração de Resultados), e avaliar os riscos financeiros da companhia. Por fim, na área de Operações, com a análise de dados, pode-se encontrar

quais são as operações gargalo do sistema, encontrar oportunidades de melhorias internas dos processos, para que assim, seja possível ter uma otimização de produção, gerando mais produtos e, conseqüentemente, aumentando a receita. “O *Data Science* é um campo em evidência e que está em alta; não requer muita investigação para encontrar prognósticos de analistas de que, nos próximos dez anos, precisaremos de bilhões e bilhões de cientistas de dados a mais do que possuímos atualmente.” (GRUS, 2016, p.8)

O setor de supermercados e hipermercados brasileiros é uma das áreas em que é possível utilizar a análise de dados de maneira eficiente. De acordo com a Expert XP (2023), tal setor movimentou mais de R\$ 600 bilhões no ano de 2021, somando mais de 92 mil lojas espalhadas pelo Brasil. Podemos observar que a magnitude monetária envolvida com esse setor é alta, movimentando mais de meio trilhão de reais no período de apenas um ano. Com os dados abertos do site da IBGE (Instituto Brasileiro de Geografia e Estatística), é possível encontrar valores reais dos índices de volume de vendas no de vendas no comércio varejista no país por mês.

Com a visualização destes números citados, é notável a grande importância deste setor para a economia e funcionamento do país. Porém, como em todos os outros setores, as tecnologias estão em constante desenvolvimento e vem evoluindo com o passar dos anos, tornando-se, quase que, obrigatório o investimento em análise de dados para tomadas de decisão estratégica nos mercados e hipermercados, já que aqueles que não evoluírem, passarão a ter dificuldades para se manter competitivo, produtivo, eficaz e pontual. Assim, com as novas tecnologias provindas da era da “Indústria 4.0”, os donos de supermercados e hipermercados passam a se perguntar de que maneira é possível utilizar as mesmas em prol da sua empresa, com o objetivo de obter vantagens, maior visibilidade e melhor reputação quando comparado aos seus concorrentes.

## 2 CONTEXTO INVESTIGADO

Para esta seção optou-se por separar a revisão técnica de Séries Temporais, discutida no item 2.1 e a discussão das pesquisas recentes relacionadas à previsão de demanda, abordadas no item 2.2. Esta seção contempla os fundamentos teóricos e empíricos (pesquisas correlatas) que sustentarão os resultados encontrados.

### 2.1 Séries temporais

Uma série temporal tem como base um agrupamento de observações sobre a trajetória de uma variável qualquer em relação a uma outra variável independente, o tempo. Uma série temporal é definida como um conjunto de observações de uma dada variável, geralmente distribuídas de maneira equidistante pelo fator tempo e, que possuem como característica central a presença de uma dependência serial entre elas (MORETTIN; TOLOI, 2004).

Dentro das análises destas séries, a variável observada depende de seu período anterior, ou seja, as variáveis estudadas em séries temporais possuem características de dependência serial. Isto facilita para o pesquisador explicar e prever tendências ou comportamentos, para auxiliar nos momentos de tomadas de decisões dentro de uma empresa.

A tentativa de criação de métodos para previsão de valores, buscando conseguir antecipar o futuro, sempre ocorreu, porém, quando os cientistas George E. P. Box e Gwilym M. Jenkins publicaram o livro *Time Series Analysis: Forecasting and control*, houve o surgimento de um novo método com base em premissas estatísticas e matemáticas para a previsão de séries temporais, a partir de valores passados da série estudada. Tal modelo para previsões foi denominado de ARIMA (modelo autorregressivo integrado de médias móveis). Desde os anos 70, o mesmo vem sendo muito utilizado em diversas áreas,

incluindo finanças, saúde, engenharia, economia, entre outras. Sua popularidade pode ser relacionada com a eficácia e acurácia do método, juntamente com a sua capacidade de compreender e lidar com complexos dados temporais, conseguindo encontrar padrões e tendências importantes para previsões de valores futuros e no auxílio para as tomadas de decisão para a resolução de problemas práticos. Entretanto, por maior que seja a eficiência do modelo, é necessário considerar que os valores dos filtros e a modelação do problema não são feitos por máquinas, então é preciso considerar a variável de erro humano na organização, estruturação e modelagem do problema. Se tal estrutura for bem feita e modelada, a probabilidade de serem fornecidos bons valores preditivos é maior.

Os modelos citados são o resultado da combinação de três filtros, o componente autorregressivo (AR), o filtro de integração (I) e o componente de médias móveis (MA), juntamente com as 4 etapas/estágios para aplicação do mesmo, que será explicado em detalhes na seção de Metodologia.

## 2.2 Previsão de demanda

As metodologias utilizadas e estudadas neste trabalho como, por exemplo, o método ARIMA, são metodologias para previsões quantitativas, baseadas em correlações e séries temporais. No cenário científico atual, existem inúmeros métodos para a previsão de demanda através de séries temporais: dentro de um período de sete anos (2015-2021), existem mais de 150 artigos científicos que utilizam métodos e modelos matemáticos/estatísticos para a previsão de demanda e criação de *Forecastings* (CAMBUÍ *et al.*, 2023).

Todas as técnicas atuais têm como premissas básicas conceitos de regressões lineares (simples ou múltiplas), com influência de variáveis independentes no comportamento de variáveis dependentes. Os métodos mais utilizados atualmente são: Método de suavização exponencial simples (sem tendências ou sazonalidade); modelo exponencial com variação de tendência (Método Holt); o modelo exponencial com variação de tendência e sazonalidade (Método Holt-Winters) e o modelo de média móvel autorregressiva (ARIMA), que identificam a existência de correlação entre as observações e com os dados de demanda sazonal repetindo um padrão de ciclo de variação ao longo do tempo, em que a suposição de autocorrelação é, como regra geral, válida (Da Veiga; Tortato, 2016).

Com essa ampla variedade de métodos e modelos de capacidade preditiva utilizando de séries temporais, dentro do setor varejista de supermercados e hipermercados, a previsão de demanda passou a assumir um papel estratégico. Por conta da intensa competitividade do setor, todos os concorrentes passaram a buscar ferramentas e métodos para otimizar seus resultados, juntamente com a capacidade de prever comportamentos e antecipá-los, auxiliando na tomada de decisão, aprimorando a produtividade e aumentando os resultados.

O presente trabalho tem aspectos em comum quando comparado ao trabalho de Borges e Landim (2018), pela Universidade do Cabo Verde, em que o modelo ARIMA foi utilizado para a previsão de vendas de uma loja estudada pelos autores. A escolha do modelo e os objetivos de antecipação do futuro com dados passados utilizando a metodologia Box-Jenkins juntamente com séries temporais são os pontos com maior similaridade, porém, existem dois principais fatores que diferenciam as duas pesquisas, o primeiro é a base de dados, em que os autores citados utilizaram uma série temporal de 14 anos (2000-2014), enquanto neste trabalho, foi utilizado uma série de apenas 6 anos (2018-2023), pois entende-se que o histórico recente esteja mais correlacionado com o futuro próximo. O outro fator é que foi escolhido o modelo ARIMA para a previsão de vendas de apenas uma loja, em contrapartida, este artigo tem o objetivo de prever a previsão de

vendas de um setor varejista por completo, neste caso, o setor de supermercados e hipermercados do Brasil. Desta forma, foi possível perceber a ampla variedade de casos e estudos específicos em que é possível utilizar métodos de previsão do futuro através de modelos como ARIMA, SARIMA e ARIMAX. Apenas para a questão de previsão de volume de vendas, tal previsão pode ser modelada para casos específicos e pequenos e para casos mais gerais, amplos e complexos.

### **3 DIAGNÓSTICO DA SITUAÇÃO-PROBLEMA**

As maiores empresas do setor de comércio varejista do Brasil fazem parte do grupo Carrefour Brasil, que teve um faturamento de R\$81,1 bilhões no ano de 2021, com 498 lojas nos 26 estados do País (SBVC, 2023). O uso de Data Science possibilita que essas empresas otimizem produções, maximizem lucros, minimizem custos de produção, de estoque e de logística.

O índice de volume de venda é um dos mais importantes índices para que seja feita uma análise da empresa no setor varejista, pois permite questionar se as tecnologias atuais podem ser úteis para otimizar produções e custos. Utilizando técnicas estatísticas e ferramentas de *Data Science*, a pergunta-problema que este trabalho busca responder é: existem maneiras de encontrar e prever padrões para o volume de vendas do setor varejista de supermercados e hipermercados brasileiros ao longo do tempo?

O objetivo principal da pesquisa é modelar o volume de vendas do setor varejista brasileiro por meio de séries temporais para fazer previsões a curto prazo. Para isso, foi analisada a série de volume de vendas do setor varejista de supermercados e hipermercados brasileiros dentro de um período de 6 anos (2018-2023) por meio de ferramentas da estatística descritiva, visando decompor seus valores em tendências, padrões e sazonalidades. Juntamente a isso, estudou-se mudanças de vendas e das relações entre vendedores e clientes ao decorrer dos últimos anos, considerando o impacto da pandemia decorrida por conta da COVID-19.

Para atingir o objetivo proposto, foi utilizado a metodologia ARIMA (modelo auto regressivo integrado de médias móveis) a fim de possibilitar aos comerciantes do setor varejista traçar e propor estratégias de mercado embasada nas previsões de vendas, auxiliando-os nas tomadas de decisões dentro de suas empresas, visando eficiências operacionais e melhorias no planejamento estratégico, além de obter vantagem competitivas dentro de seu mercado de atuação.

A análise de dados combinado com o modelo ARIMA, que utiliza de dados de séries temporais para que sejam feitas previsões, pode trazer melhorias e otimizações nas gestões de custos e demandas dos setores varejistas de hipermercados e supermercados dentro do território brasileiro. Esta análise preditiva fornecerá subsídios para tomadas de decisões estratégicas, seja sobre quantidade de produtos a serem feitas e/ou estocadas, ou sobre a escolha de quais produtos focar no marketing e na produção.

### **4 INTERVENÇÃO PROPOSTA**

Através da análise de séries temporais, mais especificamente, do método ARIMA, o presente trabalho fornece um estudo quantitativo e uma proposta de modelo para a previsão do volume de vendas do setor varejista de supermercados e hipermercados.

A metodologia Box-Jenkins se resume em ajustar modelos autorregressivos integrados de média móvel a um conjunto de dados de séries temporais e foi popularizada por Gwilym Jenkins e George Box no início dos anos 70. A notação padrão uso desta metodologia terá a representação de: ARIMA (p,d,q). Onde o “p” indica o filtro 1 do

modelo, o componente auto regressivo (AR). O “d” indica a segunda ordem ou, o filtro de integração. Por fim, o terceiro filtro, que na notação padrão está denominado por “q”, é o componente de média móvel (MA) (MORETTIN; TOLOI, 2006). A Equação 1 apresenta o modelo ARIMA (p,d,q):

$$Z_t = \Delta^d X_t$$

$$Z_t = \beta_1 Z_{t-1} + \dots + \beta_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-p} \quad (1)$$

sendo,  $X_t$  a série observada,  $Z_t$  o valor que a série assume no tempo t,  $\beta_1, \dots, \beta_p$  e  $\theta_1, \dots, \theta_q$  coeficientes numéricos e  $a_t$  o termo de erro aleatório.

A aplicação da metodologia de Box-Jenkins é composta de 4 estágios ou etapas:

**i) Identificação:** A identificação do modelo parte da determinação dos valores dos três filtros ou componentes p, d e q (BOX & JENKINS, 1991), buscando a estacionaridade da série temporal (os dados são estacionários quando sua média e variância são constantes ao longo do tempo).

**ii) Estimação:** Durante a etapa de estimar os parâmetros do modelo, utiliza-se algum procedimento iterativo de estimação de mínimos quadrados não lineares. Desta forma, é possível selecionar o melhor modelo através de alguns critérios de decisão.

**iii) Validação ou verificação:** Verifica-se se os resíduos do modelo apresentam algum tipo de dependência com o tempo. Nesta metodologia, os dados não podem apresentar tal dependência. Para um número de amostras de dados suficientemente grande, verifica-se que os resíduos são normalmente distribuídos com média zero.

**iv) Previsão:** Após passar pelas fases de identificação, estimação e validação do modelo, é realizado a aplicação do mesmo para a previsão de valores futuros com dados do passado. Esta é uma das etapas mais importantes da modelagem matemática, pois o principal objetivo do método ARIMA é de fazer previsões futuras (*Forecast*). Após feita as previsões, será feita a verificação de acuracidade do modelo.

Com a etapa da identificação, em que os parâmetros tem seus valores anexados, passamos a observar os valores de erro médio e erro percentual absoluto médio, RMSE e MAPE, respectivamente. Com a finalidade de prever o volume de vendas do setor varejista de supermercados e hipermercados brasileiros, foi necessário escolher o modelo que retorna os menores valores destes erros.

Os critérios RMSE e MAPE são definidas como:

$$RMSE = \frac{1}{N} \sum_{h=1}^N (y_{T+h} - \hat{y}_{T+h})^2 \quad (3)$$

$$MAPE = \frac{1}{N} \sum_{h=1}^N |y_{T+h} - \hat{y}_{T+h}| \times 100 \quad (4)$$

em que T é o último valor da série utilizado na modelagem e  $\hat{y}_{T+h}$  a previsão feita pelo modelo para os dias  $T + h$ .

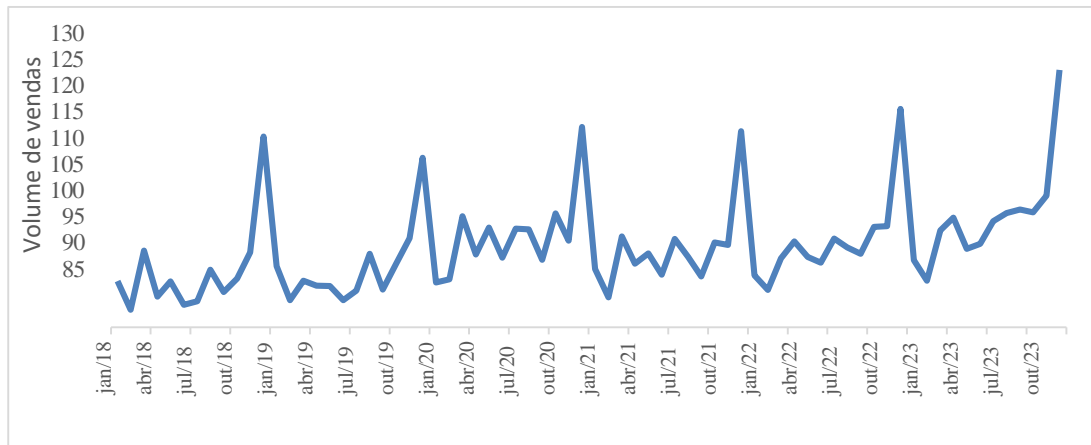
Para este trabalho, foram utilizado os dados de volumes de vendas mensais de janeiro de 2018 até dezembro de 2023, disponibilizados no site da IBGE através do endereço <https://sidra.ibge.gov.br/tabela/8883> (IBGE, 2024). Para a análise dos dados e construção do modelo foram utilizados o Excel e o software estatístico livre R.

## 5 RESULTADOS OBTIDOS

### 5.1 Análise descritiva

A Figura 1 apresenta a série temporal do volume de vendas do setor varejista de supermercados e hipermercados brasileiros no período estudado.

Figura 1: Série Temporal de volume de vendas do setor varejista de supermercados e hipermercados Brasileiros



Fonte: Autoria própria (2024).

A série temporal do setor estudado possuiu duas principais características: a primeira refere-se ao notável crescimento contínuo do volume de vendas ao passar dos anos, ou seja, independente de variações mensais e sazonalidades, a tendência é de que, ao final de cada ano, o volume médio será maior quando comparado ao do ano anterior. Tal comportamento pode ter relação com o crescimento populacional do nosso país, de acordo com a IBGE (2023), de 2010 a 2022, a população brasileira cresceu mais de 6,5%, chegando ao valor de aproximadamente 203 milhões de pessoas, com este aumento populacional, é necessário de maiores quantidades de produtos e insumos para suprir as demandas e necessidades da população. Outro fator que pode explicar este comportamento é o crescimento econômico do Brasil a partir do ano de 2017, em que houve um crescimento do PIB em mais de 1% em cada ano quando comparado ao período anterior, com a exceção do ano de 2020, por conta do colapso que a economia sofreu com o início da pandemia devido a COVID-19. No entanto, no ano de 2023, o PIB obteve um aumento significativo quando comparado com o aumento dos últimos 10 anos, crescendo quase 3% (IBGE, 2024). Desta maneira, pode-se relacionar e supor que, com o crescimento econômico do País, o poder aquisitivo da população tende a aumentar, o que resulta em um consequente crescimento de consumo no setor varejista de supermercados e hipermercados.

A Tabela 1 apresenta as medidas descritivas do volume de vendas do setor varejista de supermercados e hipermercados brasileiros por ano. Nota-se que o valor médio do volume de vendas aumentou gradativamente com o passar dos anos, com a exceção do ano de 2021, já que este comportamento estava diretamente ligado com os impactos provenientes da pandemia do COVID-19. Estes dados reforçaram e corroboraram com o comportamento visualizado na Figura 1, discutido anteriormente.

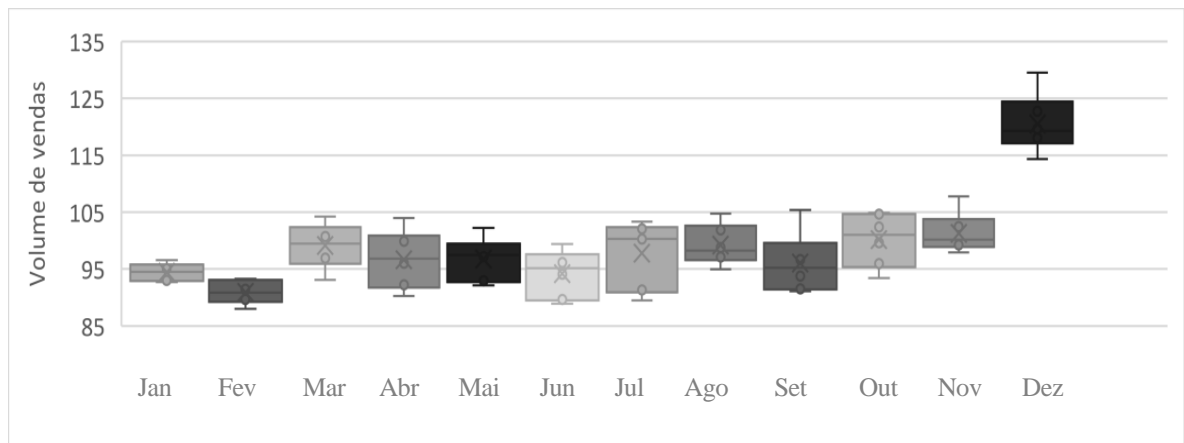
Tabela 1: Medidas descritivas do volume de vendas do setor varejista por ano.

Métrica \ Ano	2018	2019	2020	2021	2022	2023
Média	94,692	95,293	100,994	98,570	100,000	104,065
Q1	89,700	91,404	96,752	94,339	96,356	98,754
Q2	92,961	92,644	100,943	97,460	98,257	103,649
Q3	97,248	97,248	103,709	100,138	101,861	105,231
Desv Pad	8,045	6,823	7,065	7,123	7,846	9,046
Coef Var	64,718	46,551	49,916	50,736	61,561	81,829

Fonte: Autoria própria (2024).

A Figura 2 apresenta os gráficos box-plot do volume de vendas por mês, compreendendo todo o período de 2018 a 2023. Observando as variações mensais a cada ano, é possível notar uma sazonalidade que ocorre no último trimestre de cada ano. Entre outubro e, principalmente, dezembro, temos picos do volume de vendas.

Figura 2: Gráfico *box-plot* dos valores mensais do volume de vendas.



Fonte: Autoria própria (2024).

No final do ano, mais especificamente em dezembro, é o período com mais festas e feriados importantes como o Natal e Ano Novo, em que a população tende a aumentar seus gastos com presentes e alimentos para as celebrações e confraternizações não apenas em família, mas também, com colegas de trabalho, estudo, esportes e afins. Além das festas, a Black Friday, dia ou semana em que são feitas grandes promoções e ofertas, tem movimentado as vendas no setor nos últimos anos e, portanto, passou a ser mais um fator que contribuiu no período de pico do volume de vendas. De acordo com a Uol, a tendência é que a Black Friday ganhe mais força e, aos poucos, tome o espaço das vendas natalinas (Uol, 2024).

Outro comportamento notável foi a queda do volume de vendas que ocorre no mês de fevereiro, o mês com piores resultados. A principal suposição para este comportamento, é o fato de fevereiro ser o mês subsequente ao período de festas, viagens e férias da grande maioria da população, ou seja, é um período onde as pessoas tendem a gastar menos por conta dos elevados gastos feitos nos meses anteriores e de despesas típicas de início de ano como IPVA e IPTU (Ecommerce Brasil, 2014). Além disso, ocorre a volta para rotina de trabalho e/ou estudos, em que acaba o período de recessos e férias, diminuindo o tempo disponível da população para comprar e consumir produtos. E, por fim, com a última suposição para que fevereiro seja o pior mês em questão de volume de



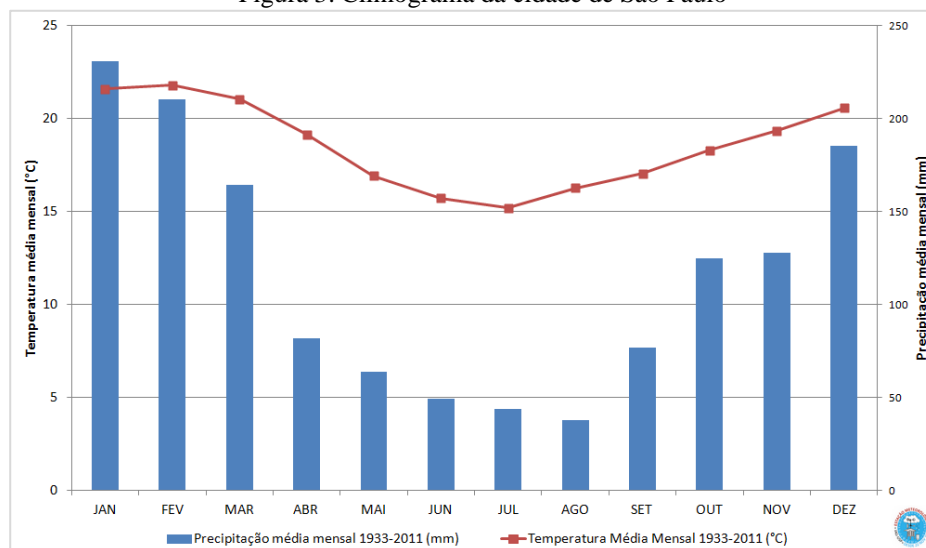
vendas, o segundo mês do ano é o com o menor número de dias úteis, que afeta no poder de compra da população, somando ainda com o extenso feriado de Carnaval, corroborando para que o mesmo seja o período mais fraco. Além disso, este mês não possui nenhum dos feriados/datas comemorativas que mais movimentam o setor, como natal, dia das mães, dia dos namorados e ano novo.

Um fator com pouco destaque, que pode impactar no comportamento dos consumidores e, conseqüentemente, afetar o volume de vendas do varejo é o clima. De acordo com a pesquisadora da ESPM, Rafaela Frankenthal, dependendo das condições meteorológicas, os clientes tendem a comprar certos itens que não são procurados com tanta frequência em outras temperaturas climáticas (Frankenthal, 2017). Por exemplo, em períodos de clima quente e seco, a procura por produtos refrescantes como sucos e sorvetes acabava sendo muito maior quando comparado com a procura destes mesmos itens em períodos de inverno.

Em entrevista à EBC (2020), Fábio Bentes afirmou que outro fator climático que afeta muito o volume de vendas é a chuva, constatando que nos meses em que há uma quantidade de chuva muito acima da média, os volumes de vendas nestas regiões tende a ter resultados negativos.

De acordo com a Figura 3, que apresenta o climograma da cidade de São Paulo no período 1933 a 2011, é possível encontrar alguns padrões, como, por exemplo, que o primeiro trimestre do ano é o período de maior chuva. Relacionando as informações de precipitação com os dados de volume de vendas nos mesmos períodos, levanta-se a hipótese de que a chuva pode ser mais um motivo que explica o fraco período de volume de vendas em fevereiro, corroborando com a possibilidade de afetar negativamente o comportamento do cliente. Ademais, quando as chuvas foram muito intensas e permaneceram por um longo período de tempo, ocorreram prejuízos em safras, plantações, construções, etc. No ano de 2020, foi registrada uma frequência maior do que a esperada de precipitações intensas, principalmente na região sudeste do país, tais chuvas resultaram em perdas de R\$200 milhões, afetando indiretamente o volume de vendas, já que tal ocorrência resultou em menos produtos no mercado, diminuindo a qualidade dos mesmos, não os deixando atrativos para os consumidores (EBC, 2020).

Figura 3: Climograma da cidade de São Paulo

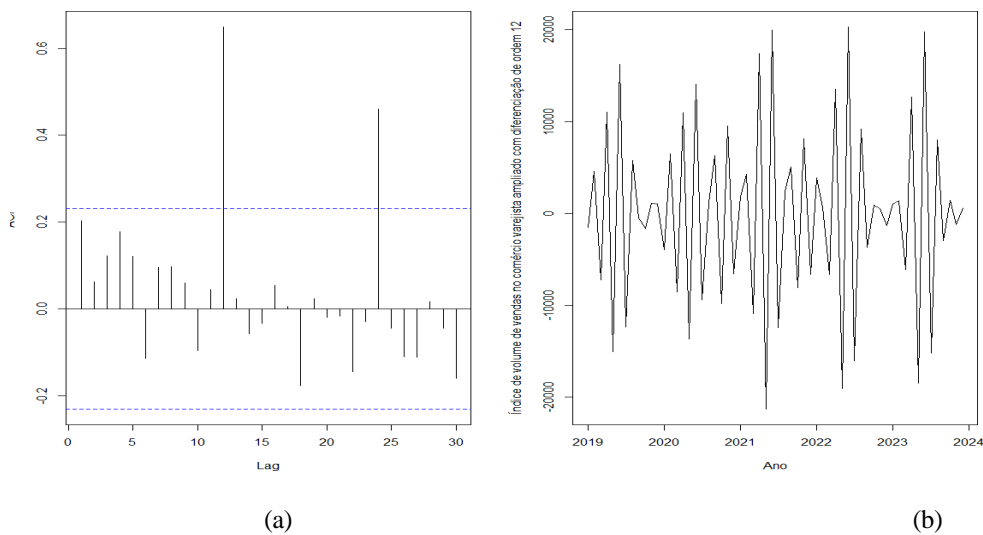


Fonte: (IAG-USP, 2014)

## 5.2 Modelagem

Durante a etapa da análise descritiva, foi observado que a série temporal estudada apresentava certos padrões e sazonalidades, comentados anteriormente, citando o exemplo dos picos de vendas no último trimestre dos anos estudados. Observou-se, na Figura 1, que a série temporal estudada não possuiu a característica de ser estacionária, que é uma premissa para a metodologia Box-Jenkins. Desta maneira, foi necessário manipular a série temporal realizando uma diferença de ordem 12, com o objetivo de que sua média e variância se tornassem constantes. Foi escolhida a diferenciação de ordem 12, justificada na Figura 4, a qual indica que os lags 12 (representa a correlação de hoje com a informação de um ano atrás) e 24 (representa a correlação de hoje com a informação de dois anos atrás) do gráfico ACF são significativos.

Figura 4: Função de autocorrelação (ACF) (a) e série temporal após diferenciação de ordem 12 (b)



Fonte: Autoria Própria (2024).

A etapa seguinte, a construção do modelo, foram utilizados dados de janeiro de 2018 a dezembro de 2022 e teve como escopo testar alterações nos valores dos parâmetros  $(p,d,q)$  e  $(P,D,Q)$  presentes no modelo SARIMA (a inclusão a letra “S” indica a presença de sazonalidade) para encontrar o melhor modelo, segundo os critérios estabelecidos e com maior acurácia para a previsão de valores futuros. Em um primeiro momento, foram ajustados apenas os parâmetros  $p$ ,  $d$ ,  $P$  e  $D$ , mantendo os valores de  $q$  e  $Q$  nulos, em que foram considerados  $p=\{0,1,2,3,4,5,6,7,8,9,10,11,12\}$ ,  $d=\{0,1,2,3,4,5,6\}$ ,  $P=\{0,1,2\}$  e  $D=\{0,1,2\}$ . Nesta etapa, a cada modificação feita nos parâmetros, foram observados os valores dos critérios AIC (*Akaike Information Criterion*) e BIC (*Bayesian Information Criterion*), quanto menor fossem os valores retornados de tais critérios com as mudanças de parâmetros, melhor (mais acurado e menos complexo) era considerado o modelo. Após 30 tentativas e mudanças nos parâmetros, foram escolhidos os 4 melhores modelos, com base nos valores de AIC e BIC, além da preferência pelos modelos com os menores valores para o parâmetro  $D$ , já que a ordem de diferença do mesmo já era consideravelmente alta, 12 (pois a sazonalidade ocorre a cada de 12 meses).

Com os 4 modelos escolhidos, a próxima etapa consistiu na mudança dos parâmetros de erro aleatório  $q$  e  $Q$ , e na verificação dos valores do RMSE (Erro médio) e MAPE (Erro percentual). Para cada um dos 4 modelos escolhidos, foram feitas 10

mudanças nos parâmetros de erro aleatório, no qual,  $q=\{0,1,2, \dots ,10\}$  e  $Q=\{0,1,2,3,4,5\}$ , observando-se o comportamento do RMSE e MAPE. Para a decisão do modelo final, foi escolhido aquele que, independente do valor dos parâmetros  $q$  e  $Q$ , retornava os menores valores de erro médio e percentual. Portanto, o modelo final SARIMA (1 1 0) (1 1 0)<sub>12</sub> foi selecionado e testou-se a significância estatística dos parâmetros, obtendo-se p-valor menor do que o nível de significância, considerado 5%, para todos eles. De acordo com a Tabela 2, é possível observar o ranqueamento dos 4 melhores modelos.

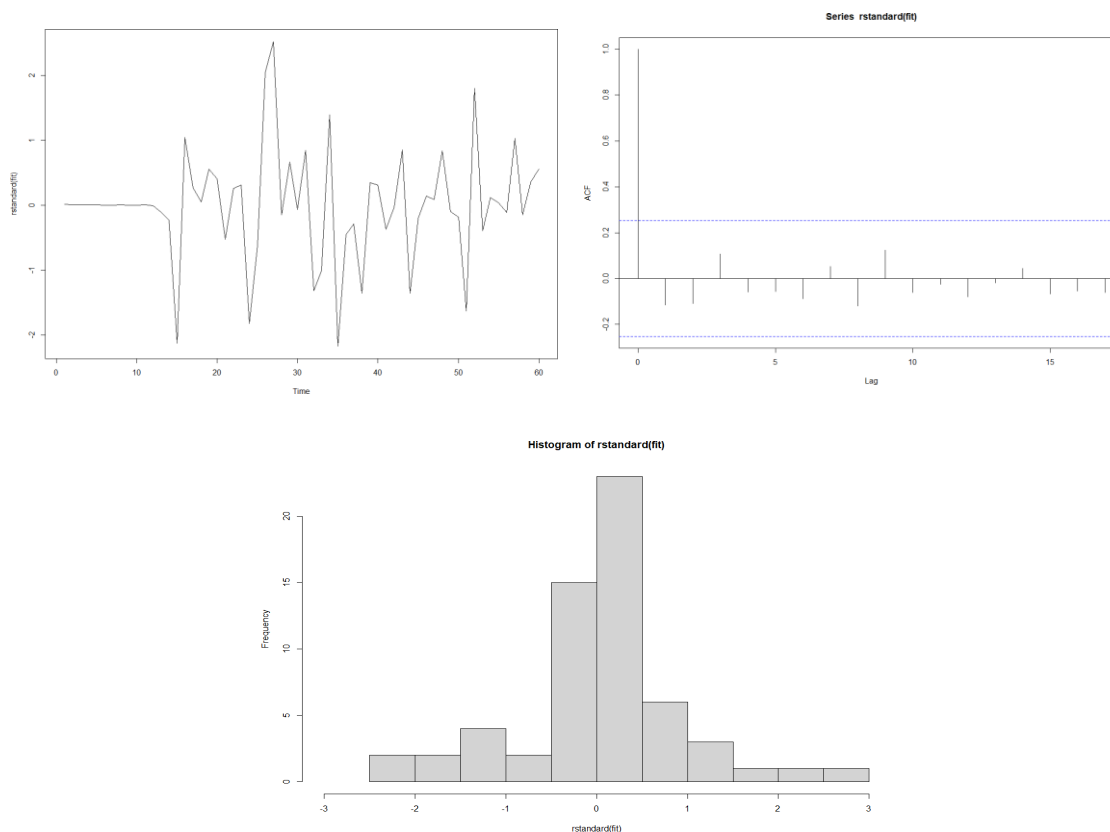
Tabela 2 – Classificação e ranqueamento de AIC, BIC, RMSE e MAPE dos melhores modelos.

Modelo	AIC	BIC	RMSE	MAPE
SARIMA (1 0 0) (1 0 0)	260,283	264,026	4,407	0,032
<b>SARIMA (1 1 0) (1 1 0)</b>	<b>247,885</b>	<b>253,395</b>	<b>2,645</b>	<b>0,021</b>
SARIMA (5 1 0) (1 1 0)	254,041	266,992	2,777	0,022
SARIMA (6 1 0) (1 1 0)	254,536	269,337	2,964	0,023

Fonte: Autoria própria (2024)

Sendo assim, o modelo final foi escolhido e logo após, realizada a análise dos resíduos. Na Figura 5, é exibido os gráficos de resíduos padronizados, ACF dos resíduos e o histograma dos mesmos.

Figura 5 – Resíduos padronizados (a), ACF dos resíduos (b) e histograma dos resíduos (c).



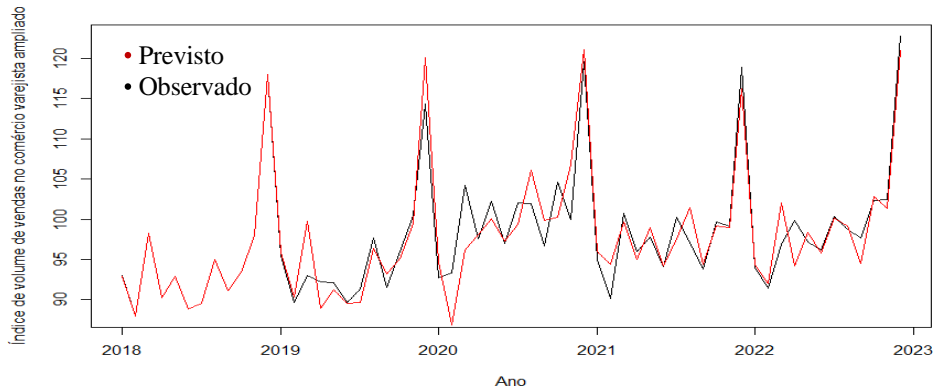
Fonte: Autoria Própria (2024).

Nota-se através do histograma apresentado na Figura 5 que não existem tendências dos resíduos, observando-se que os resíduos possuem comportamento próximo à uma distribuição normal e admite-se que o modelo atende aos pressupostos do método.

### 5.3 Validação do modelo proposto

Após a escolha do modelo final, SARIMA (1 1 0) (1 1 0)<sub>12</sub>, foi feita a previsão do volume de vendas de Supermercados e Hipermercados do setor varejista brasileiro. Na Figura 6, a linha de cor preta representa valores observados (reais) e a linha vermelha representa os valores previstos pelo modelo SARIMA (1 1 0) (1 1 0)<sub>12</sub>, englobando todo o período estudado (janeiro de 2018 a dezembro de 2023).

Figura 6 – Valores observados x previstos pelo modelo do volume de vendas do setor varejista brasileiro.



Fonte: Autoria Própria (2024).

Nota-se que, na maior parte do período considerado, o modelo acompanhou o comportamento da série observada, porém, no ano de 2020, ocorreu uma queda na acurácia do modelo, ano em que a pandemia do COVID-19 impactou as vendas consideravelmente, como já discutido anteriormente.

Após a conclusão da etapa de treino do modelo, em um segundo momento, foi feito a etapa de validação *out-of-time*, utilizando os dados que não foram utilizados na construção do modelo (de janeiro a dezembro de 2023). A Tabela 3, apresenta os valores mensais do volume de venda previstos para o ano de 2023, juntamente com os limites inferiores e superiores do intervalo de confiança de 95%.

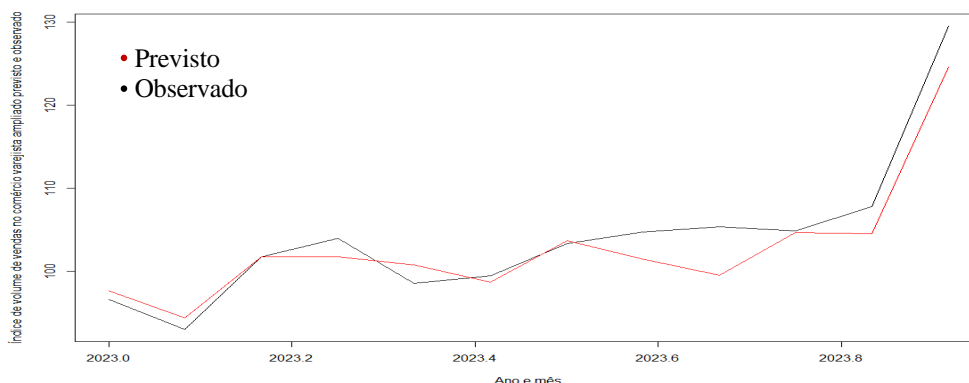
Tabela 4 – Valores mensais previstos pelo modelo e valores observados no ano de 2023 e os limites de confiança de 95%.

Mês/23	Valor previsto	Valor observado	Limite inferior do intervalo de confiança de 95%	Limite superior do intervalo de confiança de 95%
1	97,654	96,584	91,458	103,849
2	94,361	93,021	87,486	101,236
3	101,743	101,723	93,443	110,042
4	101,72	103,947	92,599	110,841
5	100,751	98,53	90,696	110,806
6	98,713	99,427	87,892	109,534
7	103,659	103,351	92,079	115,238
8	101,497	104,716	89,227	113,766
9	99,521	105,358	86,588	112,455
10	104,655	104,849	91,095	118,214
11	104,543	107,755	90,382	118,703
12	124,589	129,515	109,853	139,325

Fonte: Autoria Própria (2024).

Observa-se que todos os valores previstos estavam dentro dos limites de intervalo de confiança de 95%, afirmando e consolidando que o modelo SARIMA (1 1 0) (1 1 0)<sub>12</sub> possuiu uma boa acurácia, retornando valores previstos próximos aos reais. A Figura 7 apresenta graficamente os valores mensais previstos (na cor vermelha) e observados (na cor preta) para o ano de 2023 (*out-of-time*), concluindo a etapa de validação do modelo.

Figura 7 – Valores mensais previstos x observados do volume de vendas no ano de 2023.



Fonte: Autoria Própria (2024).

Foi notável que o modelo previu uma queda no volume de vendas no mês de fevereiro, em que o mesmo foi o com menor valor mensal, além de projetar um aumento elevado nas vendas varejistas no último trimestre do ano, em especial o mês de dezembro, sendo o mês com o maior valor do volume de vendas de supermercados e hipermercados brasileiros, podendo conectar essas informações com as explicações feitas na análise descritiva.

## 6 CONTRIBUIÇÃO TECNOLÓGICA-SOCIAL

O estudo realizado trouxe a compreensão e a análise de comportamento, tendências, padrões e sazonalidades do volume de vendas do setor varejista de supermercados e hipermercados entre os anos de 2018 e 2023, com o objetivo de modelar e prever os valores do volume de vendas. Para isso foi utilizado o modelo ARIMA, no qual após análises e mudanças de parâmetros de modelo, chegou-se na modelagem final SARIMA (1 1 0) (1 1 0)<sub>12</sub>.

Com os resultados obtidos no presente estudo, verificou-se que o modelo retornou valores previstos muito próximo aos valores reais e, todos os valores ajustados dentro do período de validação, estavam dentro dos limites do intervalo de confiança de 95%. Tais previsões podem ser subsídios para varejistas direcionarem as melhores estratégias operacionais e financeiras, a fim de maximizar seu volume de vendas, tendo como consequência positiva, um aumento significativo de receitas e lucros.

Além disso, o presente estudo trouxe a possibilidade de contribuir com os donos de supermercados e hipermercados brasileiros a terem um melhor controle de estoque, conseguindo antecipar suas necessidades e demandas de seus clientes, ajudando diretamente em pautas de desenvolvimento sustentáveis, citando como exemplo a ODS 12 de consumo e produção responsáveis (Objetivos de Desenvolvimento Sustentável da ONU), em que a presente pesquisa feita conseguia trazer como consequência positiva

uma maior eficiência e aproveitamento de alimentos e produtos perecíveis, combatendo os desperdícios de insumos, alimentos e produtos de grande importância para a população brasileira, já que a modelagem realizada teve como escopo supermercados e hipermercados brasileiros, os quais englobaram os principais insumos de necessidade básica da população.

## 7 REFERÊNCIAS

DA VEIGA, Cassia; TORTATO, Ubiratã. Previsão de Demanda no Varejo de Alimentos: Uma Comparação entre os Modelos Holt-Winters e ARIMA. **Transações WSEAS sobre Negócios e Economia**. 1. ed., 2014.

DOLLE, Camilla; RODRIGUES, Mayara; MOURA, Natalia. **O setor do varejo supermercadista**. São Paulo, 2023. Disponível em: <https://conteudos.xpi.com.br/renda-fixa/relatorios/o-setor-de-varejo-supermercadista/>. Acesso em: 23 out. 2023.

ECOMMERCE BRASIL. **Estudo revela o melhor e o pior mês do ano para o varejo brasileiro nos últimos 14 anos**. São Paulo, 2014 Disponível em: <https://www.ecommercebrasil.com.br/noticias/estudo-revela-o-melhor-e-o-pior-mes-ano-para-o-varejo-brasileiro-nos-ultimos-14-anos>. Acesso em: 12 abr. 2024.

FRANKENTHAL, R. **Qual é o impacto do tempo no comportamento do consumidor**. Rio de Janeiro, 2017. Disponível em: <https://mindminers.com/blog/qual-o-impacto-do-tempo-no-comportamento-do-consumidor/>. Acesso em: 5 abr. 2024.

GANDRA, Alana. **Chuvas de fevereiro provocam perdas de R\$ 203 milhões no Sudeste**. Disponível em: <https://agenciabrasil.ebc.com.br/economia/noticia/2020-02/chuvas-de-fevereiro-provocam-perdas-de-r-203-milhoes-no-sudeste>. Acesso em: 5 abr. 2024.

GRUS, J. **Ciência de Dados do zero: Princípios Fundamentais com Python**. 2. ed. Sebastopol, CA, USA: O'Reilly Media, 2019.

HILL, Carter; GRIFFITHS, William, JUDGE, George. **Econometria**. 4.ed. Rio de Janeiro: Editora Saraiva, 2008, 471p. Disponível em: <https://www.estantevirtual.com.br/livros/carter-hill-william-griffiths-george-judge/econometria/3335690585>. Acesso em: 24 out. 2023.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **De 2010 a 2022, população brasileira cresce 6,5% e chega a 203,1 milhões**. São Paulo, 2023. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37237-de-2010-a-2022-populacao-brasileira-cresce-6-5-e-chega-a-203-1-milhoes>. Acesso em: 27 mar. 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **PIB cresce 2,9% em 2023 e fecha o ano em R\$ 10,9 trilhões. São Paulo, 2024**. Disponível em: [https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/39303-pib-cresce-2-9-em-2023-e-fecha-o-ano-em-r-10-9-trilhoes#:~:text=Em%202023%2C%20o%20PIB%20\(Produto,Servi%C3%A7os%20\(2%2C4%25\).&text=O%20PIB%20totalizou%20R%24%2010%2C9%20trilh%C3%B5es%20em%202023](https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/39303-pib-cresce-2-9-em-2023-e-fecha-o-ano-em-r-10-9-trilhoes#:~:text=Em%202023%2C%20o%20PIB%20(Produto,Servi%C3%A7os%20(2%2C4%25).&text=O%20PIB%20totalizou%20R%24%2010%2C9%20trilh%C3%B5es%20em%202023). Acesso em: 10 jun. 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Tabela 8883 - Índice e variação da receita nominal e do volume de vendas no comércio varejista ampliado, por atividades.** São Paulo, 2024. Disponível em: <https://sidra.ibge.gov.br/tabela/8883>. Acesso em: 20 mar. 2023.

INSTITUTO DE ASTRONOMIA, GEOFÍSICA E CIÊNCIAS ATMOSFÉRICAS DA USP (IAG). **Climograma de São Paulo.** São Paulo, 2014 Disponível em: <<http://www.estacao.iag.usp.br/seasons/index.php>>. Acesso em: 29 abr. 2024.

JENKINS, B. **Time Series Analysis.** 1 ed. Londres, England: Prentice-Hall, 1991.

MORENTTIN, Pedro ; TOLOI, Clélia. **Análise de Séries Temporais: Modelos lineares univariados.** 3.ed. São Paulo: Blucher, 2018, 474 p. Disponível em: <[https://books.google.com.br/books/about/An%C3%A1lise\\_de\\_s%C3%A9ries\\_temporais.html?id=UwC5DwAAQBAJ&redir\\_esc=y](https://books.google.com.br/books/about/An%C3%A1lise_de_s%C3%A9ries_temporais.html?id=UwC5DwAAQBAJ&redir_esc=y)>. Acesso em: 15 set. 2023.

NOVOVAREJO. **O papel do varejo na economia.** Rio de Janeiro, 2023. Disponível em: <https://novovarejo.com.br/o-papel-do-varejo-na-economia/>. Acesso em: 27 nov.2023.

SATO, Renato Cesar. Gerenciamento de doenças utilizando séries temporais com o modelo ARIMA. **SciELO.**v.1, p. 1-4, 2013. Disponível em: <<https://www.scielo.br/j/eins/a/XqhY6VzmzbyZ9yRXLyZjCct/?format=pdf&lang=pt>>. Acesso em: 25 mai. 2023.

SEBRAE MINAS. **Varejo no Brasil: cenário atual, futuro e oportunidades.** Minas Gerais, 2023. Disponível em: <https://sebraeplay.com.br/content/varejo-no-brasil-cenario-atual-futuro-e-oportunidades>. Acesso em: 25 out. 2023.

SOCIEDADE BRASILEIRA DE VAREJO E CONSUMO. **Ranking 300 maiores empresas do varejo brasileiro.** São Paulo, 2023. Disponível em: [https://sbvc.com.br/wp-content/uploads/2023/08/RankingSBVC2023\\_VersaoDigital.pdf](https://sbvc.com.br/wp-content/uploads/2023/08/RankingSBVC2023_VersaoDigital.pdf). Acesso em: 10 jun. 2024.

VIGA, Rodrigo; MOREIRA, Camila. **Vendas no varejo do Brasil têm maior queda do ano em dezembro, mas fecham 2023 com ganhos.** Disponível em: <http://economia.uol.com.br/noticias/reuters/2024/02/07/vendas-no-varejo-do-brasil-recuam-13-em-dezembro-diz-ibge.htm>. Acesso em: 5 abr.