

ANÁLISE DE DADOS DE VENDAS UTILIZANDO SÉRIES TEMPORAIS, ALGORITMO APRIORI E PROPHET

SALES DATA ANALYSIS USING TIME SERIES, APRIORI ALGORITHM, AND PROPHET

ANÁLISIS DE DATOS DE VENTAS UTILIZANDO SERIES TEMPORALES, ALGORITMO
APRIORI Y PROPHET

Vanessa Velasco Cazeiro¹
André Lucio de Oliveira²

RESUMO: Esse artigo apresenta um estudo em dados de venda de uma loja do setor de informática, utilizando três ferramentas: (1) aplicação do algoritmo Apriori³ para identificar regras de associação e padrões de consumo; (2) análise de séries temporais para compreender como o dado se comporta ao longo do tempo, em diversas escalas; e (3) aplicação do algoritmo Prophet⁴ para fazer a decomposição da série histórica de vendas e realizar uma predição. A aplicação dessas ferramentas de forma conjunta permitiu uma análise mais aprofundada sobre as vendas, incluindo a detecção de comportamentos de consumo e a identificação de padrões temporais ao longo do tempo. Os resultados obtidos fornecem informações valiosas para a tomada de decisões estratégicas da empresa.

2053

Palavras-chave: Apriori. Prophet. Regra de associação.

ABSTRACT: This article presents a study on sales data from a computer store, using three tools: (1) application of the Apriori algorithm to identify association rules and consumption patterns; (2) time series analysis to understand how the data behaves over time, at different scales; and (3) application of the Prophet algorithm to decompose the historical sales series and make a prediction. The application of these tools together allowed for a more in-depth analysis of sales, including the detection of consumption behaviors and the identification of temporal patterns over time. The results provide valuable information for the company's strategic decision-making.

Keywords: Apriori. Prophet. Association rule.

¹ Bacharel em Engenharia da Computação. Universidade Veiga de Almeida. Pós-graduação em Banco de Dados / MBA em Big Data. Universidade UniBF.

² Mestrado - COPPE/PESC - UFRJ.

³ Biblioteca em Python: <https://github.com/codedheartinside/php-apriori>. Acesso em 21/11/2022.

⁴ Biblioteca em Python: <https://pypi.org/project/prophet/>. Acesso em 21/11/2022.

RESUMEN: Este artículo presenta un estudio sobre los datos de ventas de una tienda de informática, utilizando tres herramientas: (1) aplicación del algoritmo Apriori para identificar reglas de asociación y patrones de consumo; (2) análisis de series temporales para comprender cómo se comportan los datos a lo largo del tiempo, en diferentes escalas; y (3) aplicación del algoritmo Prophet para descomponer la serie histórica de ventas y realizar una predicción. La aplicación de estas herramientas en conjunto permitió un análisis más profundo de las ventas, incluyendo la detección de comportamientos de consumo y la identificación de patrones temporales a lo largo del tiempo. Los resultados proporcionan información valiosa para la toma de decisiones estratégicas de la empresa.

Palabras clave: A priori. Prophet. Regla de asociación.

1 INTRODUÇÃO

Com o avanço da tecnologia e das ferramentas computacionais, a organização e o gerenciamento dos dados se tornaram componentes essenciais na sociedade moderna (ELMASRI; NAVATHE, 2005). Esse advento possibilitou o tratamento de enormes quantidades de dados, de forma que as empresas podem ser auxiliadas nas tomadas de decisão. Essas informações são comumente armazenadas em bancos de dados, e podem ser de diversas naturezas, como registros históricos de vendas. Entretanto, as boas práticas de gerenciamento, tratamento, condicionamento e processamento dos dados para convertê-los em conhecimento não são uma realidade tão comum em pequenas empresas, devido à falta de recursos humanos dedicados a essa tarefa específica (AVIS, 2021).

Considerando o interesse geral das empresas em maximizar suas vendas, diversos autores dissertam sobre o papel dos dados como ferramenta para compreender a necessidade e o perfil do cliente. Por exemplo, Matheus Ribeiro defende a ideia de que as lojas têm clientes que variam ao longo do tempo, e que é responsabilidade dos administradores elaborar estratégias em decorrência desses perfis (RIBEIRO, 2014). Kotler e Keller (2005) afirmam que os dados são essenciais para a definição das estratégias de marketing, e correspondem às informações mais completas possíveis coletadas sobre os clientes. O conhecimento aprofundado do público-alvo sofisticou os métodos de publicidade, comercialização e tratamento personalizado para cada cliente (KOTLER; KELLER, 2005).

Como ressaltado pelo ex-diretor de marketing da Amazon, quando você entra em uma loja de departamentos como a Macy's, ninguém sabe quem você é. Entretanto, quando você entra na Amazon, é cumprimentado pelo nome, recebe sugestões personalizadas com base em suas compras

anteriores e uma série de críticas honestas de outros clientes a respeito dos produtos sugeridos. Ao sair do site, também lhe pedem permissão para enviar e-mails com promoções (KOTLER; KELLER, 2005).

Diante da relevância que o registro, tratamento, condicionamento e processamento dos dados têm nas estratégias de vendas, foi motivado um estudo detalhado acerca das vendas de uma loja do setor de informática. A loja gentilmente cedeu um dataset correspondente a uma série histórica de pedidos de venda ao longo de quase cinco anos. Seguindo uma ordenação lógica, os dados foram tratados e condicionados para servir de base para uma série de análises que visam compreender padrões e comportamentos de vendas e clientes. Em outras palavras, o cerne deste trabalho é realizar uma revisão e aplicação geral de diversos métodos, utilizando a linguagem de programação Python, que abarque a conversão de um dado bruto, inicialmente extraído do PostgreSQL⁵ - sistema gerenciador de banco de dados - para análises que constatem, por exemplo, o comportamento de compra em relação ao perfil do consumidor da referida loja. Dessa forma, os proprietários podem ter ciência do fluxo de caixa, dos produtos que são comprados juntos com frequência, e assim, os administradores da loja poderão, então, basear-se nessas informações para ajustar o mix de produtos, promoções e até mesmo treinar a equipe de vendas com base no comportamento e nas preferências do cliente. Além disso, a análise dos dados poderá indicar a necessidade de melhorias em áreas específicas da loja, como aprimoramento do atendimento ao cliente, agilidade na entrega dos produtos ou investimento em tecnologia para tornar a experiência de compra mais eficiente.

2055

Nesse sentido, é fundamental destacar que o tratamento e análise de dados se tornaram uma estratégia indispensável para as empresas que desejam se manter competitivas em um mercado cada vez mais acirrado. Por meio da coleta e análise de informações sobre seus clientes, é possível identificar oportunidades de negócio, antecipar tendências e desenvolver produtos e serviços mais adequados às necessidades do público-alvo.

Portanto, investir em tecnologias e profissionais capacitados para gerenciar e analisar dados pode ser um grande diferencial para as empresas que buscam se destacar no mercado. E este trabalho de análise de dados realizado na loja de informática pode ser um grande exemplo para outras empresas que desejam aproveitar ao máximo as informações que possuem para melhorar seus negócios e fidelizar seus clientes.

⁵ Biblioteca: <https://www.postgresql.org/>. Acesso em 21/11/2022.

2 METODOLOGIA

Esse trabalho propõe explorar algumas das diversas ferramentas e boas práticas disponíveis no contexto da análise de dados, séries temporais e aplicação do algoritmo Apriori. Com isso, neste capítulo serão apresentadas as ferramentas utilizadas. Em que, o dado que será estudado foi gentilmente cedido por uma loja do setor de informática, correspondente aos registros de vendas entre 2018 e 2022.

2.1 Séries temporais e Prophet

O Prophet é uma biblioteca desenvolvida pelo *Facebook's Core Data Science team*⁶. Sua proposta é realizar a previsão a partir de séries temporais e a decomposição em suas componentes.

Nesse contexto, serão criadas as matrizes para visualização e implementação da referida biblioteca. Os dados compreendem um período de 26/05/2018 até 12/09/2022, totalizando 4 anos, 3 meses e 20 dias de registros. Inicialmente, o *input* do dado será tratado a fim de se adequar no formato específico em que o Prophet exige. Dessa forma, os dados temporais têm que estar necessariamente no formato *datetime* em um Pandas Data Frame⁷, como YYYY-MM-DD ou YYYY-MM-DD HH:MM:SS, em uma coluna denominada *ds*. Já os dados quantitativos, ou seja, o valor que será previsto, deverão estar no em uma coluna chamada *y* e em um formato numérico (*integer* ou *float*). A Tabela 8 exemplifica o dado de entrada, tratado nesse trabalho, que o Prophet admite.

2056

Tabela 8: Dado de Entrada do Prophet
Fonte: autor

| | <i>ds</i> | <i>y</i> |
|---|---------------------|----------|
| 0 | 2019-01-02 12:59:40 | 210.0 |
| 1 | 2019-01-02 13:04:40 | 0.0 |
| 2 | 2019-01-02 13:09:40 | 0.0 |
| 3 | 2019-01-02 13:14:40 | 0.0 |
| 4 | 2019-01-02 13:19:40 | 0.0 |

⁶ Site oficial: <https://research.facebook.com/research-areas/data-science/>. Acesso em 21/11/2022.

⁷ Biblioteca: <https://pandas.pydata.org/>. Acesso em 21/11/2022.

O Prophet é implementado em dois conjuntos de dados de série temporal: treinamento e validação. Os dados sazonais são preferidos e, portanto, os dados de 2018 foram excluídos e os meses de 2022 foram usados para validação. As lacunas nos horários faltantes foram preenchidas e o modelo foi treinado e previsto com base em 1587800 registros históricos, com os últimos 9 meses contendo 2838 registros.

Além disso, o Prophet tem diversos parâmetros⁸ que podem ser personalizados e informados antes da etapa de treinamento. Assim, optou-se por utilizar alguns componentes, nomeadamente:

- **Holiday Effects:** parâmetro utilizado para levar em consideração os feriados que não sejam americanos. Essa informação é importante para que o modelo entenda o calendário relacionado ao dado;
- **Add_seasonality:** atribuído quando um dado já possui uma sazonalidade específica. No caso, o presente trabalho considerou uma sazonalidade diária com o período de um dia;
- **N_changepoints:** esse parâmetro identifica mudanças abruptas na série temporal. Por padrão, o Prophet utiliza 25 marcações para o dado, a começar a partir da primeira linha deste. Foi optado por utilizar 60 *change points*. A Figura x representa as devidas identificações do Prophet;
- **Changepoint_prior_scale:** parâmetro utilizado para modular a flexibilidade dos pontos de mudanças;
- **Seasonality_prior_scale:** utilizado para modular a força da sazonalidade.

2057

Para os dois últimos parâmetros, foi utilizado a validação cruzada para determinar os melhores valores. Além disso, houve a necessidade do uso da biblioteca Dask⁹, que oferece suporte a computação paralela em Python. Dessarte, os parâmetros utilizados para os testes de validação cruzada, foram:

- **Changepoint_prior_scale:** [0.001, 0.01, 0.1, 0.5];
- **Seasonality_prior_scale:** [0.01, 0.1, 1.0, 10.0].

O teste apresentou como resultado para os parâmetros, os valores 0.001 e 0.01, respectivamente.

⁸ Documentação disponível em: https://facebook.github.io/prophet/docs/saturating_forecasts.html. Acesso em 20/11/2022.

⁹ Biblioteca: <https://www.dask.org/get-started>. Acesso em 26/11/2022.

2.2 Algoritmo Apriori

O algoritmo Apriori tem como objetivo determinar regras de associação entre itens. Nesse contexto, como o nosso *dataset* é sobre vendas de produtos de informática, cada item corresponde a um único e distinto produto.

Foi utilizada a biblioteca *mxltend*¹⁰ para o cálculo do suporte, confiança e lift. O suporte mínimo considerado foi de 0.03%, o menor possível, sendo somente limitado pela memória RAM da máquina utilizada em questão (8Gb). Já a confiança mínima considerada foi de 10% para o estabelecimento das regras de associação. Dessa forma, foram calculadas um total de 2198 regras de associação.

3 RESUSTADOS

3.1 Séries temporais e prophet

Para entender como ocorre a distribuição do dado ao longo do tempo, inúmeras séries temporais, com escalas diferentes, foram realizadas. Preliminarmente, as séries das vendas diárias foram plotadas, nomeadamente:

1. Série Histórica de Vendas Diárias (Figura 15)
 2. Série Histórica de Vendas Diárias – Logarítmica (Figura 16)
 3. Comparação Mensal das Vendas Diárias (Figura 17)
 4. Comparação Mensal das Vendas Diárias – Logarítmica (Figura 18)
 5. Média Semanal das Vendas Diárias (Figura 19)
 6. Média Mensal das Vendas Diárias (Figura 20)
 7. Média Diária das Vendas para Cada Semana do Mês (Figura 21)
 8. Média Diária das Vendas para Cada Dia do Mês (Figura 22)
 9. Média Diária das Vendas para Cada Dia da Semana (Figura 23)
- Adicionalmente às séries de vendas, foram criados gráficos de frequência de clientes por dia:
10. Frequência Diária de Clientes - Série Histórica (Figura 24)
 11. Frequência Diária de Clientes ao Longo do Ano (Figura 25)

¹⁰ Biblioteca: https://rasbt.github.io/mlxtend/api_subpackages/mlxtend.frequent_patterns/. Acesso em 21/11/2022.

12. Frequência Diária de Clientes ao Longo do Ano - Média Móvel Centrada 7 Dias (Figura 26)
13. Distribuição de Frequências da Quantidade de Clientes Diários (Figura 27)
14. Frequência Diária Média de Clientes ao Longo da Semana (Figura 28)
15. Frequência Diária Média de Clientes ao Longo da Semana (Figura 29)

O gráfico 1 da Figura 15 mostra as vendas diárias da empresa de informática com presença de outliers e comportamento estável ao longo do tempo. O gráfico em escala logarítmica da Figura 16 evidencia uma possível sazonalidade ou tendência de longo prazo. Destaca-se a venda do dia 03/02/2021 no valor de 46 mil reais. A loja teve seu menor faturamento em 2018, período de adaptação e conquista de espaço no mercado.



Figura 14: Série Histórica de Vendas Diárias
Fonte: Autor



Figura 15: Série Histórica de Vendas Diárias - Logarítmica
Fonte: Autor

Os gráficos 3, 4, 5 e 6 comparam as vendas nos anos de 2018, 2019, 2020, 2021 e 2022 em diferentes escalas (diário, diário logarítmico, semanal e mensal). O gráfico logarítmico 4 foi necessário para facilitar a visualização. Os gráficos diários são ruidosos e não mostram correlações

claras entre anos e sazonalidades, enquanto os gráficos semanais e mensais são inconclusivos, exceto por um pequeno aumento nas vendas em janeiro, fevereiro e dezembro de 2021 e 2022. Os gráficos 4, 5 e 6 mostram que em 2018 houve um baixo faturamento devido à abertura da loja, período de adaptação e conquista de mercado.

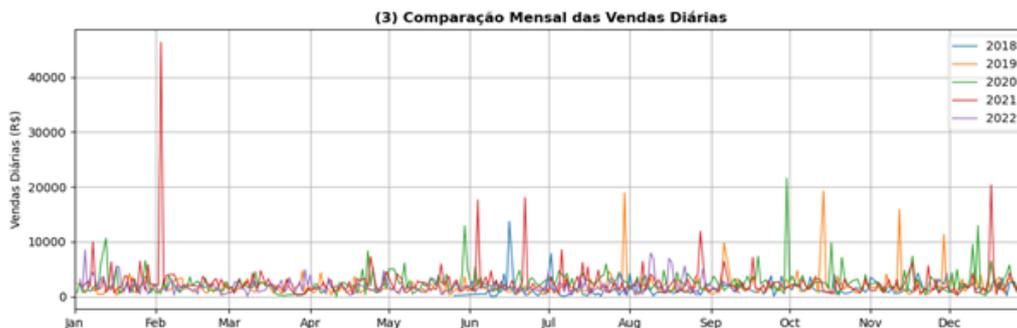


Figura 16: Comparação Mensal das Vendas Diárias

Fonte: Autor

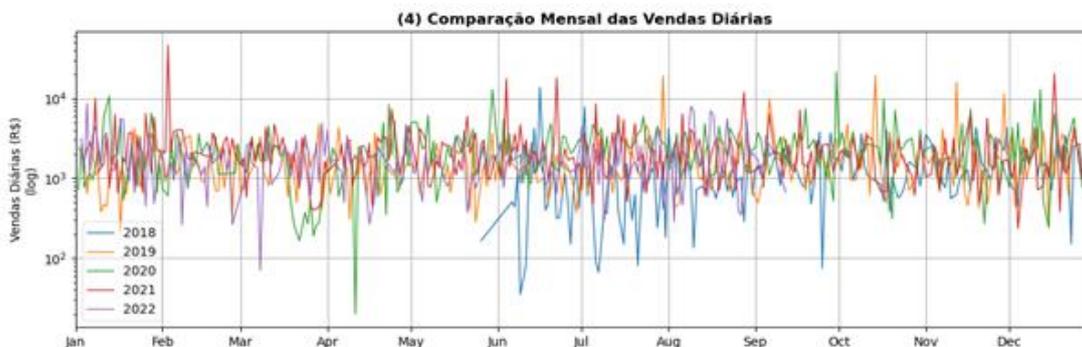


Figura 17: Comparação das Vendas Diárias

Fonte: Autor

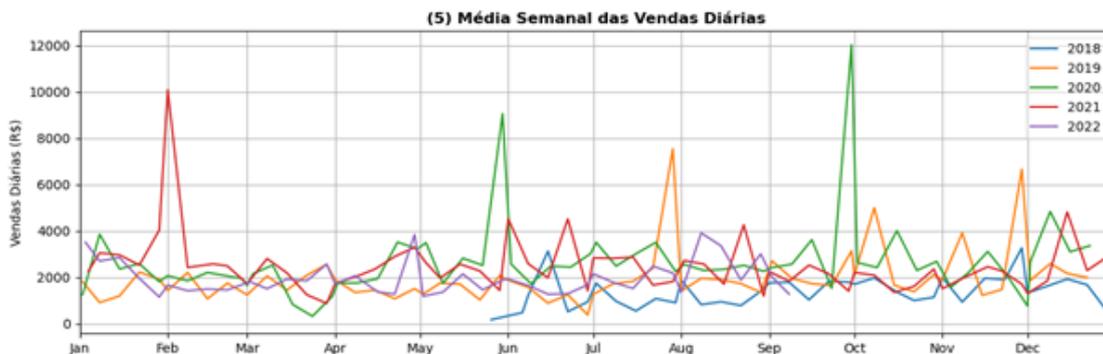


Figura 18: Média Semanal das Vendas Diárias

Fonte: Autor

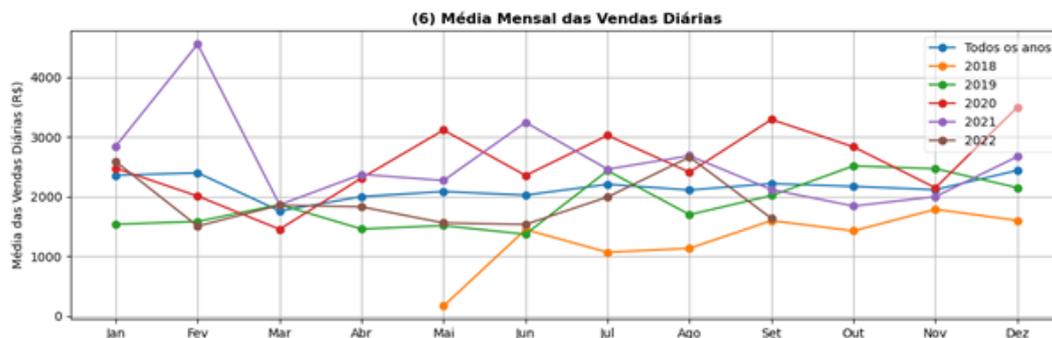


Figura 19: Média Mensal das Vendas Diárias

Fonte: Autor

Os gráficos 7, 8 e 9 mostram o comportamento dos clientes a curto prazo. Os gráficos 7 e 8 mostram a influência da semana e do dia do mês no fluxo de caixa da empresa, com as vendas mais fortes na quinta semana do mês e a mais fraca na quarta semana. O gráfico 9 mostra como as vendas ocorrem ao longo da semana, com a quinta-feira sendo o dia menos movimentado e a sexta-feira tendo as vendas de maior valor.

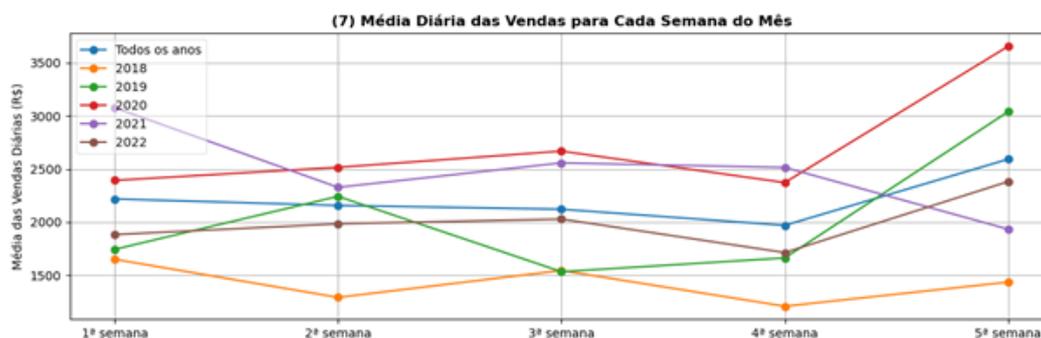


Figura 20: Média Diária das Vendas para Cada Semana do Mês

Fonte: Autor

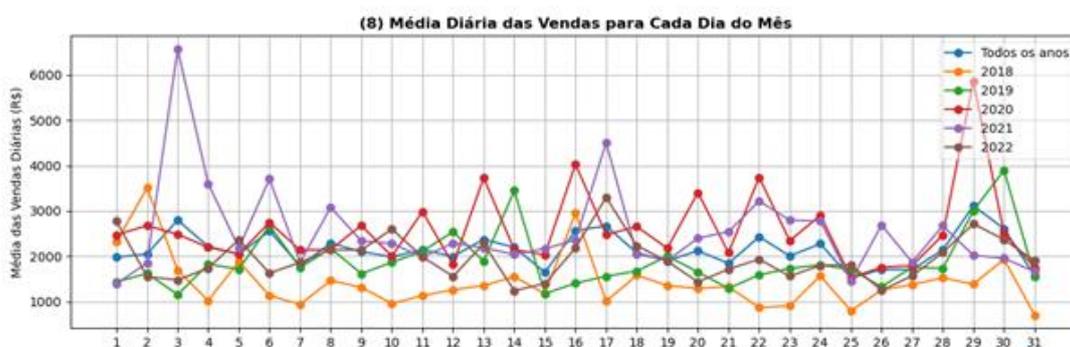


Figura 21: Média Diária das Vendas para Cada Dia do Mês

Fonte: Autor

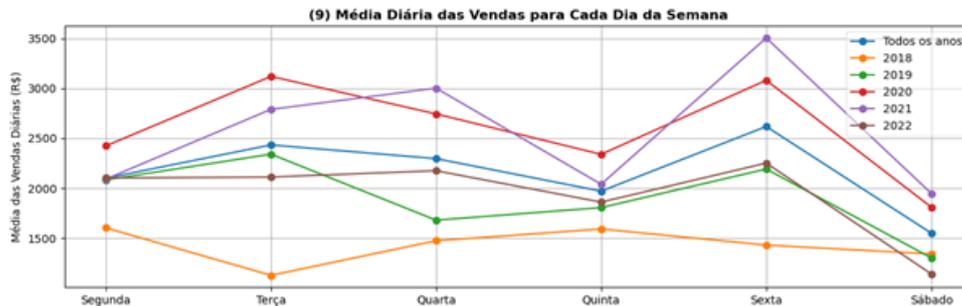


Figura 22: Média das Vendas para Cada Dia da Semana

Fonte: Autor

Os gráficos 10, 11 e 12 mostram a distribuição da quantidade diária de clientes da loja, tanto historicamente quanto ao longo do ano. Foi aplicada uma média móvel nos dados do gráfico 11 para torná-lo mais claro, e é possível identificar comportamentos semelhantes nos anos de 2019 e 2022, e 2020 e 2021. O gráfico 13 mostra a distribuição em frequência da quantidade de clientes que compram com a empresa, e todas as distribuições são semelhantes no intervalo de ocorrência, com desvio padrão alto, o que indica a ruídosidade do gráfico 11.



Figura 23: Frequência Diária de Clientes – Série Histórica

Fonte: Autor

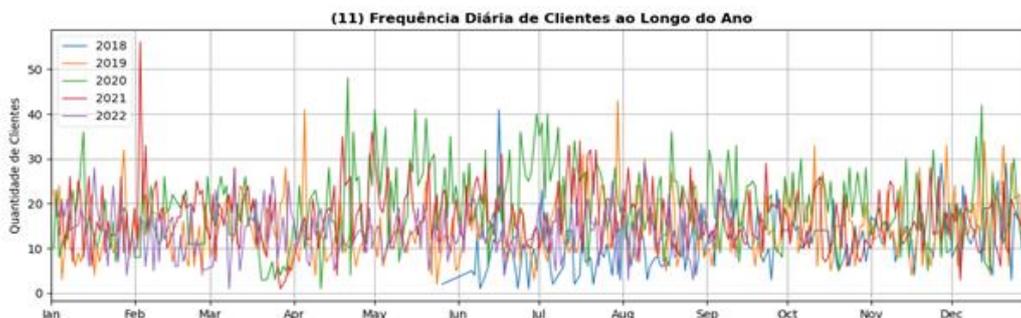


Figura 24: Frequência Diária de Clientes ao Longo do Ano

Fonte: Autor

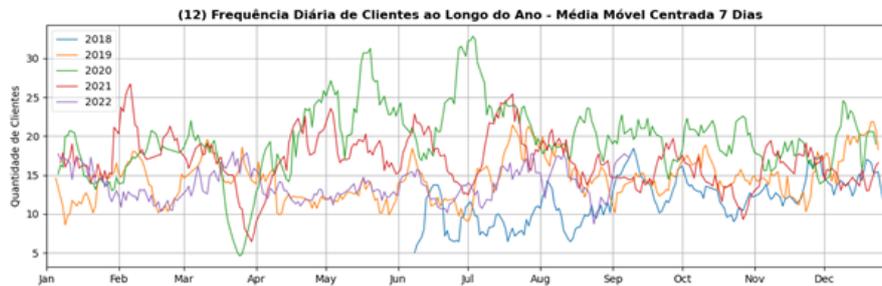


Figura 25: Frequência Diária de Clientes ao Longo do Ano – Média Móvel Centrada 7 Dias

Fonte: Autor

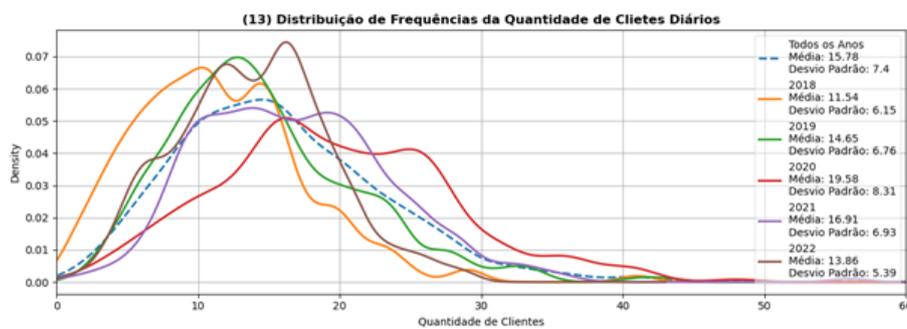


Figura 26: Distribuição de Frequência de Quantidade de Clientes Diários

Fonte: Autor

Os gráficos 14 e 15 mostram a frequência média diária de clientes em curtos intervalos de tempo, dias da semana e semanas do mês, respectivamente. O gráfico 14 apresentou comportamento homogêneo ao longo da semana, com queda no sábado, enquanto o gráfico 15 evidencia que não há diferenças relevantes no fluxo de clientes ao longo das semanas do mês. Embora a frequência diária de clientes não varie muito, o gráfico 9 mostra uma diferença significativa nas vendas entre quinta e sexta-feira, indicando que os clientes compram produtos mais caros na sexta-feira e mais baratos na quinta-feira.

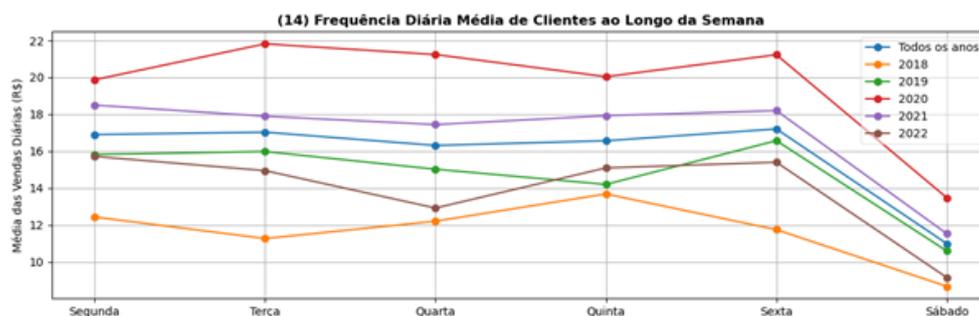


Figura 27: Frequência Diária Média de Clientes ao Longo da Semana

Fonte: Autor

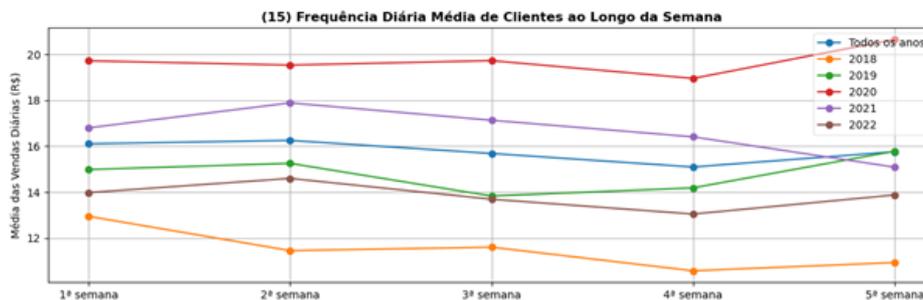


Figura 28: Frequência Diária Média de Clientes ao Longo da Semana

Fonte: Autor

3.2 Resultados do Prophet

Aplicando o Prophet com os dados da loja do setor de informática já tratados, esse modelo apresentou os seguintes resultados:

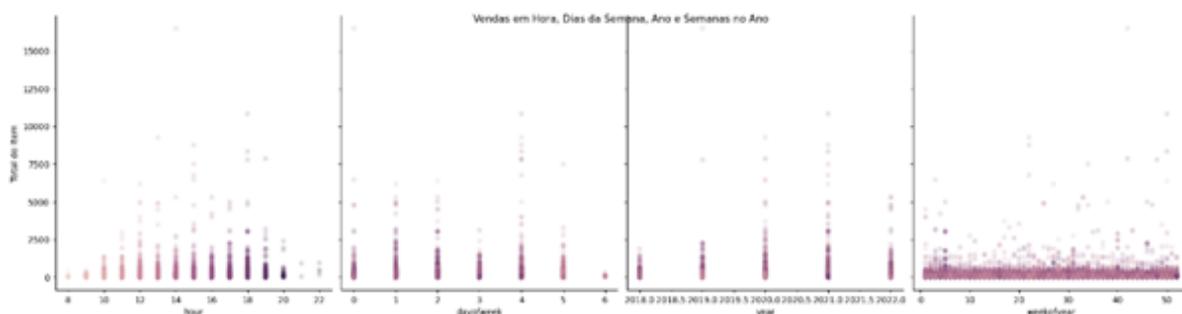


Figura 29: Gráficos de Tendências

Fonte: Autor

A Figura 29 apresenta gráficos gerados a partir dos dados de entrada do Prophet, com o objetivo de compreender as tendências dos dados e avaliar a precisão dos dados de saída. A figura contém quatro gráficos, mostrando as vendas por hora, por dia da semana, por ano e por semana ao longo dos anos. Observou-se que os horários de pico de vendas estão entre 12h e 18h, enquanto o horário de funcionamento da loja é das 10h às 19h/20h durante a semana e das 10h às 15h aos sábados. As maiores vendas foram registradas nas terças, quartas e sextas-feiras. Houve um crescimento nas vendas entre 2018 e 2020, mas em 2021 e 2022 as vendas aparentemente foram similares ou caíram. A previsão do Prophet para os últimos 9 meses do dataset foi exemplificada na Tabela 9, mas comparando-se com os dados reais da Tabela 10, pode-se observar que os valores previstos diferem dos valores reais.

Tabela 9: Atributos da Variável de Saída Forecast
Fonte: Autor

| | ds | yhat | yhat_lower | yhat_upper |
|---|---------------------|-----------|------------|------------|
| 0 | 2022-01-03 11:29:00 | 10.955219 | -84.324241 | 114.083036 |
| 1 | 2022-01-03 12:45:27 | 14.399817 | -84.520816 | 114.170417 |
| 2 | 2022-01-03 13:15:24 | 15.324874 | -89.888438 | 119.846130 |
| 3 | 2022-01-03 13:21:28 | 15.485131 | -78.896398 | 118.137518 |
| 4 | 2022-01-03 13:21:33 | 15.487271 | -85.499275 | 122.482345 |
| 5 | 2022-01-03 14:46:20 | 16.887620 | -83.059648 | 117.977212 |
| 6 | 2022-01-03 15:24:35 | 17.050340 | -85.115617 | 119.435788 |
| 7 | 2022-01-03 15:24:44 | 17.050398 | -80.435793 | 132.755167 |
| 8 | 2022-01-03 15:26:56 | 17.050713 | -82.629802 | 120.433610 |
| 9 | 2022-01-03 15:28:29 | 17.050325 | -83.474085 | 113.986528 |

A Tabela 9 representa os principais atributos da variável de saída forecast, em que:

- Ds representa o elemento temporal;
- Yhat os dados previstos pelo modelo;
- yhat_lower o limite inferior;
- yhat_upper o limite superior.

Tabela 10: Atributos Reais da Loja de Informática
Fonte: Autor

| | ds | y |
|---|---------------------|-------|
| 0 | 2022-01-03 11:29:00 | 65.0 |
| 1 | 2022-01-03 12:45:27 | 28.0 |
| 2 | 2022-01-03 13:15:24 | 109.9 |
| 3 | 2022-01-03 13:21:28 | 59.9 |
| 4 | 2022-01-03 13:21:33 | 149.0 |
| 5 | 2022-01-03 14:46:20 | 239.9 |
| 6 | 2022-01-03 15:24:35 | 80.0 |
| 7 | 2022-01-03 15:24:44 | 55.0 |
| 8 | 2022-01-03 15:26:56 | 69.9 |
| 9 | 2022-01-03 15:28:29 | 180.0 |

Para melhores visualizações, a Figura 31 mostra a representação gráfica da predição realizada. Em suma, os pontos pretos são os dados reais, enquanto a linha azul representa a previsão. Já a faixa azul clara evidencia o intervalo de incerteza.

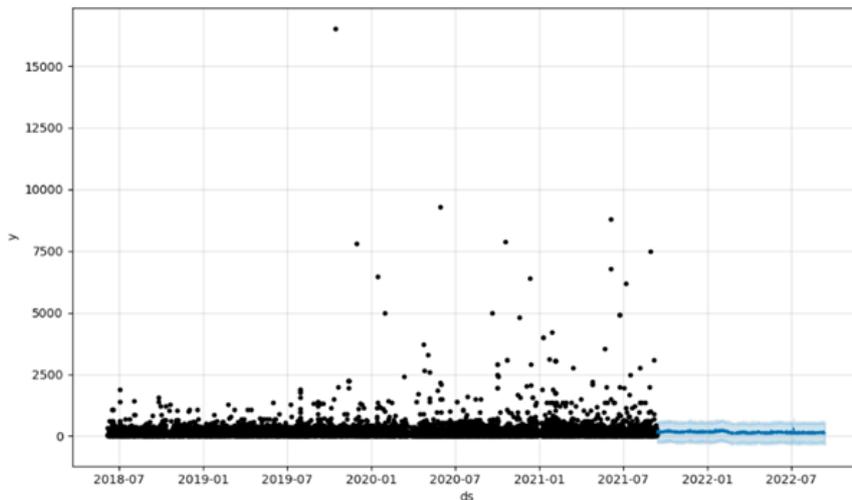


Figura 30: Representação da Predição do Prophet

Fonte: Autor

A figura 31 mostra que a predição segue o comportamento do dado, sem picos de venda aleatórios. As figuras 32 a 35 representam a tendência e sazonalidade anual, semanal e diária. O dado previsto tem uma tendência decrescente semelhante aos dados reais e a sazonalidade apresenta comportamentos válidos. A sazonalidade anual destaca um comportamento elevado em fevereiro, o que não foi evidenciado no último gráfico da Figura 29.

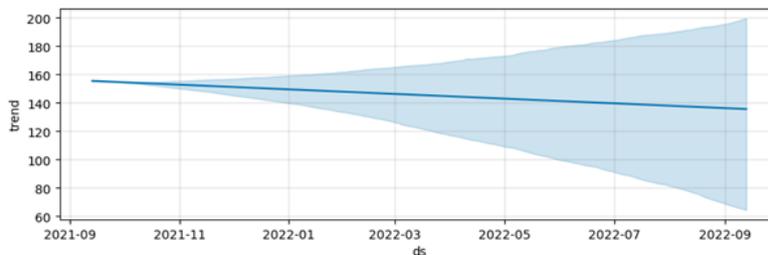


Figura 31: Tendência Inferida pelo Prophet

Fonte: Autor

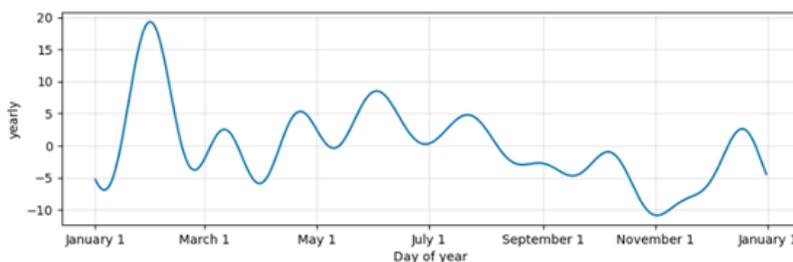


Figura 32: Sazonalidade Anual Inferida Pelo Prophet

Fonte: Autor

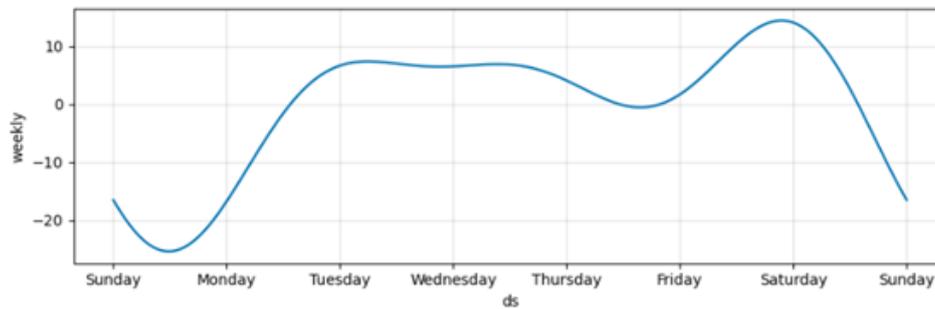


Figura 33: Sazonalidade Semanal Inferida Pelo Prophet

Fonte: Autor

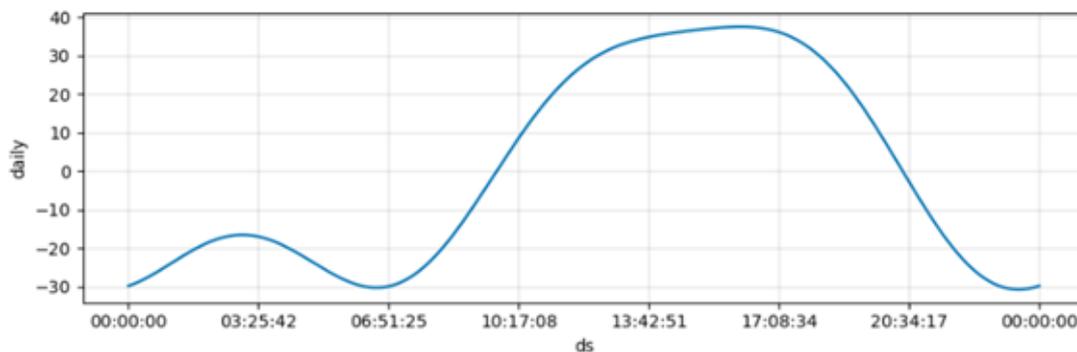


Figura 34: Sazonalidade Diária Inferida Pelo Prophet

Fonte: Autor

Por fim, a Figura 35 representa o resultado do cálculo, através do RSME, de validação aplicado na predição do Prophet. Em suma, o RSME utiliza o desvio padrão para determinar o intervalo de dispersão dos valores inferidos em relação ao dado real — dado reservado inicialmente para validação. Como é possível observar, o erro médio foi de 325.611. Ou seja, o modelo erra, em média, 325.611 para cima ou para baixo em relação aos valores de vendas.

Teste RSME: 325.611

Figura 35: RSME

Fonte: Autor

3.3 ALGORITMO APRIORI

Aplicando a ferramenta exposta na metodologia, será realizado a análise das regras de associação apresentadas pela aplicação do algoritmo Apriori. Assim, é possível estabelecer os 14 tipos de associações mais relevantes (Tabela 25).

Tabela 25: Resultado das Regras de Associação

Fonte: Autor

| Regra | Perfis de compra | Lift Médio | Confiança Média | Suporte Médio |
|-------|--|------------|-----------------|---------------|
| 1 | Montagem de computadores | 1068.28 | 66.29% | 82.76% |
| 2 | Headfones e adaptadores P2-P3 | 168.76 | 26.68% | 0.08% |
| 3 | Capa e película nova para seu celular | 71.9 | 35.16% | 0.16% |
| 4 | Base e cabo para carregador de celular | 53.84 | 19.32% | 1.49% |
| 5 | Cartuchos de impressora nas cores preta e colorida | 108.13 | 25.07% | 0.11% |
| 6 | Pilhas alcalinas AAA e AA | 98.58 | 33.58% | 0.22% |
| 7 | Upgrade HD SSD via CADDY ou CASE | 59.25 | 21.74% | 0.67% |
| 8 | Teclado <i>bluetooth</i> e pilhas respectivas | 86.25 | 23.94% | 0.11% |
| 9 | Ring Light e tripé | 51.96 | 38.58% | 0.6% |
| 10 | Tela nova para o celular e a respectiva película | 18.49 | 16.66% | 0.54% |
| 11 | Tela e bateria novas para o celular | 22.24 | 13.48% | 0.13% |
| 12 | Base do carregador e película para o celular | 5.64 | 30.19% | 0.11% |
| 13 | Limpa telas e película para o celular | 2,75 | 14.75% | 0.11% |
| 14 | Pilha alcalina AAA e película para celular | 2,67 | 14.29% | 0,04% |

2068

4 DISCUSSÕES

Em relação ao algoritmo Apriori, foram determinados 15 comportamentos de compra mais recorrentes (Tabela 1). Esses comportamentos foram encontrados reunindo uma ou mais regras de associação, de forma que generalizasse um perfil de consumo. Por exemplo, a montagem de computador contou com mais de 2000 regras de associação, todas relacionadas com o mesmo padrão de consumo: um freguês montando um computador.

Sob o ponto de vista da respectiva loja, essas regras têm a utilidade de auxiliar tanto a exposição dos produtos, quanto promoções e o oferecimento de outros produtos durante a compra de um

antecedente. Dessa forma, foi encontrado que é sempre conveniente, por exemplo, instruir o vendedor a oferecer capas de celular para clientes que queiram comprar películas, e vice e versa (regra 3 da Tabela 25). Na exposição dos produtos, por exemplo, é interessante colocar junto dos headfones os adaptadores P2-P3 (regra 2 da Tabela 25).

Em linhas gerais, todas as regras de associação encontradas refletem padrões reais de consumo. Nesse momento, cabe à referida loja realizar ações que considerem esses perfis para buscar a melhoria da oferta de produtos.

Além disso, foram plotadas diversas séries temporais durante essa pesquisa. Elas evidenciam a distribuição dos dados em inúmeras escalas, o que permitiu a investigação de padrões. Nessa etapa os resultados foram analisados qualitativamente, sob a perspectiva da interpretação gráfica. As principais conclusões foram:

- Os dados são extremamente ruidosos, apresentando grandes oscilações diárias (Figuras 24, 25, 26 e 27), tanto nas vendas quanto no fluxo de clientes (Figura 27). Isso significa que o montante de venda diário, por exemplo, é muito variável, o que dificulta a percepção dos lojistas. Nesse caso, trabalhar com médias móveis/semanais/mensais e escalas logarítmicas é recomendado, para melhorar a visualização das séries e seus padrões. Outra consequência negativa relacionada com a alta variação diária é a presença de outliers muito discrepantes. Como são valores que destoam em demasia, eles deslocam as médias de forma que possa induzir padrões irreais nos dados. Por exemplo, a Figura 19 apresenta picos que equivalem aos outliers da Figura 18;
- De uma forma generalista, não é possível detectar visualmente padrões muito claros e evidentes, pois a série histórica de vendas permanece relativamente constante ao longo do tempo. Somado ao agravante do elevado ruído, poucos e sutis padrões foram encontrados nessa metodologia de análise para esse set de dados;
- A quinta semana do mês (dias 29, 30 e 31) é a semana que mais vende (Figura 21). Entretanto, o fluxo de clientes é bem constante ao longo das semanas do mês (Figura 29). Logo, entende-se que é na quinta semana do mês que os clientes resolvem comprar produtos mais caros;
- As vendas mais caras ocorrem na sexta-feira, enquanto as de menor montante ocorrem na quinta-feira. Mas, como mencionado no tópico anterior, o fluxo de clientes não varia muito ao longo da semana. Esse fenômeno é interessante e abre espaço para especulações: talvez os clientes comprem mais na sexta para poderem aproveitar o lazer do final de semana;
- Uma possível reorganização das vitrines da loja, tendo como base os resultados expostos, poderia ser: (1) apresentar produtos práticos e com custo reduzido em alguns dias da semana, como na quinta-feira; (2) investir na exposição de produtos mais caros na sexta-feira e no sábado. Isso poderia despertar o interesse do consumidor para que eles entrem na loja, visto que a empresa estaria investindo no perfil de compra do cliente.

Complementarmente, agrupando a análise dos gráficos com os contextos externos, foi possível constatar que os dados de 2018 (Figuras 18, 28 e 29) apresentaram um comportamento de vendas reduzido. Esse desempenho se dá devido a data de inauguração da loja, que aconteceu nesse mesmo ano. Além disso, essa loja de informática não possui franquia, dependendo, portanto, de marketing próprio e do conhecimento das pessoas do bairro para sua propagação. Já o ano de 2019 é caracterizado pelo crescimento significativo em relação ao momento de sua abertura (Figura 20). Além disso, vale ressaltar que a pandemia da Covid-19, declaração realizada no dia 11 de março de 2020, impactou consideravelmente as vendas de 2020. A Figura 51 evidencia um aumento em relação à média mensal de vendas diárias desse ano. Esse comportamento ocorreu porque houve um investimento em vendas online e os proprietários fecharam as duas lojas que possuem, deixando apenas a loja que foi estudada neste trabalho para a realização de entregas nas residências. Ou seja, apesar do fechamento das lojas, o aumento das vendas pode ser explicado devido a concentração das duas lojas em uma e a abertura de vendas online. Em 2021, as lojas físicas foram abertas novamente, com isso, as vendas aumentaram, principalmente nos meses de janeiro, fevereiro e junho (Figura 20). Por fim, o ano de 2022 apresentou uma queda nos valores das vendas, o que pode ser explicado através da diminuição de compras online em relação aos anos de pandemia (2022 e 2021), visto que as pessoas voltaram a frequentar as ruas.

2070

Outra observação realizada, através do Prophet, é que a tendência analisada por esse modelo confirmou a constatação das séries temporais vistas em primeiro momento, em que o ano de 2022 está tendo seu fluxo reduzido. Ao passo que as sazonalidades apresentadas foram de grande valia, porque o Prophet utiliza a transformada de Fourier para converter uma função de tempo e sinal (valores das vendas) em uma função de potência e frequência. Com isso, é visto quais frequências o sinal tem e se apresentam uma potência forte ou fraca. Dessa forma, esse modelo apresentou novas informações e confirmou constatações anteriores, como:

- Na sazonalidade diária, os picos de vendas se encontram entre 12h e 18h;
- Na sazonalidade semanal, sábado é o dia em que a empresa possui maior fluxo de caixa.
- Na sazonalidade anual, fevereiro é mês com maior fluxo de caixa;
- Na sazonalidade da tendência, a inferência de 2022 tem uma curva decrescente de vendas.

Dessa forma, os proprietários, ao receber o feedback das informações extraídas, se atentaram primordialmente para as vendas online, pois foi observado que existe um potencial de

saída alto quando esse subterfúgio de venda está ativo (anos de 2020 e 2021). Com isso, a empresa pôde realizar algumas tomadas de decisões em decorrência dos dados coletados, uma delas foi a contratação de uma empresa especializada em marketing. O objetivo é ampliar as vendas digitalmente no decorrer dos próximos anos.

REFERÊNCIAS

- Aalst, W. M. (2011). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*.
- Agrawal, R., & Srikant, R. (1998). Fast Algorithms for Mining Association Rules. pp. 1-13.
- Agrawal, R., Imielinski, T., & Swami, A. (1998). Mining Association Rules between Sets of Items in Large Databases. pp. 1-10.
- Almeida, M. (2014). O varejo virtual na realidade do consumidor e lojas físicas. pp. 1-19.
- Anson, R. J. (1994). *Systemic Reform: Perspectives on Personalizing Education*. ERIC.
- De Gooijer, J. G. (2006). 25 years of time series forecasting.
- De Livera, A. M. (2011). A state space framework for automatic forecasting using exponential smoothing methods.
- Ehlers, R. a. (2002). *Efficient Construction of Reversible Jump MCMC Proposals for Autoregressive Time Series Models*. 2071
- Elmasri, R., & Navathe, S. (2005). *Sistemas de Baco de Dados*. pp. 1-744.
- Fayyad, U., Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*.
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. pp. 1-740.
- Harvey, A. C. (1993). *Structural time series models* (Vol. 11).
- Hyndman, R. J. (2002). A state space framework for automatic forecasting using exponential smoothing methods.
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting Principles and Practice* (2 ed.). Editora Texts.
- Kotler, P., & Keller, K. (2005). *Administração de Marketing: A Bíblia do Marketing* (12 ed.). Pearson.
- Lopes, H. a. (1997). *Bayesian Forecasting (the Levels) of Vector Autoregressive Log-transformed Time Series*.
- Maria Carolina, A. (2021). *Marketing Digital Baseado em Dados: Métricas e Performance* (1 ed.). Inter Saberes.
- Morettin, P. (2017). *Econometria financeira: um curso em séries temporais financeiras* (1 ed.). Edgard Blücher.
- Morettin, P., & Toloi, C. (2009). *Análise de séries temporais*. Blucher.
- Navathe, S., & Elmasri, R. (2018). *Sistemas de Banco de Dados* (7 ed.). Pearson.

- Oliveira, A. L. (n.d.). Do code refactorings influence the merge effort? pp. 1-12.
- Schaedler, A., & Mendes, G. (2021). *Business Intelligence* (1 ed.). Intersaberes.
- Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: an analysis and review.
- Taylor S, L. B. (2017). *Forecasting at Scale*.
- Witten, I., Frank, E., & Hall, M. (2011). *Data Mining: Practical Machine Learning Tools and Techniques* (3 ed.).
doi:<https://doi.org/https://doi.org/10.1016/B978-0-12-374856-0.00023-7>