

UNIVERSIDADE FEDERAL FLUMINENSE – UFF
CAMPUS RIO DAS OSTRAS
GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

ANDRÉ AUGUSTO DA SILVA FERNANDES
LENILDO MACHADO RIBEIRO LIMA JÚNIOR

RELATÓRIO FINAL DE ANÁLISE PREDITIVA

Rio das Ostras

2025

ANDRÉ AUGUSTO DA SILVA FERNANDES
LENILDO MACHADO RIBEIRO LIMA JÚNIOR

RELATÓRIO FINAL DE ANÁLISE PREDITIVA

Relatório de Análise Preditiva, apresentado à Universidade Federal Fluminense – Campus Rio das Ostras, como requisito obrigatório para a obtenção de conceito na disciplina de Análise Preditiva Obrigatória do Curso de Graduação em Engenharia de Produção, sob a orientação do professor Dalton Borges.

Rio das Ostras

2025

Sumário

| | |
|-----------------------------------|----------|
| 1. CONTEXTUALIZAÇÃO | 4 |
| 2. INTRODUÇÃO | 4 |
| 3. METODOLOGIA CRISP-DM | 5 |
| 4. ENTENDIMENTO DO NEGÓCIO | 7 |
| 5. ENTENDIMENTO DOS DADOS | 9 |

1. CONTEXTUALIZAÇÃO

A previsão de demanda é uma metodologia utilizada para estimar valores futuros de uma variável de interesse, como as vendas de um produto. Essa ferramenta é essencial para o gerenciamento eficiente das operações e para apoiar a tomada de decisões estratégicas em diversos setores, inclusive no varejo de moda (LUO; CHANG; XU, 2022).

Modelos de previsão inadequados podem levar a uma gestão ineficaz dos estoques, resultando em perdas financeiras e impactos negativos em toda a cadeia de suprimentos. Por isso, grandes empresas investem constantemente em técnicas de previsão cada vez mais precisas, buscando identificar padrões de comportamento nas séries históricas. Um modelo eficiente garante não apenas uma melhor organização dos estoques, mas também oferece vantagem competitiva (GIRI; CHEN, 2022).

No cenário atual, marcado pela globalização e pela aceleração do consumo digital, o comportamento do mercado se torna ainda mais volátil. Produtos podem ter variações bruscas de demanda ao longo dos anos, reforçando a necessidade do uso de dados históricos aliados a métodos de previsão para orientar as estratégias comerciais (SOUSA; LOUREIRO; MIGUÉIS, 2025).

Particularmente no setor de moda, a sazonalidade, a volatilidade e a sensibilidade a tendências são ainda mais intensas (GIRI; CHEN, 2022). Dentro desse contexto, o presente estudo foca na rede *Segrob Notlad*, uma empresa de fast fashion reconhecida pelo design acessível de seus produtos. Com mais de 80 lojas no Brasil e três unidades na Europa, a empresa busca aprimorar seu processo de reposição de camisetas básicas, um item-chave em seu portfólio.

O objetivo deste estudo é estimar o volume diário de vendas de camisetas básicas ao longo do mês de dezembro de 2024, a fim de apoiar o planejamento estratégico da empresa. A previsão de demanda permitirá um abastecimento mais eficiente das lojas, reduzindo excessos e rupturas de estoque, além de otimizar os custos logísticos e operacionais. Por fim, ao alcançar esses resultados, a empresa poderá obter uma vantagem competitiva em relação aos seus concorrentes, fortalecendo sua posição no mercado.

2. INTRODUÇÃO

3. METODOLOGIA CRISP-DM

O CRISP-DM (Cross-Industry Standard Process for Data Mining) é um processo utilizado com frequência na área de ciência de dados. Essa metodologia tem como intuito garantir que os dados sejam tratados de forma confiável e os modelos construídos sejam avaliados de forma adequada (SCHRÖER; KRUSE; GÓMEZ, 2021).

Desse modo, o CRISP-DM é um modelo independente para executar projetos de mineração de dados, tendo sido desenvolvido no final da década de 1990, mas que permanece relevante dentro do cenário tecnológico, apesar das diversas mudanças ao longo dos anos (SCHRÖER; KRUSE; GÓMEZ, 2021).

O processo é constituído de seis fases, que são: Entendimento do Negócio (Business Understanding), Entendimento dos Dados (Data Understanding), Preparação dos Dados (Data Preparation), Modelagem (Modeling), Avaliação (Evaluation) e Implantação (Deployment) (RAMOS et al., 2020).

1. Entendimento do Negócio (Business Understanding)

O entendimento do negócio possui como foco entender os objetivos do projeto com uma perspectiva empresarial, convertendo esse conhecimento em uma definição de um problema de mineração de dados. Sendo assim, um planejamento pode ser desenvolvido a fim de atingir esses objetivos (SHEARER, 2000).

2. Entendimento dos Dados (Data Understanding)

O entendimento dos dados inicia com a coleta dos dados, para que assim o analista consiga verificar a qualidade das informações. Diante desse quadro, nessa fase observa-se possíveis problemas encontrados (como por exemplo, dados faltantes ou espaços em branco), além de percepções e hipóteses sobre o assunto abordado (SHEARER, 2000).

3. Preparação dos Dados (Data Preparation)

A preparação dos dados envolve todas as atividades necessárias para construir o conjunto de dados final que será utilizado na modelagem. Nessa fase, está inclusa a seleção, limpeza, construção, integração e a formatação dos dados, mudando de forma a ficar com a maior eficiência possível para a próxima etapa do CRISP-DM (SHEARER, 2000).

4. Modelagem (Modeling)

Nessa fase, algumas técnicas de modelagem são selecionadas e aplicadas visando possuir uma boa modelagem para o problema. Diante disso, são construídos e avaliados

modelos com base nas técnicas escolhidas (SHEARER, 2000).

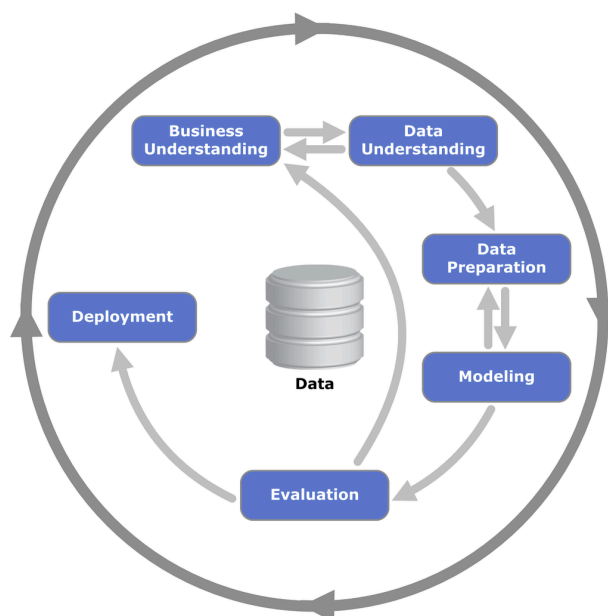
5. Avaliação (Evaluation)

Antes de implementar o modelo, é necessário avaliar de forma mais detalhada e revisar sua construção com o intuito de garantir que ele atinja os objetivos do negócio. Ademais, é importante determinar se alguma questão comercial não foi considerada anteriormente. Por último, o líder do projeto deve nessa fase, decidir como utilizar os resultados da mineração de dados, além de decidir os próximos passos do projeto (SHEARER, 2000).

6. Implantação (Deployment)

Nessa última fase, é necessário desenvolver e documentar um plano para implementar o modelo construído nas fases anteriores do CRISP-DM. Sendo assim, esse plano deve conter um resumo do que foi elaborado em todas as etapas do projeto, além de realizar uma revisão do que foi bem implementado e do que poderia ter sido diferente para melhorar no futuro. Esse documento é importante para monitorar os resultados do projeto e manter um processo de melhoria contínua (SHEARER, 2000).

Figura 1 - Processo CRISP-DM



Fonte: Shearer, 2000

4. ENTENDIMENTO DO NEGÓCIO

Segundo a metodologia, tem-se como primeiro marco para uma análise de demanda assertiva o entendimento do negócio, fase na qual as atividades projetuais estarão focadas na determinação dos reais objetivos do projeto, ponderando a todo tempo as premissas do objeto de estudo, assim como suas restrições e características que o diferenciam dos demais concorrentes da indústria em que se insere. Nesse contexto, o objetivo central do desafio proposto é a captação de talentos na área de análise de dados, a fim de que estes possam auxiliar a empresa *Segrob Notlad* no desenvolvimento de um modelo preditivo capaz de estimar a demanda diária por camisetas básicas ao longo do mês de dezembro de 2024.

Entre as premissas estabelecidas no desafio está a flexibilidade do escopo a ser trabalhado, o que pode representar obstáculos adicionais à equipe envolvida com as adições de dados e mudanças oriundas do conglomerado estratégico organizacional. Tal complexidade decorre, também, da própria natureza do mercado de *fast fashion*, que equivale a uma cadeia de suprimentos internacional altamente dinâmica, composta por uma ampla rede de fornecedores, distribuidores e clientes conectados por fluxos de material, informação e capital (OLIVEIRA, 2017). Diante disso, concretizar o objetivo da organização representa uma vantagem estratégica significativa, possibilitando uma gestão de estoques mais eficiente, alinhamento da produção à demanda real e redução de custos operacionais. Além disso, o sucesso na previsão fortalece a tomada de decisão baseada em dados, promovendo maior competitividade e inovação frente ao mercado (GIRI; CHEN, 2022).

O setor de *fast fashion*, por sua vez, é caracterizado pela produção acelerada e em grande escala de peças que acompanham as últimas tendências, com o intuito de disponibilizar novos produtos nas lojas em ciclos curtos e a preços acessíveis. Esse modelo demanda elevada agilidade nos processos de criação, produção e distribuição, sendo altamente sensível a fatores externos como sazonalidade, comportamento do consumidor, tendência e eventos pontuais (LUO; CHANG; XU, 2022). Diante desse cenário, a construção de um modelo preditivo robusto torna-se essencial, exigindo o uso de métricas de performance reconhecidas na literatura. Serão empregadas, neste projeto, três principais métricas: o MAPE (Erro Percentual Médio Absoluto), que indica a precisão em termos percentuais; o RMSE (Raiz do Erro Médio Quadrático), que enfatiza grandes desvios ao penalizá-los com maior severidade; e o MAD (Erro Médio Absoluto), que oferece uma medida clara do erro médio, independente da direção.

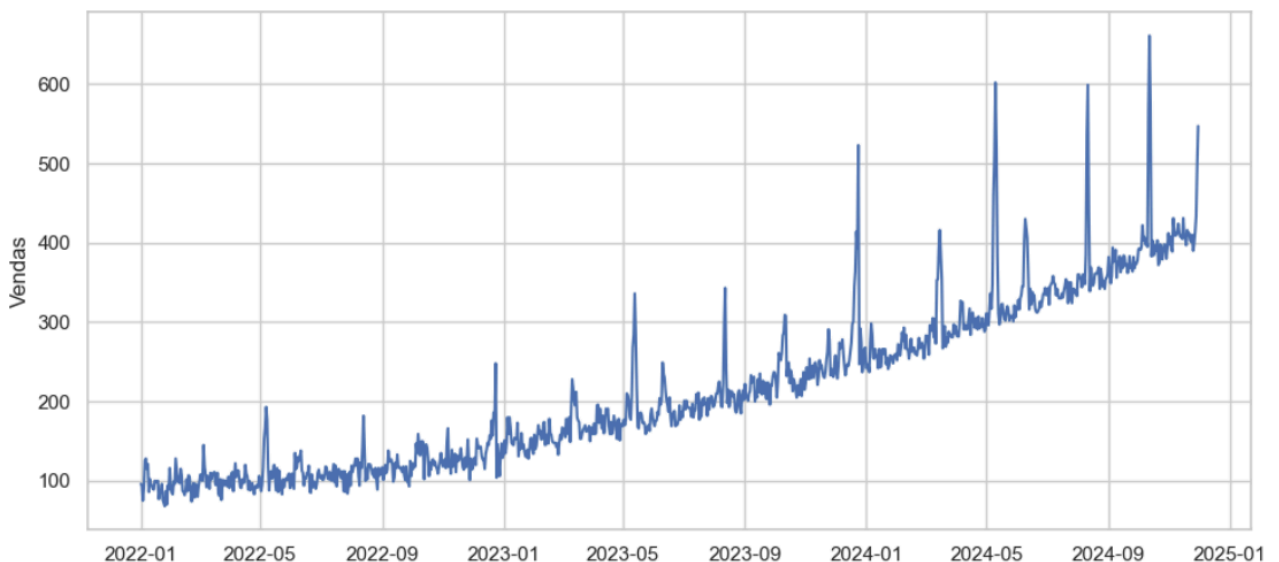
Contudo, uma das restrições mais relevantes do projeto diz respeito à limitação do conjunto de dados disponível, que, inicialmente, contempla apenas as variáveis de data e quantidade vendida de camisetas básicas. A ausência de atributos complementares, como indicadores promocionais, dados meteorológicos, feriados e localização das vendas, compromete a capacidade do modelo de capturar com precisão os padrões sazonais e as variações externas que influenciam o comportamento de compra. Essa carência de informações pode representar um desafio à qualidade das previsões, exigindo, portanto, abordagens cuidadosas e, se possível, a incorporação futura de variáveis adicionais que enriquecem a análise.

5. ENTENDIMENTO DOS DADOS

Conforme mencionado anteriormente, os dados utilizados neste projeto são provenientes do *dataset* de vendas de camisetas pretas disponibilizado pela empresa *Segrob Notlad*. Inicialmente, foi realizada uma análise exploratória com o objetivo de compreender melhor os dados e embasar uma previsão de demanda mais assertiva. Todos os dados foram considerados relevantes nessa etapa inicial, e constatou-se que não há valores ausentes ou zerados nas colunas referentes ao período de 2022 a 2024. Para facilitar a visualização e identificação de possíveis tendências, foram elaborados gráficos que oferecem uma visão mais clara dos padrões presentes nos dados.

No gráfico 1, a série temporal de vendas de camisetas básicas masculinas revela uma tendência clara de crescimento na demanda ao longo do tempo. Os dados mostram um comportamento consistente, com média de 214 vendas e picos recorrentes que sugerem a influência de sazonalidades ou ações promocionais. Esses picos se intensificam nos anos mais recentes, indicando uma possível ampliação da base de clientes ou maior eficiência nas estratégias comerciais. A presença de tendência e padrões sazonais colabora para ela ser aplicável para modelagem preditiva.

Gráfico 1 - Dados Série Temporal

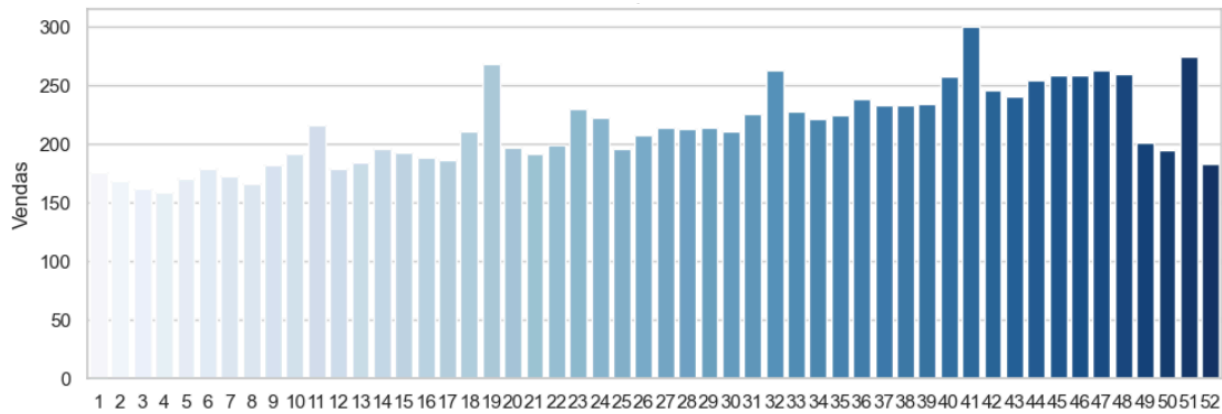


Fonte: Autoria Própria

No gráfico 2, a análise da média de vendas por semana do ano revela uma distribuição sazonal bem definida ao longo das 52 semanas. Observa-se um aumento gradual nas vendas a partir da semana 22, com picos mais acentuados nas semanas 32 e 51. A elevação consistente

no volume médio semanal ao longo do ano reforça a tendência de crescimento da demanda. Algo importante é se observar que o ano de 2024 não teve dados de dezembro o que justifica os números mais baixos nas últimas semanas do gráfico.

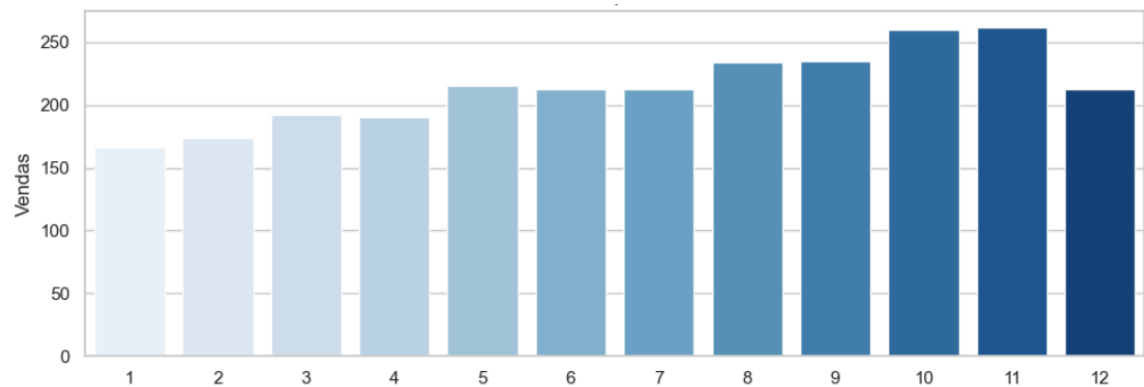
Gráfico 2 - Vendas por Semana



Fonte: Autoria Própria

No gráfico 3, é revelado uma tendência geral de crescimento ao longo do ano, com destaque para o aumento consistente a partir do mês 5 e atingindo picos nos meses 10 e 11, onde as vendas ultrapassam 250 unidades. Nos primeiros quatro meses, os valores permanecem abaixo de 200 unidades, sugerindo um início de ano mais modesto. No entanto, em dezembro (mês 12), há uma queda perceptível nas vendas em comparação aos picos anteriores, o que está relacionado ao fato do último ano não ter dados desse mês o que gerou uma queda desse número.

Gráfico 3 - Vendas por Mês

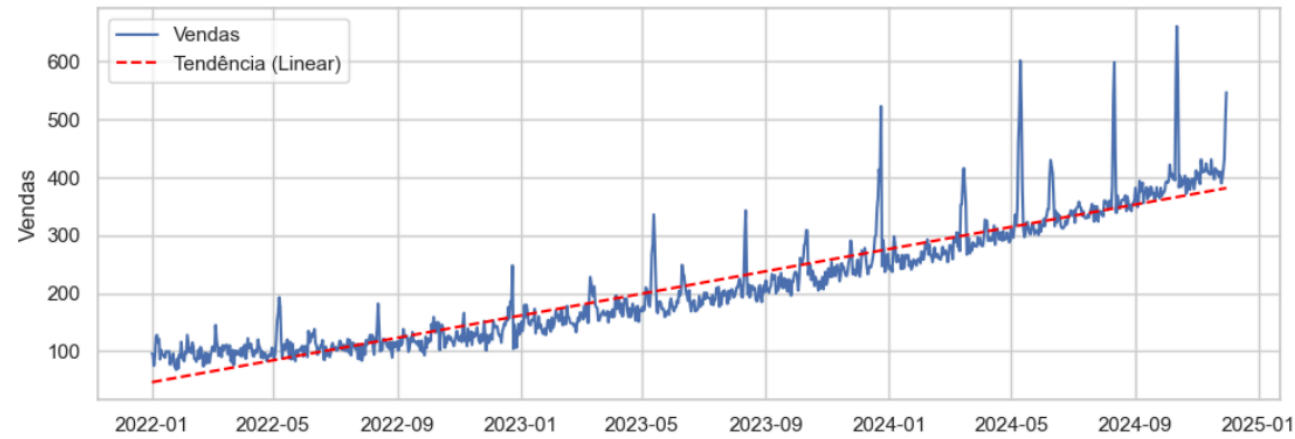


Fonte: Autoria Própria

No gráfico 4, a análise da tendência das vendas ao longo do período de 2022 a 2024 demonstra um crescimento consistente, evidenciado tanto pela linha azul, que representa as vendas diárias, quanto pela linha de tendência linear (vermelha tracejada), que confirma o

aumento gradual. A tendência ascendente indica que o negócio está em expansão, o que reforça a importância de planejar capacidade operacional, estoque e recursos.

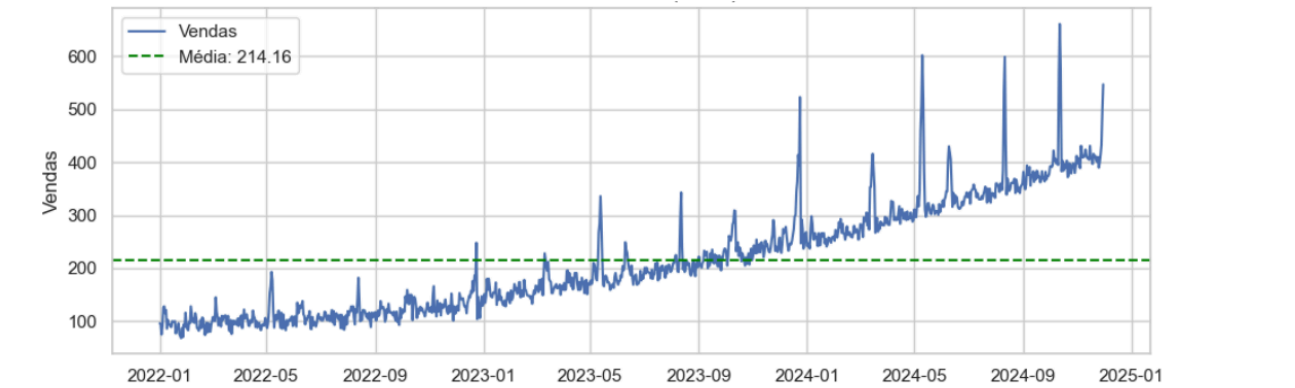
Gráfico 4 - Vendas por Semana (Tendência)



Fonte: Autoria Própria

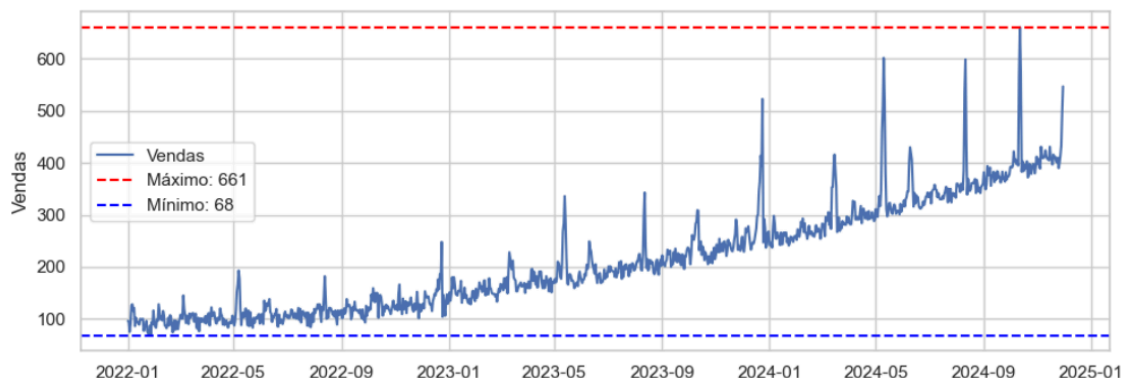
Posteriormente foram gerados gráficos e uma tabela para obter informações descritivas dos dados. A média das vendas ficou em 214,16 unidades, enquanto o desvio padrão foi de 103,57, indicando uma variação considerável em torno da média e reforçando a presença de oscilações no volume de vendas. O valor máximo registrado foi de 661 unidades, enquanto o valor mínimo observado foi de 68 unidades.

Gráfico 5 - Média de Vendas



Fonte: Autoria Própria

Gráfico 6 - Máximo e Mínimo de Vendas



Fonte: Autoria Própria

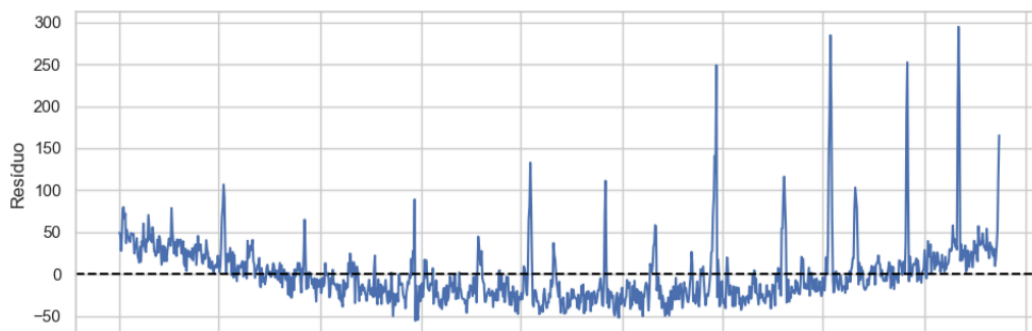
Imagem 2 - Informações Descritivas

Média das vendas: 214.16
Desvio padrão das vendas: 103.57
Valor máximo: 661.00
Valor mínimo: 68.00

Fonte: Autoria Própria

O gráfico de resíduos apresentado mostra a variação das vendas em relação à tendência linear esperada, permitindo identificar desvios e verificar se os erros se comportam como ruído branco. Observa-se que, ao longo do período analisado, os resíduos flutuam em torno de zero, como indicado pela linha preta tracejada, confirmando que, em média, os erros não apresentam viés. Contudo, destacam-se picos esporádicos positivos e negativos, que representam eventos atípicos. Esse comportamento é importante para avaliar a qualidade do modelo, já que a predominância de resíduos próximos de zero sugere um bom ajuste, mas os picos apontam oportunidades de refinamento, seja incorporando variáveis explicativas adicionais ou uma outra alternativa.

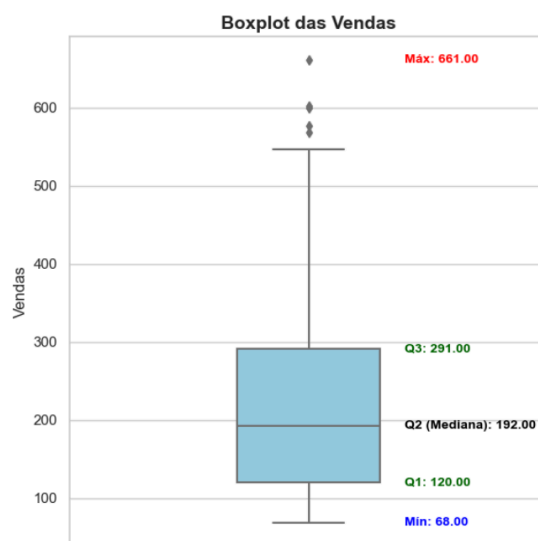
Gráfico 7 - Ruído Branco



Fonte: Autoria Própria

O boxplot das vendas abaixo permite visualizar a distribuição dos dados e identificar a presença de outliers, além de fornecer informações sobre os quartis e a dispersão das vendas ao longo do período analisado. O primeiro quartil (Q1) foi de 120,00, enquanto a mediana (Q2) foi de 192,00 e o terceiro quartil (Q3) foi 291. O valor mínimo registrado foi de 68,00 e o máximo, de 661,00, evidenciando uma ampla variação nos dados. Observa-se também a presença de outliers acima do limite superior, representando picos de vendas fora do padrão esperado.

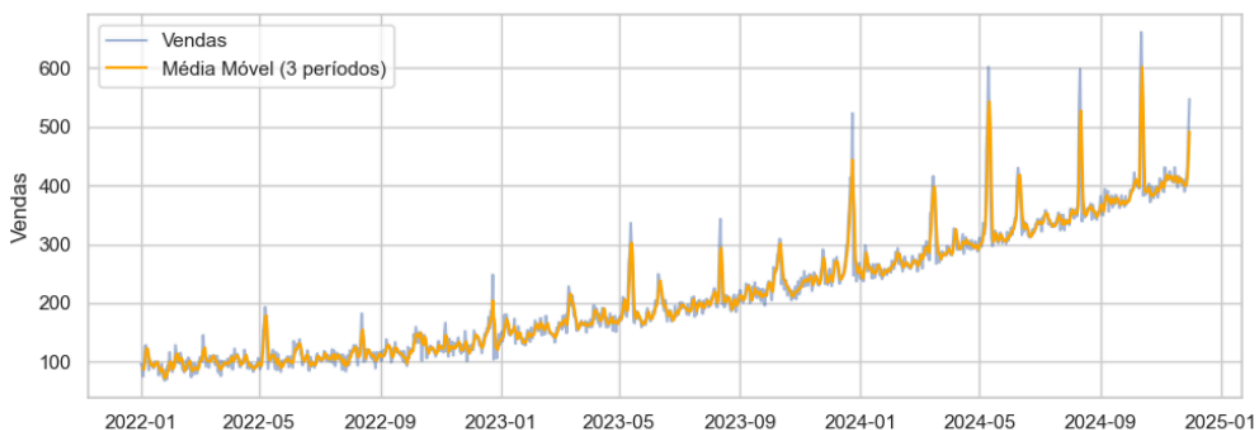
Gráfico 8 - Boxplot



Fonte: Autoria Própria

Por fim, a imagem abaixo apresenta a média móvel das vendas ao longo do tempo, suavizando as variações diárias e permitindo identificar tendências gerais. A linha laranja, correspondente à média móvel de 3 períodos, acompanha de perto os dados reais, destacando os padrões de elevação gradual e suavizando os ruídos. Esse tipo de análise é útil para prever comportamentos futuros.

Gráfico 9 - Média Móvel



Fonte: Autoria Própria

Bibliografia

GIRI, C.; CHEN, Y. Deep Learning for Demand Forecasting in the Fashion and Apparel Retail Industry. **Forecasting**, v. 4, n. 2, p. 565–581, 20 jun. 2022.

LUO, T.; CHANG, D.; XU, Z. Research on Apparel Retail Sales Forecasting Based on xDeepFM-LSTM Combined Forecasting Model. **Information**, v. 13, n. 10, p. 497, 15 out. 2022.

OLIVEIRA, H. B. de J. *O mercado das empresas fast fashion: um estudo de caso da cadeia de suprimentos da H&M e Zara*. **SIMPÓSIO DE ENGENHARIA DE PRODUÇÃO DE SERGIPE**, v; 9, 2017, São Cristóvão.

RAMOS, J. L. C. et al. **CRISP-EDM: uma proposta de adaptação do Modelo CRISP-DM para mineração de dados educacionais**. Anais do XXXI Simpósio Brasileiro de Informática na Educação (SBIE 2020). **Anais...** Em: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO. Brasil: Sociedade Brasileira de Computação, 24 nov. 2020.

SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A Systematic Literature Review on Applying CRISP-DM Process Model. **Procedia Computer Science**, v. 181, p. 526–534, 2021.

SOUSA, M. S.; LOUREIRO, A. L. D.; MIGUÉIS, V. L. Predicting demand for new products in fashion retailing using censored data. **Expert Systems with Applications**, v. 259, p. 125313, jan. 2025.