



# BOOTCAMP DE DATA SCIENCE PROJETO FINAL – CASE VAI STORE

Gustavo Morin de Sousa | João Ferraz

# Agenda

- Contexto
- Modelagem e Análises
- Conclusões
- Recomendações

# Agenda

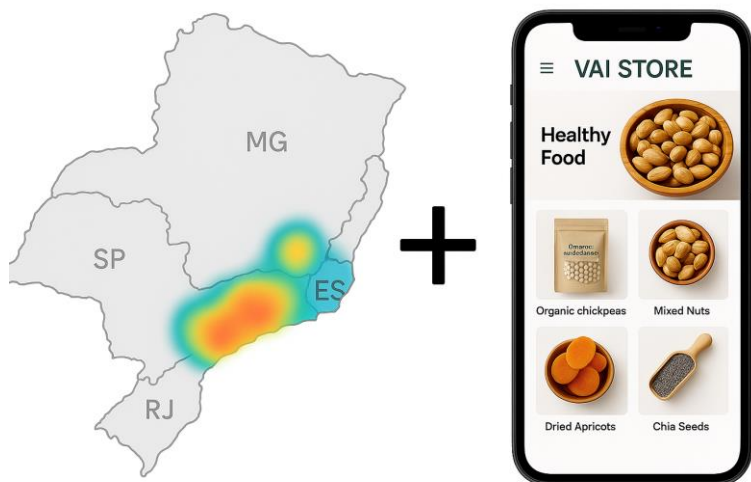
- Contexto
- Modelagem e Análises
- Conclusões
- Recomendações



# Contexto

Existem dois grandes desafios para o time de dados da VAI Store: (1) compreender os padrões de demanda e (2) utilizar previsões para decisões operacionais e estratégicas

Atualmente, a rede cobre diversas regiões com produtos saudáveis como grãos, frutas secas e especiarias, mas ainda há oportunidades não exploradas no uso de dados.



...o que representa oportunidades tanto no entendimento da sazonalidade da demanda quanto na aplicação de modelos preditivos para otimização da operação e aumento de vendas.

1 Oportunidade de identificação de padrões de demanda

Análise do histórico permite identificar sazonalidade e anomalias que impactam vendas e planejamento de estoque.

2 Oportunidade de uso preditivo para decisões práticas

Com um modelo robusto, é possível antecipar a demanda e melhorar decisões de compras, promoções e alocação de recursos.

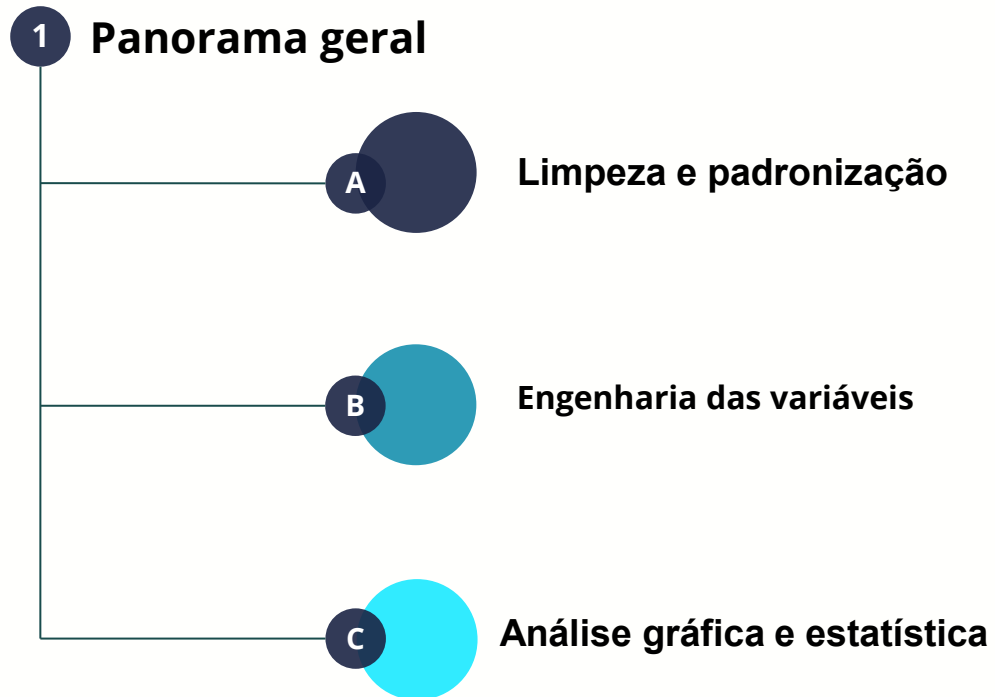


# Agenda

- Contexto
- Modelagem e Análises
- Conclusões
- Recomendações

# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

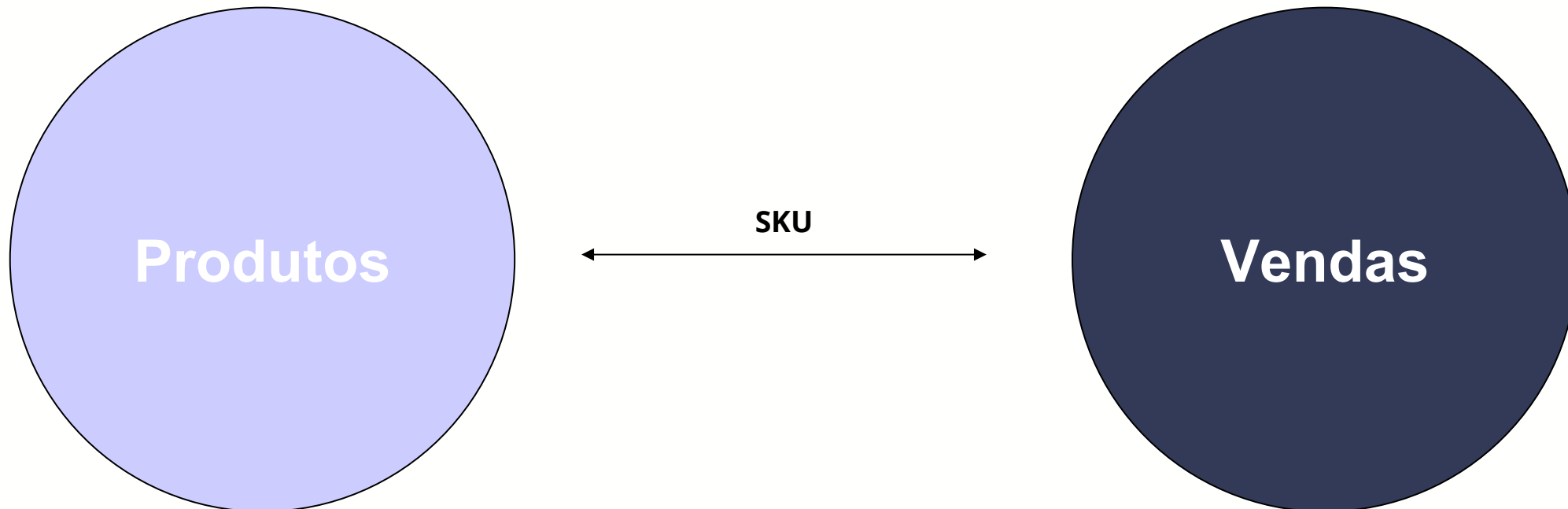


# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

A

**Limpeza e padronização:** Padronizamos os dados e reunimos as bases por meio do SKU, criando uma visão consolidada por produto, data, filial e tipo de unidade.



# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

A

**Limpeza e padronização:** Padronizamos os dados e reunimos as bases por meio do SKU, criando uma visão consolidada por produto, data, filial e tipo de unidade.



Dataset





# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

B

**Engenharia de variáveis:** estruturando sinais relevantes para previsão



## Médias Móveis

Criamos variáveis de médias móveis (MM2, MM3, MM7...) por categoria, suavizando oscilações e capturando tendências de consumo ao longo do tempo.



## Lags de Vendas

Inserimos variáveis de *lag* (ex: vendas dos últimos 1 a 7 dias) para refletir o impacto do comportamento recente na previsão de demanda atual.



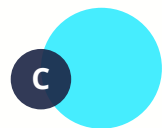
## Referenciais Temporais

Derivamos variáveis temporais como semana do ano, mês, trimestre, dia da semana e feriado nacional. Com isso, o modelo capta padrões sazonais e variações comportamentais ao longo do calendário.



# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

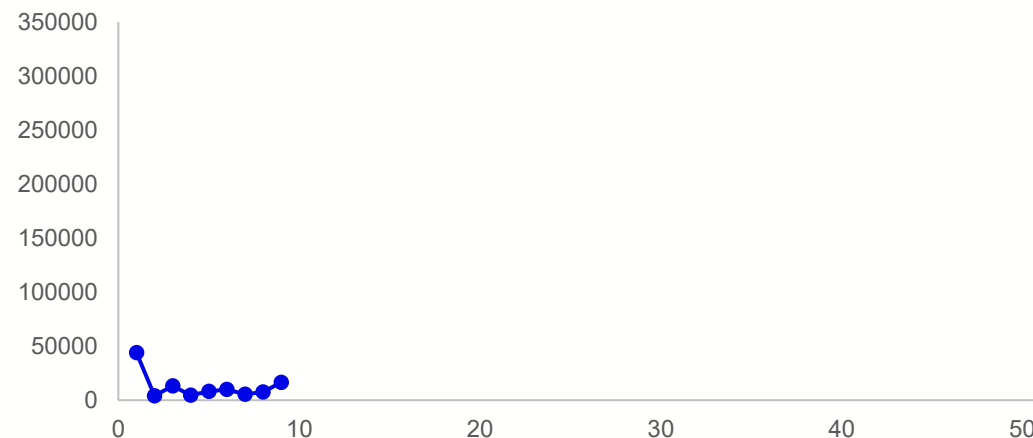


**Análise gráfica e estatística:** Identificamos padrões sazonais em determinados SKUs, especialmente o SKU 1813 (Bacalhau) ao fazer uma análise semanal durante o ano de 2024.

Quantidade Vendida (Kg) X Semanas (Bacalhau & Pescados)

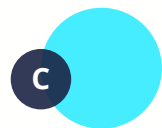


Faturamento (R\$) X Semanas (Bacalhau & Pescados)



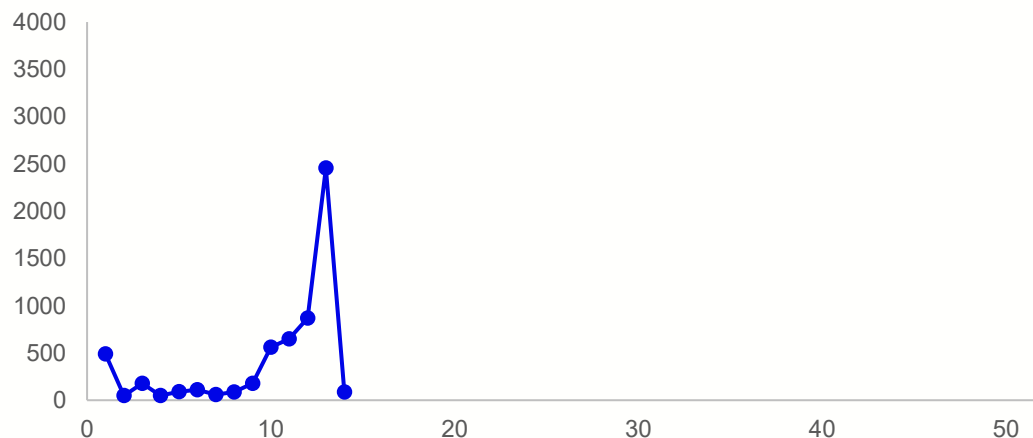
# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

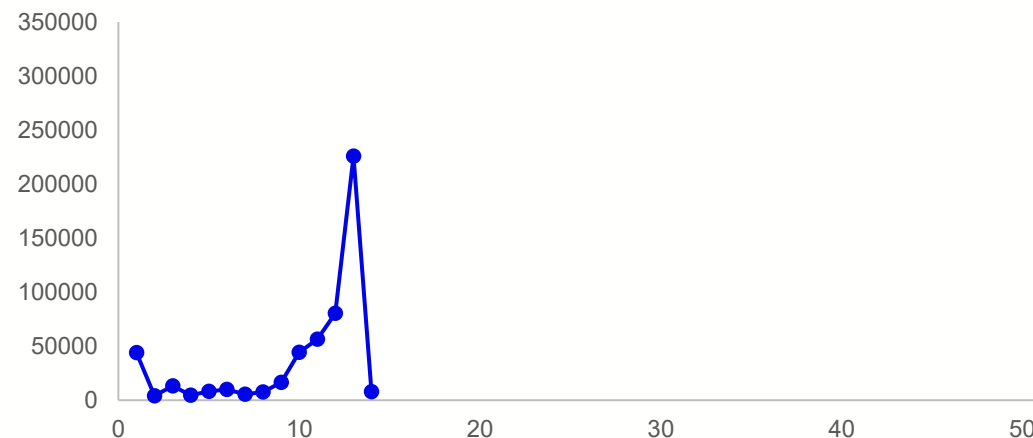


**Análise gráfica e estatística:** Identificamos padrões sazonais em determinados SKUs, especialmente o SKU 1813 (Bacalhau) ao fazer uma análise semanal durante o ano de 2024.

Quantidade Vendida (Kg) X Semanas (Bacalhau & Pescados)

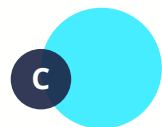


Faturamento (R\$) X Semanas (Bacalhau & Pescados)



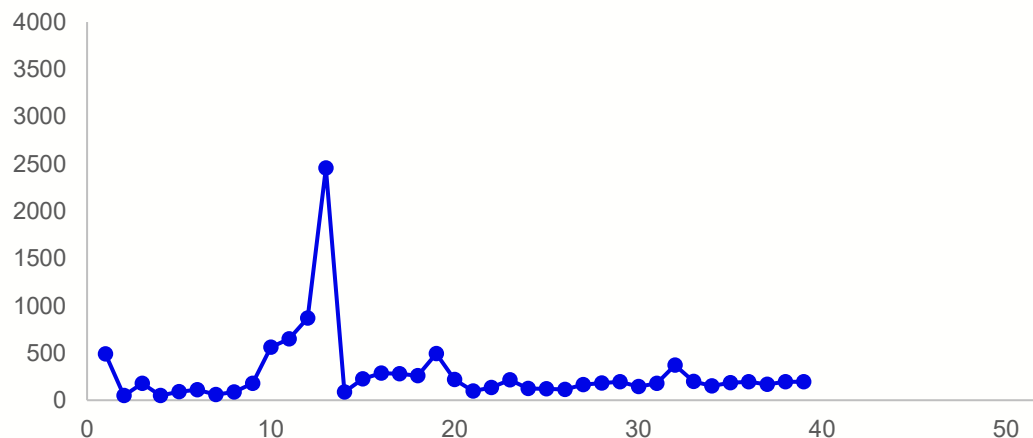
# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

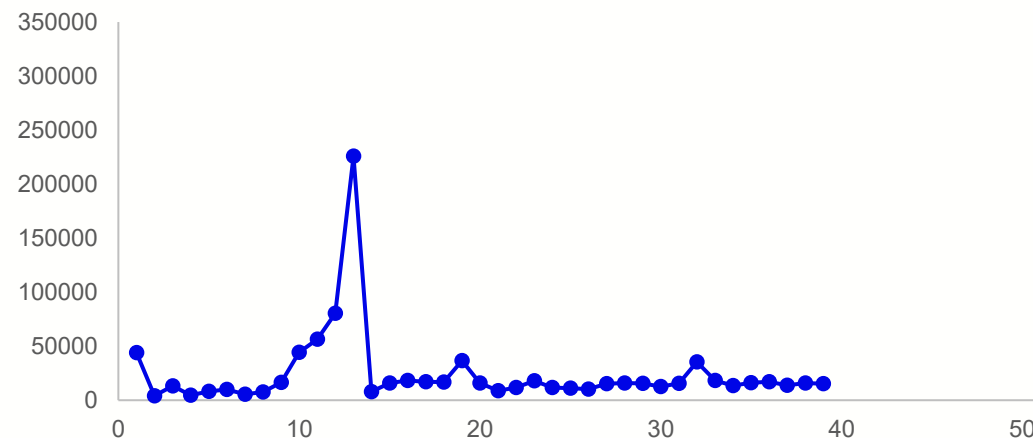


**Análise gráfica e estatística:** Identificamos padrões sazonais em determinados SKUs, especialmente o SKU 1813 (Bacalhau) ao fazer uma análise semanal durante o ano de 2024.

Quantidade Vendida (Kg) X Semanas (Bacalhau & Pescados)

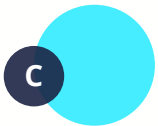


Faturamento (R\$) X Semanas (Bacalhau & Pescados)



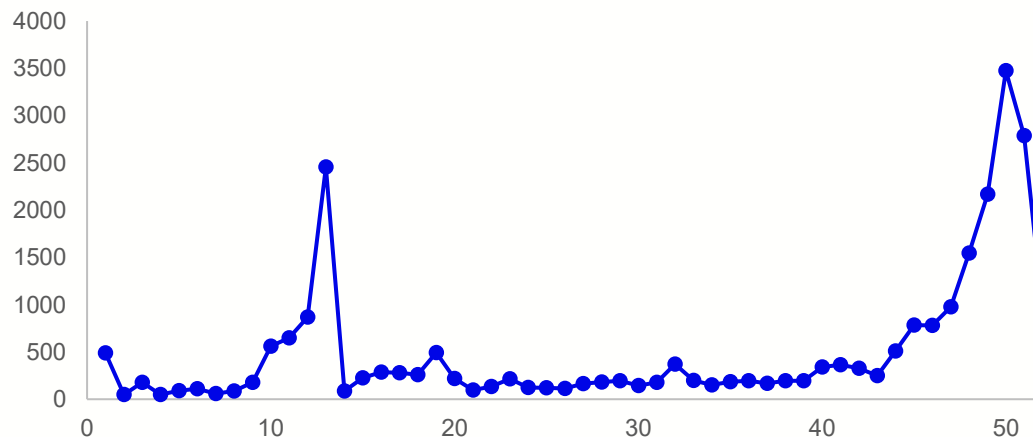
# Modelagem e Análises | Análise Exploratória

Começamos o projeto com duas bases principais: uma com o histórico de vendas por atendimento e outra com o cadastro dos produtos. Era necessário entender se esses dados, muitas vezes desestruturados no varejo, poderiam revelar padrões úteis para uma previsão robusta.

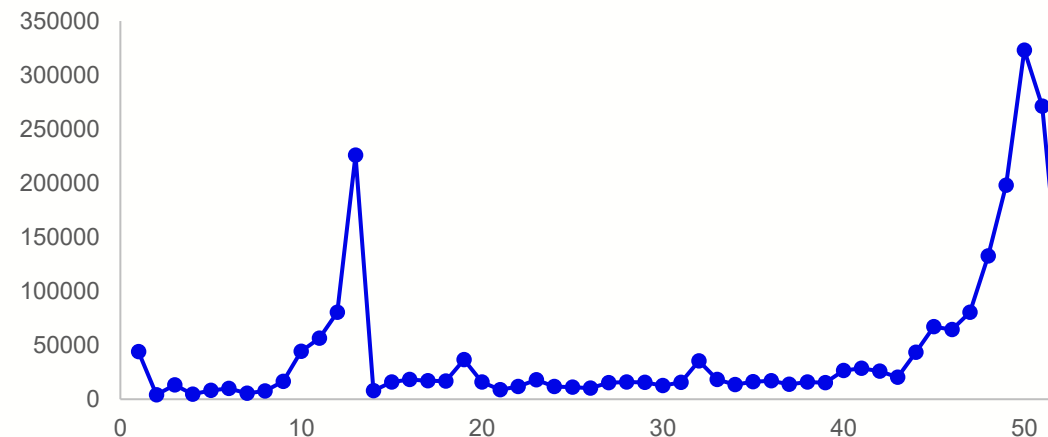


**Análise gráfica e estatística:** Identificamos padrões sazonais em determinados SKUs, especialmente o SKU 1813 (Bacalhau) ao fazer uma análise semanal durante o ano de 2024.

Quantidade Vendida (Kg) X Semanas (Bacalhau & Pescados)



Faturamento (R\$) X Semanas (Bacalhau & Pescados)



# Modelagem e Análises | Modelo

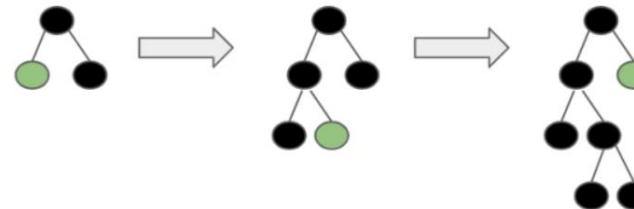
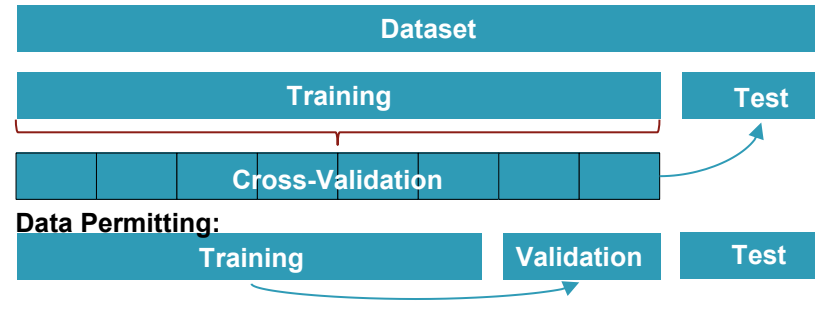
Aplicamos o algoritmo LightGBM para gerar previsões eficientes e escaláveis de demanda com múltiplas variáveis derivadas, mantendo rapidez e alta performance.

## Separação dos Dados

**Treino:** onde o modelo aprende os padrões.

**Validação:** usado para testar diferentes algoritmos, features e hiperparâmetros.

**Teste:** reservado para uma avaliação final e não enviesada do modelo treinado.

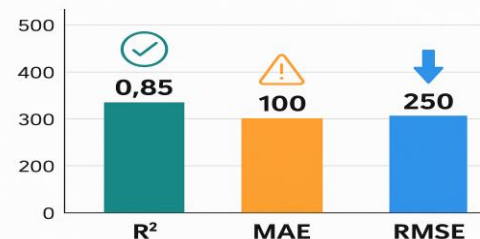


## Aplicação do LightGBM

Treinamos o LGBMRegressor, um modelo de boosting que utiliza histogramas para acelerar o aprendizado, mesmo com grande volume de SKUs, múltiplas variáveis temporais e features derivadas.

## Desempenho do Modelo

Utilizamos o modelo final no conjunto de teste para avaliar sua performance com  $R^2$ , MAE e RMSE, assegurando previsões robustas e realistas para os dados futuros..



## LGBM Regressor

Capturou bem relações entre variáveis derivadas como médias móveis, lags e datas..

Utiliza histogramas em vez de busca exaustiva para dividir os dados, o que reduz drasticamente o tempo de treino sem comprometer a acurácia.

Apresentou baixo tempo de treinamento, ideal para múltiplas iterações e ajustes de hiperparâmetros ao longo do projeto.





# Modelagem e Análises | Resultados

Após treinamento e validação com os dados históricos da VAI Store, o modelo LGBM apresentou desempenho robusto em diferentes métricas de erro e mostrou-se consistente em capturar os padrões sazonais e flutuações de demanda.

**01****Avaliação do Modelo**

O LGBMRegressor foi treinado e validado com divisão temporal, respeitando a cronologia dos dados. O modelo apresentou alta performance tanto no conjunto de treino quanto no de validação.

**02****GridSearch**

Realizamos um grid search com diversas combinações de hiperparâmetros. As melhores configurações foram identificadas com base na média e no desvio padrão dos scores.

**03****Métricas**

Avaliamos o modelo com as seguintes métricas no conjunto de validação:

Score médio (validação cruzada): 0.794 +/- 0.022

R<sup>2</sup> Score: 0.897

MAE (Erro Absoluto Médio): 1.119

MSE (Erro Quadrático Médio): 6.161



Alta acurácia ( $R^2 = 0.897$ ), explicando quase 90% da variância.

Erro médio muito baixo, indicando previsões consistentes até mesmo em dias com grandes flutuações.

Estabilidade nas validações cruzadas, reforçando a confiança na generalização do modelo.

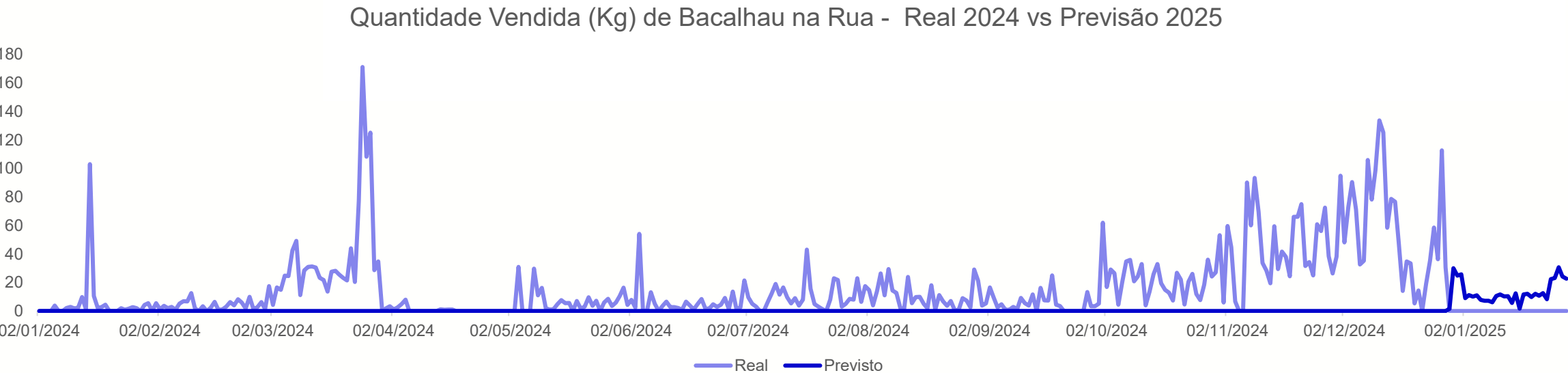
Modelo apto para ser utilizado na fase de previsão operacional com segurança.



# Modelagem e Análises | Previsões

A partir dos dados históricos da VAI Store, o modelo LGBM foi treinado para estimar a demanda futura de bacalhau, resultando em previsões consistentes que refletem o padrão de consumo típico do início do ano.

## 1) Previsão para Jan/25



## 2) Análise das previsões

O modelo LGBM prevê um total de 415 kg de bacalhau vendidos em janeiro de 2025, com distribuição suave ao longo do mês e ausência de picos extremos.

Essa projeção indica uma expectativa de demanda controlada e alinhada ao histórico sazonal do produto, oferecendo uma base segura para o planejamento operacional.



# Agenda

- Contexto
- Modelagem e Análises
- Conclusões
- Recomendações

# Conclusões

A partir das análises e previsões realizadas, foi possível obter aprendizados relevantes para a tomada de decisão estratégica do time de compras.

**1**

**Diagnóstico dos Dados Históricos:** A demanda por bacalhau apresenta forte padrão sazonal e concentração em canais específicos.

1. É possível modelar a sazonalidade com clareza a partir dos gráficos semanais e diários de vendas.
2. O bacalhau mantém volumes baixos ao longo do ano, com picos expressivos na Semana Santa e no Natal.
3. Em janeiro, o produto só é comercializado em lojas de rua, não havendo registros de vendas em shoppings nesse período.

**2**

**Previsão para Janeiro de 2025:** A previsão estimada para o mês é de 415 kg de bacalhau vendidos na modalidade "Rua".

1. Um lote muito pequeno pode resultar em perda de vendas, ruptura de estoque e insatisfação do cliente.
2. Um lote muito grande pode gerar sobrecusto de armazenagem, risco de desperdício e capital imobilizado.
3. A previsibilidade do comportamento da demanda permite que a compra seja fracionada por semana, garantindo flexibilidade e resposta mais precisa ao giro real.






# Agenda

- Contexto
- Modelagem e Análises
- Conclusões
- Recomendações

# Recomendações

Para transformar os aprendizados em ações operacionais, sugerimos um conjunto de iniciativas com base nos dados e nas previsões para janeiro de 2025.

	Ações	Próximos Passos
<div></div> <div>Ações Imediatas</div>	<ul style="list-style-type: none"><li>Realizar a compra de lote de bacalhau para janeiro de 2025, com base na previsão do modelo escolhido (415 kg do LGBM).</li><li>Estabelecer <b>limites mínimo e máximo de estoque</b> para o período, considerando margens de segurança.</li></ul>	<ul style="list-style-type: none"><li>Avaliar compras semanais para reduzir riscos logísticos e financeiros.</li><li>Integrar as previsões ao sistema de compras e abastecimento.</li></ul>
<div></div> <div>Ações de curto/médio prazo</div>	<ul style="list-style-type: none"><li>Incentivar a venda em shoppings durante janeiro, onde não houve registro em 2024.</li><li>Explorar ações promocionais em janeiro, como descontos ou combos (ex: 1 kg de bacalhau), para estimular vendas.</li></ul>	<ul style="list-style-type: none"><li>Planejar campanhas de comunicação específicas para o período de janeiro.</li><li>Escolher filiais com bom fluxo e mensurar resultados de venda e aceitação.</li></ul>
<div></div> <div>Ações Estruturantes</div>	<ul style="list-style-type: none"><li>Implementar um <b>sistema contínuo de previsão de demanda</b> com atualização mensal.</li><li>Desenvolver um <b>painel de apoio à decisão para o time de compras</b>, integrando previsão, histórico e alertas.</li></ul>	<ul style="list-style-type: none"><li>Automatizar a coleta e análise dos dados para alimentar os modelos com frequência.</li><li>Garantir que o dashboard esteja acessível aos decisores com interface clara e atualizações automáticas.</li></ul>





# Q&A

Breakfast  
Skin & Sun Care  
Coffee & Tea  
Biscuits

4

Cereal

PRICES  
DROPPED

Was \$4.29  
**\$3.50**



Coffee

LOW PRICE  
ALWAYS

**\$11**

LOW PRICE  
ALWAYS

**\$4**

PRICES  
DROPPED

**\$1.60**

Muesli

LOW PRICE  
ALWAYS

**\$14**



# Anexo



# Modelagem e Análises | Modelo

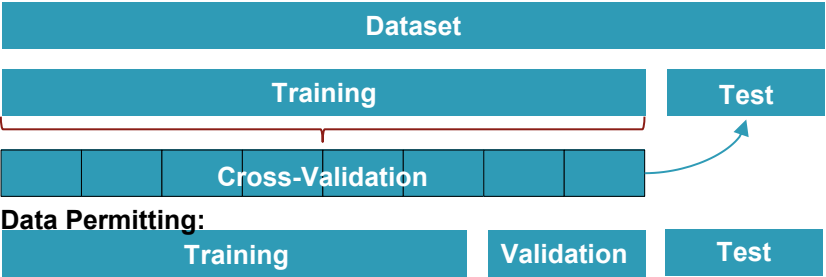
Na etapa de modelagem preditiva, foi utilizado o algoritmo Random Forest para construir previsões robustas da demanda a partir de múltiplas variáveis históricas e sazonais.

Separação dos Dados

**Treino:** onde o modelo aprende os padrões.

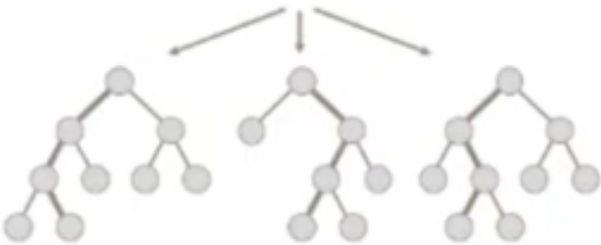
**Validação:** usado para testar diferentes algoritmos, features e hiperparâmetros.

**Teste:** reservado para uma avaliação final e não enviesada do modelo treinado.



Aplicação do Random Forest

Ajustamos o RandomForestRegressor com os dados de treino, utilizando validação cruzada para seleccionar a melhor configuração do modelo e evitar overfitting.



Desempenho do Modelo

Utilizamos o modelo final no conjunto de teste para avaliar sua performance com  $R^2$ , MAE e RMSE, assegurando previsões robustas e realistas para os dados futuros..



## Random Forest

Captura relações não lineares entre variáveis como vendas históricas, médias móveis, lags, tipo de loja e calendário.

Combina diversas árvores para reduzir variância e gerar previsões mais estáveis, mesmo com alta oscilação.

Demonstrou bom desempenho em SKUs com sazonalidade, lidando bem com a diversidade do portfólio.

Permite identificar as variáveis mais relevantes, como semana do ano e tipo de unidade, apoiando decisões estratégicas.





# Modelagem e Análises | Resultados

Após treinamento e validação com os dados históricos da VAI Store, o modelo Random Forest apresentou desempenho preditivo de alta qualidade, com métricas muito favoráveis e excelente capacidade de capturar padrões complexos de demanda.

**01****Avaliação do Modelo**

O RandomForestRegressor foi treinado e validado com divisão temporal, respeitando a ordem cronológica dos dados. Obteve alta performance no conjunto de validação, com bom equilíbrio entre variância e generalização

**02****GridSearch**

Realizamos uma busca por hiperparâmetros, ajustando profundidade das árvores, número de estimadores e critério de divisão. As melhores combinações foram selecionadas com base na média e no desvio padrão dos scores de validação.

**03****Métricas**

Avaliamos o modelo com as seguintes métricas no conjunto de validação:

Score médio (validação cruzada):  $0.791 \pm 0.023$

$R^2$  Score: 0.966

MAE (Erro Absoluto Médio): 0.597

MSE (Erro Quadrático Médio): 2.054

Excelente acurácia ( $R^2 = 0.966$ ), indicando que o modelo explica quase toda a variação observada nos dados.

Erro absoluto médio muito baixo, o que garante previsões altamente confiáveis para apoio à operação.

Alta estabilidade nas validações cruzadas, reforçando a consistência do modelo em diferentes subconjuntos.

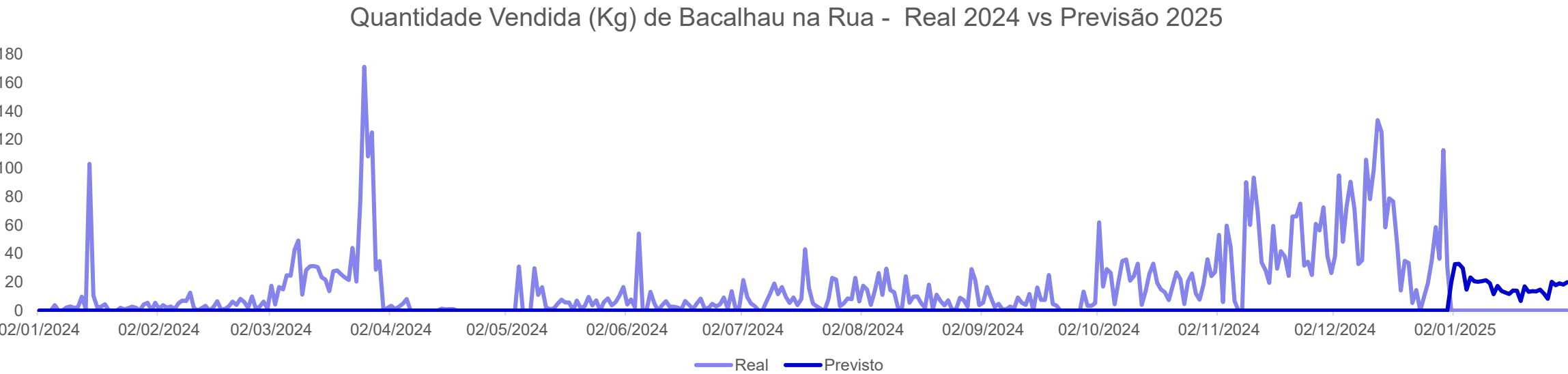
Ótima opção para previsão de SKUs com comportamento não linear e volumes médios a altos de venda.



# Modelagem e Análises | Previsões

A partir dos dados históricos da VAI Store, o modelo Random Forest foi capaz de antecipar a demanda de bacalhau com alto grau de aderência, oferecendo uma visão clara do comportamento esperado para o início de 2025.

## 1) Previsão para Jan/25



## 2) Análise das previsões

O modelo Random Forest prevê um total de 543 kg de bacalhau vendidos em janeiro de 2025, com distribuição consistente ao longo do mês e picos moderados de demanda após a primeira semana.

Em relação a janeiro de 2024, observa-se a manutenção da sazonalidade e uma leve elevação no volume médio diário, o que sugere um comportamento estável e confiável para planejamento de compras



# Modelagem e Análises | Modelo

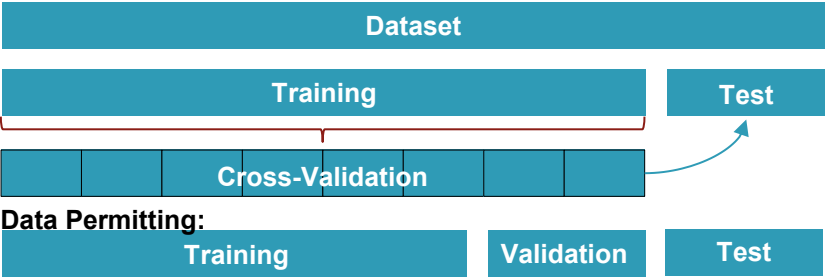
Utilizamos o algoritmo XGBoost como modelo complementar para validar a robustez das previsões, devido à sua alta estabilidade, regularização e performance preditiva.

Separação dos Dados

**Treino:** onde o modelo aprende os padrões.

**Validação:** usado para testar diferentes algoritmos, features e hiperparâmetros.

**Teste:** reservado para uma avaliação final e não enviesada do modelo treinado.



Aplicação do Random Forest

Treinamos o XGBRegressor, que realiza boosting de árvores com regularização e controle detalhado dos parâmetros. Seu mecanismo de penalização ajuda a evitar overfitting mesmo com dados ruidosos.



Desempenho do Modelo

Utilizamos o modelo final no conjunto de teste para avaliar sua performance com  $R^2$ , MAE e RMSE, assegurando previsões robustas e realistas para os dados futuros..



## XGBRegressor

Captura interações complexas entre features temporais, lags e categorias com alta acurácia.

Possui mecanismos de regularização L1 e L2, que ajudam a controlar overfitting em dados com muitos SKUs e padrões irregulares.

Sua estratégia de divisão e aprendizado sequencial garante previsões sólidas em produtos com forte variabilidade.





# Modelagem e Análises | Resultados

O modelo XGBoost apresentou performance sólida e estável, com bons resultados nas métricas de erro e validação cruzada. Foi eficaz na captura de padrões complexos e oscilantes, sendo uma alternativa confiável para previsão de demanda na VAI Store.

**01****Avaliação do Modelo**

O XGBRegressor foi treinado e validado com divisão temporal, respeitando a cronologia dos dados. A performance foi consistente, com bom desempenho mesmo em produtos com maior variação de vendas.

**02****GridSearch**

Testamos diversas configurações envolvendo profundidade máxima, taxa de aprendizado e número de estimadores. A escolha dos melhores hiperparâmetros foi guiada pelo score médio e sua estabilidade nas validações.

**03****Métricas**

Avaliamos o modelo com as seguintes métricas no conjunto de validação:

Score médio (validação cruzada):  $0.784 \pm 0.017$

$R^2$  Score: 0.872

MAE (Erro Absoluto Médio): 1.191

MSE (Erro Quadrático Médio): 7.657



Boa acurácia geral ( $R^2 = 0.872$ ), com previsões confiáveis para a maioria dos SKUs.

Capacidade de capturar variações bruscas, útil para categorias com sazonalidade e comportamento irregular.

Erro médio absoluto abaixo de 1,2 unidades por dia, o que é aceitável operacionalmente.

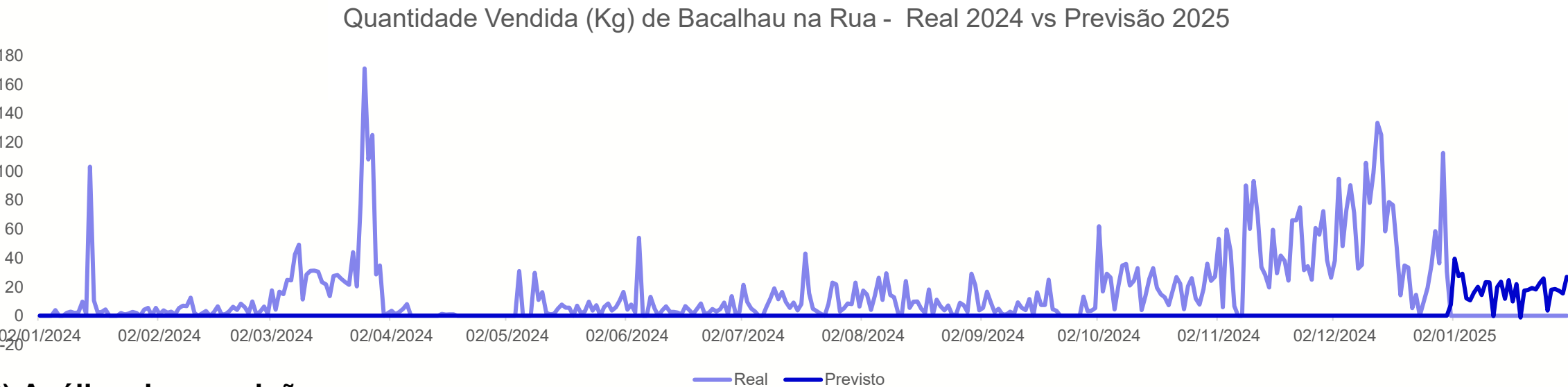
Modelo robusto e flexível, ideal para validar e comparar resultados com outras abordagens..



# Modelagem e Análises | Previsões

Com base nos dados históricos da VAI Store, o modelo XGBoost foi treinado para prever a demanda futura de bacalhau, apresentando resultados que mantêm coerência com o padrão comportamental observado no início do ano.

## 1) Previsão para Jan/25



## 2) Análise das previsões

O modelo XGBoost prevê um total de 555 kg de bacalhau vendidos em janeiro de 2025, com variações leves ao longo do mês e poucos picos acentuados.

A previsão sugere um cenário estável e confiável, refletindo os padrões sazonais do produto e oferecendo uma boa base para suporte ao planejamento de estoque e abastecimento.



# Modelagem e Análises | Métricas

## 1. $R^2$ (Coeficiente de Determinação)

Mede o quanto da variação dos dados reais é explicada pelas previsões do modelo.

Vai de 0 a 1: quanto mais próximo de 1, melhor.

Interpretação: um  $R^2$  de 0,90 significa que 90% da variação na demanda foi explicada pelo modelo.

Utilidade: ajuda a entender o quão “aderente” o modelo está ao comportamento real.

## 2. MAE (Erro Absoluto Médio)

Calcula a média das diferenças absolutas entre valores previstos e reais.

Mantém as unidades originais (ex: kg vendidos).

Interpretação: um MAE de 1,2 indica que, em média, o modelo erra por 1,2 kg.

Utilidade: mostra o erro médio esperado, sendo fácil de interpretar operacionalmente.

## 3. MSE (Erro Quadrático Médio)

Calcula a média dos erros elevados ao quadrado.

Penaliza erros grandes de forma mais severa.

Interpretação: quanto menor o MSE, melhor o desempenho do modelo.

Utilidade: bom para detectar modelos que erram pouco na média, mas cometem grandes desvios pontuais.

