

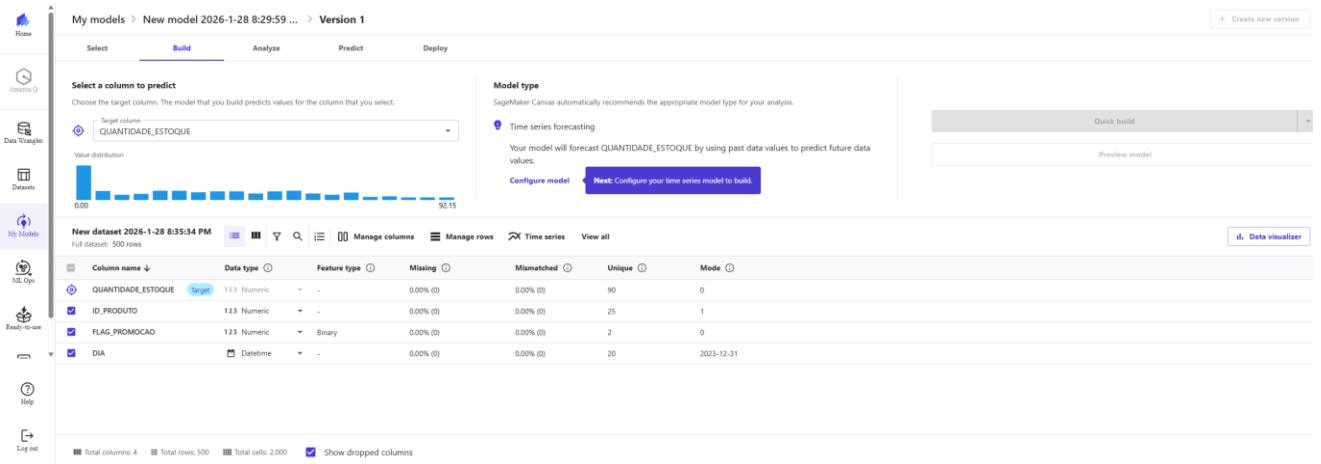
DESAFIO AWS SAGEMAKER CANVAS

Previsão de Estoque Inteligente na AWS com SageMaker Canvas.

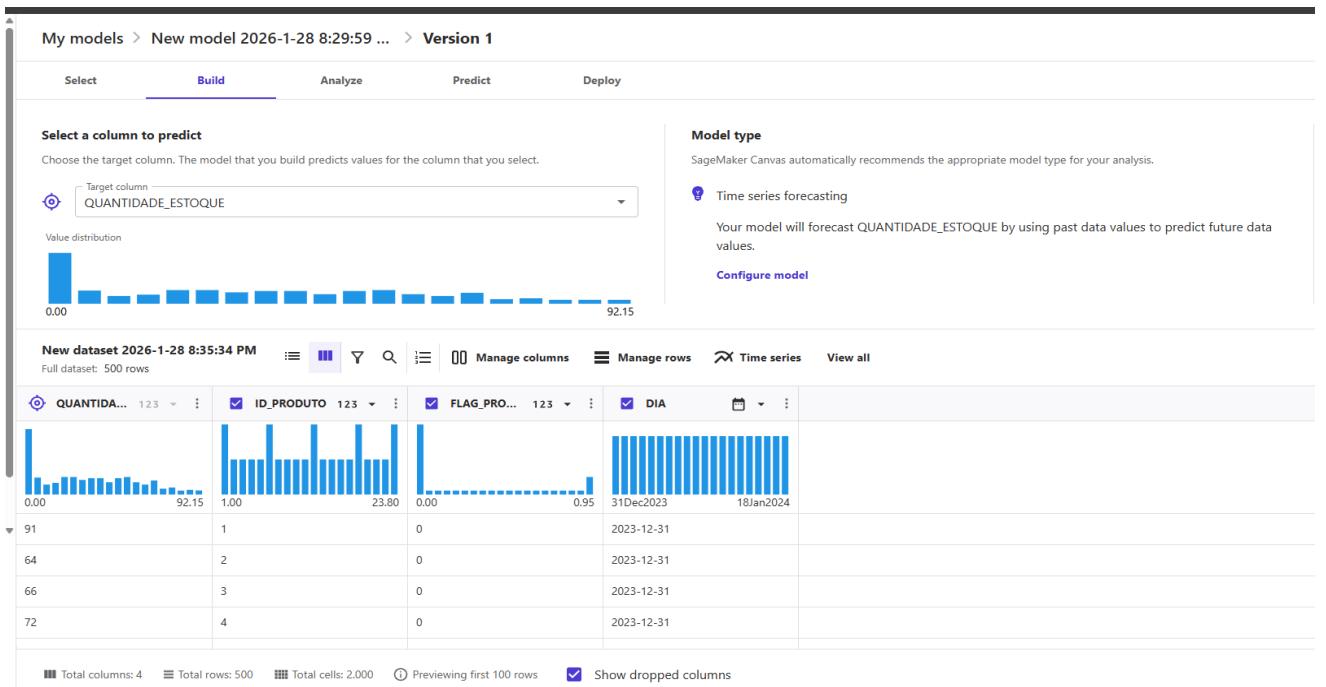
O desafio consiste em criar um modelo de machine learning No-Code para prever o estoque de uma empresa. Vamos ao passo a passo do projeto!

PASSO 1 - Selecionar Dataset

Configurei meu ambiente no AWS Sage Maker, criando um domínio para organizar meu projeto. Escolhi o dataset “dataset-500-curso-sagemaker-canvas-dio.csv” para treinar o modelo e fiz o upload dele no AWS SageMaker Canvas.



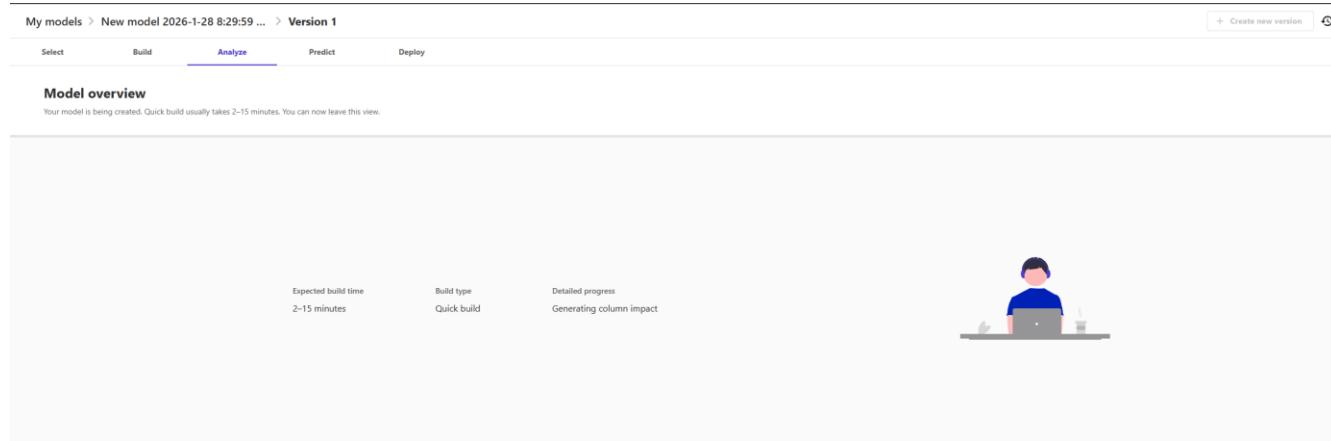
The screenshot shows the AWS SageMaker Canvas interface. On the left, there's a sidebar with icons for Home, Amazon QuickSight, Data Wrangler, Datasets, My Models, ML Opt., Ready-to-use, Help, and Log out. The main area has tabs for Select, Build, Analyze, Predict, and Deploy. The Build tab is selected. A section titled "Select a column to predict" shows a dropdown set to "Target column: QUANTIDADE_ESTOQUE" and a histogram of its value distribution. To the right, a "Model type" section recommends "Time series forecasting" for predicting QUANTIDADE_ESTOQUE. Below that is a "Configure model" button. The central part of the screen displays a table of data columns: QUANTIDADE_ESTOQUE (target), ID_PRODUTO, FLAG_PROMOCAO, and DIA. The table includes columns for Column name, Data type, Feature type, Missing, Mismatched, Unique, and Mode. At the bottom, it shows "Total columns: 4", "Total rows: 500", "Total cells: 2,000", and a checkbox for "Show dropped columns".



This screenshot shows the same AWS SageMaker Canvas interface as the previous one, but with more data columns selected in the table. The columns QUANTIDADE_ESTOQUE, ID_PRODUTO, FLAG_PROMOCAO, and DIA are now checked. Each column has its own histogram below it. The table headers remain the same: Column name, Data type, Feature type, Missing, Mismatched, Unique, and Mode. The bottom of the screen shows the same summary statistics and "Show dropped columns" option.

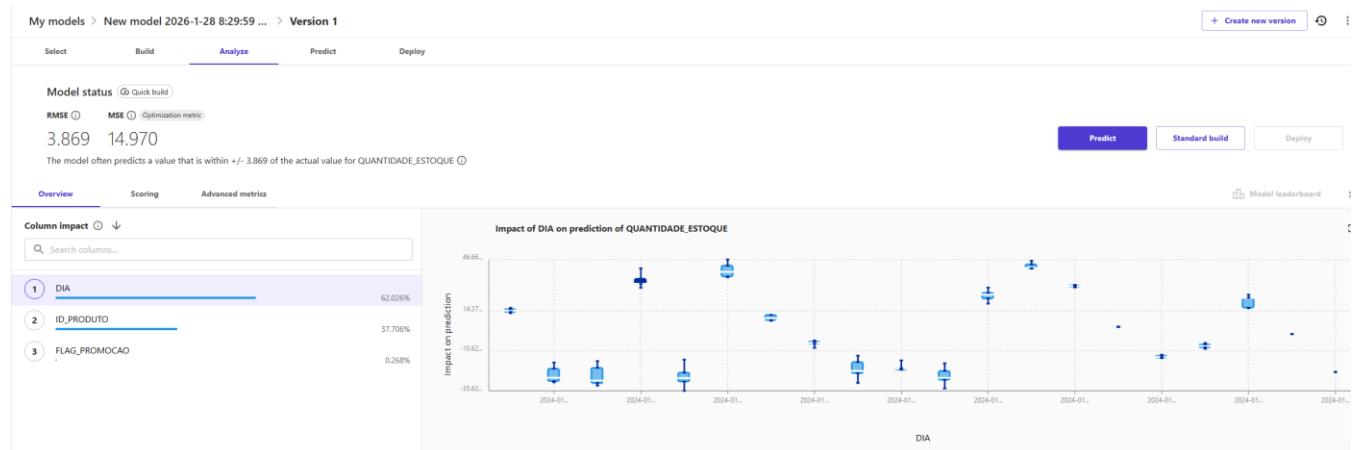
PASSO 2 – Construir/Treinar

Configure as variáveis de entrada e saída de acordo com os dados. O objetivo foi prever a quantidade de estoque, então escolhi a coluna **ID_Estoque**. Inicie o treinamento do modelo. Escolhi o ‘Quick Build’ para treinar o modelo, isso levou cerca de 10 minutos.



PASSO 3 – Analisar

Esse foi o momento mais importante do processo. Após o modelo estar treinado examinei as métricas do modelo, verifiquei as principais características que influenciaram as previsões e as métricas que o modelo forneceu.



Para entender melhor as métricas fornecidas pelo modelo, pesquisei sobre elas e quero mostrar aqui conforme elas apareceram. Vou destacá-las individualmente como forma de ampliar o conhecimento sobre o projeto.

1 RMSE e MSE (e por que eles importam)

- ◆ **MSE — Mean Squared Error** - É o erro médio ao quadrado entre o valor real e o valor previsto. Em termos simples:

- O modelo erra
- Esse erro é elevado ao quadrado
- Depois tira-se a média

Resultado ➡ **MSE = 14.970** (Isso significa que, em média, o **quadrado do erro** é ~15)

📌 O problema do MSE é que ele fica em **unidade ao quadrado**, então não é muito intuitivo para o negócio.

- ◆ **RMSE — Root Mean Squared Error** - É a **raiz quadrada** do **MSE**.

Ele traz o erro **de volta para a mesma unidade do estoque** (quantidade).

Resultado ➡ **RMSE = 3.869**

O próprio Canvas explica: “*The model often predicts a value that is within +/- 3.869 of the actual value for QUANTIDADE_ESTOQUE*”

📦 Interpretação prática para estoque:

- Em média, o modelo **erra cerca de 4 unidades**
- Se o estoque real for 50, a previsão costuma ficar entre **46 e 54**

📊 Esse RMSE é bom ou ruim?

Depende totalmente do **seu contexto**:

Cenário **RMSE ≈ 3,9** é...

Estoque médio de 10 unidades ✗ Ruim

Estoque médio de 50 unidades ⚠ Aceitável

Estoque médio de 200+ unidades ✓ Bom

👉 Uma boa prática é calcular:

RMSE / Estoque médio

Se der algo como:

- até **5%** → muito bom
- entre **5% e 10%** → ok
- acima de **15%** → precisa melhorar o modelo

2 Impacto das variáveis: DIA, ID_PRODUTO e FLAG_PROMOÇÃO

“O que mais influencia a previsão do estoque?”

- ◆ **DIA — 62,026%**

📅 É disparado o fator mais importante.

👉 Significa que:

- O modelo aprendeu **padrões temporais**
- Dias específicos (data, dia da semana, sazonalidade, tendência) explicam **mais de 60%** da variação do estoque

💡 Isso é comum em:

- Varejo
- Estoque com giro
- Demanda sazonal

✓ **Sinal muito positivo:** indica que existe padrão previsível no tempo.

- ◆ **ID_PRODUTO — 37,706%**

ID Cada produto tem um comportamento próprio.

👉 O modelo entende que:

- Produtos diferentes têm níveis de estoque e giro diferentes
- Isso explica quase **38%** da previsão

📌 Observação importante: Se o ID_PRODUTO for apenas um **código numérico**, o Canvas trata isso como categórico internamente — então está tudo certo.

- ◆ **FLAG_PROMOÇÃO — 0,268%**

💡 Impacto **quase irrelevante** no modelo atual.

👉 Pode significar:

- Promoções não afetam muito o estoque **nesse histórico**
- Ou a variável está mal representada (ex: poucos dias com promoção)
- Ou o efeito da promoção já está “embutido” no comportamento por dia

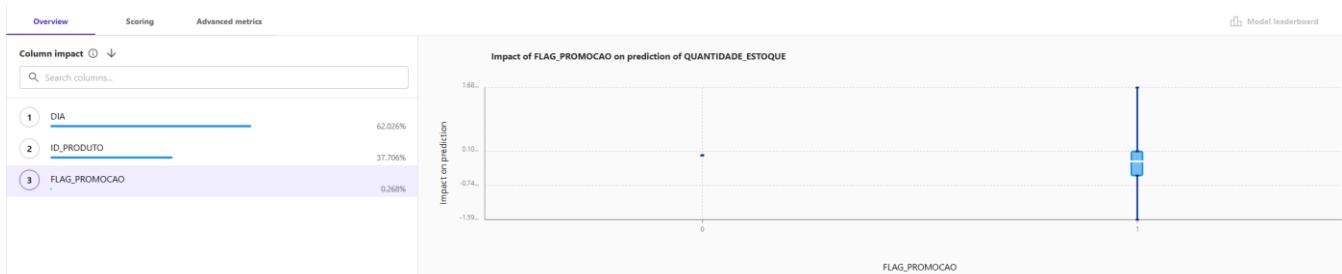
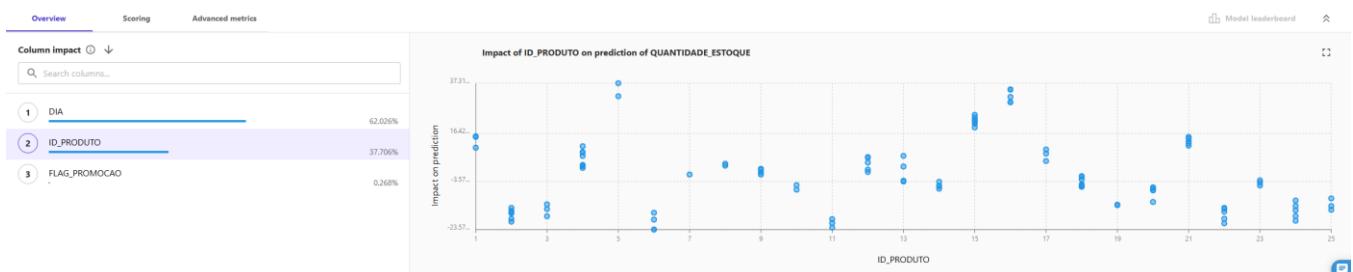
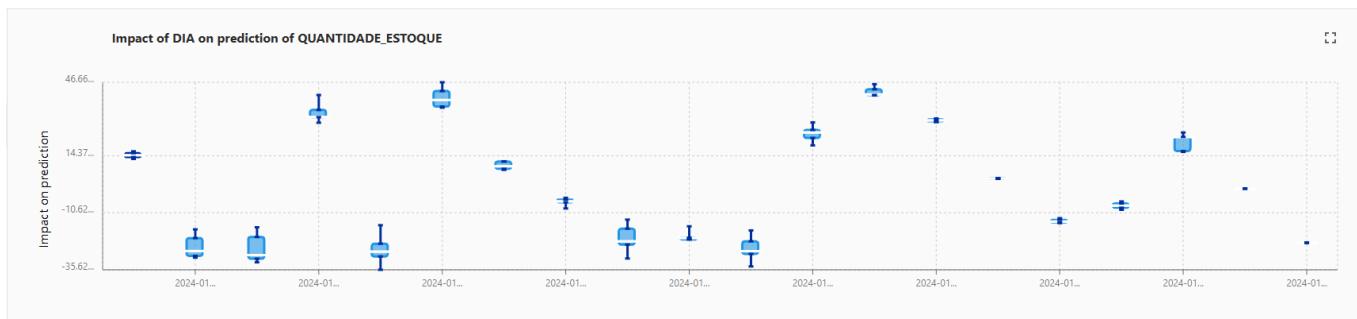
3 Leitura geral do modelo

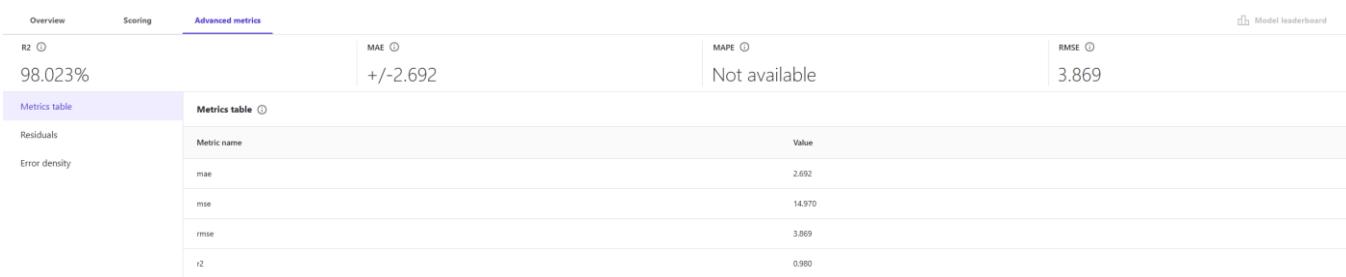
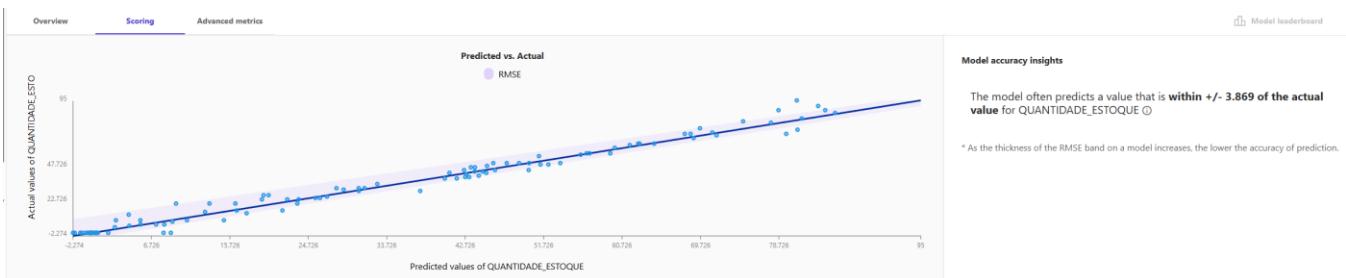
✓ **Pontos positivos**

- RMSE relativamente baixo
- Forte padrão temporal
- Produtos bem diferenciados
- Modelo aparentemente estável

⚠️ Pontos de atenção

- Promoção quase não influencia (vale revisar)
- Talvez falte:
 - preço
 - vendas passadas
 - estoque mínimo
 - dia da semana / feriado explícito





Seguindo o raciocínio quero destacar ainda: **R², MAE, MAPE** e os números que aparecem no Canvas:

1 R² (Coeficiente de Determinação) - O R² mede quanto da variação do estoque real o modelo consegue explicar.

- Vai de **0 a 1** (ou 0% a 100%)
- Quanto mais perto de 1, melhor
- ◆ **R² = 98,023% (~0,98)**

💡 Isso significa que:

- O modelo explica **98% do comportamento do estoque**
- Apenas **~2%** fica sem explicação (ruído, eventos inesperados)

📌 Para problemas de estoque e demanda:

- **> 90%** → excelente
- **95%+** → nível produção
- **98%** → muito forte

2 MAE (Mean Absolute Error) - É o erro médio absoluto, sem elevar ao quadrado.

👉 Em termos simples: “Em média, o modelo erra **X unidades**”

- ◆ **MAE = ±2,692**

💡 Isso significa que:

- Em média, a previsão erra **cerca de 2,7 unidades**
- Esse é o erro **mais fácil de explicar para o time de negócio**

📌 Comparação com RMSE:

- MAE é sempre **menor ou igual** ao RMSE
 - Como:
 - MAE = 2,69
 - RMSE = 3,87
- 👉 indica que **não há muitos erros extremos** (outliers)

✓ Isso é um ótimo sinal de estabilidade.

3 MAPE (Mean Absolute Percentage Error) - MAPE mede o erro em percentual.

👉 Exemplo:

- MAPE = 5%
→ “O modelo erra, em média, 5% do valor real”

◆ MAPE = Not available

❓ Por que não está disponível?

Normalmente acontece quando:

- Existem valores reais **zero ou muito próximos de zero**
- Estoque zerado em alguns dias
- O Canvas desativa o MAPE para evitar divisão por zero

📌 Isso é **normal em modelos de estoque** e não é um problema grave.

💡 Alternativa: já temos MAE e RMSE, que são **mais confiáveis** nesse cenário.

📌 Agora como interpretar a **Metrics Table** (linha por linha)?

Vamos ler exatamente o que ela está dizendo 👇

Métrica Valor O que isso diz

MAE 2.692 Erro médio ≈ 3 **unidades**

MSE 14.970 Erro médio ao quadrado (técnico)

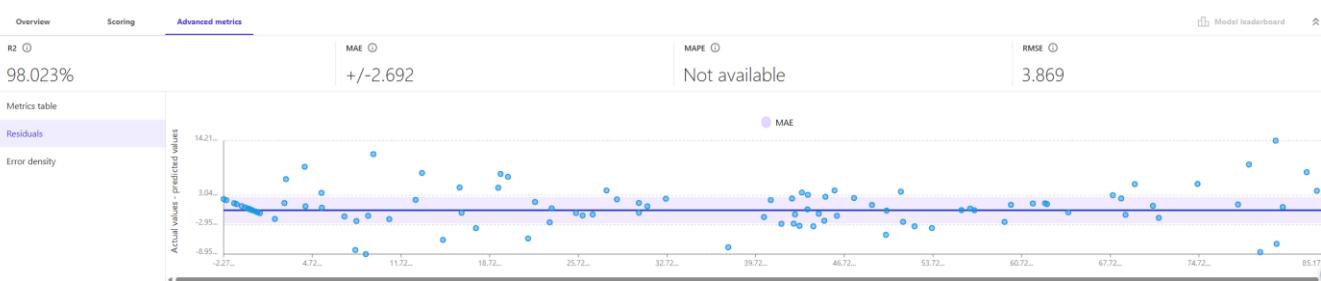
RMSE 3.869 Erro típico ≈ 4 **unidades**

R² ~0.98 **98% do estoque explicado**

💡 **Resumo executivo do modelo:** “O modelo explica **98% da variação do estoque** e costuma errar entre **2 e 4 unidades** por previsão, sem apresentar grandes erros extremos.”

✓ **Isso indica:**

- Modelo **confiável**
- Baixa volatilidade nos erros
- Bom para:
 - planejamento de reposição
 - alertas de ruptura
 - sugestão de compra



Scoring — A aba **Scoring** mostra **como o modelo se comporta ao prever dados que ele não viu** (validação/teste).

👉 Pense assim:

- **Build / Analyze** → como o modelo foi treinado
- **Scoring** → como ele se sai na vida real

Aqui você responde a pergunta: “Posso confiar nas previsões?”

2 **Model Accuracy Insights** - “*The model often predicts a value that is within +/- 3.869 of the actual value for QUANTIDADE_ESTOQUE*”

💡 **Interpretação prática:** Isso é o **RMSE traduzido para linguagem de negócio**:

- Em média, o modelo:
 - erra **~4 unidades**
 - tanto para mais quanto para menos

👉 Isso é extremamente útil para decisão:

- Se o estoque mínimo for 10 → ⚠️ cuidado
- Se o estoque mínimo for 30+ → ✅ seguro

◆ **Observação importante** - “*As the thickness of the RMSE band on a model increases, the lower the accuracy of prediction.*”

👉 A faixa azul em volta da linha é a **banda de erro (RMSE)**:

- Mais fina → modelo mais preciso
- Mais grossa → maior incerteza

No gráfico: ✅ A banda está **estreita e consistente**, ótimo sinal.

3 Predicted vs. Actual — como ler o gráfico principal

Essa é a parte mais importante visualmente.

◆ **Eixos**

- **Eixo X (horizontal)** → valor **previsto**
- **Eixo Y (vertical)** → valor **real**

Cada ponto azul = **uma previsão**

◆ **Linha diagonal (linha perfeita)**

A linha central representa:

Previsão perfeita (Predicted = Actual)

Se o modelo fosse perfeito:

- Todos os pontos estariam **exatamente sobre essa linha**

◆ **Pontos azuis (seus dados)**

No caso:

- Pontos bem **alinhados à diagonal**
- Pouca dispersão
- Sem padrão estranho (curva, funil, inclinação)

➡ Isso indica:

- Pouco viés
- Modelo equilibrado
- Não está sempre superestimando nem subestimando

4 A banda azul (RMSE band)

Essa “faixa” em volta da linha mostra:

Onde a maioria dos erros acontece

🔍 O que observar:

- **Espessura constante** → erro uniforme
 - **Sem abrir muito nas pontas** → modelo generaliza bem
 - **Sem deslocamento para cima ou para baixo** → sem viés sistemático
- ✓ O gráfico passa nesses três critérios.

5 Sinais de alerta que NÃO aparecem (ótimo!)

O gráfico **não mostra**:

- ✗ Nuvem espalhada
- ✗ Curva em “S”
- ✗ Erro crescendo muito em estoques altos
- ✗ Padrão de sub ou superestimação
- 👉 Se algum desses aparecesse, seria sinal de problema estrutural.

6 Leitura final do Scoring (em linguagem de negócio)

📦 **Resumo executivo:** “Quando o modelo prevê um estoque, o valor real costuma estar a no máximo **±4 unidades** do previsto, com comportamento estável em diferentes níveis de estoque.”

Isso significa que o modelo é:

- ✓ Estável
- ✓ Bem calibrado
- ✓ Pronto para apoiar decisões reais

7 Como usar isso na prática 🚀

- Definir **estoque de segurança = RMSE × 1,5**
- Criar alertas do tipo:
 - “Se previsão < estoque mínimo + 4 → comprar”
- Simular impacto de erro sem risco operacional

🚀 PASSO 4 - Prever

Depois de analisar o modelo e comprovar que ele está ok, usei o modelo treinado para fazer previsões de estoque. Escolhi o “Single prediction” para gerar 5 previsões de diferentes ID.

My models > New model 2026-1-28 8:29:59 ... > Version 1

Select Build Analyze Predict Deploy

Predict target values

Batch prediction Single prediction

Generate predictions for an entire dataset. ⓘ

Manual Automatic

Predictions

All Jobs Configuration

Filter by configuration name: All

My models > New model 2026-1-28 8:29:59 ... > Version 1

Select Build Analyze Predict Deploy + Create new version ⌂

Predict target values

Batch prediction Single prediction

Modify values to predict QUANTIDADE_ESTOQUE in real time.

Filter columns

Column	Value
ID_PRODUTO	10
DIA	12/31/2023
FLAG_PROMOCAO	0

QUANTIDADE_ESTOQUE Prediction

You made 1 change to the values. Update for a new prediction and feature importance metrics.

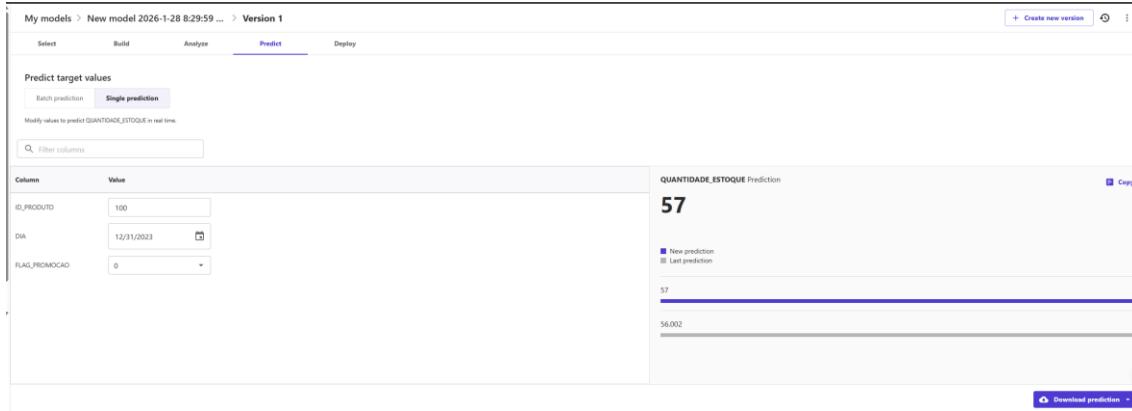
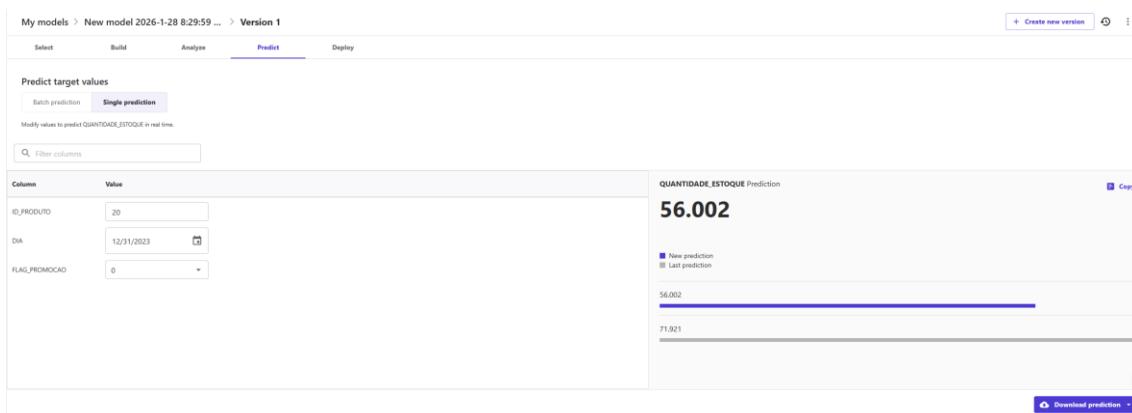
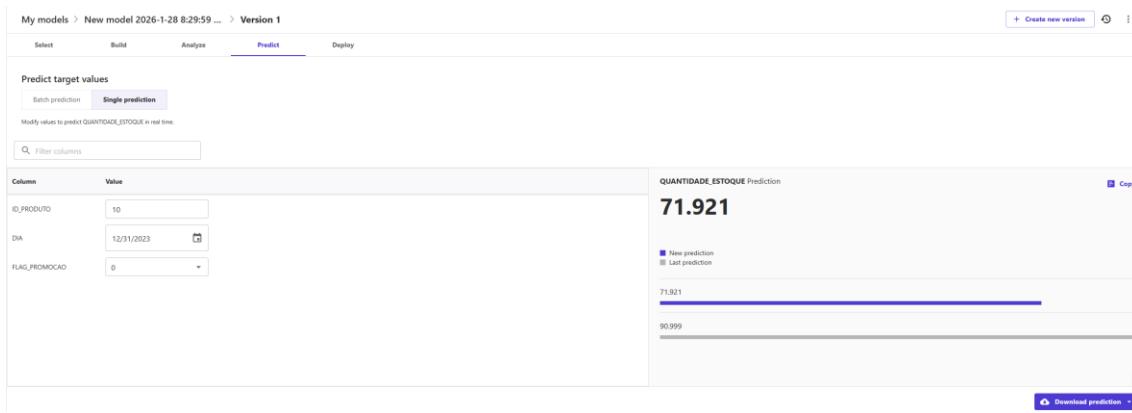
New prediction Last prediction

90.999

Update prediction

Update prediction with feature importance

Copy



RESULTADO – modelo criado e testado ✓

The screenshot shows the AWS SageMaker Canvas interface. At the top, it says "My models > New model 2026-1-28 8:29:59 PM". Below that is a "Versions" section with a table:

Version	Status	Build type	Created ↑	Dataset	RMSE	Model Registry
V1	Ready	Quick	01/28/2026 8:30 PM	New datas...	3.869	Not Registered ⓘ ::

Below the table is a "My models" section with a grid view. One model card is visible:

New model 2026-1-28 8:29:59...

Ready

Version 1

Target QUANTIDADE_ESTOQUE

Problem type Numeric prediction

Updated 2026-1-28 8:56:11 PM

[View](#) ::

📌 CONSIDERAÇÕES FINAIS

Este desafio teve como objetivo a construção de uma análise preditiva de estoque utilizando o AWS SageMaker Canvas para estimar a quantidade futura de produtos com base em padrões temporais, comportamento por item e variáveis de contexto, alcançando alta precisão e baixo erro médio. A solução permite reduzir rupturas e excessos de estoque, melhorar o planejamento de compras e reposição e apoiar decisões estratégicas com dados confiáveis, substituindo achismos por previsões quantitativas. Além disso, a análise antecipa tendências de demanda, aumenta a eficiência operacional e contribui diretamente para a redução de custos e melhoria do nível de serviço ao cliente.