

Desafio Técnico

Analytics & Machine Learning Commerce



**mercado
livre**

THIAGO AUGUSTO SANTOS LIMA
CIENTISTA DE DADOS SÊNIOR

ESTUDOS DE CASO



Ofertas Relâmpago – Uma análise exploratória



Previsão de Falha - Modelagem

Ofertas Relâmpago

Análise Exploratória

CONTEXTO & HIPÓTESES

Este estudo analisa resultados de ofertas do tipo *relâmpago* realizadas em um período e país específicos, buscando entender quais fatores mais influenciam o desempenho das promoções.

Formas de avaliar os resultados:

- **Número de vendas** – volume total de itens vendidos.
- **Aproveitamento de estoque** – proporção vendida em relação ao estoque disponível.
- **Receita gerada** – valor total obtido com as vendas.

Hipóteses analisadas:

- O frete grátis aumenta o volume de vendas.
- O horário de início da promoção impacta o resultado.
- Promoções com mais tempo no ar geram mais vendas.
- O desempenho varia de acordo com a vertical do produto.

VALIDAÇÃO E TRATAMENTO DE DADOS

Após definidas as hipóteses e o contexto geral, aprofundou-se a análise para entender o comportamento das promoções e a relação entre as variáveis.



7,8% das promoções duraram menos de 1 minuto e não registraram vendas;



6,1% dos casos não seguiram a fórmula esperada:
 $SOLD_QUANTITY = INVOLVED_STOCK - REMAINING_STOCK_AFTER_END$;



2,6% apresentaram estoque ao final da promoção negativo;



49,6% não possuíam informação das vendas e seus valores totais arrecadados.



Dentre as promoções sem informações, apenas 2,9% tiveram vendas.



Valores faltantes devidamente tratados (NA e none).

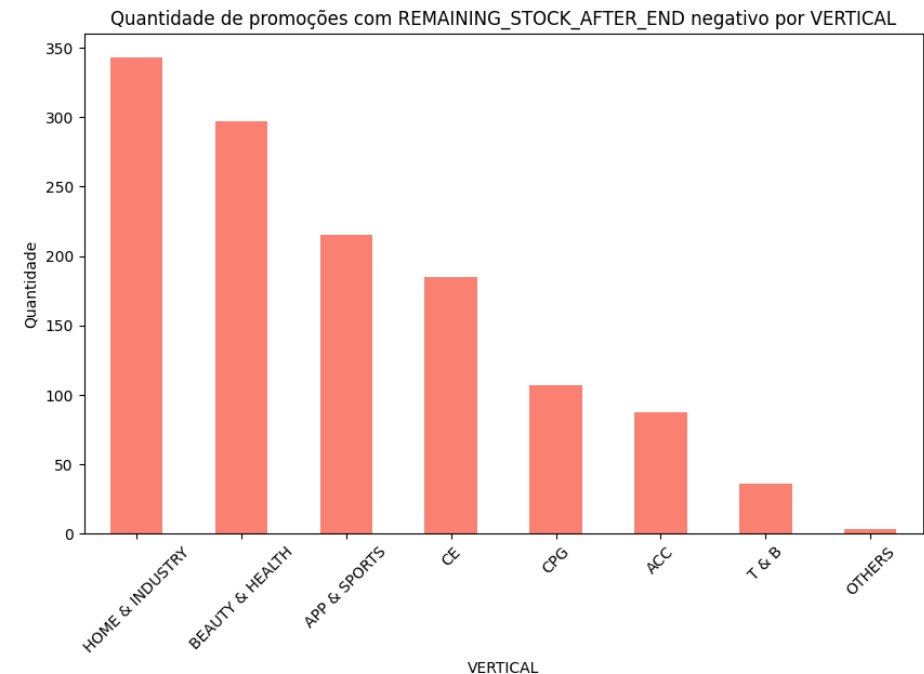


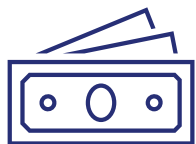
Gráfico de Estoque Pós Promoção Negativo por Vertical

NOVAS MÉTRICAS



Aproveitamento de Estoque:

$QTDE\ DE\ VENDAS / ESTOQUE\ ENVOLVIDO$



Preço Médio por Produto

$ARRECADAÇÃO\ VENDAS / QTDE\ DE\ VENDAS$

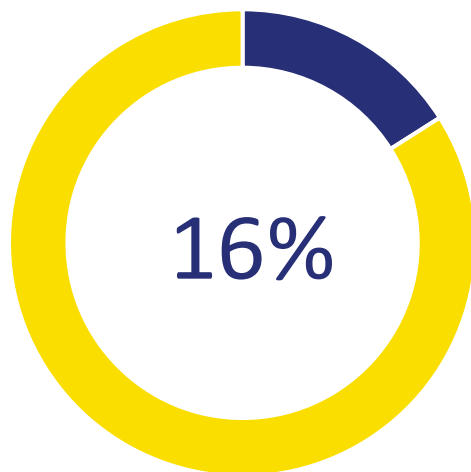
OFERTAS RELÂMPAGOS

VISÃO GERAL

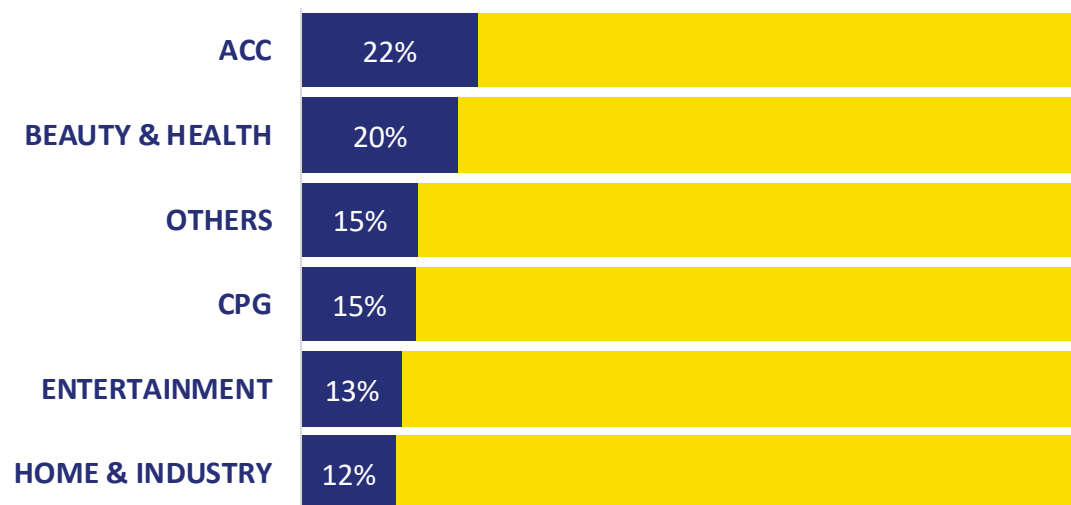
+44 Mil ofertas realizadas

+264 Mil itens vendidos

+1.2 MM de receita gerada



De aproveitamento de estoque



Top 6 verticais com maiores aproveitamentos de estoque.

TOP OFERTAS

OFERTAS COM 100% DE APROVEITAMENTO

3.8 mil ofertas realizadas tiveram 100% de aproveitamento. (8,7%)

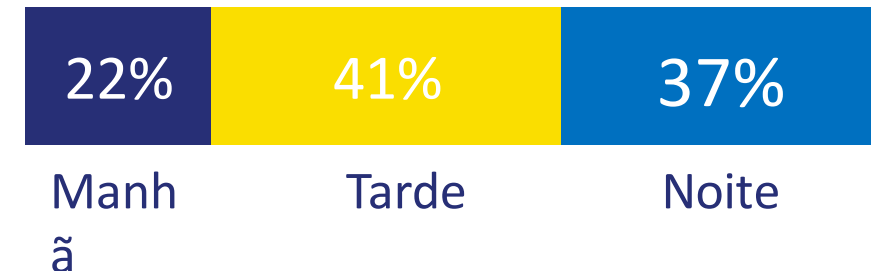
Dessas, 75% possuíam estoques de **até 15 itens.**

264.2 Mil de receita total gerada. (22%)

32,4% das ofertas foram com frete grátis.

Em média, a duração das ofertas foram de **3h e 48min.**

VERTICAL	(%)
APP & SPORTS	23%
BEAUTY & HEALTH	22%
HOME & INDUSTRY	22%



TOP OFERTAS

VENDAS & RECEITAS

Itens de Destaque:

- **Luvras Descartáveis** (Relevância em vendas e receita)
- **Máscaras Industriais** (Relevância em vendas e receita)
- **Celulares** (Relevância em Receita)

Vertical	Vendas	Receita
BEAUTY & HEALTH	177.3 Mil	547.4 Mil
HOME & INDUSTRY	26.5 Mil	155.5 Mil
APP & SPORTS	25.1 Mil	131.1 Mil
CE	21.6 Mil	362.7 Mil
CPG	5.4 Mil	16.7 Mil
ACC	4.9 Mil	26.8 Mil
T & B	2.2 Mil	14.3 Mil
ENTERTAINMENT	357	1.9 Mil
OTHERS	354	1.9 Mil

VENDAS E RECEITAS POR VERTICAL

FRETE GRÁTIS

24.4 mil ofertas realizadas foram com Frete Grátis. (54%)

Frete Grátis foi um fator que impulsionou as seguintes métricas?

- Vendas Médias ✗
- Aproveitamento de Estoque ✗
- Receita Média ✓

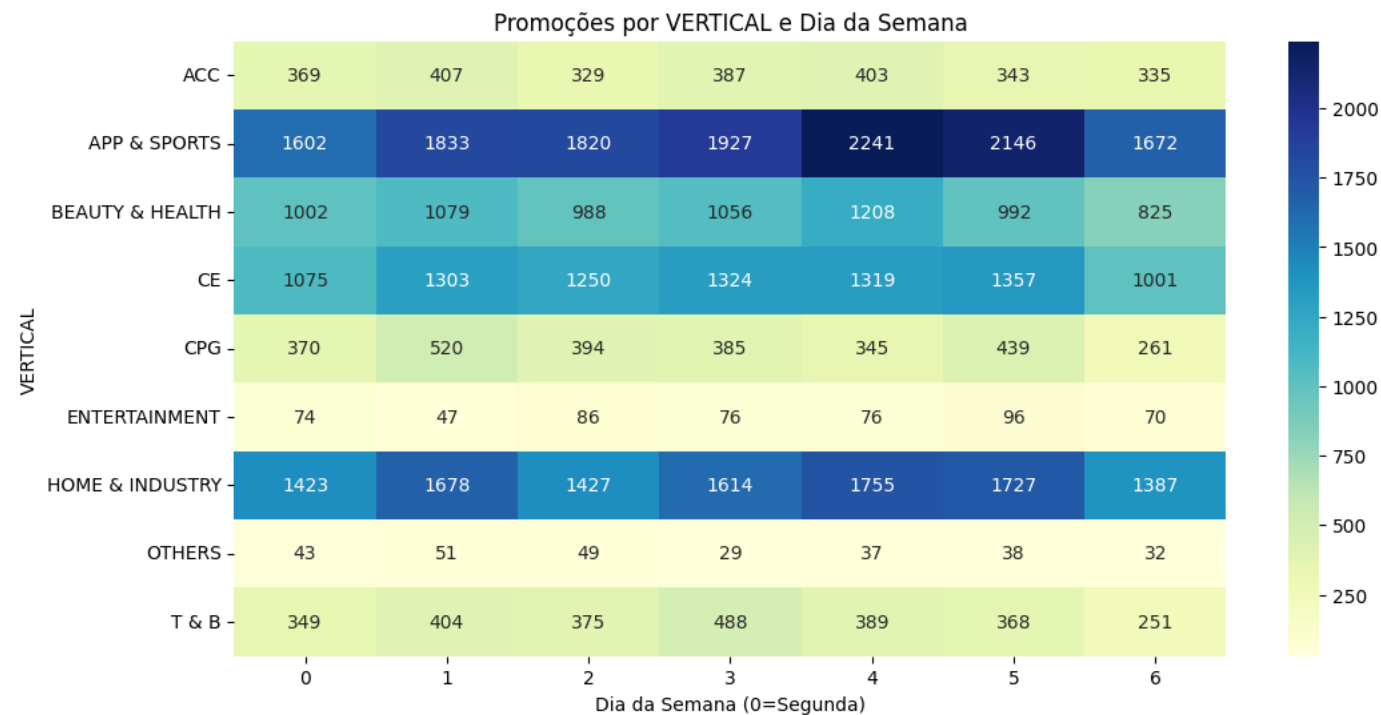
IMPACTO DO PERÍODO DA OFERTA

Alocação das Ofertas Relâmpago

Como alocar as ofertas nos dias e períodos corretos?

Nota-se uma boa distribuição das ofertas entre os diferentes dias de semana.

- App & Sports com presenças em Sextas e Sábados.
- Beauty & Health com presenças em Terças, Quintas e Sextas



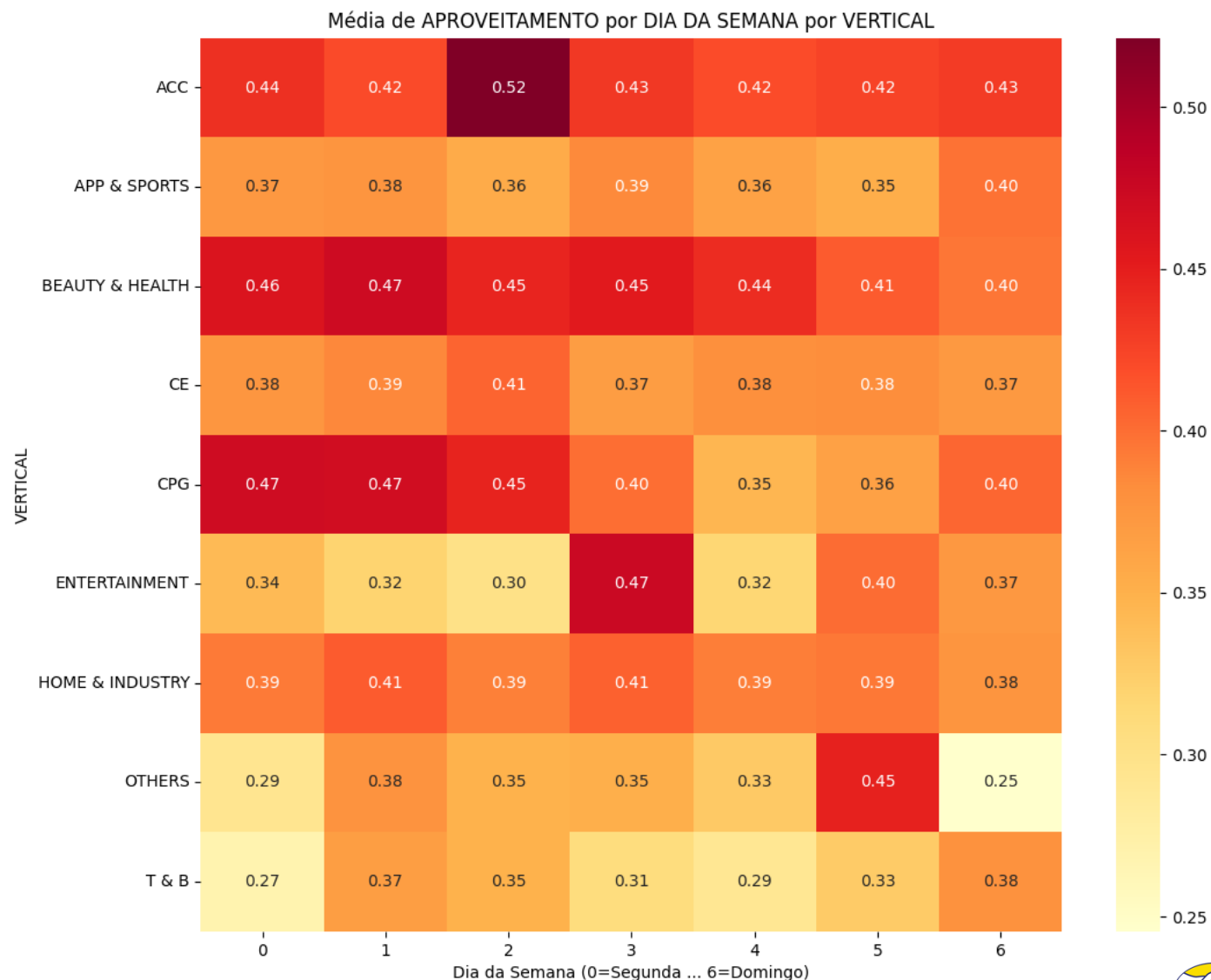
IMPACTO DO PERÍODO DA OFERTA

Aproveitamento Ofertas Relâmpago por Vertical

Ao olhar o aproveitamento do estoque, temos dias que se destacam para cada vertical:

- ACC: Quartas
- APP & SPORTS: Domingos
- ENTERTAINMENT: Quintas

Ou seja, cada vertical pode ter performances diferentes dependendo do dia de semana. Mas será que o item vendido interfere?



IMPACTO DO PERÍODO DA OFERTA

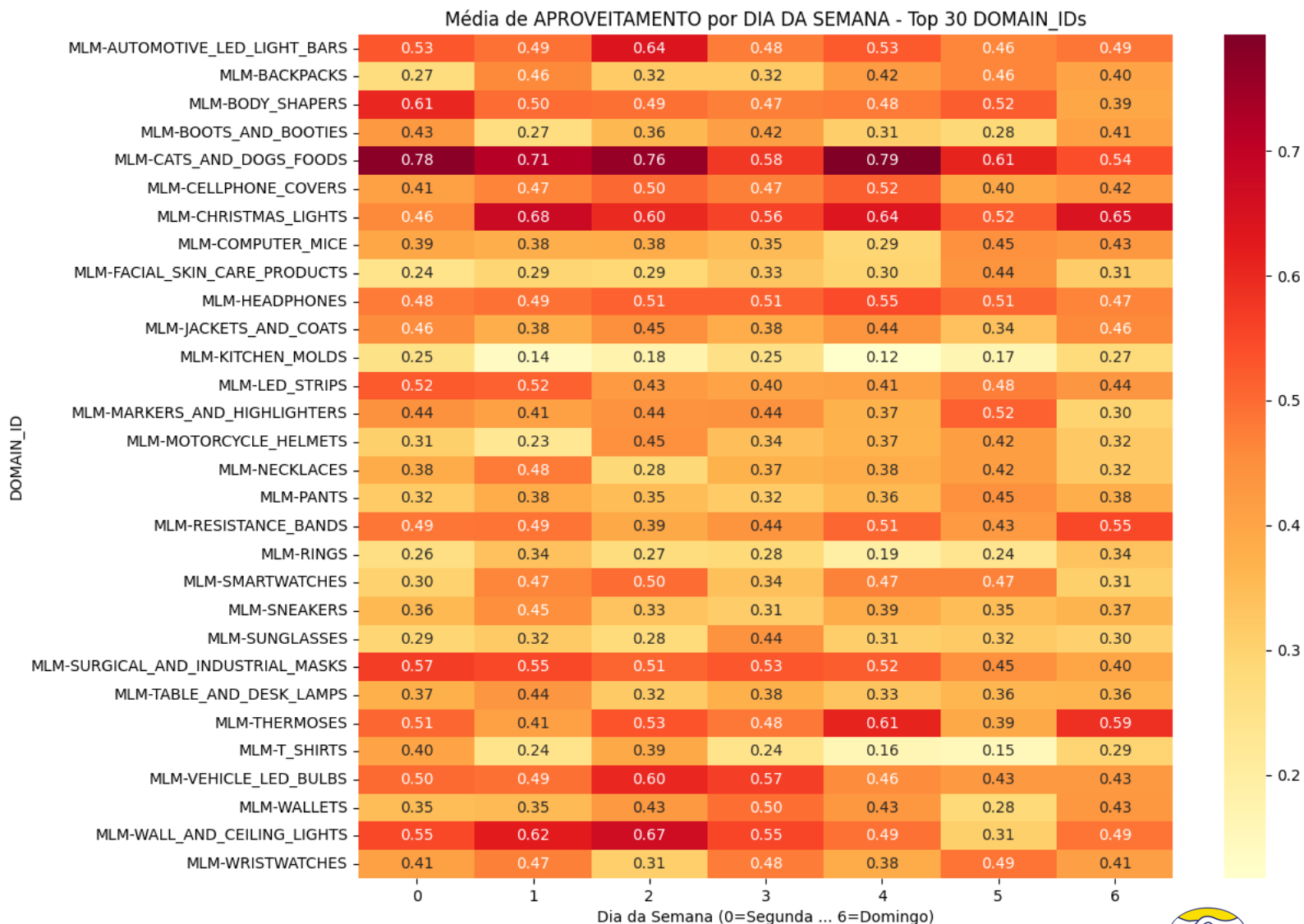
Aproveitamento Ofertas Relâmpago por DomainID

Podemos olhar também item a item:

- Body-Shapers: Segundas
- Headphones: Sextas

Ou seja, é possível entender em que dia da semana cada item performa melhor.

Podemos ter essa visão por dia de semana e por período do dia.



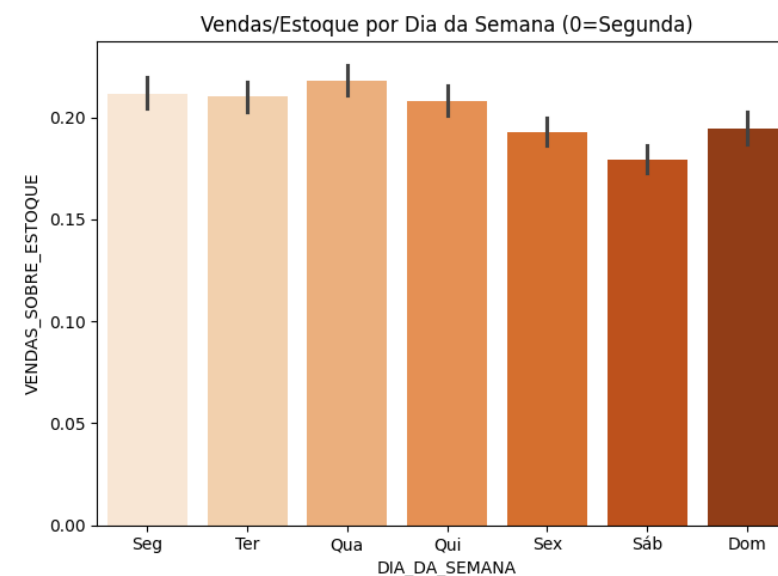
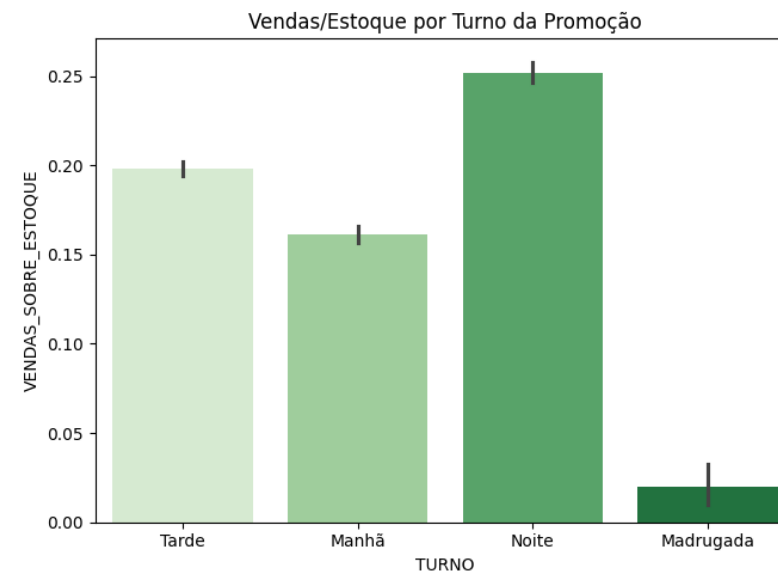
IMPACTO DO PERÍODO DA OFERTA

Performance por Dia de Semana e Turno

Foram avaliados de um olhar macro a influência de fatores temporais nos desempenhos das ofertas realizadas.

Através de aplicação de testes de hipóteses, temos as seguintes conclusões:

- Promoções lançadas no turno da Noite tem maior aproveitamento médio.
- Sábado é o dia que mais se difere em relação ao aproveitamento do estoque. Os demais não possuem diferenças estatisticamente significantes.



PROJEÇÕES DE RECEITA

+1.25 MM de receita
gerada

Porém, temos 1,1% de ofertas que, sabemos a quantidade de vendas mas não sabemos o valor de receita arrecadado.

Baseado no preço médio de um item do DomainID, podemos construir um Intervalo de confiança de 95% para o total de receita gerada.

- Pessimista: 1.29MM
- Esperado: 1.30MM
- Otimista: 1.31MM

Previsão de Falha

Modelagem

CONTEXTO

Manutenção Preditiva

- Avalia a **condição do equipamento** para prever quando a manutenção é necessária.
- Uma **interrupção por falha** custa, em média, **o dobro de uma interrupção preventiva**.
- Permite **prever a probabilidade de falha** antes que ocorra.
- Possibilidade de **calcular a economia** gerada pelo modelo.

Objetivo: prever falha **com 1 dia de antecedência**.

Alinhamento temporal: y_{t+1} a partir de $X_t, X_{t-1}, X_{t-2}, \dots$

PRÉ-PROCESSAMENTO

ENTENDIMENTO DAS VARIÁVEIS E NOVAS FEATURES

Preparação das Variáveis para Modelo

- **Janela de observação:** Optou-se por avaliar por máquina e aprender apenas com dados até 4 dias antes da possível falha.
- **Variáveis Lag:** criadas para alguns atributos (Atributo1 e Atributo6) com lags de **d-1, d-2 e d-3**.
- **Tendência:** variável de tendência construída para o **Atributo6, porém descartada**.
- **Redução de redundância:** Atributos 7 e 8 são iguais → **um foi descartado**.
- **Variáveis únicas:** Atributo 9 tinha um valor diferente para cada máquina.
- **Normalização dos valores dos atributos.**
- **Balanceamento de dados:** mesmo após limpeza, os dados ficaram **desbalanceados** → uso de **SMOTE** para gerar **amostras sintéticas**.

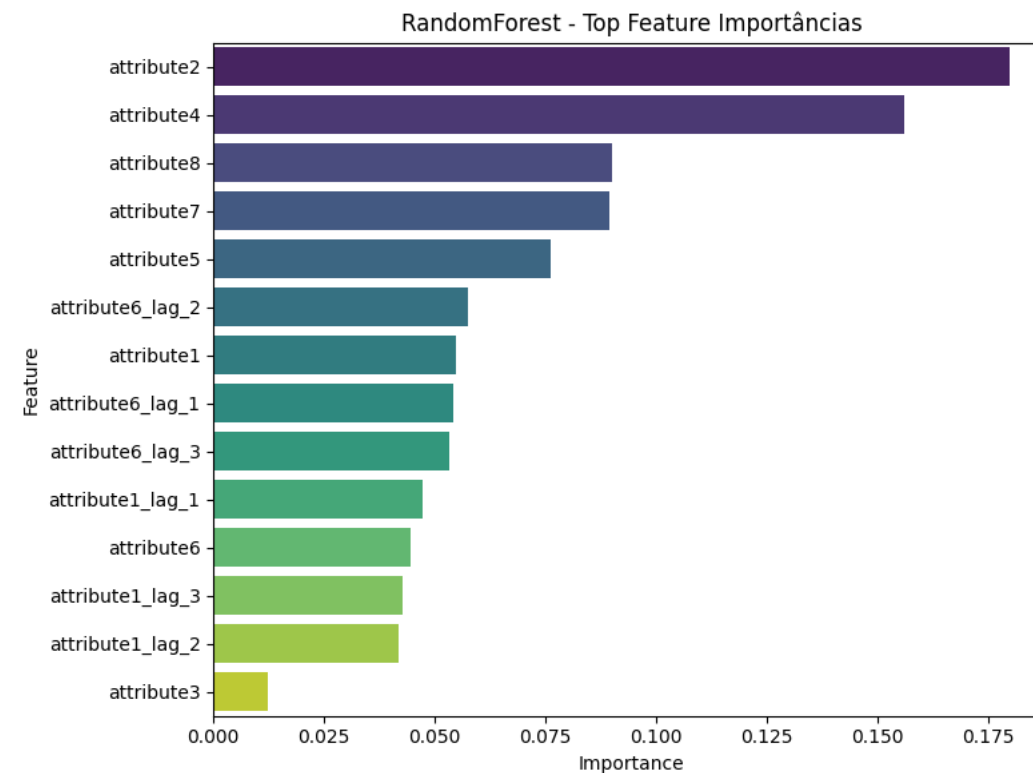
MODELAGEM

TREINAMENTO E AVALIAÇÃO DE MODELOS

- Treino (70%) e Teste (30%)
- Curva Roc para otimização do ponto de corte.

O modelo Random Forest apresentou os melhores resultados na amostra teste:

Modelo	(%Acurácia)	(%Precisão)	(%Recall)	CUSTO
REGRESSÃO LOGÍSTICA	83%	25%	50%	29,5
RANDOM FOREST	86%	34%	73%	18
XGBOOST	83%	30%	73%	23



Importância de atributos

PROJEÇÃO DE ECONOMIA

EM MANUTENÇÃO PREVENTIVA

Regras de custo por máquina:

- Prevê falha & falha ocorre → economiza **50%** do custo.
- Prevê falha & falha não ocorre → gasta **50%** à toa.
- Não prevê falha & falha ocorre → gasta **100%** do custo.
- Não prevê falha & falha não ocorre → não gasta nada.

- Se esperasse falhar, o custo seria de: **30**;
- Se usasse o modelo o custo seria de: **18**;
- Economia de **40%**

Matriz de Confusão Random Forest

278	42 +50%
8 +100%	22 -50%

Obrigado!