



SUPERSTORE

modelo de mitigação de prejuízos

Por: Pedro Kaneko



Conteúdo

▶ <u>Introdução</u>	3
▶ <u>Objetivo</u>	4
▶ <u>Descrição do problema</u>	5
▶ <u>Análise</u>	6
▶ <u>Proposta</u>	11
▶ <u>Resultados</u>	15
▶ <u>Recomendações</u>	19
▶ <u>Apêndice</u>	21





Introdução

A "Superstore" é um eCommerce dos EUA que vende produtos de tecnologia, mobília e material de escritório. Atende os públicos consumidor, corporativo e home-office.

A base de dados abrange informações sobre transações, clientes, produtos, categorias, localização, datas e métricas financeiras como receita e lucro.

Objetivo

Mitigar prejuízos causados por descontos excessivos em produtos através de um modelo preditivo que estima a margem de lucro antes dos produtos serem anunciados na plataforma.

Metodologia

Análise Exploratória de Dados (EDA)

Análise Estatística

Análise Preditiva (Machine Learning)

Experimentação Científica (RCT)

Validação Estatística (Teste de Fisher)



Descrição do Problema

A empresa está em fase de expansão e apresenta um aumento anual nos lucros. No entanto, constatamos que o potencial de rentabilidade está sendo significativamente comprometido por perdas expressivas, que têm sido recorrentes e que se intensificaram no último ano. Esses prejuízos derivam de transações que não apenas deixam de gerar lucro, mas também resultam em perdas, principalmente devido aos descontos elevados aplicados nos produtos.

Impacto Financeiro

No último ano de 2017 esse prejuízo chegou a 36,55% do lucro bruto que corresponde ao valor financeiro de \$ -53,836
(Crescimento de 42,2% em relação ao ano anterior)



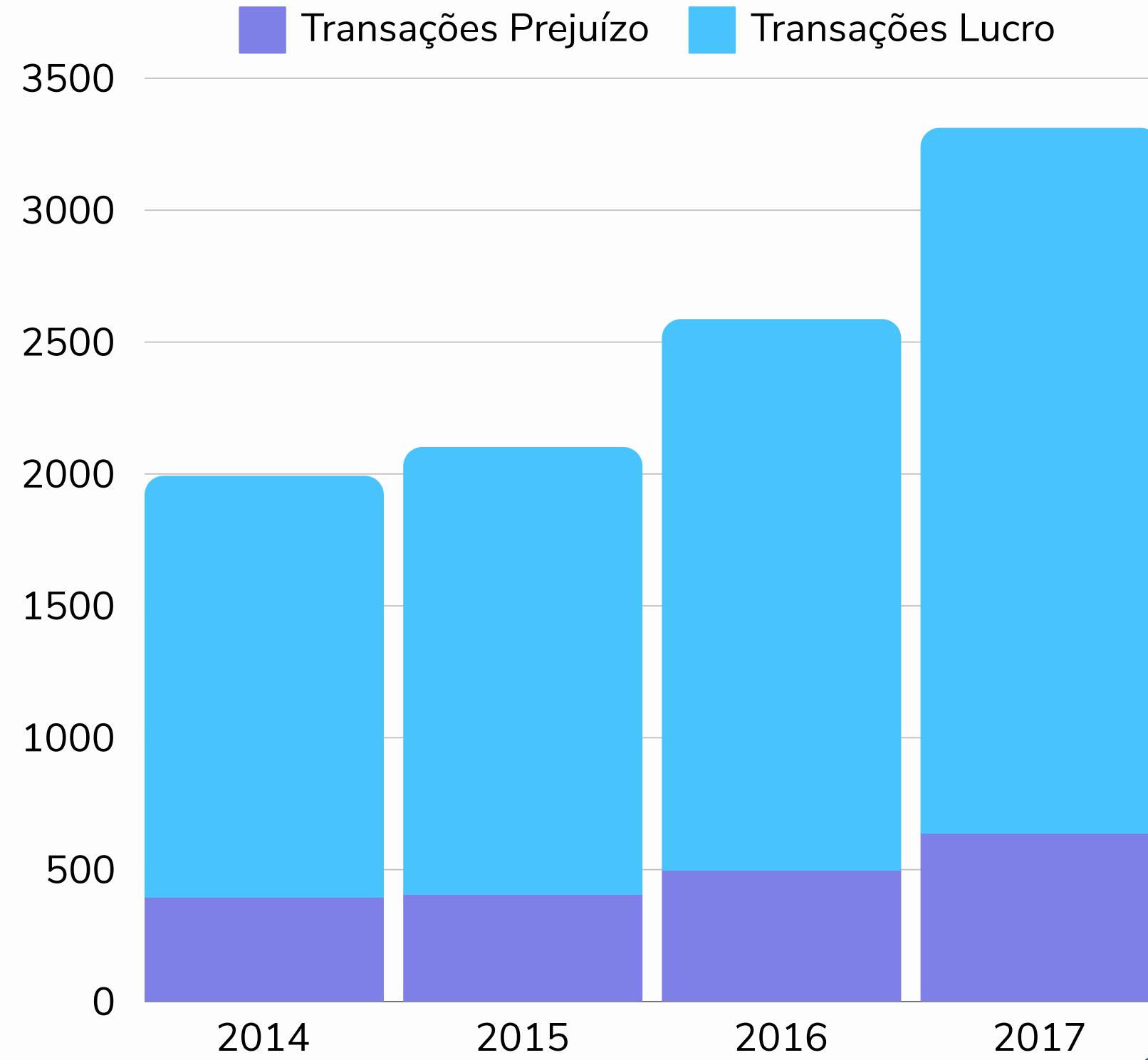
ANÁLISE

Avaliação do desempenho geral das vendas e lucros e identificação de áreas problemáticas.

Transações

O volume de transações aumentou, assim como o número de transações que resultaram em prejuízos.

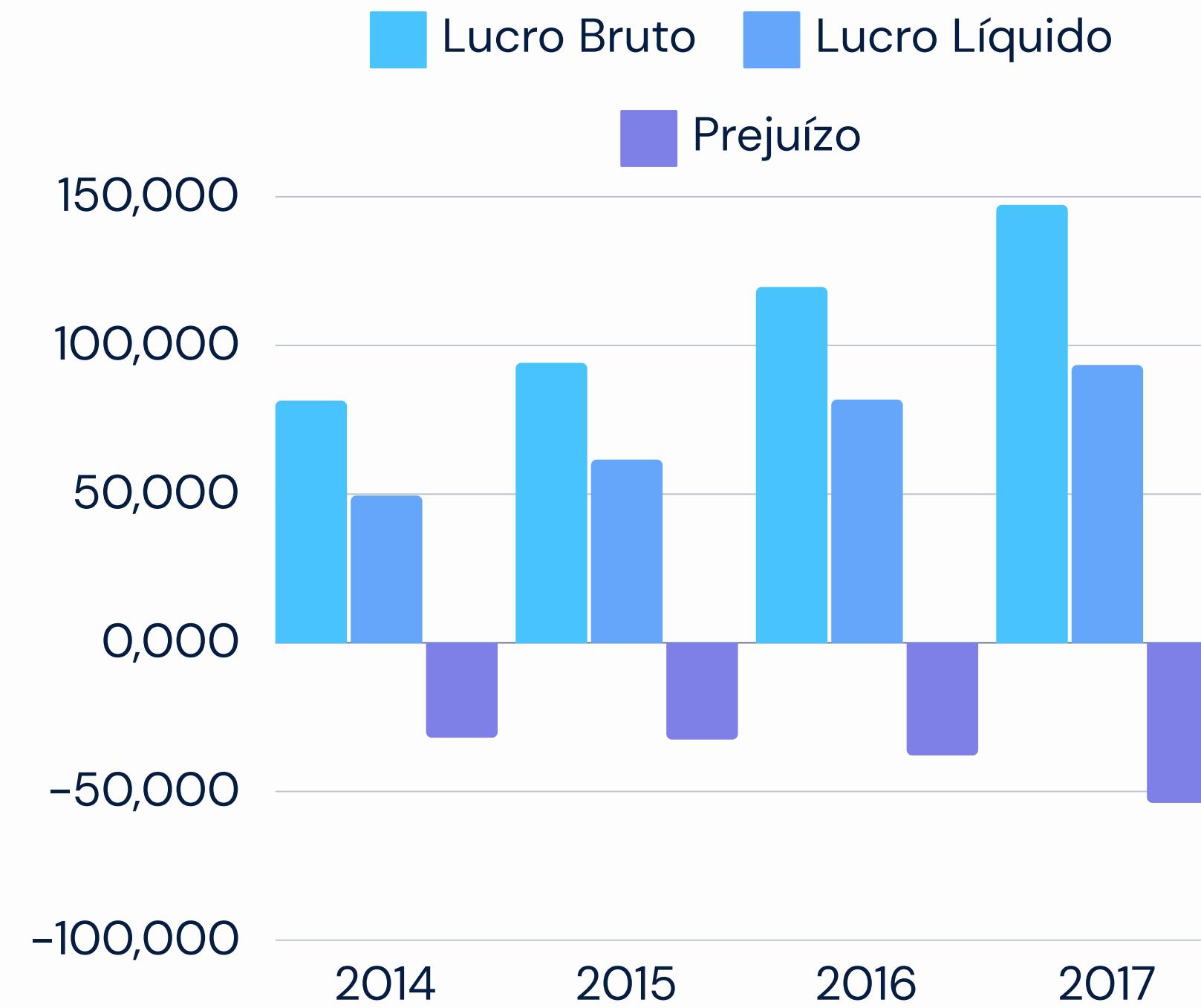
Volume de Transações	2014	2015	2016	2017
Total	1,993	2,102	2,587	3,312
Lucrativas	1,599	1,696	2,090	2,673
Prejudiciais	394	406	497	639
Prejuízo (%)	19.77	19.31	19.21	19.29



Lucro

Embora os lucros tenham aumentado ano após ano, é evidente que esses ganhos estão sendo prejudicados por perdas contínuas e crescentes.

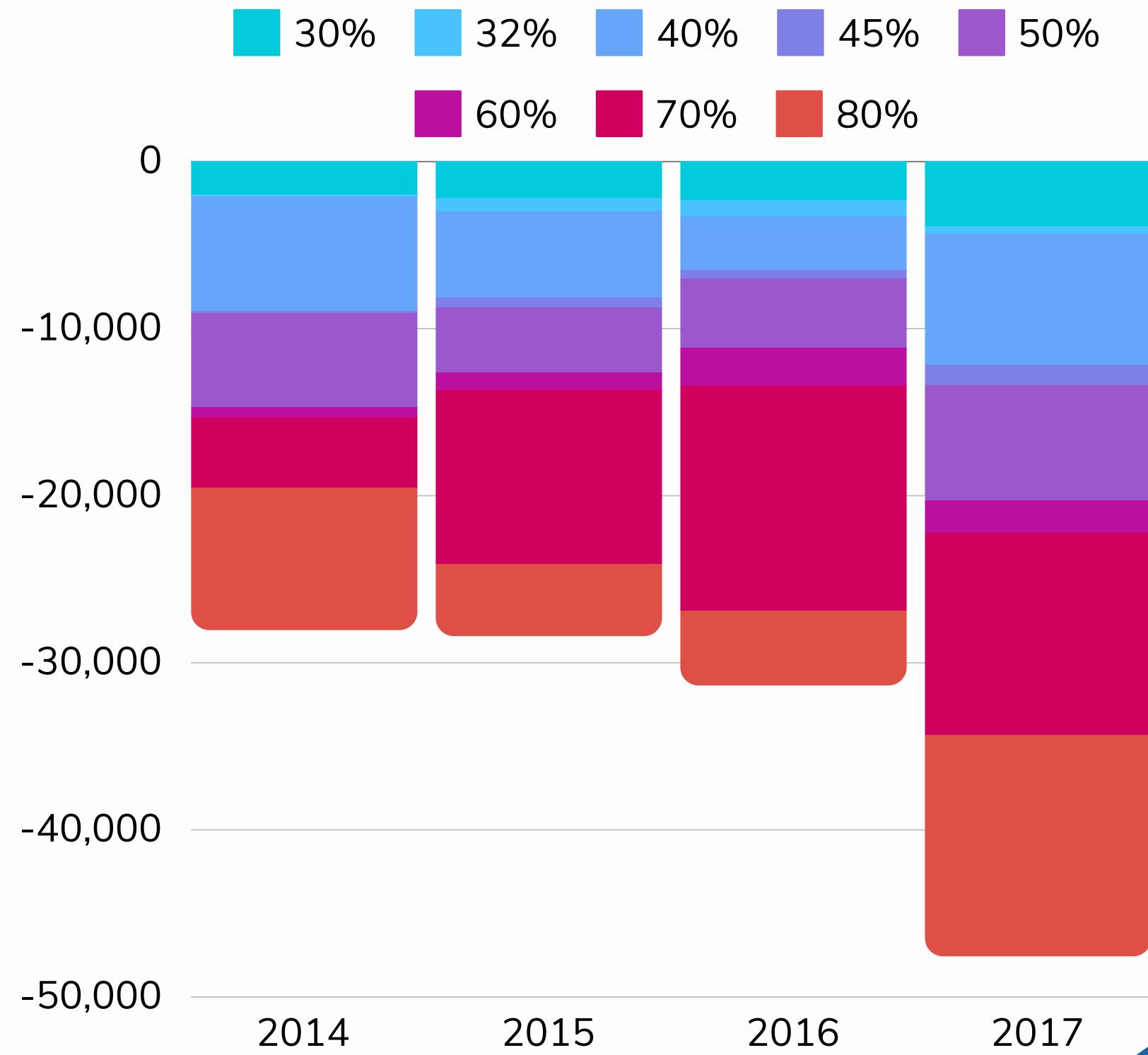
Lucro	2014	2015	2016	2017
Bruto (\$)	81,436	94,147	119,668	147,275
Líquido (\$)	49,543	61,618	81,795	93,439
Prejuízo (\$)	-31,892	-32,529	-37,872	-53,836
Prejuízo (%)	-39.16	-34.55	-31.65	-36.55



Descontos

A seguir temos o valor do prejuízo acumulado por ano e faixa de desconto. É notável o aumento do prejuízo em 2017 em especial para as faixas de desconto de 70% e 80%.

Prejuízo por Faixa de Desconto (\$)	2014	2015	2016	2017
30%	-1,981	-2,200	-2,320	-3,867
40%	-6,869	-5,154	-3,199	-7,833
50%	-5,640	-3,887	-4,114	-6,864
60%	-627	-1,051	-2,296	-1,968
70%	-4,167	-10,379	-13,457	-12,070
80%	-8,516	-4,318	-4,460	-13,244



Pedidos

Ao analisarmos os pedidos que mais geraram prejuízo em 2017, notamos uma forte ligação entre altos descontos concedidos e o prejuízo gerado. A magnitude da margem de lucro negativa evidencia a gravidade da situação e a necessidade de ajustar as políticas de desconto.

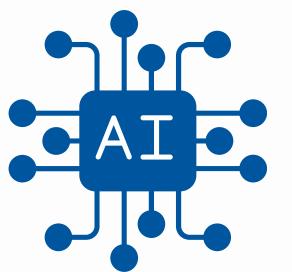
Data	Categoria	Subcategoria	Nome do produto	Quantidade	Desconto	Prejuízo	Margem
2017-11-04	Technology	Machines	Cubify CubeX 3D Printer Triple Head Print	4	50%	\$-3,839.99	-48%
2017-04-17	Technology	Machines	Lexmark MX611dhe Monochrome Laser Printer	5	70%	\$-3,399.98	-133%
2017-12-07	Office Supplies	Binders	Ibico EPK-21 Electric Binding System	5	80%	\$-2,929.48	-155%
2017-11-19	Office Supplies	Binders	Fellowes PB500 Electric Punch Plastic Comb Bin...	6	80%	\$-2,287.78	-150%
2017-10-19	Office Supplies	Binders	GBC DocuBind P400 Electric Binding System	4	70%	\$-1,306.55	-80%

PROPOSTA

Mitigar prejuízos causados por descontos excessivos em produtos através de um modelo preditivo.

Solução Proposta

Estimar a expectativa de margem de lucro dos produtos com base nos preços e descontos por meio de um modelo preditivo de Machine Learning, antes dos produtos de serem anunciados no e-commerce.



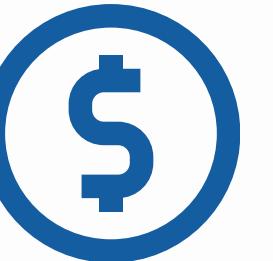
Previsibilidade

Estimar a margem de lucro dos produtos ajuda a estabelecer preços e descontos de forma precisa e evita prejuízos.



Segurança

Considerar que a previsão pode ter variações criando um intervalo de confiança garante uma camada extra de segurança.



Rentabilidade

Validar apenas as estimativas que resultem em 5 % de margem de lucro ou mais, evita transações arriscadas de baixa lucratividade.



Como Funciona

O modelo foi desenvolvido para ser utilizado tanto em lote como para consultas individuais. Seu funcionamento é descrito a seguir.

01

O modelo foi desenvolvido para **estimar a margem de lucro** de um produto com base em sua categoria, subcategoria, valor e desconto atribuído.

02

O **intervalo de confiança** considera a variação mínima das previsões garantido uma camada extra de segurança.

03

Com o **indicador de risco** o modelo evita transações com baixa lucratividade e arriscadas. Validando apenas estimativas mínimas com 5% ou mais de margem de lucro.

04

A saída do modelo inclui a projeção da margem de lucro mínima e prevista, e o indicador de risco classifica a transação como segura ou arriscada.

Consulta

Exemplo de uma consulta realizada utilizando um sistema de interação desenvolvido no notebook técnico com o modelo preditivo treinado.

Dados de entrada para realizar a consulta:

- Categoria (Seleção)
- Subcategoria (Seleção)
- Valor do produto (Inserir valor)
- Valor do desconto (Inserir valor)

▶ estimar_margem_lucro()

→ Escolha uma opção:
1. Furniture
2. Office Supplies
3. Technology
Número da opção: 1
Categoria selecionada: Furniture

Escolha uma opção:
1. Bookcases
2. Chairs
3. Tables
4. Furnishings
Número da opção: 3
Subcategoria selecionada: Tables

Insira o valor do produto: 208.70
Insira o percentual de desconto (0 à 99): 50

PREVISÃO DE LUCRO E RISCO	VALOR
Margem de Lucro Prevista	-58.25 %
Margem de Lucro Mínima	-64.03 %
Risco	Arriscado
DETALHES DA CONSULTA	
Valor do Produto	\$ 208.7
Desconto Aplicado	50.0 %
Categoria	Furniture
Sub-categoria	Tables

* Previsões de margem de lucro mínima inferiores a 5% são classificadas como arriscadas.

Consulta

Saída

Margem de lucro real do produto consultado: -80%

RESULTADOS

O sucesso do modelo preditivo é medido pela redução percentual de transações prejudiciais.

Resultado do Teste

O modelo proposto foi capaz de reduzir 98.6% das transações que resultam em prejuízos, levando a taxa atual de transações que resultam em prejuízos de 20% para 0.61% em ambiente de teste.



**Redução de 98.6 %
nas transações que
geram prejuízo**

Resultados Práticos

Estimativa do número de transações prejudiciais 2018

Número estimado médio de transações potencialmente prejudiciais para 2018

Mínimo: 653
Máximo: 854

754

Número de transações prejudiciais reduzido

Número estimado reduzido de transações potencialmente prejudiciais para 2018 utilizando o modelo preditivo.

*Redução: **99.2%**

6

Resultados Práticos

Prejuízo médio estimado (2018)

Prejuízo médio estimado para 2018 com base no crescimento médio histórico.

Mínimo: \$ -51,545

Máximo: \$ -77,867

\$ -64,706

Prejuízo Reduzido (2018)

Prejuízo médio estimado para 2018 utilizando o modelo preditivo

Redução: 99.2%

Mínimo: \$ -412

Máximo: \$ -623

\$ -518

Prejuízo médio por transação

Prejuízo médio derivado do valor do prejuízo médio por transação.

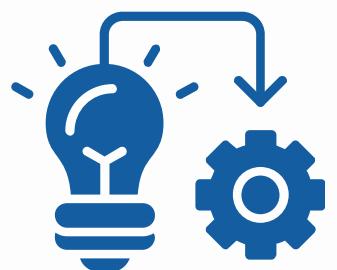
Mínimo: \$ -78.8

Máximo: \$ -91.1

\$ -85.8

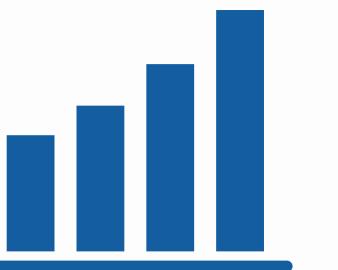
Recomendações

Sugerimos a realização de experimentos controlados para validar o modelo em um ambiente de produção, analisar os resultados e avaliar o impacto na redução de prejuízos (passo a passo disponível no notebook técnico).



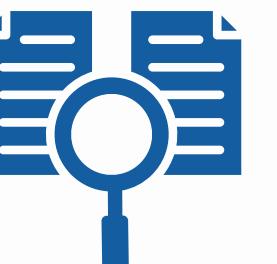
Experimento

Antes de implementar o modelo, é recomendável realizar um Teste Randomizado Controlado (RCT) em um ambiente de produção.



Validação

Realize um teste estatístico, como o Teste de Fisher, para comparar as amostras do experimento e garantir validade estatística ao resultado.



Monitoramento

Monitore e ajuste continuamente o modelo com base em novos dados e feedback do ambiente de produção.





OBRIGADO



Pedro Kaneko

Analista & Cientista de Dados



pedrokaneko.ds@gmail.com



www.linkedin.com/in/pedrokaneko



São Paulo, Brasil



APÊNDICE

Informações de apoio e esclarecimentos técnicos

Regressão

Verificando o desempenho do modelo de regressão nos conjuntos de validação e teste para garantir uma boa generalização e evitar underfitting e overfitting.

Métrica	Validação	Teste
R ²	0.976	0.979
Cross Val R ²	0.949	0.951
MAE	0.035	0.035
Cross Val MAE	0.058	0.064
RMSE	0.074	0.073
Cross Val RMSE	0.106	0.106

Coeficiente de determinação (R²):

Alto, indica boa capacidade explicativa do modelo.

Erro absoluto médio (MAE) e raiz do erro quadrático médio (RMSE):

Baixos, indicam que as previsões são precisas e os erros são pequenos. Essas métricas possuem a mesma grandeza da variável alvo podendo ser facilmente interpretáveis e por esse motivo foram escolhidas.

Validações Cruzadas:

Confirmam a robustez e a consistência do modelo através de diferentes subconjuntos de dados.

Observação:

O intervalo de confiança aplicado no conjunto de testes e para projeções futuras, é calculado com base no valor do erro absoluto médio (MAE) obtido no conjunto de validação. Além disso, o RMSE pode ser empregado para resultados mais conservadores, e a margem de lucro mínima pode ser ajustada conforme as necessidades.

Classificação

Transformamos as previsões contínuas do modelo em categorias como "margem de lucro positiva" ou "margem de lucro negativa" para facilitar a análise.

Métrica	Validação	Teste
Precisão	97.85 %	99.2 %
FPR	1.74 %	0.61 %
FNR	1.29 %	4.95 %

O modelo de regressão foi ajustado para prever cenários de lucro ou prejuízo, dessa maneira o sucesso do modelo preditivo é medido pela redução percentual de transações prejudiciais. As principais métricas utilizadas neste caso são:

- **Precisão:** Mede a proporção de previsões corretas de lucro em relação ao total de previsões de lucro, ajudando a entender quantas dessas previsões são realmente benéficas.
- **Taxa de Falsos Positivos (FPR):** Indica a frequência de previsões de lucro que, na realidade, resultam em prejuízo.
- **Taxa de Falsos Negativos (FNR):** Indica a frequência com que o modelo falha em identificar oportunidades de lucro, classificando-as incorretamente como prejuízo. Vale ressaltar que o modelo descarta também previsões com baixa lucratividade por conta do indicador de risco criado.

Performance das previsões

