

# Segmentação de Risco de Margem no Varejo

## *Relatório executivo para tomada de decisão*

Autor: Matheus Khatib do Amaral | Data: 2026-01-20

### Resumo Executivo

Este projeto analisa pedidos de um varejista com o objetivo de identificar, antes da venda, quais pedidos apresentam maior risco de prejuízo e orientar políticas de desconto e envio mais eficientes.

A análise encontrou dois grupos de pedidos com comportamento financeiro distinto:

- Pedidos Estáveis (68,6%): margens consistentemente positivas (margem média: 23,5%).
- Pedidos de Alto Risco (31,4%): concentram 40% das margens negativas (margem média: -13,2%).

No período analisado, pedidos de alto risco somaram \$677 em perdas. Uma política direcionada a esse grupo tem potencial de reduzir aproximadamente 67,86% dessas perdas.

### Problema de Negócio

No varejo, regras de desconto e envio muitas vezes são aplicadas de forma uniforme. Isso pode gerar perdas quando descontos elevados e custos operacionais se combinam em pedidos com baixa capacidade de sustentar margem. A pergunta central foi: é possível antecipar pedidos com alta probabilidade de margem negativa?

### Abordagem

Foi desenvolvido um modelo de segmentação de risco que estima, para cada pedido, a margem esperada e a probabilidade de margem negativa, usando variáveis operacionais (desconto, categoria/subcategoria, região e envio).

Fluxo de trabalho:

- Limpeza e estruturação dos dados; criação de variáveis (localização, categoria, envio e descontos).
- Benchmark de modelos (Lasso/Ridge, GAMLSS t, XGBoost/LightGBM).
- Modelo final: mistura de regressão com dois clusters e validação por bootstrap (1000 reamostragens).

### Resultados e Impacto

Grupo	% dos pedidos	Margem média	Participação nas perdas
Estável	68,6%	23,5%	Residual
Alto Risco	31,4%	-13,2%	40% das margens negativas

Recomendação executiva: aplicar intervenções apenas no grupo de alto risco (por exemplo, revisão de descontos e regras de envio), preservando a maior parte das vendas saudáveis.

## Principais Drivers de Alto Risco

Os principais fatores associados ao cluster de alto risco incluem: desconto, subcategorias Mesas e Estantes, e a região Mountain. O risco emerge da combinação desses fatores, não de um único indicador isolado.

## Recomendações (30 dias)

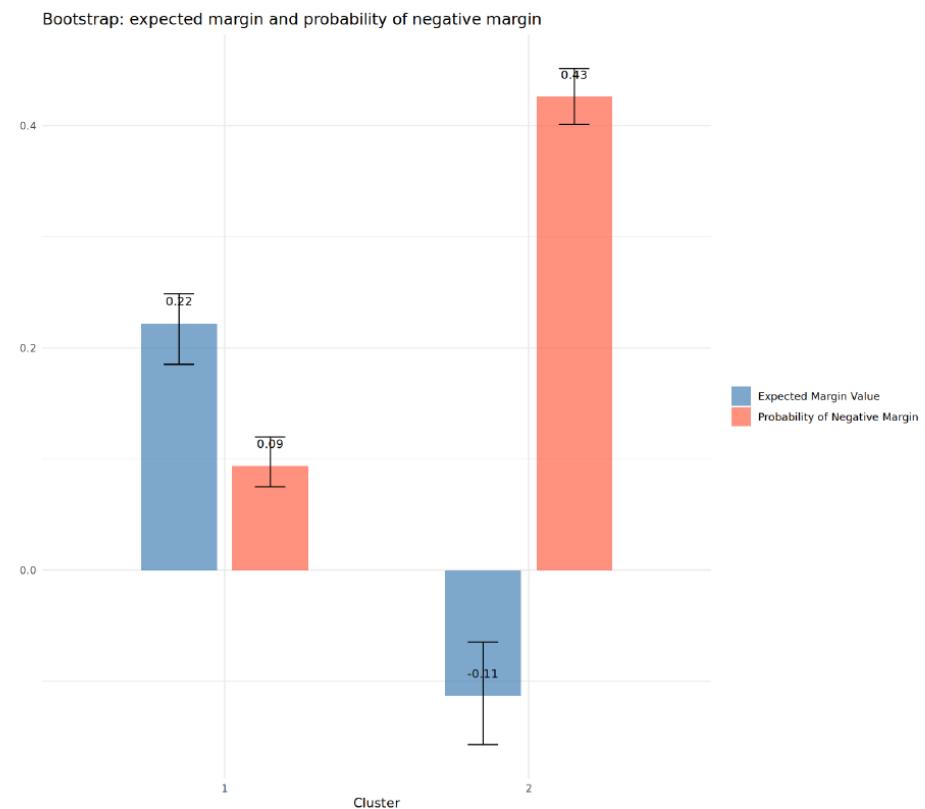
1. Implementar alerta/flag de pedidos com alta probabilidade de margem negativa.
2. Revisar política de descontos (prioridade: descontos elevados em subcategorias e regiões de alto risco).
3. Monitorar impacto em margem e volume, ajustando regras com base no desempenho por cluster.

## Visualizações e Evidências

As figuras a seguir resumem o trade-off risco vs retorno, os fatores operacionais por cluster e a validação do modelo.

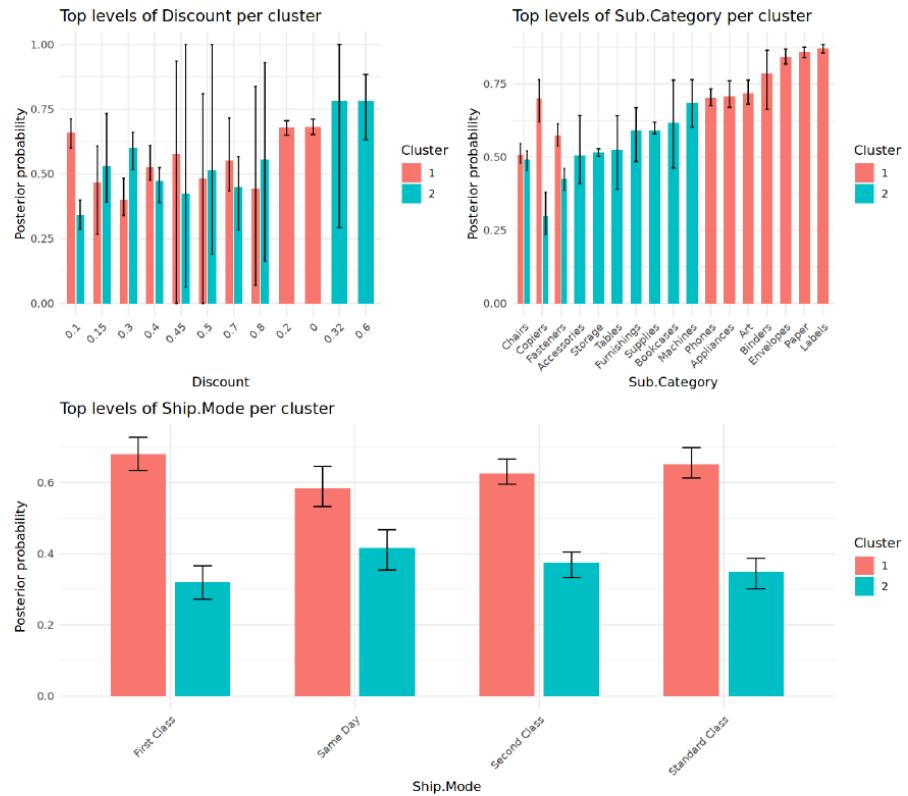
## Visualizações

### 1. Trade-off risco vs. retorno entre clusters



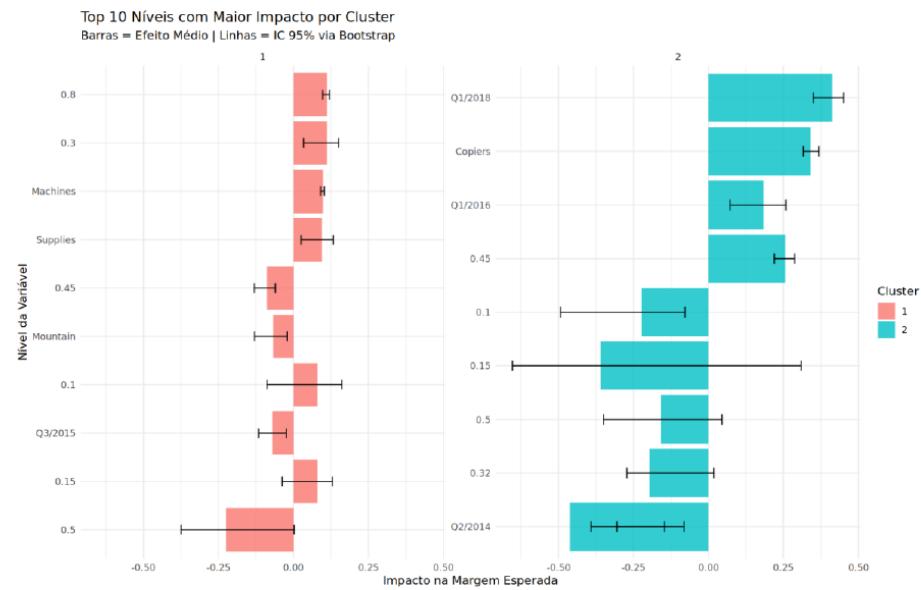
### 2. Fatores operacionais de alta performance

Figura 1. Trade-off risco vs retorno (margem esperada vs probabilidade de margem negativa).



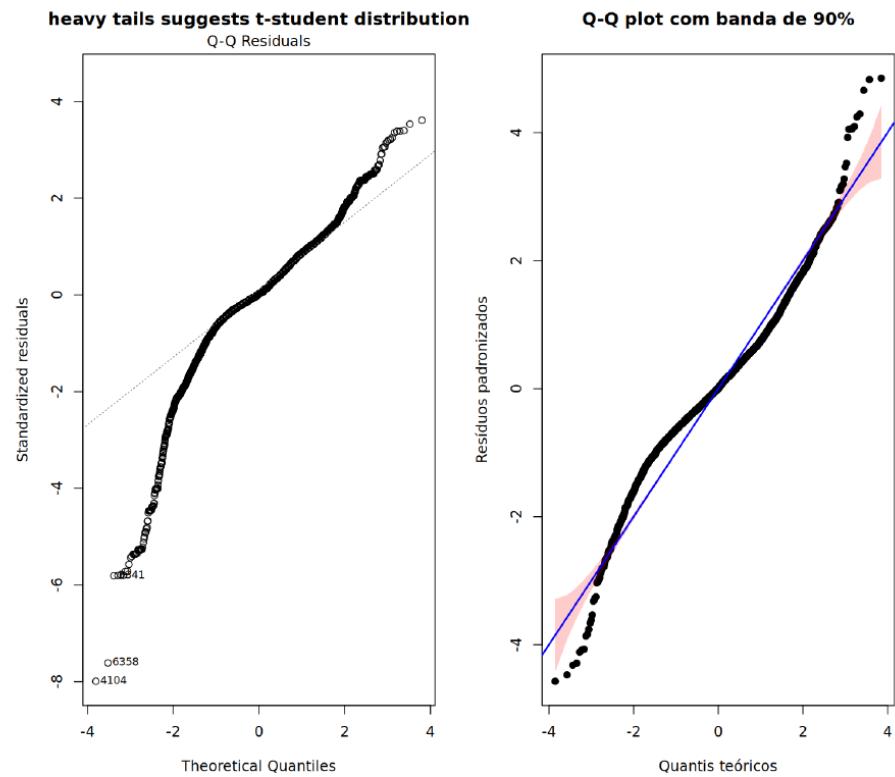
3. 10 níveis mais impactantes por cluster

Figura 2. Fatores operacionais e níveis mais impactantes por cluster.



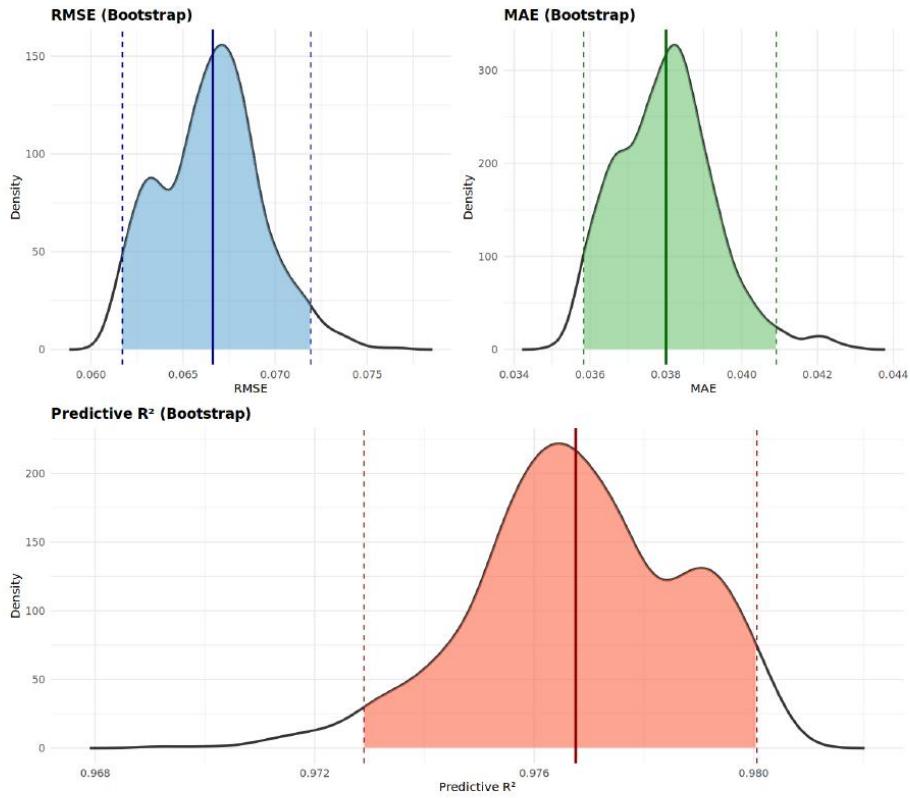
4. Comparação de resíduos entre modelos

*Figura 3. Comparação de resíduos entre modelos.*



5. Validação via bootstrap

*Figura 4. Validação via bootstrap (Q-Q plot e banda).*



6. Principais fatores de custo no Cluster 1

Figura 5. Distribuições de métricas no bootstrap (RMSE, MAE e  $R^2$ ).

	Variavel	Nivel	Lucro_Medio	Contagem
	<chr>	<chr>	<dbl>	<int>
1	Discount	0.5	-340.	53
2	Discount	0.45	-206.	9
3	Discount	0.8	-111.	259
4	Discount	0.4	-108.	178
5	Discount	0.7	-92.7	353
6	Discount	0.32	-92.6	24
7	Discount	0.3	-44.7	194
8	Sub.Category	Tables	-44.6	264
9	Discount	0.6	-40.2	123
10	division	West South Central	-14.7	977
11	Sub.Category	Bookcases	-13.9	190
12	Sub.Category	Supplies	-8.83	164
13	Sub.Category	Machines	-4.46	100
14	division	Mountain	-3.55	498

Figura 6. Principais combinações/níveis associados a perdas (ranking).