Retail Margin Risk Segmentation

Executive report for decision-making

Author: Matheus Khatib do Amaral | Date: 2026-01-20

Executive Summary

This project analyzes a retailer's orders to identify, before the sale, which orders have a higher risk of losses and to guide more effective discount and shipping policies.

The analysis identified two groups of orders with distinct financial behavior:

- Stable Orders (68,6%): consistently positive margins (average margin: 23,5%).
- High-Risk Orders (31,4%): concentrate 40% of negative margins (average margin: -13,2%).

In the analyzed period, high-risk orders totaled $677 in losses. A policy targeted at this group has the potential to reduce approximately 67,86% of these losses.

Business Problem

In retail, discount and shipping rules are often applied uniformly. This can generate losses when high discounts and operational costs combine in orders with low ability to sustain margin. The central question was: is it possible to anticipate orders with a high probability of negative margin?

Approach

A risk segmentation model was developed to estimate, for each order, the expected margin and the probability of negative margin, using operational variables (discount, category/subcategory, region, and shipping).

Workflow:

- Data cleaning and structuring; creation of variables (location, category, shipping, and discounts).
- Model benchmarking (Lasso/Ridge, GAMLSS t, XGBoost/LightGBM).
- Final model: regression mixture with two clusters and bootstrap validation (1000 resamples).

Results and Impact

Group	% of orders	Average margin	Share of losses
Stable	68,6%	23,5%	Residual
High Risk	31,4%	-13,2%	40% of negative margins

Executive recommendation: apply interventions only to the high-risk group (e.g., reviewing discounts and shipping rules), preserving most healthy sales.

Key High-Risk Drivers

The main factors associated with the high-risk cluster include discount, the Tables and Bookcases subcategories, and the Mountain region. Risk emerges from the combination of these factors, not from a single isolated indicator.
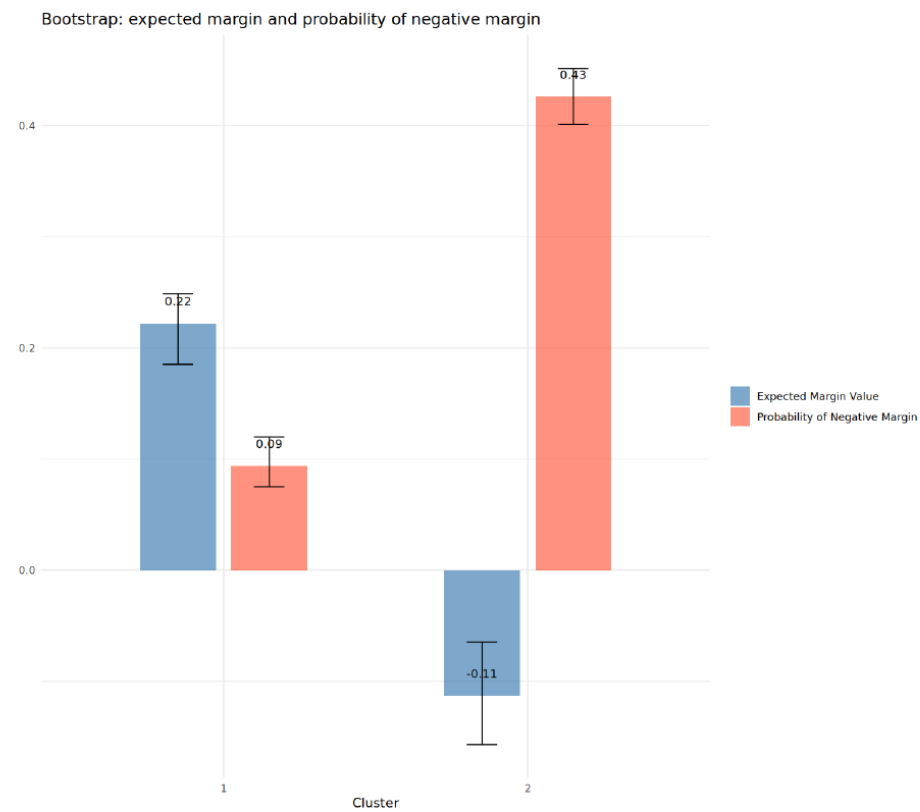
Recommendations (30 days)

1. Implement an alert/flag for orders with a high probability of negative margin.
2. Review the discount policy (priority: high discounts in high-risk subcategories and regions).
3. Monitor margin and volume impact, adjusting rules based on cluster-level performance.

Visualizations and Evidence

The figures below summarize the risk vs. return trade-off, operational factors by cluster, and model validation.
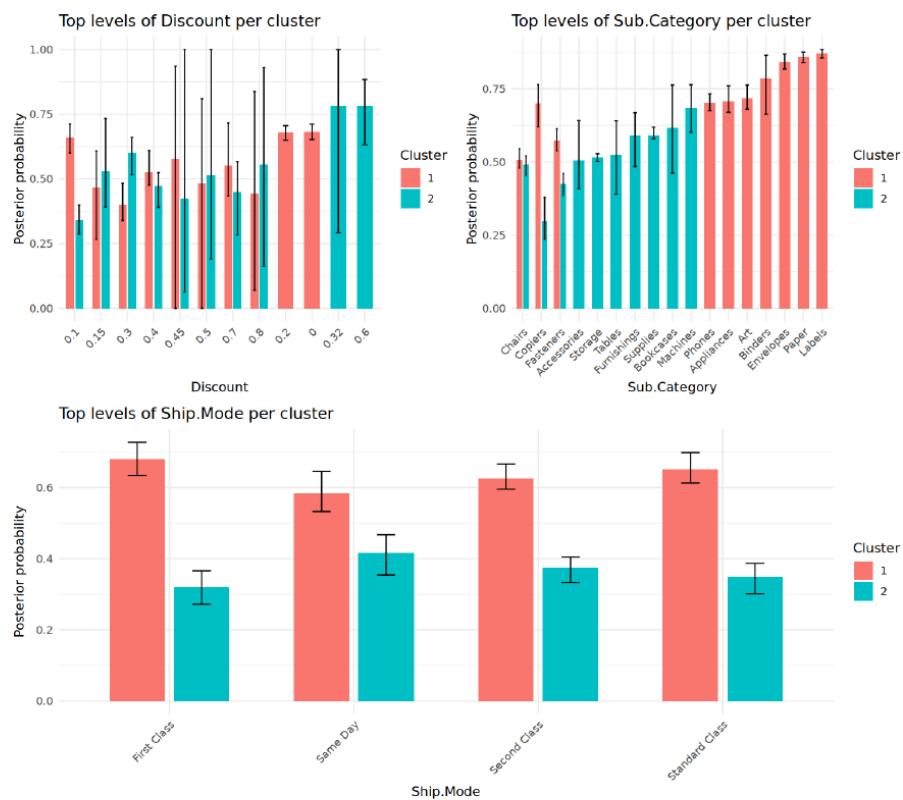
Visualizações

1. Trade-off risco vs. retorno entre clusters

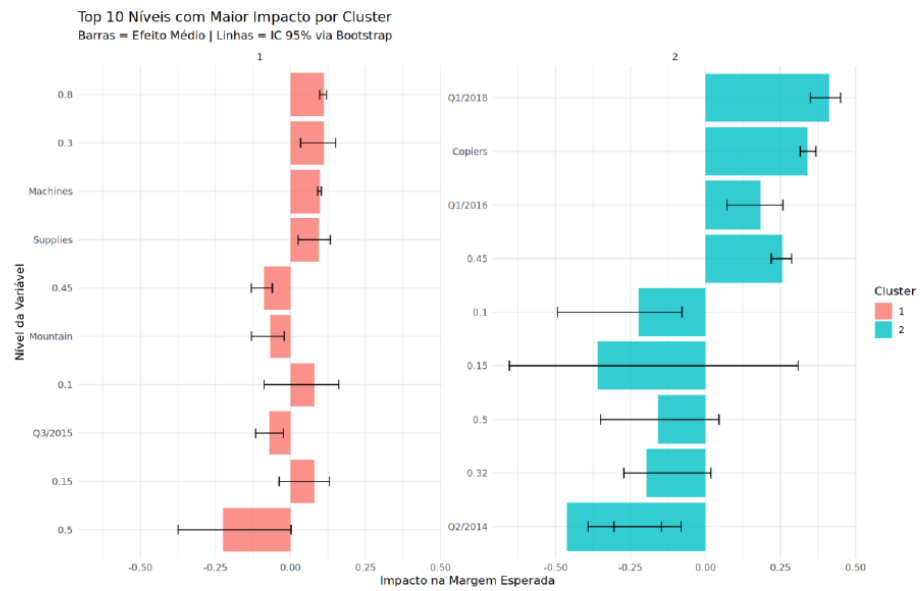Bootstrap: expected margin and probability of negative margin

2. Fatores operacionais de alta performance

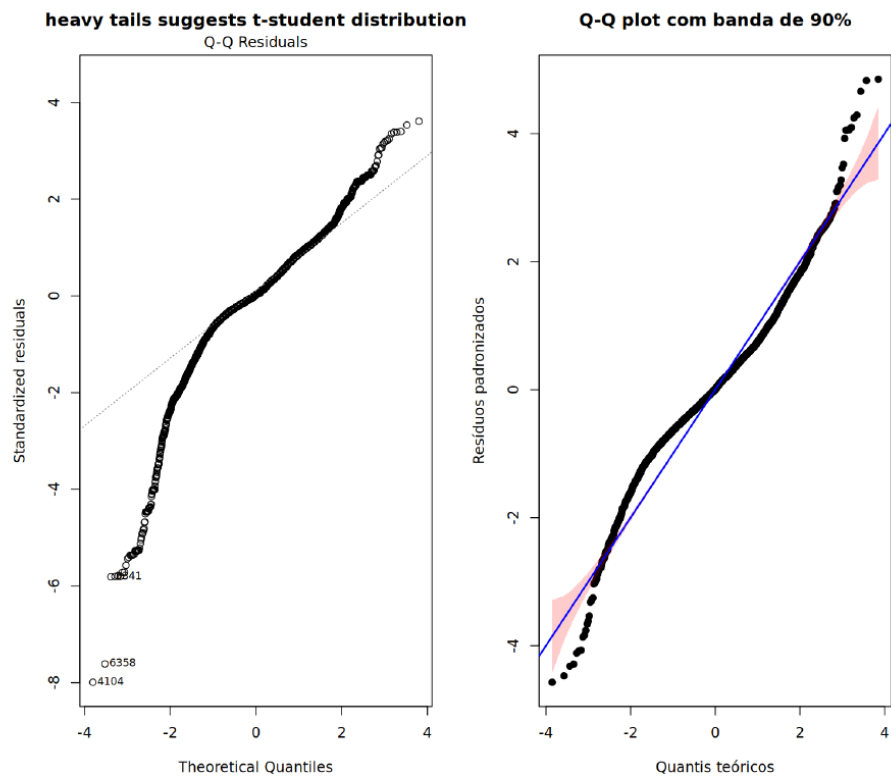Figure 1. Risk vs. return trade-off (expected margin vs. probability of negative margin).

3. 10 níveis mais impactantes por cluster

Figure 2. Operational factors and the most impactful levels by cluster.
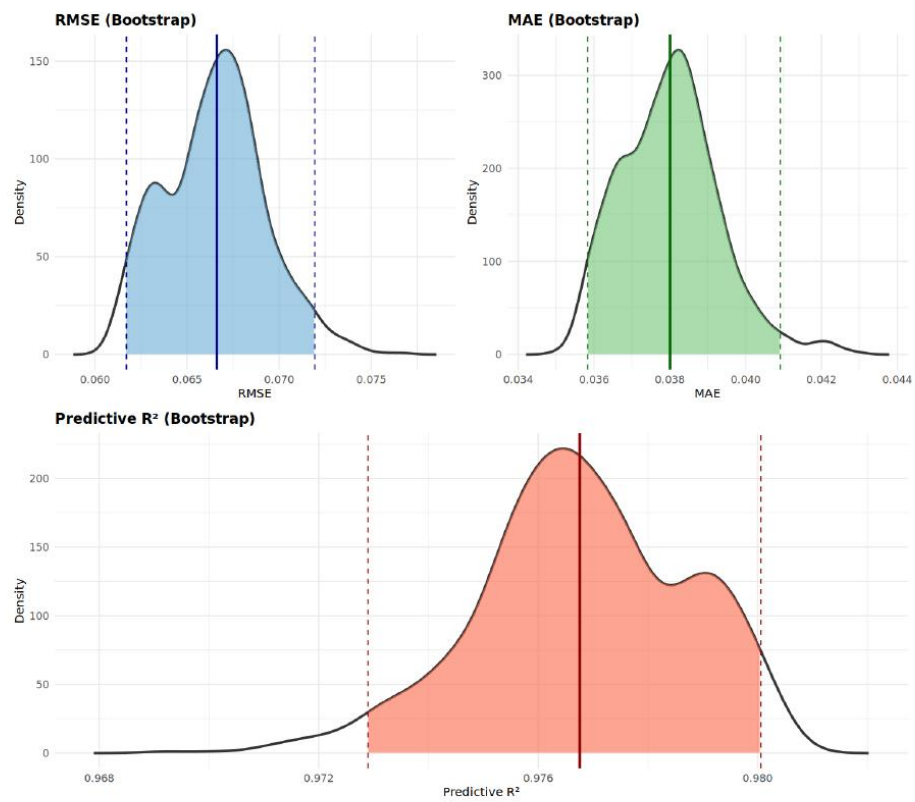
Top 10 Níveis com Maior Impacto por Cluster
Barras = Efeito Médio | Linhas = IC 95% via Bootstrap

4. Comparação de resíduos entre modelos

Figure 3. Residual comparison across models.

Figure 4. Bootstrap validation (Q-Q plot and band).

5. Validação via bootstrap

6. Principais fatores de custo no Cluster 1

Figure 5. Bootstrap metric distributions (RMSE, MAE, and R²).

	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

Figure 6. Top combinations/levels associated with losses (ranking).