

TEMPLATE PARA ENTREGA DO PROJETO DA DISCIPLINA

Projeto em Business Intelligence e Analytics

Fase 2: Apresentação Executiva

NOME DO ESTUDANTE:

Rafael de Menezes Ehlers

Predição de Avaliações Negativas no Ecossistema Olist

Uma Abordagem Integrada de Business Intelligence e Machine Learning

Contexto e Problema

O e-commerce brasileiro e o desafio da retenção

R\$ 204,3 bi

Faturamento do e-commerce brasileiro em 2024 (ABComm, 2025)

+10,5%

Crescimento ano contra ano em 2024

414,9 mi

Pedidos realizados no período

O Problema

A expansão acelerada do varejo digital pressiona os Custos de Aquisição de Clientes (CAC) e torna a retenção crítica para a sustentabilidade financeira do setor.

Em marketplaces como a Olist, a insatisfação do cliente representa um risco direto ao Lifetime Value (LTV), via avaliações negativas e churn tácito.

Identificar os fatores que determinam essa insatisfação e antecipar pedidos de risco torna-se uma vantagem operacional mensurável.

Pergunta de Pesquisa e Hipóteses

Três mecanismos investigados

Pergunta de pesquisa

Quais fatores operacionais, logísticos e financeiros possuem maior peso estatístico na determinação de uma avaliação negativa no e-commerce brasileiro?

H1

Atrasos na entrega

Pedidos cujo tempo real de entrega ultrapassa o prazo estimado apresentam maior probabilidade de avaliação negativa.

H2

Frete em baixo ticket

Em produtos de baixo ticket (< R\$ 50), o valor do frete tem impacto negativo maior na satisfação do que o preço do produto.

H3

Pedidos multi-vendedor

Pedidos com itens de múltiplos vendedores geram maior tempo de consolidação logística e taxas mais altas de avaliações negativas.

Metodologia

Abordagem quantitativa e aplicada, framework CRISP-DM

Framework

CRISP-DM (Cross-Industry Standard Process for Data Mining):
Entendimento do Negócio, Entendimento dos Dados,
Preparação dos Dados, Modelagem, Avaliação e Implantação.

Dataset

Brazilian E-Commerce Public Dataset by Olist (Kaggle): 9 arquivos CSV relacionais, ~100 mil pedidos realizados em marketplaces brasileiros entre 2016 e 2018.

Modelos

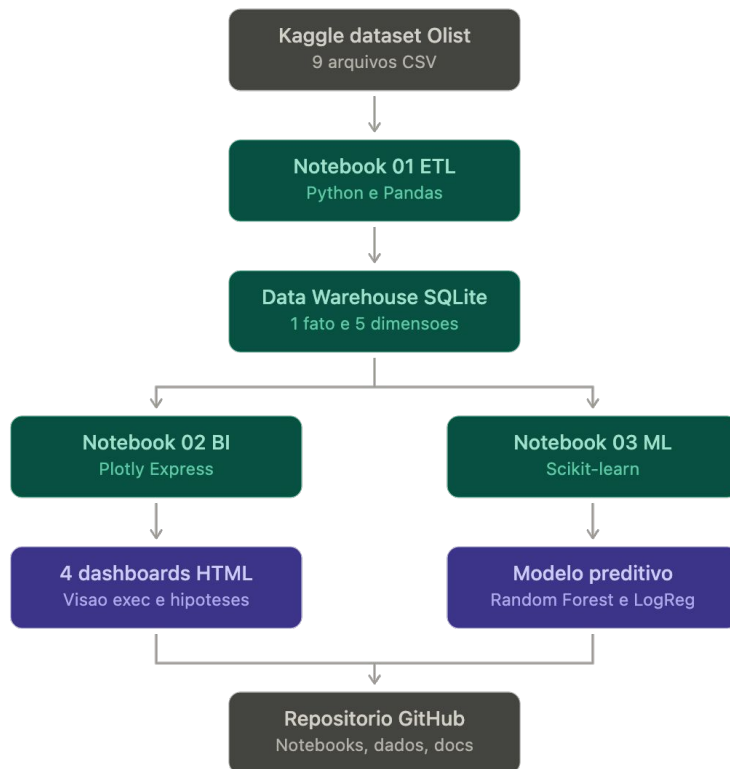
Classificação binária supervisionada para o alvo `avaliacao_negativa` (`review_score <= 3`): Regressão Logística como baseline e Random Forest como modelo principal.

Avaliação

Métricas de desempenho: acurácia, precisão, recall, F1-score e ROC-AUC. Divisão treino/teste 80/20 estratificada com `class_weight balanced`.

Arquitetura da Solução

Pipeline em seis camadas, cloud-native, gratuito



KPIs do Ecossistema

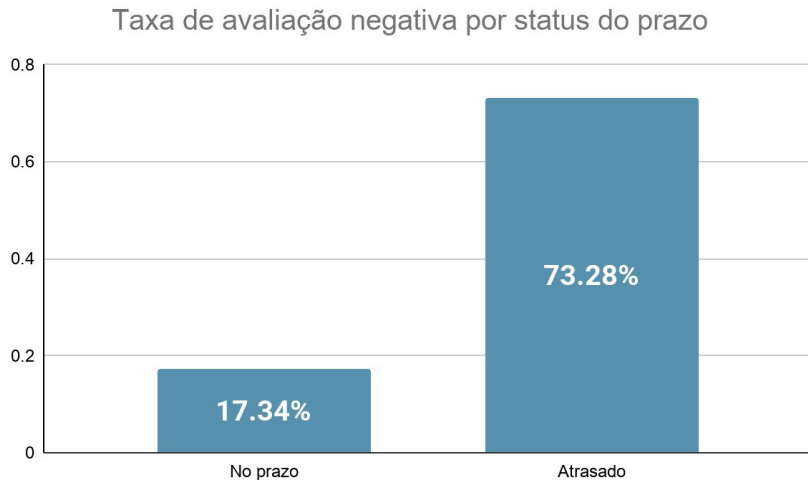
Indicadores estratégicos definidos para o monitoramento

Faturamento Total Sigma(price + freight_value) <i>Financeiro</i>	Ticket Médio Faturamento / Pedidos <i>Financeiro</i>	Prazo Médio de Entrega Média(entrega - compra) <i>Operacional</i>	Taxa de Atraso % entregas > prazo estimado <i>Operacional</i>
NPS Estimado % nota 5 - % notas 1-3 <i>Satisfação</i>	Taxa de Aval. Negativa % review_score <= 3 <i>Satisfação (alvo)</i>	Razão Frete/Preço Média(frete / preco) <i>Diagnóstico</i>	Taxa Multi-Vendedor % pedidos com > 1 seller <i>Diagnóstico</i>

Todos os KPIs são calculáveis diretamente do Data Warehouse SQLite, garantindo reprodutibilidade analítica.

Resultado: Hipótese 1 (Atrasos)

Confirmada com larga margem



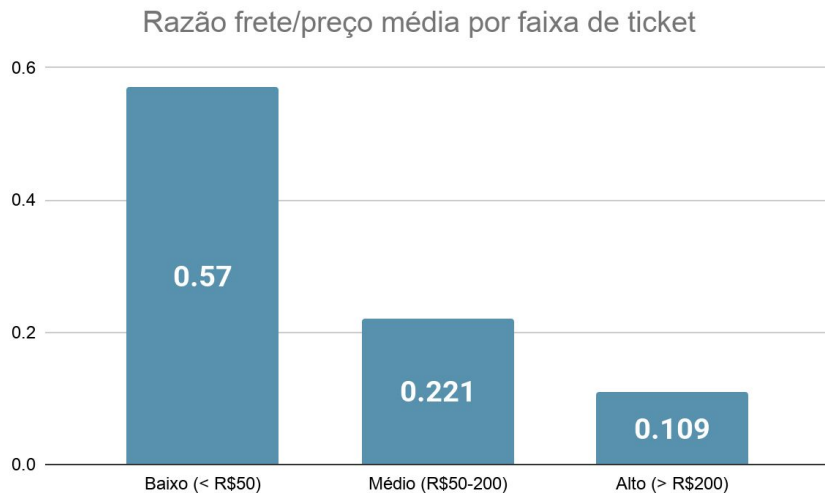
Status estatístico

CONFIRMADA

- Três das features mais importantes do modelo (delta de entrega, tempo total de entrega e atraso) endereçam diretamente a Hipótese 1.
- Juntas, respondem por aproximadamente 52% da importância total do Random Forest.
- A curva de severidade evidencia efeito dose-resposta robusto: a taxa de avaliação negativa cresce de forma praticamente monotônica com o tamanho do atraso, saindo de 17% (no prazo) e atingindo 89% na faixa de 8 a 15 dias.
- Achado convergente com Zaghloul, Barakat e Rezk (2024).

Resultado: Hipótese 2 (Frete em Baixo Ticket)

Confirmada como driver secundário relevante



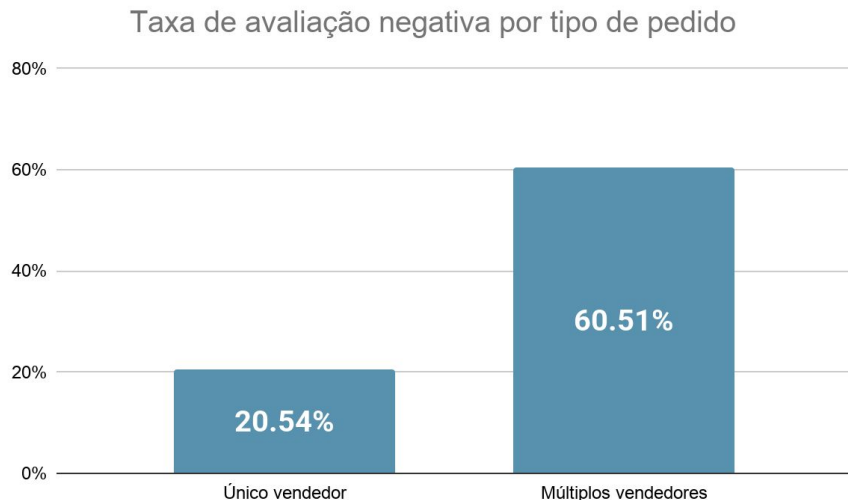
Status estatístico

CONFIRMADA

- Valor de frete e razão frete/preço aparecem em 4º e 5º lugares no ranking de importância do modelo.
- A razão frete/preço em ticket baixo pode superar 100%, situação em que o frete é maior que o produto.
- O mecanismo proposto é validado: não é o frete absoluto, mas o frete relativo ao preço, que afeta a percepção de valor.
- Heatmap valida o efeito dose-resposta dentro de cada faixa de ticket: aumentar a razão frete/preço sempre eleva a taxa de avaliação negativa.

Resultado: Hipótese 3 (Múltiplos Vendedores)

Associação forte na descritiva, mecanismo causal refutado



Status estatístico

PARCIALMENTE CONFIRMADA

- Pedidos multi-vendedor são apenas 1,3% do volume (1.261 de 95.824), mas têm taxa de avaliação negativa 3x maior: 60,5% contra 20,5% do single-seller.
- Tempo de consolidação logística aparece em 6º lugar no ranking de importância do modelo ($\approx 0,06$), enquanto a flag multi_vendedor em si fica bem atrás ($\approx 0,02$).
- Achado contraintuitivo: o tempo médio de consolidação é menor para multi-vendedor (1,38 vs 2,30 dias), refutando o canal causal originalmente proposto (multi-vendor $\rightarrow$ consolidação lenta $\rightarrow$ insatisfação).
- Hipótese 3 fica parcialmente confirmada: a associação multi-vendor vs avaliação negativa é forte e real, mas o mecanismo causal exato (entregas fragmentadas, dessincronia entre sellers, mix de categoria) permanece em aberto para investigação futura.

Análise Preditiva: Métricas Comparativas

Regressão Logística (baseline) versus Random Forest (principal)

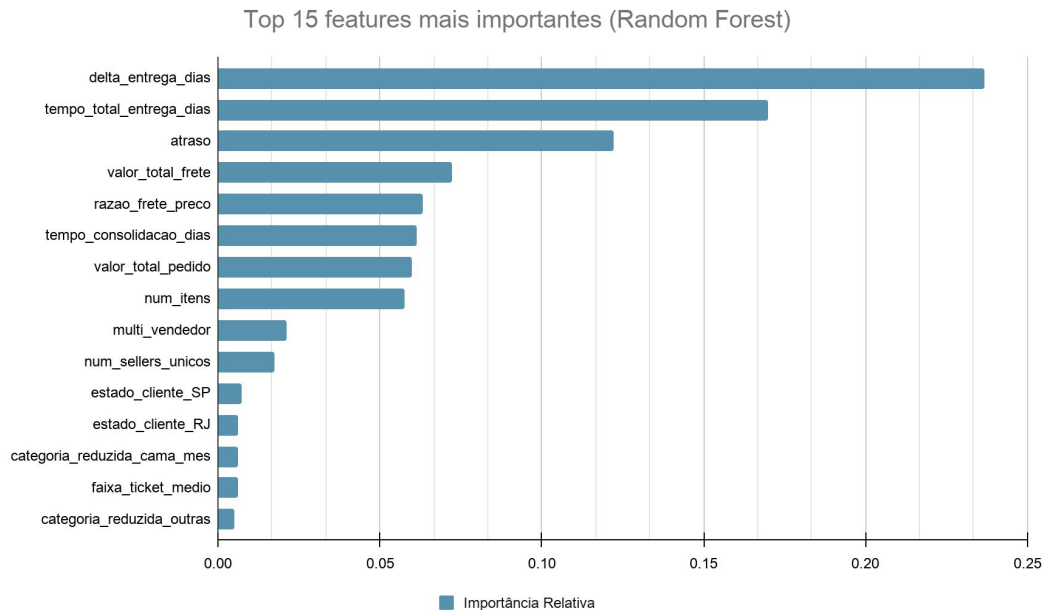
Métrica	Regressão Logística	Random Forest	Diferença (RF - LR)
Acurácia	0,7550	0,7887	+0,0337
Precisão	0,4257	0,4975	+0,0718
Recall	0,4704	0,3975	-0,0730
F1-score	0,4469	0,4419	-0,0050
ROC-AUC	0,6983	0,7024	+0,0041

Trade-off operacional

- Random Forest: maior acurácia e precisão.
- Regressão Logística: maior recall (captura mais negativas).
- F1 e AUC praticamente empatados.
- Escolha do modelo depende do uso: monitoramento de churn (LR) ou priorização de revisão (RF).

Importância das Features (Random Forest)

Resposta direta à pergunta de pesquisa da Fase 1



Três clusters identificados

Cluster H1 (Atrasos):

delta_entrega, tempo_total e atraso. Aprox. 52% da importância total.

Cluster H2 (Frete):

valor_frete e razao_frete_preco em 4º e 5º lugares (~0,07 e 0,06).

Cluster H3 (Multi-vendor):

tempo_consolidacao em 6º lugar (0,06); flag multi_vendedor com importância menor (0,02). H3 parcialmente confirmada (mecanismo refutado, ver slide H3).

Aplicação Prática: Inferência

Dois cenários contrastantes demonstram o uso operacional do modelo

CENÁRIO A - Baixo risco

- Entrega 5 dias adiantada
- Ticket alto (R\$ 350)
- Frete leve (R\$ 15)
- Single-vendedor
- Cliente em SP

37,3%

probabilidade de avaliação negativa

CENÁRIO B - Alto risco

- Atraso de 12 dias
- Ticket baixo (R\$ 30)
- Frete pesado (R\$ 25, razão 0,83)
- Multi-vendedor
- Cliente em AM

74,7%

probabilidade de avaliação negativa

Diferença de 37,5 pontos percentuais entre os dois cenários demonstra a capacidade do modelo de distinguir perfis operacionais, viabilizando monitoramento preventivo de churn em tempo real.

Conclusões

Resposta à pergunta de pesquisa

Quais fatores operacionais, logísticos e financeiros possuem maior peso estatístico na determinação de uma avaliação negativa no e-commerce brasileiro?

01

Tempo de entrega

Principal determinante, responsável por aproximadamente metade da capacidade preditiva do modelo. Hipótese 1 confirmada com grande margem.

02

Razão frete/preço

Segundo driver de insatisfação, especificamente em produtos de baixo ticket. Hipótese 2 confirmada: o vetor crítico é o frete relativo, não o absoluto.

03

Pedidos multi-vendedor

Pedidos multi-vendedor têm taxa de avaliação negativa 3x maior (60,5% vs 20,5%), mas o canal causal originalmente proposto (consolidação logística mais lenta) foi refutado pelos dados. Hipótese 3 parcialmente confirmada, com o mecanismo real em aberto.

Limitações e Trabalhos Futuros

Honestidade metodológica e próximos passos

LIMITAÇÕES

- ROC-AUC modesto ($\sim 0,70$): há sinal não capturado pelas features atuais.
- Recall do Random Forest de 39,75%: o modelo deixa passar mais da metade dos pedidos de risco.
- Sem análise de texto dos comentários das reviews (perda de sinal qualitativo).
- Dataset estático de 2016-2018: não captura tendências pós-pandemia.

TRABALHOS FUTUROS

- Sentiment analysis sobre o texto das reviews via NLP, ampliando o conjunto de features.
- Análise geoespacial: distância entre vendedor e cliente como preditor adicional.
- Modelagem em janelas temporais móveis para monitoramento dinâmico de churn.
- Generalização da abordagem para outros marketplaces brasileiros (transferibilidade).

Referências Bibliográficas

Associação Brasileira de Comércio Eletrônico. “E-commerce fatura R\$ 204,3 bilhões no Brasil em 2024”. Disponível em: <https://www.ecommercebrasil.com.br/noticias/e-commerce-resultados-2024-brasil-abcomm>, Maio 2026.

Kimball, R.; Ross, M. “The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling”. Indianapolis: John Wiley & Sons, 2013, 3 ed., 600.

Luizon, B.F.G. “Análise preditiva de churn em um e-commerce”, Trabalho de conclusão de curso (graduação em gestão da informação), Universidade Federal do Paraná, Curitiba, Brasil, 2019, disponível em: <https://hdl.handle.net/1884/64983>. Acesso em: maio 2026.

Olist; Sionek, A. “Brazilian E-Commerce Public Dataset by Olist”. Disponível em: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>, Maio 2026.

Provost, F.; Fawcett, T. “Data Science para Negócios: o que você precisa saber sobre mineração de dados e pensamento analítico de dados”. Rio de Janeiro: Alta Books, 2016, 408.

Wangkiat, P.; Polprasert, C. “Machine Learning Approach to Predict E-commerce Customer Satisfaction Score”. In: 2023 8th International Conference on Business and Industrial Research (ICBIR), 2023, p. 1176–1181.

Zaghloul, M.; Barakat, S.; Rezk, A. “Predicting e-commerce customer satisfaction: Traditional machine learning vs. deep learning approaches”, Journal of Retailing and Consumer Services, v. 79, 2024, p. 103865.

