

Título - Análise de Atrasos de Entrega e Satisfação no E-commerce Brasileiro com Frameworks de ML

Gustavo Scalabrini Sampaio¹, Pedro Gabriel Marotta Silva^{1,2}

¹Faculdade de Computação e Informática (FCI)
Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

²Programa de pós-graduação em Computação Aplicada – Faculdade de Computação e Informática (FCI) – Universidade Presbiteriana Mackenzie – São Paulo, SP – Brasil

{gustavo.sampaio, 10418073@mackenzista.com.br}@mackenzie.br

Resumo. Este relatório apresenta a etapa N1 do projeto de Inteligência Artificial com a Opção Framework, cujo objetivo é preparar e analisar o Brazilian E-Commerce Public Dataset by Olist para entender o impacto dos atrasos de entrega na satisfação do cliente e produzir um conjunto de dados processado para modelagem no segundo bimestre, conforme diretrizes da disciplina e estrutura do template institucional. A análise integrou registros de pedidos e avaliações, calculou métricas de entrega e atrasos, e consolidou 96.353 pedidos entregues com notas de review, resultando em tempo médio de entrega de 12,1 dias, atraso médio de -11,9 dias e nota média geral de 4,16. A correlação entre atraso e satisfação foi negativa (-0,267), e as categorias “Antecipado” e “No Prazo” apresentaram notas médias substancialmente maiores que “Atrasado” e “Muito Atrasado”, evidenciando a relevância operacional do cumprimento de prazos. Como entregáveis da N1, disponibilizam-se tabelas, figura com visualizações e o dataset_processado_atrasos.csv, que será a base para modelos de classificação na N2 com scikit-learn, seguindo a lista de artefatos exigida para o primeiro bimestre.

Palavras-chave: e-commerce; atraso de entrega; satisfação do cliente; análise exploratória; classificação supervisionada.

Abstract. This report describes the N1 stage of an AI project using the Framework option to prepare and analyze the Olist public e-commerce dataset, aiming to assess the impact of delivery delays on customer satisfaction and to produce a processed dataset for modeling in the second term, following course guidelines and the institutional template. The analysis merged orders and reviews, computed delivery and delay metrics, and consolidated 96,353 delivered orders with review scores, yielding an average delivery time of 12.1 days, an average delay of -11.9 days versus the estimate, and an overall average review score of 4.16. Delay correlated negatively with satisfaction (-0.267), and “Early” and “On time” categories received substantially higher scores than “Late” and “Very Late”, highlighting the operational relevance of meeting delivery estimates. The N1 deliverables include tables, a figure with visualizations, and the dataset_processado_atrasos.csv, which will be used for supervised models in N2 with scikit-learn as required by the first-term checklist.

Keywords: *e-commerce; delivery delay; customer satisfaction; exploratory analysis; supervised learning.*

1. Introdução

Contextualização

Operações de e-commerce dependem fortemente de logística de última milha, e o cumprimento da data estimada de entrega é um determinante direto da satisfação do cliente e da reputação do serviço, o que torna o tema adequado a uma aplicação de IA em cenário real de negócios conforme o escopo da disciplina. Neste projeto, utiliza-se o Brazilian E-Commerce Public Dataset by Olist, que reúne nove arquivos com cerca de 99 mil pedidos e mais de 1 milhão de registros auxiliares, possibilitando medir tempos de entrega, atrasos e avaliações de clientes de forma integrada para análise quantitativa reprodutível. Na execução desta N1, consolidaram-se 96.353 pedidos entregues com reviews, com tempo médio de 12,1 dias e atraso médio de -11,9 dias frente ao estimado, fornecendo base objetiva para relacionar pontualidade e experiência do cliente.

Justificativa

A disciplina solicita um problema de negócio em que técnicas de IA possam melhorar funcionamento ou gerar previsões, e os resultados exploratórios mostram impacto substancial do atraso nas notas, com correlação atraso-satisfação de -0,267 e queda acentuada de avaliações quando há atraso, o que demonstra relevância prática e potencial de ganho operacional com predição e mitigação de atrasos. Por categoria, 87,1% dos pedidos ficaram “Antecipado” com nota média 4,30, enquanto “Muito Atrasado” concentrou as piores avaliações com média 1,70, reforçando que atuar preventivamente no prazo tem efeito direto na satisfação e no risco de review ruim. Essa evidência empírica configura a lacuna a ser endereçada por modelos na N2: identificar, antes do envio ou durante o fluxo logístico, pedidos com maior propensão a atraso e a avaliações 1–3 para priorização e comunicação proativa.

Objetivo

Objetivo geral: preparar e analisar o dataset da Olist para quantificar o efeito dos atrasos na satisfação, consolidar variáveis-chave e entregar o `dataset_processado_atrasos.csv` como insumo imediato para modelagem preditiva no segundo bimestre. Objetivos específicos: (i) integrar pedidos e reviews, calcular `delivery_time_days` e `delay_days` e classificar atrasos; (ii) produzir estatísticas e visualizações que fundamentem o problema de negócio; (iii) definir o rótulo supervisionado de satisfação (review ruim 1–3 vs boa 4–5) e documentar as distribuições de classe para orientar técnicas de balanceamento na N2. Tais objetivos atendem ao pedido do template para explicitar contexto, variáveis estudadas e resultados esperados de forma clara e acionável para as próximas etapas do projeto.

Opção do projeto

Foi escolhida a Opção Framework, que prevê o emprego de bibliotecas de Machine Learning (por exemplo, scikit-learn) para resolver um problema de classificação/regressão de negócio, o que neste caso se traduzirá, na N2, em modelos para prever risco de atraso e probabilidade de review ruim usando as features derivadas na N1. A N1 entrega os artefatos exigidos: relatório no template, notebook de análise exploratória e arquivos de dados processados, incluindo o `dataset_processado_atrasos.csv` que será consumido diretamente pelos classificadores e métricas de

avaliação na etapa seguinte. Essa opção mantém aderência às diretrizes do curso e ao formato do relatório acadêmico previsto pelo template institucional.

Estrutura do documento

A seção de Metodologia descreve o tipo de pesquisa, os dados utilizados, o processo de integração e preparação das variáveis de atraso e satisfação, além dos critérios de reprodutibilidade do notebook, em conformidade com o template institucional. Em Resultados e discussão, apresentam-se as estatísticas gerais, as tabelas por categoria de atraso e a figura com distribuições e correlação atraso–review, seguidas de interpretações alinhadas ao problema de negócio e aos resultados esperados para a N2, como solicita o documento da disciplina. Por fim, incluem-se aspectos éticos, materiais e repositório, e referências, atendendo à lista de itens obrigatórios da N1 e preparando a continuidade do projeto no segundo bimestre.

Descrição do problema

O problema de negócio consiste em quantificar como a diferença entre a data estimada e a data real de entrega altera a distribuição das avaliações, identificando padrões e fornecendo insumos práticos para reduzir insatisfação e priorizar ações logísticas. As variáveis centrais são dias de atraso (`delay_days`), tempo total de entrega (`delivery_time_days`), status do pedido e `review_score`, derivadas de ordens e reviews e unificadas em um dataset processado para uso posterior em modelos de classificação.

Aspectos éticos e responsabilidade

Os dados utilizados são públicos e anonimizados, sem identificação direta de pessoas, e a finalidade é exclusivamente acadêmica, com reprodutibilidade por meio de código e artefatos disponibilizados em repositório público, conforme orientação institucional. As análises são reportadas em nível agregado por categoria de atraso, mitigando risco de vieses contra indivíduos ou grupos específicos na etapa N1 e preparando terreno para uso responsável na N2.

2. Referencial Teórico

E-commerce e última milha

A “última milha” é a etapa final que leva o pedido até o cliente e influencia diretamente a experiência e a satisfação, pois pontualidade, flexibilidade e comunicação durante a entrega determinam a percepção de qualidade e a probabilidade de recompra no varejo online. Otimizar a última milha é estratégia central de marcas de e-commerce, pois entregas no tempo prometido e com visibilidade (rastreamento/avisos) aumentam confiança e lealdade, enquanto atrasos e falhas geram reviews negativos e perda de reputação.

Qualidade em serviços: SERVQUAL

Para entender “qualidade do serviço” do ponto de vista do cliente, o modelo SERVQUAL propõe cinco dimensões lembradas pelo acrônimo RATER: confiabilidade, segurança, tangíveis, empatia e responsividade, que cobrem desde cumprir o prometido até atender rápido e com atenção personalizada. Essas dimensões ajudam a traduzir “cumprimento de prazo” e “comunicação de

entrega” em atributos de serviço mensuráveis, conectando logística de última milha com satisfação expressa em avaliações de clientes no e-commerce.

Satisfação e avaliações do cliente

No contexto digital, a satisfação costuma aparecer em notas e comentários, e o cumprimento de prazos está entre os fatores mais fortes que explicam variações de avaliação em operações de última milha, pois atrasos repetidos reduzem confiança e elevam a chance de notas baixas. Assim, medir “atraso em dias” e cruzar com “review_score” é uma forma direta de avaliar o impacto da logística na experiência do cliente usando dados reais como os do conjunto Olist utilizado no projeto.

IA supervisionada para risco de atraso/insatisfação

Em problemas de negócio, modelos supervisionados podem prever risco de atraso e de review ruim a partir de variáveis como atraso histórico, tempo total de entrega e contexto operacional, permitindo ações proativas de priorização e comunicação antes que a experiência se deteriore. Esta abordagem está alinhada à opção Framework pedida na disciplina, na qual a N1 prepara variáveis e rótulos, e a N2 treina classificadores com bibliotecas como scikit-learn sobre o dataset processado.

Métricas: precisão, recall e F1

Para avaliar classificadores sob desbalanceamento (ex.: poucos “reviews ruins”), usa-se precisão e recall, que medem respectivamente a proporção de acertos entre os positivos previstos e a proporção de positivos verdadeiros recuperados, sendo a F1 a média harmônica que equilibra ambas. Em cenários onde falsos negativos e falsos positivos têm custos distintos, a F1 é preferível à acurácia por refletir melhor o compromisso entre “achar casos ruins” e “evitar alarmes falsos”, prática comum em projetos de classificação de risco em operações.

Conexão com o projeto Olist

Este projeto usa o Brazilian E-Commerce Public Dataset by Olist para calcular “delay_days” e relacionar com “review_score”, operacionalizando conceitos de qualidade de serviço e última milha em variáveis que explicam satisfação do cliente. O referencial acima justifica por que focar em atraso e pontualidade, e embasa as escolhas de preparação de dados na N1 e das métricas de modelagem que serão aplicadas na N2, conforme diretrizes do professor e estrutura do template institucional.

3. Metodologia

O fluxo de preparação converteu timestamps, calculou delivery_time_days (compra → entrega) e delay_days (entrega real – estimada), e classificou o atraso em quatro categorias: Antecipado, No Prazo, Atrasado e Muito Atrasado, padronizando o conjunto para análise. Em seguida, os dados de pedidos foram conectados às avaliações para estimar o impacto do atraso na nota e produzir tabelas e visualizações usadas no relatório, com exportação de arquivos CSV e imagem para reprodutibilidade. Por fim, foi criado o alvo binário de satisfação para uso futuro em modelos (review_ruim para 1–3 e review_boa para 4–5) e exportado o dataset consolidado para a N2 (dataset_processado_atrasos.csv) a ser versionado e utilizado com scikit-learn.

Dataset, origem e análise exploratória

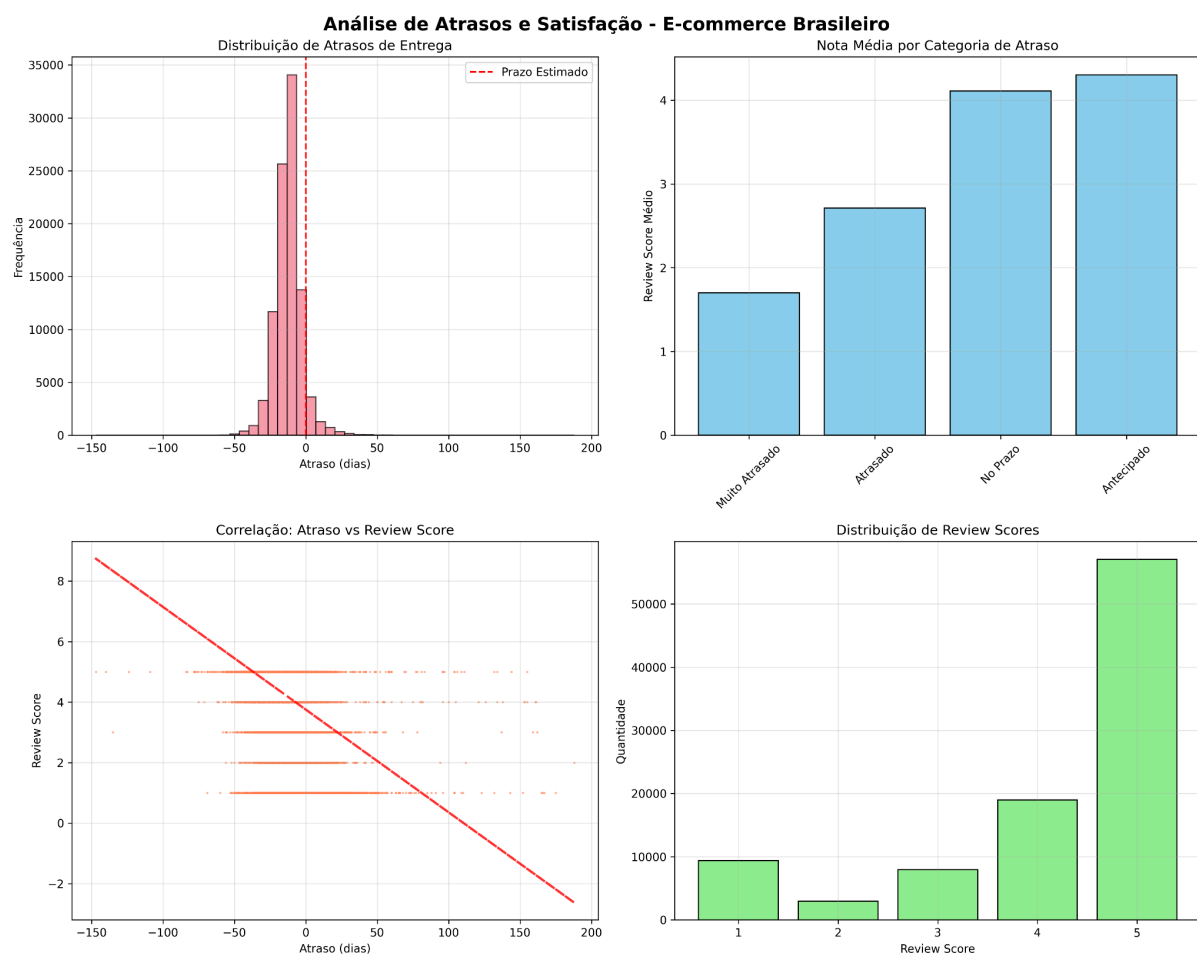
Foram empregados nove arquivos do Olist com clientes, geolocalização, itens, pagamentos, reviews, pedidos, produtos, sellers e tradução de categorias, totalizando mais de 1 milhão de registros brutos e 99 mil pedidos, conforme resumo com contagens e percentual de faltantes. A auditoria indica que “order_reviews” concentra a maior parte dos faltantes (21,0%), enquanto os demais arquivos apresentam baixa ou nenhuma ausência, o que orientou a seleção de colunas e filtros na preparação. Para a análise final, consolidaram-se 96.353 pedidos entregues com dados completos de avaliação após conversões de data, cálculos de métricas e junção com reviews.

| Dataset | Registros | Colunas | Missing (%) |
|----------------------|-----------|---------|-------------|
| customers | 99.441 | 5 | 0,0 |
| geolocation | 1.000.163 | 5 | 0,0 |
| order_items | 112.650 | 7 | 0,0 |
| order_payments | 103.886 | 5 | 0,0 |
| order_reviews | 99.224 | 7 | 21,0 |
| orders | 99.441 | 8 | 0,6 |
| products | 32.951 | 9 | 0,8 |
| sellers | 3.095 | 4 | 0,0 |
| category_translation | 71 | 2 | 0,0 |

4. Resultados e discussão

| Categoria | Pedidos | Review médio | Desvio | Atraso médio (dias) | Mín | Máx | % |
|-----------------------|----------------|---------------------|---------------|----------------------------|-------------|------------|-------------|
| Antecipado | 83.922 | 4,30 | 1,14 | -14,37 | -147 | -4 | 87,1 |
| No Prazo | 6.022 | 4,11 | 1,23 | -1,62 | -3 | 0 | 6,2 |
| Atrasado | 3.612 | 2,71 | 1,66 | 3,62 | 1 | 7 | 3,7 |
| Muito Atrasado | 2.797 | 1,70 | 1,23 | 19,44 | 8 | 188 | 2,9 |

| Categoria | Nota 1 (%) | Nota 2 (%) | Nota 3 (%) | Nota 4 (%) | Nota 5 (%) |
|------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| Antecipado | 6,5 | 2,6 | 7,8 | 20,2 | 62,9 |
| Atrasado | 41,4 | 8,0 | 12,3 | 14,4 | 23,9 |
| Muito Atrasado | 69,7 | 9,5 | 9,0 | 4,8 | 7,0 |
| No Prazo | 8,1 | 3,4 | 11,8 | 22,6 | 54,0 |



Métricas gerais: 96.353 pedidos analisados, tempo médio de entrega de 12,1 dias, atraso médio de -11,9 dias em relação à previsão e nota média 4,16, indicando prevalência de entregas antecipadas e boa satisfação média. Por categoria de atraso, observou-se distribuição de 87,1% Antecipado, 6,2% No Prazo, 3,7% Atrasado e 2,9% Muito Atrasado, com notas médias 4,30, 4,11, 2,71 e 1,70 respectivamente, evidenciando degradação acentuada com atrasos. A correlação atraso vs satisfação foi -0,267 e a distribuição de notas por categoria confirma forte queda das avaliações 4-5 quando há atraso relevante, com “Muito Atrasado” concentrando avaliações 1-2.

5. Conclusão

A etapa N1 cumpriu a proposta de estruturar o problema, inspecionar e preparar os dados e demonstrar quantitativamente o impacto dos atrasos na satisfação, produzindo insumos claros para modelagem preditiva com frameworks de ML no segundo bimestre, alinhada à Opção Framework da disciplina. Os resultados mostram que o cumprimento do prazo estimado é determinante para notas altas, e o dataset_processado_atrasos.csv entregue nesta fase permitirá treinar classificadores supervisionados na N2 para reduzir risco de insatisfação por meio de decisões operacionais informadas.

Metodologia e resultados esperados (para a N2)

A N2 empregará frameworks de ML (por exemplo, regressão logística, Random Forest e Gradient Boosting) para prever atraso e probabilidade de review ruim usando o dataset_processado_atrasos.csv com features como delay_days, delivery_time_days e dimensões derivadas, com avaliação por F1, precisão e revocação considerando desbalanceamento de classes. Espera-se obter modelos com poder discriminativo suficiente para acionar comunicações proativas e prioridades operacionais, com análise de importância de variáveis e calibragem de limiar para cenários de negócio.

6. Referências bibliográficas

Universidade Presbiteriana Mackenzie (UPM). Inteligência Artificial – 7ºJ SI – Noite: Projeto da Disciplina. São Paulo: Faculdade de Computação e Informática, 2025. Documento institucional.

Faculdade de Computação e Informática (FCI). Template de artigo: estilo e normas de formatação. São Paulo: FCI, 2025. Documento institucional.

Olistbr. Brazilian E-Commerce Public Dataset by Olist. Kaggle, 2016–2018. Conjunto de dados público.

Parasuraman, A.; Zeithaml, V.; Berry, L. Modelo de qualidade de serviços (SERVQUAL): uso e aplicações. Estudo de revisão/relato de uso do modelo. SciELO Brasil.

SERVQUAL. Verbete enciclopédico sobre o modelo RATER e dimensões de qualidade de serviço. 2025.

Assessing the e-commerce last-mile logistics' hidden risk. Artigo acadêmico sobre riscos na última milha em operações de e-commerce. ScienceDirect.

Impact of Last-Mile Delivery on Customer Satisfaction. Texto técnico sobre como a última milha afeta a satisfação e recompra. 2024.

E-commerce Last-Mile Delivery: Trends and Strategies. Revisão prática de tendências e estratégias de última milha no varejo online. 2025.

Classification: Accuracy, recall, precision, and related metrics. Guia técnico de métricas para classificação supervisionada. Google Developers. 2025.

Precision and recall. Definições formais e relações com métricas de classificação. Enciclopédia técnica.

F-score. Definição e propriedades da F1 (média harmônica entre precisão e recall). Enciclopédia técnica.

Understanding and Applying F1 Score: AI Evaluation. Guia prático sobre interpretação e uso da F1 em cenários com desbalanceamento. 2024.