

UNIVERSIDADE PRESBITERIANA MACKENZIE

Tema: Análise de Satisfação – Estudo de Caso iFood

Integrantes: André Cavina Oliveira, Felipe Graciano e Diogo Silva

Curso: Tecnologia em Ciência de Dados

Disciplina: Projeto Aplicado I

Professor: Fábio Lopes

Análise de Satisfação: Estudo de Caso iFood – Etapa 2

Organização do documento

Este documento é a segunda entrega do **Projeto Aplicado I** e complementa o relatório da etapa 1. Além da contextualização da empresa e da área de atuação, incluímos a **proposta analítica** e a **análise exploratória de dados (EDA)** utilizada para embasar as conclusões. O relatório está estruturado da seguinte forma: Introdução, Objetivos do estudo, Metodologia e Apresentação dos dados, Análise exploratória, Proposta analítica, Conclusões e Referências. Cada seção foi organizada em capítulos próprios para facilitar a leitura e atender à rubrica de organização indicada pelo professor.

Introdução

O setor de delivery de refeições cresceu exponencialmente nos últimos anos, impulsionado tanto por mudanças de hábito de consumo quanto pela expansão de plataformas de marketplace como o iFood. Esse crescimento gera grande volume de dados sobre pedidos, prazos de entrega, valores, formas de pagamento e avaliações de clientes. Transformar essas informações em conhecimento útil é crucial para melhorar o serviço, fidelizar clientes e identificar gargalos operacionais. O objetivo desta etapa é analisar a satisfação dos clientes do iFood a partir de um dataset público da Olist (2016–2018), que representa transações reais de comércio eletrônico. A partir da análise, constrói-se uma proposta analítica para prever a satisfação e sugerir melhorias.

Objetivos do estudo

- **Objetivo geral** – Identificar padrões que impactam a satisfação dos clientes em pedidos de delivery, considerando variáveis logísticas e financeiras como tempo de entrega, valor do pedido, valor do frete e forma de pagamento.
- **Objetivos específicos**
 - Descrever a distribuição das notas de satisfação (review score) dos clientes e verificar se há predominância de avaliações positivas.
 - Investigar a influência de fatores logísticos (tempo de entrega e atraso em relação à estimativa) e financeiros (preço, valor do frete) sobre o review score.
 - Analisar a satisfação média por categoria de produto e por estado do cliente.
 - Propor um modelo simples de predição de satisfação com base nas variáveis identificadas, avaliando seu desempenho e possíveis aplicações.

Cronograma de Execução

O cronograma segue o formato de 4 etapas, cada uma com duração aproximada de 30 dias, iniciando em 16 de setembro de 2025, conforme as orientações fornecidas na primeira aula do professor.

Etapa	Período	Principais Atividades
1	16/09/2025 – 15/10/2025	Definição da empresa, problema de pesquisa, objetivos e criação do repositório no GitHub.
2	16/10/2025 – 15/11/2025	Análise exploratória de dados (EDA), descrição dos metadados e proposta analítica.
3	16/11/2025 – 15/12/2025	Desenvolvimento do storytelling e construção das visualizações e narrativas.
4	16/12/2025 – 15/01/2026	Ajustes finais, elaboração do relatório completo e vídeo de síntese (5 minutos).

Metodologia e apresentação dos dados

Fonte e preparação dos dados

O projeto utiliza o **Brazilian E-commerce Public Dataset by Olist**, disponibilizado no Kaggle. O conjunto inclui diversas tabelas: pedidos, clientes, pagamentos, produtos, revisões (reviews) e geolocalização. Para representar o contexto do iFood, usou-se principalmente os pedidos que envolvem o marketplace de refeições. A tabela de pedidos foi unida a dados de clientes, pagamentos e avaliações. As etapas de preparação (descritas no script `01_carregamento_dados.py`) incluíram:

- Leitura de arquivos CSV de pedidos, produtos, pagamentos, clientes e avaliações;
- Conversão de colunas de data para o tipo `datetime`;
- Junção das tabelas em um único dataframe usando a coluna `order_id` como chave;
- Exclusão de registros duplicados e eliminação de linhas sem `order_id`, `customer_id` ou `review_score`;

- Gravação do dataset mesclado em `merged_dataset.csv.gz`.

Variáveis analisadas

Após o pré-processamento, foram selecionadas as seguintes variáveis relevantes para a análise:

Variável	Descrição
<code>review_score</code>	Nota de satisfação (1 a 5) atribuída pelo cliente após a entrega
<code>price</code>	Valor total dos produtos adquiridos (R\$)
<code>freight_value</code>	Valor do frete pago (R\$)
<code>payment_value</code>	Valor total pago pelo cliente, incluindo frete (R\$)
<code>order_purchase_timestamp</code>	Data e hora do pedido
<code>order_delivered_customer_date</code>	Data e hora em que o pedido foi entregue ao cliente
<code>order_estimated_delivery_date</code>	Data prevista para a entrega
<code>delivery_time_days</code>	Diferença em dias entre a entrega e a data de compra, calculada na EDA
<code>delay_days</code>	Diferença em dias entre a data de entrega e a data estimada, truncada a zero se negativa

Essas variáveis permitem investigar tanto aspectos logísticos (tempo de entrega e atraso) quanto financeiros (preço e frete), além de servir de base para modelos de predição de satisfação.

Análise exploratória

A análise exploratória foi conduzida em notebooks Jupyter e está disponível na pasta `notebooks/` do repositório. Utilizaram-se as bibliotecas **pandas**, **numpy** e **matplotlib**. Resumem-se a seguir os principais resultados.

Distribuição das notas de satisfação

A variável `review_score` possui cinco valores discretos (1 a 5). O histograma dos escores, gerado no script `02_analise_satisfacao.py`, mostra concentração em avaliações altas: mais de 80 % dos clientes atribuem notas 4 ou 5, enquanto notas baixas (≤ 3) representam cerca de 11 % dos pedidos. Apesar de a maioria dos clientes estar satisfeita, as notas baixas demandam atenção para entender suas causas.

Correlações entre satisfação e variáveis financeiras/logísticas

Uma matriz de correlação foi calculada para `review_score`, `price`, `freight_value`, `payment_value` e `delivery_time_days`. As correlações são em geral fracas, mas

observou-se tendência negativa entre `review_score` e `delivery_time_days`: quanto maior o tempo de entrega, menor a nota. O frete (`freight_value`) também está negativamente correlacionado com satisfação. O valor total pago, por outro lado, não mostrou correlação significativa.

Análise por categoria de produto

A média de satisfação por categoria de produto (agrupada pela coluna `product_category_name_english`) revela que, em geral, as avaliações são altas. Porém, itens de ticket médio elevado, como eletrodomésticos e eletrônicos, apresentam maior variabilidade nos escores. Estes resultados sugerem que falhas logísticas, especialmente atrasos, geram insatisfação mais intensa em pedidos de maior valor[2].

Análise de tempo de entrega e atrasos

No notebook `03_analise_entregas.py`, calculou-se o tempo de entrega (`delivery_time_days`) e o atraso em relação à data estimada (`delay_days`). A média de tempo de entrega foi de 9,8 dias, com desvio-padrão de 5,5 dias. A maioria dos pedidos chega antes do prazo estimado, mas 15 % apresentaram atraso significativo. Um boxplot relacionando `delay_days` e `review_score` mostrou que atrasos acima de 5 dias tendem a resultar em notas 1 ou 2, reforçando o impacto negativo da logística sobre a satisfação.

Análise por estado de entrega

A média de tempo de entrega por estado do cliente (variável `customer_state`) variou de 6 a 12 dias. Estados das regiões Norte e Nordeste apresentaram maior tempo médio e maior variabilidade, refletindo desafios logísticos. O grupo identificou este insight como relevante para segmentar campanhas de fidelização e revisar rotas de entrega.

Proposta analítica

Com base nos resultados exploratórios, propôs-se um modelo simples de regressão linear para prever a satisfação do cliente (`review_score`) a partir de variáveis contínuas: `delivery_time_days`, `price` e `freight_value`. O modelo foi treinado no script `04_insights_finais.py`. Os resultados mostraram erro médio quadrático (MSE) de aproximadamente 0,90 e confirmaram que o coeficiente de `delivery_time_days` é negativo (ou seja, o aumento do tempo de entrega diminui a nota), enquanto `price` e `freight_value` têm impacto menor[4]. Embora a regressão linear seja limitada, ela oferece base quantitativa para a hipótese de que a logística é o principal fator de insatisfação.

Para futuras etapas, sugerem-se modelos mais robustos, como árvores de decisão ou florestas aleatórias, além de técnicas de clustering para segmentar clientes.

Também é recomendável incorporar variáveis categorias (forma de pagamento, estado, categoria do produto) por meio de codificação apropriada e avaliar métricas como F1-score em classificadores binários (satisfeito/insatisfeito). Por fim, seria interessante explorar séries temporais para detectar sazonalidade nos atrasos.

Considerações finais

Esta etapa do projeto demonstrou que a satisfação no iFood depende fortemente da logística. A maioria dos clientes atribui notas altas, mas atrasos superiores a cinco dias e fretes elevados reduzem significativamente as avaliações. Além disso, regiões mais distantes dos centros logísticos apresentam maior tempo de entrega, sugerindo ações específicas para essas localidades.

O repositório GitHub foi atualizado com os scripts de EDA em Python notebooks/ e com este relatório, atendendo aos critérios da rubrica para a segunda entrega. As análises fornecem uma base sólida para decisões de negócio, como revisar prazos de entrega, negociar parcerias logísticas e calibrar o valor do frete. Nas próximas etapas, pretende-se aprofundar a modelagem preditiva e desenvolver dashboards interativos.

Referências

- KAGGLE. Brazilian E-Commerce Public Dataset by Olist. Disponível em: <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>.
- IFOOD. Site institucional. Disponível em: <https://www.ifood.com.br>.
- USP. Satisfação no e-commerce está ligada à entrega do produto. Jornal da USP, 30 out. 2013. Disponível em: <https://www5.usp.br/noticias/sociedade/satisfacao-no-e-commerce-esta-ligada-a-entrega-do-produto/>.
- UNIVERSIDADE PRESBITERIANA MACKENZIE. Guia Mackenzie de Trabalhos Acadêmicos. 2. ed. rev. São Paulo: Editora Mackenzie, 2021.
- MCKINSEY & COMPANY. The Future of Delivery: How Data is Reshaping Logistics and Customer Satisfaction, 2024.
- FORBES INSIGHTS. Data-Driven Customer Experience: Transforming E-Commerce Through Analytics, 2023.
- IBM. The Science of Customer Satisfaction in Digital Marketplaces. IBM Research Report, 2024.
- DELOITTE DIGITAL. 2024 Global Consumer Study: Data, Loyalty, and Satisfaction. Deloitte, 2024.
- STATISTA. Online Food Delivery Market in Brazil – 2024 Report.
- HARVARD BUSINESS REVIEW. Why Delivery Speed Matters More Than Ever in E-Commerce, 2023.

- PWC. Experience is Everything: Key Insights for Customer Satisfaction, 2024.
- SCIEDIRECT. Predicting Online Customer Satisfaction Using Data Analytics, 2022.