

Autor: Felipe Cidade Soares

ANÁLISE DE DADOS: SISTEMA DE E-COMMERCE

Fevereiro de 2026

SUMÁRIO

1. INTRODUÇÃO	1
2. CONTEXTO DOS DADOS	1
3. METODOLOGIA.....	2
4. RESULTADOS E ANÁLISES	4
5. PRINCIPAIS INSIGHTS	8
6. CONCLUSÃO.....	9
APÊNDICES.....	10

1. INTRODUÇÃO

O crescimento do e-commerce brasileiro tornou a logística um fator crítico para a competitividade, onde a eficiência na entrega e o cumprimento de prazos impactam diretamente a satisfação do cliente e indicadores estratégicos, como o faturamento e a frequência de compra. Diante dessa complexidade, o presente trabalho utiliza um conjunto de dados reais para investigar como o desempenho operacional, com foco especial nos atrasos de entrega, influencia o comportamento do consumidor e as métricas de negócio.

A metodologia foi dividida em duas fases complementares: primeiramente, utilizou-se scripts em Python para a exploração dos dados de pedidos, pagamentos e logística. Em seguida, os dados foram exportados para o Power BI, onde foi desenvolvido um dashboard interativo para permitir a visualização clara de gargalos e facilitar a interpretação de tendências de mercado.

A estrutura da análise está organizada em três eixos: visão geral do negócio, evolução histórica das vendas e análise técnica de logística (SLA). O objetivo é demonstrar como a transformação de dados transacionais em *insights* estratégicos pode orientar decisões mais precisas, permitindo que as empresas otimizem suas operações e melhorem a experiência do cliente em um cenário de alta competitividade.

2. CONTEXTO DOS DADOS

Os dados analisados neste estudo são da Olist, empresa brasileira de e-commerce, e foram obtidos na plataforma *Kaggle*, amplamente reconhecida por disponibilizar datasets públicos de diversas áreas para pesquisa e desenvolvimento. O conjunto de dados reúne informações reais do comércio eletrônico no Brasil e permite explorar aspectos relevantes do desempenho logístico, como pedidos, pagamentos, envios, reviews, clientes e vendedores. O material pode ser acessado em <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce/data>.

Esse conjunto de dados é bastante robusto, reunindo mais de cem mil registros, o que oferece uma base sólida para análises com boa representatividade. A amplitude do volume permite observar padrões, tendências e variações com maior segurança, enriquecendo as conclusões do estudo.

Os dados abrangem o período de 2016, 2017 e 2018, possibilitando uma visão evolutiva do comportamento do e-commerce brasileiro ao longo de três anos consecutivos. Essa janela temporal facilita comparações entre ciclos e ajuda a identificar mudanças relevantes no desempenho logístico.

3. METODOLOGIA

Nesta seção, apresentam-se os principais blocos do trabalho: análise, modelagem, visualização e *storytelling*, destacando como cada etapa contribui para a construção e comunicação dos resultados.

3.1. Análise dos Dados (Python)

A análise por meio de scripts em Python revelou que os dados de 2016 são pouco representativos. De um universo superior a 99 mil pedidos, apenas 312 pertencem a esse ano. Por essa razão, optou-se por considerar apenas os dados de 2017 e 2018, com aproximadamente 45 mil e 54 mil pedidos, respectivamente, garantindo maior consistência estatística.

O uso da linguagem Python foi essencial nessa etapa, pois oferece agilidade e precisão no tratamento de grandes volumes de dados. Para as análises, utilizaram-se principalmente as bibliotecas Pandas, para manipulação e limpeza das tabelas, e Plotly, para visualizações interativas e mais informativas. Os scripts em questão podem ser acessados no repositório do GitHub <https://github.com/cidade-felipe/ecommerce-shipping-analytics>.

3.2. Modelagem Analítica

A modelagem analítica teve como objetivo transformar dados transacionais em métricas consistentes de negócio, capazes de sustentar análises comparáveis ao longo do tempo e entre diferentes recortes. Foram definidas métricas comerciais como GMV (*Gross Merchandise Value*), calculado a partir da soma do valor dos itens e do frete, além do ticket médio por pedido, permitindo distinguir crescimento por volume de crescimento por valor. Essas métricas foram estruturadas para respeitar a

granularidade dos dados e os filtros aplicados, garantindo coerência nas análises por período, região e status do pedido.

Em paralelo, foram construídas métricas operacionais voltadas ao desempenho logístico e ao cumprimento de SLA, como percentual de pedidos entregues, percentual de pedidos atrasados, tempo médio de entrega e quantidade média de dias de atraso. O atraso foi definido como a diferença entre a data real de entrega e a data estimada informada ao cliente, permitindo avaliar tanto a ocorrência quanto a intensidade do problema. Para viabilizar análises temporais consistentes, utilizou-se uma tabela calendário dedicada, com comparações mês a mês (MoM) e ano contra ano (YoY), separando análises descritivas do “o que está acontecendo” de análises analíticas que investigam relações e impactos, como o efeito de atrasos no ticket médio e na satisfação do cliente.

3.3. Visualização e *Storytelling* (Power BI)

A etapa de visualização foi concebida com foco em *storytelling* analítico e apoio à tomada de decisão, utilizando dashboards interativos para transformar métricas e análises em leituras claras e acionáveis. O desenho priorizou a organização da informação em camadas, evitando sobrecarga visual e permitindo que o leitor avance do entendimento macro do negócio para recortes mais específicos. Cada visual foi selecionado com base na pergunta que precisava responder, e não apenas por estética ou conveniência de gráfico.

O *dashboard* foi estruturado em três páginas com papéis analíticos distintos: uma visão geral com indicadores-chave (GMV, pedidos entregues, *ticket* médio, taxa de atraso, tempo médio de entrega e avaliação de clientes), uma página temporal com análises MoM e YoY e relações entre volume de pedidos e *ticket* médio, e uma página operacional focada em SLA, impactos financeiros e experiência do cliente, incluindo recortes geográficos. Essa arquitetura separa análises descritivas do “o que está acontecendo” de análises analíticas sobre evolução e impacto, enquanto insights textuais estratégicos reforçam a interpretação dos dados e orientam conclusões relevantes. As páginas deste *dashboard* podem ser vistas na seção de apêndices.

4. RESULTADOS E ANÁLISES

Nesta seção, apresentam-se os insights extraídos de cada uma das três páginas do dashboard: a primeira traz o panorama do negócio, a segunda explora o comportamento ao longo do tempo e a terceira foca no desempenho operacional e no nível de serviço, considerando 2017 e 2018 de forma consolidada para garantir maior robustez e comparabilidade das análises.

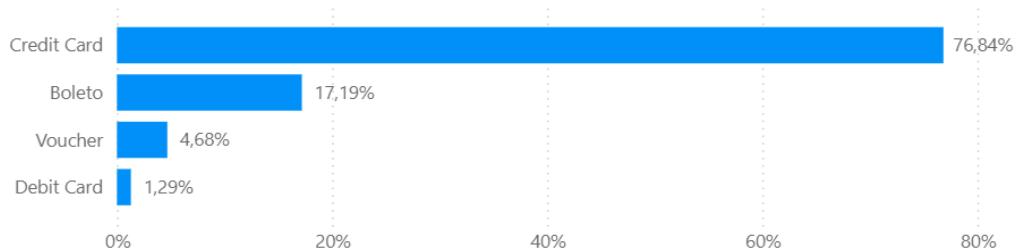
4.1. Visão geral do negócio

Nesta página, avaliaram-se os dados gerais dos pedidos, com foco nos entregues. Relacionaram-se a quantidade de pedidos às taxas de entrega e atraso, ao GMV e às avaliações dos clientes.

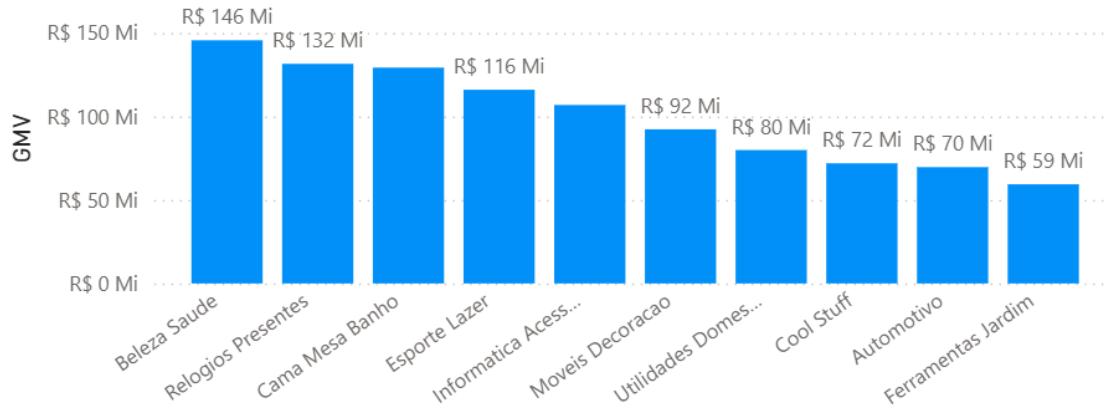
Foram realizados aproximadamente 98,7 mil pedidos, dos quais 97,78% foram entregues (cerca de 96,5 mil). Desses, 93,23% chegaram dentro do prazo e 6,77% chegaram atrasados.

O GMV dos pedidos entregues alcançou 1,61 bilhão de reais. As categorias de beleza e saúde foram as principais contribuidoras, com GMV beirando os 146 milhões de reais. Entre os meios de pagamento, o cartão de crédito lidera com cerca de 1,24 bilhão de reais (76,84%), enquanto o cartão de débito apresenta o menor volume, em torno de 20,86 milhões de reais (1,29%).

% GMV Entregues por Meio de Pagamento



Top 10 Categorias



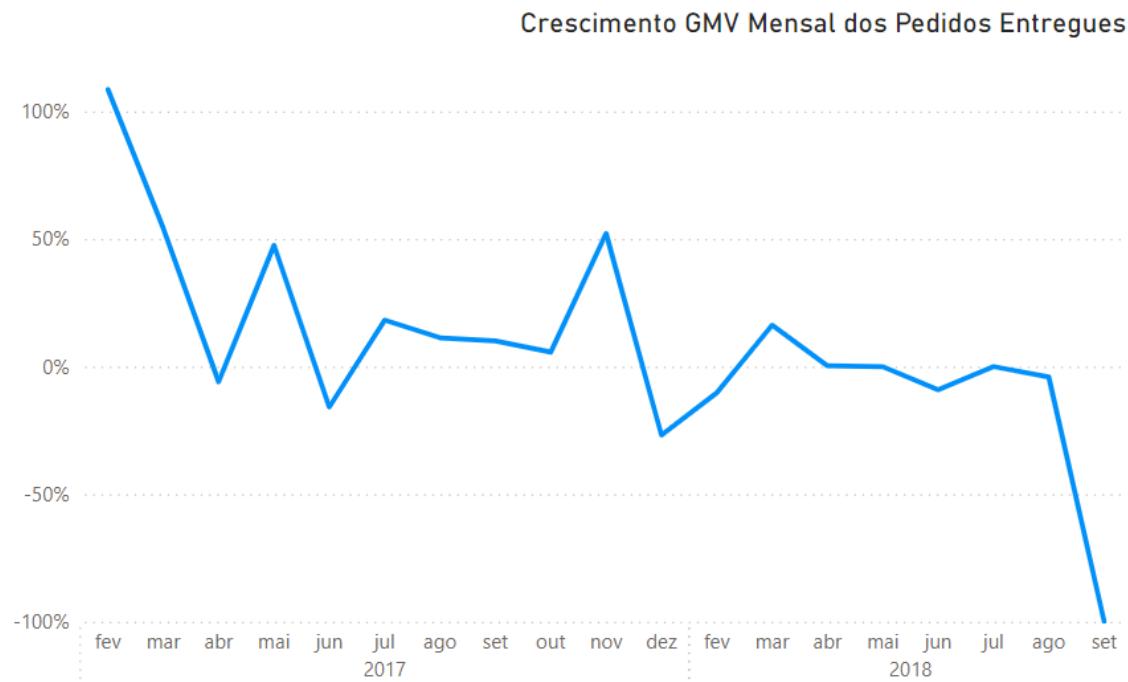
A avaliação média dos produtos entregues é 4,16 (escala de 0 a 5). Para entregas no prazo, a média sobe para 4,29; para entregas atrasadas, cai para 2,27, uma redução de 47,1%.

4.2. Evolução e Tendência

Apesar de o GMV de 2018 ser superior ao de 2017, observa-se desaceleração no ritmo de crescimento. Na comparação entre os três últimos meses e o trimestre imediatamente anterior, o avanço do GMV foi 69,07% menor do que o registrado no ano precedente, indicando perda de fôlego. Essa tendência também aparece nos picos de crescimento mensal, que se tornam mais contidos ao longo do período: em fevereiro, o crescimento alcança 108,6%, enquanto em setembro ocorre retração de -100,0%, sinalizando inversão na trajetória.

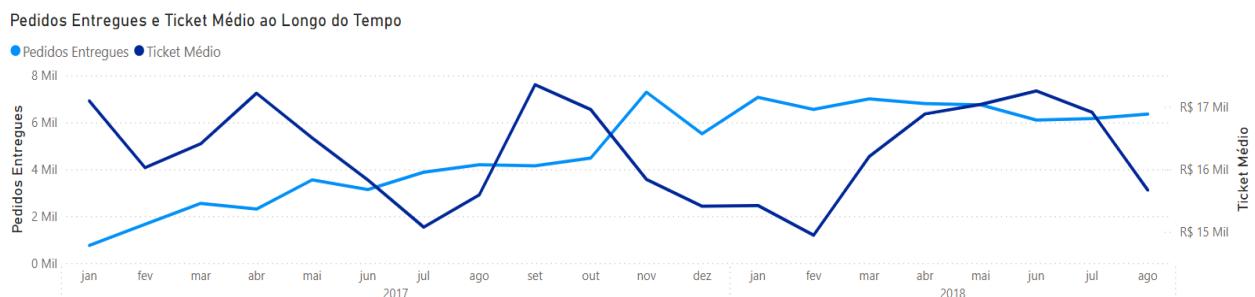
Últimos 3 Meses vs 3 Meses Anteriores





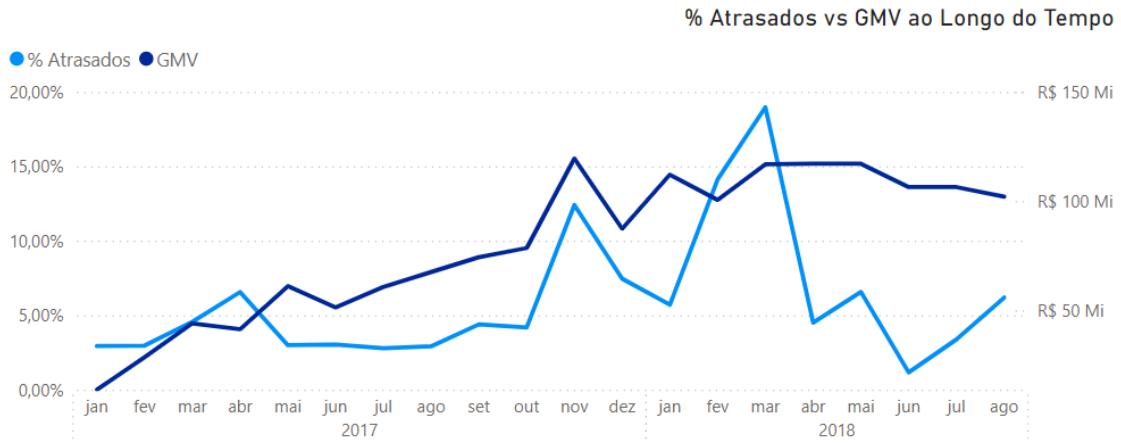
A análise evidencia correlação direta entre GMV e volume de pedidos: conforme o número de compras aumenta, o valor total transacionado também cresce. Esse comportamento confirma que a expansão do GMV está ancorada no aumento da base transacional, e não apenas em variações pontuais de preço ou de mix de produtos.

Entretanto, o ticket médio não acompanha o mesmo ritmo de expansão, o que sugere que a elevação do GMV é impulsionada principalmente pela maior quantidade de pedidos, e não por um acréscimo proporcional no valor médio por compra. Em termos práticos, o crescimento observado reflete escala de demanda, com menor contribuição do aumento de gasto por pedido.



4.3. Operação e SLA

O crescimento do GMV mantém-se relativamente estável, independentemente dos picos de atraso logístico. Ainda assim, observa-se que aumentos de atraso tendem a anteceder quedas no ticket médio, indicando possível sensibilidade do valor gasto ao desempenho de entrega.



As maiores médias de atraso concentram-se na região Norte, com o Amapá apresentando a maior média de dias de atraso (72,5 dias), seguido de Roraima (36,4 dias) e Amazonas (30,25 dias). Na sequência, a região Nordeste aparece com destaque para o Sergipe (16,2 dias).

A região Sul registra as menores médias de atraso, com o Paraná em 8,21 dias. O menor indicador individual é o Distrito Federal, com média de 7,3 dias. Todos esses resultados consideram o estado do consumidor como referência.

Média de Dias Atrasados por Estado Consumidor



5. PRINCIPAIS INSIGHTS

O negócio apresentou alto volume de pedidos e elevada taxa de entrega, com GMV robusto e concentração de receita em categorias-chave e no cartão de crédito. A satisfação do cliente é sensível ao prazo: entregas no prazo têm avaliação significativamente superior às atrasadas, reforçando o impacto direto da logística na experiência.

No recorte temporal, o GMV de 2018 supera 2017, mas o ritmo de crescimento desacelera, com picos mensais menos intensos e sinais de retração em alguns meses.

O aumento do GMV acompanha o volume de pedidos, porém o ticket médio não cresce no mesmo ritmo, indicando expansão por escala e não por aumento do valor médio por compra.

Na operação, picos de atraso não derrubam o GMV de forma imediata, mas antecedem quedas no ticket médio, sugerindo efeito indireto na disposição de gasto. Regionalmente, o Norte concentra as maiores médias de atraso (Amapá, Roraima e Amazonas), enquanto Sul e Distrito Federal apresentam os menores índices, apontando diferenças logísticas relevantes por localidade.

6. CONCLUSÃO

As análises mostram um negócio com alto volume e boa performance de entrega, mas com sinais claros de desaceleração do crescimento e sensibilidade do cliente ao cumprimento de prazos. O GMV cresce principalmente por escala de pedidos, enquanto o ticket médio não acompanha o mesmo ritmo, indicando espaço para iniciativas que elevem valor por compra. Além disso, as diferenças regionais de atraso evidenciam que eficiência logística não é homogênea e impacta diretamente a satisfação.

Esse tipo de análise transforma dados transacionais em evidências acionáveis, permitindo decisões com impacto real: priorizar melhorias logísticas onde o atraso é mais crítico, calibrar metas comerciais com base na dinâmica de crescimento e orientar estratégias de retenção e experiência do cliente a partir do efeito do SLA no ticket médio e nas avaliações.

Como próximos passos, recomenda-se aprofundar por seller e por categoria para identificar onde o desempenho logístico mais afeta receita, além de segmentar por região e tipo de produto para priorização operacional. Também é útil investigar elasticidades entre atraso, ticket médio e avaliação, e evoluir para análises preditivas de atraso e de satisfação, apoiando ações preventivas e ganhos sustentáveis no negócio.

APÊNDICES

Páginas do dashboard

