

Relatório: Análise Exploratória e Estratégica do Dataset Olist

Como Executar o Código

Este documento explica como executar o código Python fornecido e entender os insights gerados a partir do dataset brasileiro de e-commerce da plataforma Olist.

Requisitos Necessários

Antes de rodar o código, você precisa ter instalado:

- **Python 3.x**
- **Jupyter Notebook** (ou Google Colab)
- As seguintes bibliotecas:
 - pandas
 - numpy
 - plotly
 - scikit-learn
 - statsmodels
 - sqlite3 (já vem com o Python)
 - shap (opcional, mas recomendado)

Instale as dependências via pip:

```
!pip install pandas numpy plotly scikit-learn statsmodels shap
```

Célula 3 - Arquivos Necessários (Markdown)

Arquivos Necessários

Você precisará dos seguintes arquivos CSV da Olist:

- olist_customers_dataset.csv
- olist_orders_dataset.csv
- olist_order_items_dataset.csv
- olist_products_dataset.csv
- olist_order_reviews_dataset.csv
- olist_order_payments_dataset.csv
- olist_sellers_dataset.csv
- olist_geolocation_dataset.csv
- product_category_name_translation.csv

💡 Certifique-se de colocar todos esses arquivos na mesma pasta onde o código será executado.

Passos para Execução

1. **Abra o Google Colab.**
2. **Crie uma nova célula e cole o código fornecido.**
3. **Execute célula por célula.**

O código faz o seguinte:

Etapa	Descrição
1. Importação de bibliotecas	Carrega ferramentas como <code>pandas</code> , <code>plotly</code> , <code>sklearn</code> etc.
2. Leitura e limpeza dos dados	Carrega os CSVs e corrige valores faltantes, duplicatas e tipos de colunas.
3. Criação de um banco SQLite	Armazena os dados em tabelas relacionais para consultas SQL.
4. Análise exploratória	Gera gráficos interativos sobre vendas, entregas, categorias e estados.
5. Modelagem preditiva e segmentação	Usa Machine Learning para prever atrasos e agrupar clientes.

Etapa	Descrição
6. Dashboards interativos	Mostra visualizações dinâmicas e mapas com insights estratégicos.

Objetivo do Projeto

O objetivo é extrair **insights estratégicos** do dataset Olist para responder perguntas como:

- Como aumentar vendas?
- Como melhorar a experiência do cliente?
- Como otimizar operações logísticas?

O código utiliza técnicas modernas de análise de dados, visualização interativa e modelagem preditiva para responder a essas questões.

Storytelling: Insights Encontrados

1. Vendas variam sazonalmente – aproveite o timing certo!

- O volume de pedidos tem picos em novembro (Black Friday) e cai em janeiro/fevereiro (pós-Natal e férias).
- Isso indica que campanhas promocionais devem ser intensificadas no final do ano.
- Recomendação: Prepare estoque e logística antes dos períodos de alta demanda.

2. Entregas demoradas afetam a satisfação do cliente

- A maioria dos pedidos é entregue em até 15 dias, mas alguns levam mais de 30 dias.
- Clientes avaliam negativamente quando a entrega ultrapassa 20 dias.
- Recomendação: Monitore rotas críticas e revise contratos com transportadoras.

3. Frete varia com a distância, mas não linearmente

- Existe uma correlação moderada entre a distância entre cliente e vendedor e o valor do frete.
- Em distâncias longas (> 1000 km), o custo do frete sobe significativamente.
- Recomendação: Ofereça frete grátis ou descontado para regiões distantes.

4. Produtos de saúde/beleza e artigos de presente lideram vendas

- Essas categorias têm alto ticket médio e grande aceitação entre consumidores.
- Produtos de baixa categoria (ex.: "cama_mesa_banho") têm menor engajamento.
- Recomendação: Invista em marketing direcionado e bundles com produtos complementares.

5. Alguns estados são mais valiosos que outros

- Estados como PB (Paraíba) e RR (Roraima) têm alto valor médio por pedido.
- SP e RJ têm maior volume de vendas.
- Recomendação: Personalize campanhas regionais e invista em logística local.

6. Taxa de retenção de clientes é baixa (< 5%)

- Poucos clientes voltam a comprar após a primeira compra.
- Isso pode indicar falta de programas de fidelidade ou insatisfação pós-compra.
- Recomendação: Crie programas de fidelidade e envie ofertas personalizadas.

7. Modelo prevê atrasos com razoável acurácia

- Usando Random Forest, o modelo conseguiu identificar pedidos com risco de atraso com F1-Score ~0.7.
- Principais fatores: valor do frete, preço do produto e estado do cliente.
- Recomendação: Use esse modelo para alertar proativamente clientes com entrega em risco.

8. Clientes se dividem em três grupos claros

- **Alta Receita:** Compram frequentemente, gastam muito e pagam frete médio.
- **Média Recorrência:** Compram esporadicamente, com gasto moderado.
- **Baixa Recorrência:** Compram uma vez, gastam pouco e pagam frete alto.

- Recomendação: Campanhas diferenciadas para cada grupo: VIP, upsell e reativação.

9. 😊 Satisfação do cliente varia por categoria

- Alguns produtos (como eletrônicos) têm forte relação entre tempo de entrega e nota.
- Outros (como livros) não mostram essa relação tão claramente.
- Recomendação: Monitore expectativas por categoria e ajuste prazos divulgados no site.

📌 Conclusões Finais

Esse projeto transformou dados brutos em **decisões inteligentes** para o negócio. Foram usadas técnicas modernas de análise de dados, visualização e machine learning para:

- ✓ Entender padrões sazonais
- ✓ Melhorar a experiência do cliente
- ✓ Segmentar clientes e vendedores
- ✓ Prever atrasos nas entregas
- ✓ Identificar oportunidades regionais e por categoria

Com isso, é possível tomar decisões embasadas e otimizar tanto a operação quanto a estratégia comercial.

🔮 Próximos Passos Recomendados

1. **Criar um painel de controle em tempo real** para monitorar entregas e satisfação.
2. **Desenvolver um sistema de recomendação** baseado nos hábitos de compra.
3. **Aprofundar a análise logística** com dados externos (clima, trânsito, rotas).
4. **Implementar um programa de fidelidade** com benefícios acumulativos.
5. **Expandir a análise de sentimentos** com processamento de texto das avaliações.