**Conjunto de dados público brasileiro de comércio eletrônico da Olist**

**100.000 pedidos com informações sobre produtos, clientes e críticas**

# Conjunto de dados público brasileiro de comércio eletrônico da Olist

Bem-vinda! Este é um conjunto de dados público de comércio eletrônico brasileiro de pedidos feitos na [Olist Store](http://www.olist.com/" \t "_blank) . O conjunto de dados possui informações de 100 mil pedidos de 2016 a 2018 feitos em vários mercados no Brasil. Seus recursos permitem visualizar um pedido de várias dimensões: do status do pedido, preço, pagamento e desempenho do frete até a localização do cliente, atributos do produto e, finalmente, revisões escritas pelos clientes. Também lançamos um conjunto de dados de geolocalização que relaciona os códigos postais brasileiros às coordenadas lat / lng.

São dados comerciais reais, foram anonimizados e as referências às empresas e parceiros no texto da revisão foram substituídas pelos nomes das grandes casas de Game of Thrones.

## Contexto

Esse conjunto de dados foi fornecido generosamente pela Olist, a maior loja de departamentos dos mercados brasileiros. O Olist conecta pequenas empresas de todo o Brasil a canais sem problemas e com um único contrato. Esses comerciantes podem vender seus produtos através da Olist Store e enviá-los diretamente aos clientes usando os parceiros de logística da Olist. Veja mais em nosso site: [www.olist.com](https://www.olist.com/)

Depois que um cliente compra o produto na Olist Store, um vendedor é notificado para atender a esse pedido. Depois que o cliente recebe o produto ou a data estimada de entrega é vencida, ele recebe uma pesquisa de satisfação por e-mail, onde pode anotar a experiência de compra e anotar alguns comentários.

### **Atenção**

1. Um pedido pode ter vários itens.
2. Cada item pode ser atendido por um vendedor distinto.
3. Todo o texto que identifica lojas e parceiros foi substituído pelos nomes das grandes casas de Game of Thrones.

## Esquema de dados

Os dados são divididos em vários conjuntos de dados para melhor compreensão e organização. Consulte o seguinte esquema de dados ao trabalhar com ele:

Tela de celular com publicação numa rede social

Descrição gerada automaticamente

## Conjunto de dados classificados

Anteriormente, lançamos um conjunto de dados classificados, mas o removemos na versão 6 . Pretendemos lançá-lo novamente como um novo conjunto de dados com um novo esquema de dados. Embora não terminemos, você pode usar o conjunto de dados classificados disponível na versão 5 ou anterior.

## Inspiração

Aqui estão algumas inspirações para possíveis resultados desse conjunto de dados.

**PNL:**  
  
este conjunto de dados oferece um ambiente supremo para analisar o texto das revisões através de suas múltiplas dimensões.

**Cluster:**  
  
alguns clientes não escreveram um comentário. Mas por que eles estão felizes ou loucos?

**Previsão de vendas:**  
  
com as informações da data da compra, você poderá prever vendas futuras.

**Desempenho da entrega:**  
  
Você também poderá trabalhar no desempenho da entrega e encontrar maneiras de otimizar os prazos de entrega.

**Qualidade do produto:**  
  
descubra as categorias de produtos mais propensas à insatisfação do cliente.

**Engenharia de recursos:**  
  
crie recursos deste rico conjunto de dados ou anexe algumas informações públicas externas a ele.

# Conjunto de dados de clientes - Customers Dataset

Este conjunto de dados tem informações sobre o cliente e sua localização. Use-o para identificar clientes únicos no conjunto de dados de pedidos e para encontrar o local de entrega dos pedidos.

No nosso sistema, cada pedido é atribuído a um único customer\_id. Isso significa que o mesmo cliente receberá IDs diferentes para pedidos diferentes. O objetivo de ter um customer\_unique\_id no conjunto de dados é permitir que você identifique os clientes que fizeram compras na loja. Caso contrário, você descobriria que cada pedido tinha um cliente diferente associado.

Por favor, consulte o esquema de dados acima.

# Customers Dataset - Colunas

Identificação do Cliente

chave para o conjunto de dados de pedidos. Cada pedido tem um customer\_id exclusivo.

customer\_unique\_id

identificador exclusivo de um cliente.

customer\_zip\_code\_prefix

primeiros cinco dígitos do CEP do cliente

customer\_city

nome da cidade do cliente

customer\_state

estado do cliente

# Conjunto de dados de geolocalização - Geolocation Dataset

Este conjunto de dados possui códigos postais brasileiros e suas coordenadas lat / lng. Use-o para plotar mapas e encontrar distâncias entre vendedores e clientes.

Por favor, consulte o esquema de dados acima.

# Geolocation Dataset - Colunas

geolocation\_zip\_code\_prefix

primeiros 5 dígitos do CEP

geolocation\_lat

latitude

geolocation\_lng

longitude

geolocation\_city

Nome da Cidade

geolocation\_state

Estado

# Conjunto de dados de itens de pedido -Order Items Dataset

Esse conjunto de dados inclui dados sobre os itens comprados em cada pedido.

## Exemplo:

O order\_id = 00143d0f86d6fbd9f9b38ab440ac16f5possui 3 itens (mesmo produto). Cada item tem o frete calculado de acordo com suas medidas e peso. Para obter o valor total do frete de cada pedido, basta somar.

**O valor total do order\_item é:** 21.33 \* 3 = 63.99

**O valor total do frete é:** 15.10 \* 3 = 45.30

**O valor total do pedido (produto + frete) é:** 45.30 + 63.99 = 109.29

# Order Items Dataset -Colunas

order\_id

identificador exclusivo do pedido

order\_item\_id

número sequencial que identifica o número de itens incluídos na mesma ordem.

ID do produto

identificador exclusivo do produto

seller\_id

identificador exclusivo do vendedor

shipping\_limit\_date

Mostra a data limite de envio do vendedor para processar o pedido até o parceiro logístico.

preço

Preço do item

shipping\_value

item valor do frete do item (se um pedido tiver mais de um item, o valor do frete será dividido entre os itens)

# Conjunto de dados de pagamentos -Payments Dataset

Este conjunto de dados inclui dados sobre as opções de pagamento de pedidos.

# Payments Dataset -Colunas

order\_id

identificador exclusivo de um pedido.

payment\_sequential

um cliente pode pagar um pedido com mais de um método de pagamento. Se ele fizer isso, uma sequência será criada para acomodar todos os pagamentos.

tipo de pagamento

método de pagamento escolhido pelo cliente.

payment\_installments

número de parcelas escolhidas pelo cliente.

payment\_value

valor da transação.

# Conjunto de dados de avaliações de pedidos - Order Reviews Dataset

Esse conjunto de dados inclui dados sobre as revisões feitas pelos clientes.

Depois que um cliente compra o produto na Olist Store, um vendedor é notificado para atender a esse pedido. Depois que o cliente recebe o produto ou a data estimada de entrega é vencida, ele recebe uma pesquisa de satisfação por e-mail, onde pode anotar a experiência de compra e anotar alguns comentários.

# Order Reviews Dataset- Colunas

review\_id

identificador de revisão exclusivo

order\_id

identificador exclusivo de pedidos

review\_score

Observe de 1 a 5, fornecido pelo cliente em uma pesquisa de satisfação.

review\_comment\_title

Título do comentário da revisão deixada pelo cliente, em português.

review\_comment\_message

Mensagem de comentário da revisão deixada pelo cliente, em português.

review\_creation\_date

Mostra a data em que a pesquisa de satisfação foi enviada ao cliente.

review\_answer\_timestamp

Mostra o carimbo de data e hora da resposta da pesquisa de satisfação.

# Conjunto de dados do pedido -Order Dataset

Este é o conjunto de dados principal. Em cada pedido, você pode encontrar todas as outras informações.

# Order Dataset - Colunas

order\_id

identificador exclusivo do pedido.

Identificação do Cliente

chave para o conjunto de dados do cliente. Cada pedido tem um customer\_id exclusivo.

status do pedido

Referência ao status do pedido (entregue, enviado, etc.).

order\_purchase\_timestamp

Mostra o carimbo de data / hora da compra.

order\_approved\_at

Mostra o carimbo de data e hora da aprovação do pagamento.

order\_delivered\_carrier\_date

Mostra o registro de data e hora do lançamento do pedido. Quando foi entregue ao parceiro logístico.

order\_delivered\_customer\_date

Mostra a data real de entrega do pedido ao cliente.

order\_estimated\_delivery\_date

Mostra a data de entrega estimada que foi informada ao cliente no momento da compra.

# Conjunto de dados de produtos-Products Dataset

Este conjunto de dados inclui dados sobre os produtos vendidos pela Olist.

# Products Dataset -Colunas

identificador exclusivo do produto

product\_category\_name

categoria raiz do produto, em português.

product\_name\_lenght

número de caracteres extraídos do nome do produto.

product\_description\_lenght

número de caracteres extraídos da descrição do produto.

product\_photos\_qty

número de fotos publicadas do produto

product\_weight\_g

peso do produto medido em gramas.

product\_length\_cm

comprimento do produto medido em centímetros.

altura\_do\_produto

altura do produto medida em centímetros.

product\_width\_cm

largura do produto medida em centímetros.

# Conjunto de dados de vendedores -Sellers Dataset

Este conjunto de dados inclui dados sobre os vendedores que atenderam aos pedidos feitos na Olist. Use-o para encontrar a localização do vendedor e identificar qual vendedor atendeu a cada produto.

# Sellers Dataset-Coluna

seller\_id

identificador exclusivo do vendedor

seller\_zip\_code\_prefix

primeiros 5 dígitos do CEP do vendedor

seller\_city

nome da cidade do vendedor

seller\_state

estado do vendedor

# Categoria Nome Tradução -Category Name Translation

Converte o product\_category\_name para inglês.

# Category Name Translation-Coluna

product\_category\_name

nome da categoria em português

product\_category\_name\_english

nome da categoria em inglês