

# Dataset Description

This is a ecommerce public dataset of orders made at an e-commerce Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/Ing coordinates.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

After a customer purchases the product from Olist Store a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

## Attention

1. An order might have multiple items.
2. Each item might be fulfilled by a distinct seller.
3. All text identifying stores and partners where replaced by the names of Game of Thrones great houses.

## Orders

This is the core dataset. From each order you might find all other information.

- **Order\_id**: unique identifier of the order.
- **Vpn\_keycustomer\_id**: key to the customer dataset. Each order has a unique customer\_id.
- **Text\_formatorder\_status**: Reference to the order status (delivered, shipped, etc).
- **Calendar\_todayorder\_purchase\_timestamp**: Shows the purchase timestamp.
- **Calendar\_todayorder\_approved\_at**: Shows the payment approval timestamp.
- **Calendar\_todayorder\_delivered\_carrier\_date**: Shows the order posting timestamp. When it was handled to the logistic partner.
- **Calendar\_todayorder\_delivered\_customer\_date**: Shows the actual order delivery date to the customer.
- **Calendar\_todayorder\_estimated\_delivery\_date**: Shows the estimated delivery date that was informed to customer at

## Products

This dataset includes data about the products sold by Olist.

- **product\_id**: unique product identifier
- **product\_category\_name**: root category of product, in Portuguese.
- **product\_name\_lenght**: number of characters extracted from the product name.
- **product\_description\_lenght**: number of characters extracted from the product description.
- **product\_photos\_qty**: number of product published photos:
- **product\_weight\_g**: product weight measured in grams.
- **product\_length\_cm**: product length measured in centimeters.
- **product\_height\_cm**: product height measured in centimeters.
- **product\_width\_cm**: product width measured in centimeters.

## Order Items

This dataset includes data about the items purchased within each order.

- **Order\_id**: order unique identifier
- **Order\_item\_id**: sequential number identifying number of items included in the same order.
- **Product\_id**: product unique identifier
- **seller\_id**: seller unique identifier
- **shipping\_limit\_date**: Shows the seller shipping limit date for handling the order over to the logistic partner.
- **price**: item price
- **freight\_value**: item freight value item (if an order has more than one item the freight value is splitted between items)

## Customers

This dataset has information about the customer and its location. Use it to identify unique customers in the orders dataset and to find the orders delivery location.

- **Customer\_id**: key to the orders dataset. Each order has a unique customer\_id.
- **Customer\_unique\_id**: unique identifier of a customer.

- **Customer\_zip\_code\_prefix:** first five digits of customer zip code
- **Customer\_city:** customer city name
- **Customer\_state:** customer state

## Geolocation

This dataset has information Brazilian zip codes and its lat/long coordinates. Use it to plot maps and find distances between sellers and customers.

- **Geolocation\_zip\_code\_prefix:** first 5 digits of zip code
- **geolocation\_lat:** latitude
- **geolocation\_lng:** longitude
- **text\_formatgeolocation\_city:** city name
- **text\_formatgeolocation\_state:** state

## Order Payments

This dataset includes data about the orders payment options.

- **order\_id:** unique identifier of an order.
- **payment\_sequential:** a customer may pay an order with more than one payment method. If he does so, a sequence will be created to accommodate all payments.
- **text\_formatpayment\_type:** method of payment chosen by the customer.
- **payment\_installments:** number of installments chosen by the customer.
- **payment\_value:** transaction value.

## Order Reviews

This dataset includes data about the reviews made by the customers.

- **review\_id:** unique review identifier
- **vpn\_keyorder\_id:** unique order identifier
- **review\_score:** Note ranging from 1 to 5 given by the customer on a satisfaction survey.
- **text\_formatreview\_comment\_title:** Comment title from the review left by the customer, in Portuguese.
- **text\_formatreview\_comment\_message:** Comment message from the review left by the customer, in Portuguese.
- **calendar\_todayreview\_creation\_date:** Shows the date in which the satisfaction survey was sent to the customer.
- **calendar\_todayreview\_answer\_timestamp:** Shows satisfaction survey answer timestamp.

## Sellers

This dataset includes data about the sellers that fulfilled orders made at Olist. Use it to find the seller location and to identify which seller fulfilled each product.

- **seller\_id:** seller unique identifier
- **seller\_zip\_code\_prefix:** first 5 digits of seller zip code
- **text\_formatseller\_city:** seller city name
- **text\_formatseller\_state:** seller state

## Categories

Category Name Translation

- **product\_category\_name:** category name in Portuguese
- **Product\_category\_name\_english:** category name in English