The Dataset

This is a Brazilian ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allow 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/lng coordinates. This is real commercial data, it has been anonymized, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

Metadata

Following are the descriptions of different columns of the dataset.

Customers Dataset

1. customer_id: key to the orders dataset. Each order has a unique customer_id.
2. customer_unique_id: unique identifier of a customer.
3. customer_zip_code_prefix: first five digits of customer zip code.
4. customer_city: customer city name
5. customer_state: customer state

Geolocation Dataset

1. geolocation_zip_code_prefix: first 5 digits of zip code
2. geolocation_lat: latitude
3. geolocation_lng: longitude
4. geolocation_city: city name
5. geolocation_state: state

Order Items

1. order_id: order unique identifier
2. order_item_id: sequential number identifying number of items included in the same order.
3. product_id: product unique identifier
4. seller_id: seller unique identifier
5. shipping_limit_date: shows seller shipping limit date for handling the order over to the logistic partner.

Order Payments

1. order_id: unique identifier of an order
2. 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.
3. payment_type: method of payment chosen by a customer
4. payment_installments: number of installments chosen by the customer
5. payment_value: transaction value

Order Reviews

1. review_id: unique review identifier
2. order_id: unique order identifier

3. review_score: note ranging from 1-5 given by the customer on a satisfaction survey.
4. review_comment_title: Comment title from the review left by the customer, in Portuguese.
5. review_comment_message: Comment message from the review left by the customer, in Portuguese.
6. review_creation_date: Shows the date in which the satisfaction survey was sent to the customer.
7. review_answer_timestamp: Shows satisfaction survey answer timestamp

Orders

1. order_id: unique identifier of the order
2. customer_id: key to the customer dataset. Each order has a unique customer_id.
3. order_status: Reference to the order status (delivered, shipped, etc)
4. order_purchase_timestamp: Shows the purchase timestamp
5. order_approved_at: Shows the payment approval timestamp
6. order_delivered_carrier_date: Shows the order posting timestamp. When it was handled to the logistic partner
7. order_delivered_customer_date: Shows the actual order delivery date to the customer
8. order_estimated_delivery_date: Shows the estimated delivery date that was informed to customer at the purchase moment.

Products

1. product_id: unique product identifier
2. product_category_name: root category of product, in Portuguese
3. product_name_length: number of characters extracted from the product name
4. product_description_length: number of characters extracted from the product description
5. product_photos_qty: number of product published photos
6. product_weigth_g: product weight measured in grams
7. product_length_cm: product length measured in centimeters
8. product_height_cm: product height measured in centimeters
9. product_width_cm: product width measured in centimeters

Seller

1. seller_id: seller unique identifier
2. seller_zip_code_prefix: first 5 digits of seller zip code
3. seller_city: seller city name
4. seller_state: seller state