

# DATA DICTIONARY

## 1. product\_category\_name\_translation.csv

- **Description:** This file contains the mapping of product category names from the original language to English.
- **Columns:**
  - product\_category\_name: Name of the product category in the original language.
  - product\_category\_name\_english: Corresponding name of the product category in English.

## 2. products\_dataset.csv

- **Description:** This dataset includes details about the products listed in the marketplace.
- **Columns:**
  - product\_id: Unique identifier for each product.
  - product\_category\_name: Category of the product (e.g., "perfumaria").
  - product\_name\_length: The length of the product name in characters.
  - product\_description\_length: The length of the product description in characters.
  - product\_photos\_qty: Number of photos associated with the product.
  - product\_weight\_g: Weight of the product in grams.
  - product\_length\_cm: Length of the product in centimeters.
  - product\_height\_cm: Height of the product in centimeters.
  - product\_width\_cm: Width of the product in centimeters.

## 3. sellers\_dataset.csv

- **Description:** Contains information about the sellers on the marketplace.
- **Columns:**
  - seller\_id: Unique identifier for each seller.
  - seller\_zip\_code\_prefix: ZIP code prefix of the seller.
  - seller\_city: City where the seller is located.
  - seller\_state: State where the seller is located.

## 4. orders\_dataset.csv

- **Description:** Provides details about customer orders, including status and timestamps for different stages of the order lifecycle.

- **Columns:**

- order\_id: Unique identifier for each order.
- customer\_id: Unique identifier for the customer who placed the order.
- order\_status: Status of the order (e.g., delivered, shipped).
- order\_purchase\_timestamp: Timestamp of when the order was placed.
- order\_approved\_at: Timestamp of when the order was approved.
- order\_delivered\_carrier\_date: Date when the order was handed over to the carrier.
- order\_delivered\_customer\_date: Date when the order was delivered to the customer.
- order\_estimated\_delivery\_date: The estimated delivery date provided to the customer.

#### 5. order\_items\_dataset.csv

- **Description:** Contains information about the individual items in each order.

- **Columns:**

- order\_id: Unique identifier for the order.
- order\_item\_id: Unique identifier for each item within an order.
- product\_id: Identifier of the product ordered.
- seller\_id: Identifier of the seller who fulfilled the item.
- shipping\_limit\_date: Latest date by which the order must be shipped.
- price: Price of the item.
- freight\_value: Shipping cost for the item.

#### 6. order\_payments\_dataset.csv

- **Description:** This dataset includes details about payments made for orders.

- **Columns:**

- order\_id: Unique identifier for the order.
- payment\_sequential: Sequential number of the payment (useful when multiple payments are made for an order).
- payment\_type: Type of payment (e.g., credit card, debit card, voucher).
- payment\_installments: Number of installments used for payment.
- payment\_value: Total value of the payment.

## 7. order\_reviews\_dataset.csv

- **Description:** Contains customer reviews for orders, including review scores and timestamps.
- **Columns:**
  - review\_id: Unique identifier for the review.
  - order\_id: Unique identifier for the order associated with the review.
  - review\_score: Numerical score of the review (e.g., 1-5).
  - review\_comment\_title: Title of the review comment.
  - review\_comment\_message: Message content of the review.
  - review\_creation\_date: Date when the review was created.
  - review\_answer\_timestamp: Timestamp when the review was answered or processed.

## 8. geolocation\_dataset.csv

- **Description:** Provides geolocation information based on ZIP code prefixes, including latitude, longitude, and city/state.
- **Columns:**
  - geolocation\_zip\_code\_prefix: ZIP code prefix for the location.
  - geolocation\_lat: Latitude of the location.
  - geolocation\_lng: Longitude of the location.
  - geolocation\_city: City of the location.
  - geolocation\_state: State of the location.

## 9. customers\_dataset.csv

- **Description:** Contains customer information, including unique identifiers and their location data.
- **Columns:**
  - customer\_id: Unique identifier for each customer.
  - customer\_unique\_id: Anonymized unique identifier for customers.
  - customer\_zip\_code\_prefix: ZIP code prefix of the customer's location.
  - customer\_city: City where the customer is located.
  - customer\_state: State where the customer is located.

This dictionary summarizes the structure of each dataset, providing an overview of the available columns and their respective descriptions.