**Olist Brazilian E-commerce Data Dictionary**

**1. Customers Table (olist\_customers\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| customer\_id | String | Unique ID assigned to each customer |
| customer\_unique\_id | String | Unique identifier for a customer across multiple orders. |
| customer\_zip\_code\_prefix | Integer | First 5 digits of the customer’s postal code. |
| customer\_city | String | City where the customer is located. |
| customer\_state | String | State where the customer is located (two-letter abbreviation, e.g., SP, RJ). |

**2. Geolocation Table (olist\_geolocation\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| geolocation\_zip\_code\_prefix | Integer | First 5 digits of the zip code (used for mapping customer and seller locations). |
| geolocation\_lat | Float | Latitude of the location. |
| geolocation\_lng | Float | Longitude of the location. |
| geolocation\_city | String | City of the location. |
| geolocation\_state | String | State of the location. |

**3. Orders Table (olist\_orders\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| order\_id | String | Unique identifier for each order. |
| customer\_id | String | ID of the customer who placed the order (linked to olist\_customers\_df). |
| order\_status | String | Current status of the order (delivered, canceled, shipped, etc.). |
| order\_purchase\_timestamp | DateTime | Timestamp when the order was placed. |
| order\_approved\_at | DateTime | Timestamp when the order was approved. |
| order\_delivered\_carrier\_date | DateTime | Date when the order was handed over to the carrier. |
| order\_delivered\_customer\_date | DateTime | Date when the order was delivered to the customer. |
| order\_estimated\_delivery\_date | DateTime | Estimated delivery date provided to the customer. |

**4. Order Items Table (olist\_order\_items\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| order\_id | String | ID of the order (linked to olist\_orders\_df). |
| order\_item\_id | Integer | Item number within the order (1st item, 2nd item, etc.). |
| product\_id | String | Unique identifier for the product (linked to olist\_products\_df). |
| seller\_id | String | Unique identifier for the seller. |
| shipping\_limit\_date | DateTime | Latest date for the seller to ship the item. |
| price | Float | Price of the product. |
| freight\_value | Float | Cost of shipping for this item. |

**5. Order Payments Table (olist\_order\_payments\_df)**

| **Column Name** | **Data Type** | | **Description** |
| --- | --- | --- | --- |
| order\_id | String | | ID of the order (linked to olist\_orders\_df). |
| payment\_sequential | Integer | | Payment attempt sequence (1st, 2nd, etc.). |
| payment\_type | String | | Payment method (credit\_card, boleto, voucher, debit\_card). |
| payment\_installments | Integer | Number of installments chosen by the customer. | |
| payment\_value | Float | Total amount paid in this payment attempt. | |

**6. Order Reviews Table (olist\_order\_reviews\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| review\_id | String | Unique identifier for the review. |
| order\_id | String | ID of the order being reviewed (linked to olist\_orders\_df). |
| review\_score | Integer | Rating given by the customer (1-5). |
| review\_comment\_title | String | Title of the review (optional). |
| review\_comment\_message | String | Full text of the customer’s review. |
| review\_creation\_date | DateTime | Date when the review was submitted. |
| review\_answer\_timestamp | DateTime | Date when the seller or platform responded to the review. |

**7. Products Table (olist\_products\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| product\_id | String | Unique identifier for the product. |
| product\_category\_name | String | Product category name (in Portuguese). |
| product\_name\_lenght | Float | Length of the product name (number of characters). |
| product\_description\_lenght | Float | Length of the product description (number of characters). |
| product\_photos\_qty | Float | Number of photos available for the product. |
| product\_weight\_g | Float | Weight of the product in grams. |
| product\_length\_cm | Float | Length of the product in centimeters. |
| product\_height\_cm | Float | Height of the product in centimeters. |
| product\_width\_cm | Float | Width of the product in centimeters. |

**8. Product Category Translation Table (product\_category\_name\_translation\_df)**

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| product\_category\_name | String | Product category name in Portuguese (matches with olist\_products\_df). |
| product\_category\_name\_english | String | Translated product category name in English. |

## ****5. Sellers Table (****olist\_sellers\_df****)****

| **Column Name** | **Data Type** | **Description** |
| --- | --- | --- |
| seller\_id | String | Unique identifier for each seller. |
| seller\_zip\_code\_prefix | Integer | First 5 digits of the seller’s postal code. |
| seller\_city | String | City where the seller is located. |
| seller\_state | String | State where the seller is located (two-letter abbreviation, e.g., SP, RJ). |