**Brazilian E-Commerce Public Dataset by Olist**

All the files are listed below:

[1. olist\_customers\_dataset.csv 1](#_Toc48573098)

[2. olist\_geolocation\_dataset.csv 2](#_Toc48573099)

[3. olist\_orders\_dataset.csv 2](#_Toc48573100)

[4. olist\_order\_items\_dataset.csv 3](#_Toc48573101)

[5. olist\_order\_payments\_dataset.csv 3](#_Toc48573102)

[6. olist\_order\_reviews\_dataset.csv 4](#_Toc48573103)

[7. olist\_products\_dataset.csv 4](#_Toc48573104)

[8. olist\_sellers\_dataset.csv 5](#_Toc48573105)

[9. product\_category\_name\_translation.csv 5](#_Toc48573106)

# olist\_customers\_dataset.csv

**Shape: (99441, 5)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| customer\_id | object | 0 | 00012a2ce6f8dcda20d059ce98491703 | ffffe8b65bbe3087b653a978c870db99 | 99441 |
| customer\_unique\_id | object | 0 | 0000366f3b9a7992bf8c76cfdf3221e2 | ffffd2657e2aad2907e67c3e9daecbeb | 96096 |
| customer\_zip\_code\_prefix | int64 | 0 | 1003 | 99990 | 14994 |
| customer\_city | object | 0 | abadia dos dourados | zortea | 4119 |
| customer\_state | object | 0 | AC | TO | 27 |

**Summary:** 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. Each order assigned to a unique customer id. So the same customer will get different ids for different orders. The purpose of having a customerunique\_id on the dataset is to allow you to identify customers.

# olist\_geolocation\_dataset.csv

**Shape: (1000163, 5)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| geolocation\_zip\_code\_prefix | int64 | 0 | 1001 | 99990 | 19015 |
| geolocation\_lat | float64 | 0 | -36.6054 | 45.06593 | 717358 |
| geolocation\_lng | float64 | 0 | -101.467 | 121.1054 | 717613 |
| geolocation\_city | object | 0 | \* cidade | Ã³leo | 8011 |
| geolocation\_state | object | 0 | AC | TO | 27 |

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

# olist\_orders\_dataset.csv

**Shape: (99441, 8)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| order\_id | object | 0 | 00010242fe8c5a6d1ba2dd792cb16214 | fffe41c64501cc87c801fd61db3f6244 | 99441 |
| customer\_id | object | 0 | 00012a2ce6f8dcda20d059ce98491703 | ffffe8b65bbe3087b653a978c870db99 | 99441 |
| order\_status | object | 0 | Approved | unavailable | 8 |
| order\_purchase\_timestamp | object | 0 | ######## | ######## | 98875 |
| order\_approved\_at | object | 0.16 |  |  | 90733 |
| order\_delivered\_carrier\_date | object | 1.79 |  |  | 81018 |
| order\_delivered\_customer\_date | object | 2.98 |  |  | 95664 |
| order\_estimated\_delivery\_date | object | 0 | ######## | ######## | 459 |

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

# olist\_order\_items\_dataset.csv

**Shape: (112650, 7)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| order\_id | object | 0 | 00010242fe8c5a6d1ba2dd792cb16214 | fffe41c64501cc87c801fd61db3f6244 | 98666 |
| order\_item\_id | int64 | 0 | 1 | 21 | 21 |
| product\_id | object | 0 | 00066f42aeeb9f3007548bb9d3f33c38 | fffe9eeff12fcbd74a2f2b007dde0c58 | 32951 |
| seller\_id | object | 0 | 0015a82c2db000af6aaaf3ae2ecb0532 | ffff564a4f9085cd26170f4732393726 | 3095 |
| shipping\_limit\_date | object | 0 | ######## | ######## | 93318 |
| price | float64 | 0 | 0.85 | 6735 | 5968 |
| freight\_value | float64 | 0 | 0 | 409.68 | 6999 |

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

# olist\_order\_payments\_dataset.csv

**Shape: (103886, 5)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| order\_id | object | 0 | 00010242fe8c5a6d1ba2dd792cb16214 | fffe41c64501cc87c801fd61db3f6244 | 99440 |
| payment\_sequential | int64 | 0 | 1 | 29 | 29 |
| payment\_type | object | 0 | Boleto | voucher | 5 |
| payment\_installments | int64 | 0 | 0 | 24 | 24 |
| payment\_value | float64 | 0 | 0 | 13664.08 | 29077 |

**Summary:**  This dataset includes data about the orders payment options.

# olist\_order\_reviews\_dataset.csv

**Shape: (100000, 7)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| review\_id | object | 0 | 0001239bc1de2e33cb583967c2ca4c67 | fffefe7a48d22f7b32046421062219d1 | 99173 |
| order\_id | object | 0 | 00010242fe8c5a6d1ba2dd792cb16214 | fffe41c64501cc87c801fd61db3f6244 | 99441 |
| review\_score | int64 | 0 | 1 | 5 | 5 |
| review\_comment\_title | object | 88.28 |  |  | 4600 |
| review\_comment\_message | object | 58.25 |  |  | 36921 |
| review\_creation\_date | object | 0 | ######## | ######## | 637 |
| review\_answer\_timestamp | object | 0 | ######## | ######## | 99010 |

**Summary:** This dataset includes data about the reviews made by the customers.

# olist\_products\_dataset.csv

**Shape: (32951, 9)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| product\_id | object | 0 | 00066f42aeeb9f3007548bb9d3f33c38 | fffe9eeff12fcbd74a2f2b007dde0c58 | 32951 |
| product\_category\_name | object | 1.85 |  |  | 73 |
| product\_name\_lenght | float64 | 1.85 | 5 | 76 | 66 |
| product\_description\_lenght | float64 | 1.85 | 4 | 3992 | 2960 |
| product\_photos\_qty | float64 | 1.85 | 1 | 20 | 19 |
| product\_weight\_g | float64 | 0.01 | 0 | 40425 | 2204 |
| product\_length\_cm | float64 | 0.01 | 7 | 105 | 99 |
| product\_height\_cm | float64 | 0.01 | 2 | 105 | 102 |
| product\_width\_cm | float64 | 0.01 | 6 | 118 | 95 |

**Summary:** This dataset includes data about the products sold by Olist.

# olist\_sellers\_dataset.csv

**Shape: (3095, 4)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| seller\_id | object | 0 | 0015a82c2db000af6aaaf3ae2ecb0532 | ffff564a4f9085cd26170f4732393726 | 3095 |
| seller\_zip\_code\_prefix | int64 | 0 | 1001 | 99730 | 2246 |
| seller\_city | object | 0 | 4482255 | Xaxim | 611 |
| seller\_state | object | 0 | AC | SP | 23 |

**Summary:** 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.

# product\_category\_name\_translation.csv

**Shape: (71, 2)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| column\_name | data\_type | null\_value\_pct | min\_value | max\_value | distinct\_value\_cnt |
| product\_category\_name | object | 0 | agro\_industria\_e\_comercio | utilidades\_domesticas | 71 |
| product\_category\_name\_english | object | 0 | agro\_industry\_and\_commerce | watches\_gifts | 71 |

**Summary:** Translates the productcategoryname from Portuguese to English.