

Brazilian E-Commerce Public Dataset by Olist

All the files are listed below:

1.	olist_customers_dataset.csv	1
2.	olist_geolocation_dataset.csv	2
3.	olist_orders_dataset.csv	2
4.	olist_order_items_dataset.csv	3
5.	olist_order_payments_dataset.csv	3
6.	olist_order_reviews_dataset.csv	4
7.	olist_products_dataset.csv	4
8.	olist_sellers_dataset.csv	5
9.	product_category_name_translation.csv	5

1. 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	0000366f3b9a7992bf8c76cfd3221e2	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.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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	ffffefe7a48d22f7b32046421062219d1	99173
order_id	object	0	00010242fe8c5a6d1ba2dd792cb16214	ffe41c64501cc87c801fd61db3f6244	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.

7. 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	ffe9eeff12fcbd74a2f2b007dde0c58	32951
product_category_name	object	1.85			73
product_name_length	float64	1.85	5	76	66
product_description_length	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.

8. 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.

9. 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.