

Brazilian E-Commerce

Data Analyst Portfolio
Project

Daniel Müller

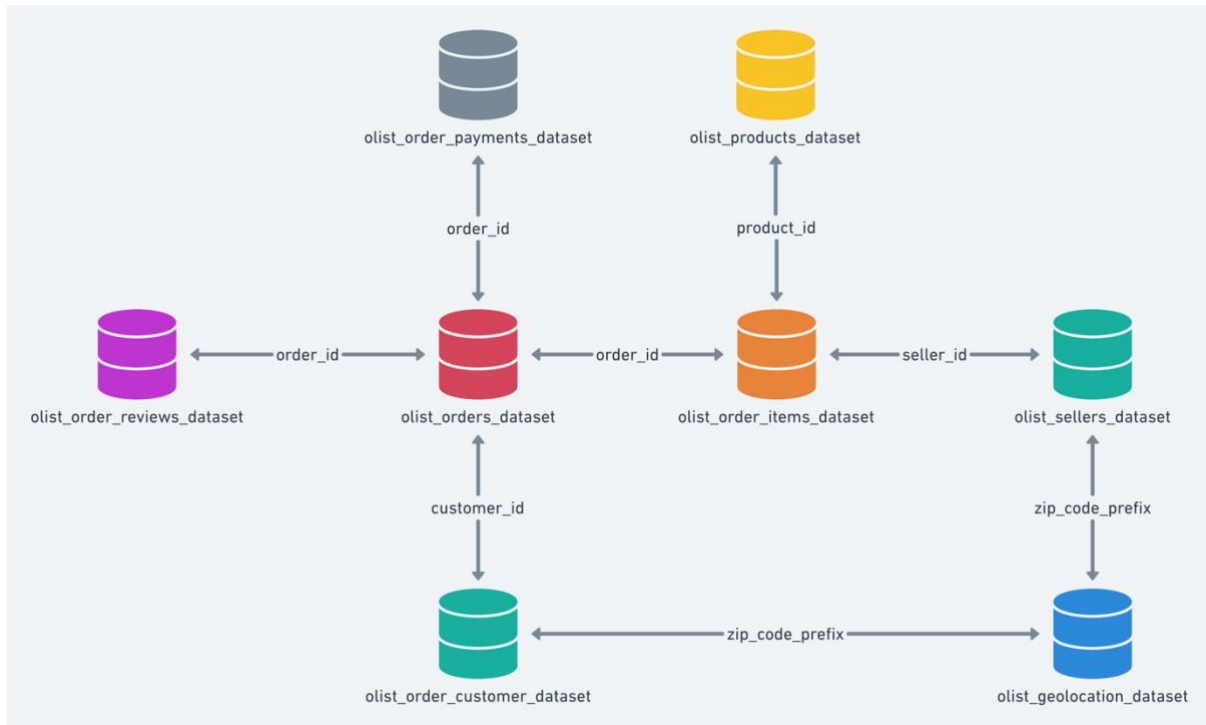


Table of Contents

1	ERD of the Brazilian E-Commerce Public Dataset by Olist	3
2	Data Sets and Source	3
2.1	Source	3
2.2	Data Sets	3
2.2.1	olist_orders_datset	3
2.2.2	olist_order_payments_dataset	5
2.2.3	olist_order_reviews_dataset	6
2.2.4	olist_customers_dataset	7
2.2.5	olist_order_itmes_dataset	8
2.2.6	olist_sellers_dataset	9
2.2.7	olist_geolocation_dataset	10
2.2.8	olist_products_dataset	11
3	Business Questions	12
3.1	Revenue Analysis	12
3.2	Customer Segmentation	12
3.3	Optimize Efficiency	12
3.4	Product Performance	12
3.5	Time-based Trends	12

Table 1:	Data Set: olist_orders_datset	4
Table 2:	Data Profile: olist_orders_datset	4
Table 3:	Data Set: olist_order_payments_dataset	5
Table 4:	Data Profile: olist_order_payments_dataset	5
Table 5:	Data Set: olist_order_reviews_dataset	6
Table 6:	Data Profile: olist_order_reviews_dataset	6
Table 7:	Data Set: olist_customers_dataset	7
Table 8:	Data Profile: olist_customers_dataset	7
Table 9:	Data Set: olist_order_itmes_dataset	8
Table 10:	Data Profile: olist_order_itmes_dataset	8
Table 11:	Data Set: olist_sellers_dataset	9
Table 12:	Data Profile: olist_sellers_dataset	9
Table 13:	Data Set: olist_geolocation_dataset	10
Table 14:	Data Profile: olist_geolocation_dataset	10
Table 15:	Data Set: olist_products_dataset	11
Table 16:	Data Profile: olist_products_dataset	11

1 ERD of the Brazilian E-Commerce Public Dataset by Olist



2 Data Sets and Source

2.1 Source

Brazilian E-Commerce Public Dataset by Olist can be accessed on [Kaggle](#)

2.2 Data Sets

2.2.1 olist_orders_datset

olist_orders_datset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about customer ID, order status, order purchase time stamp, and additional insight into when the order was delivered.

Data Relevance	The data is relevant for estimating the shipping performance and developing improvements as well as conducting predictive analysis of future sales
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 1: Data Set: olist_orders_datset

2.2.1.1 Data Profile

Variables	Time-variant/ -invariant	Structured/ Unstructured	Qualitative / Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
order_id	Time-invariant	Structured	Qualitative	Ordinal
Customer_id	Time-invariant	Structured	Qualitative	Ordinal
Order_status	Time-invariant	Structured	Qualitative	Ordinal
Order_purchase_timestamp	Time-invariant	Structured	Qualitative	Ordinal
Order_approved_at	Time-invariant	Structured	Qualitative	Ordinal
Order_delivered_carrier_date	Time-invariant	Structured	Qualitative	Ordinal
Order_estimated_delivery_date	Time-invariant	Structured	Qualitative	Ordinal

Table 2: Data Profile: olist_orders_datset

2.2.2 olist_order_payments_dataset

olist_order_payments_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about payment type, payment installment, payment value
Data Relevance	The data is relevant to calculate the customer live time value
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 3: Data Set: olist_order_payments_dataset

2.2.2.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Order_id	Time-invariant	Structured	Qualitative	Ordinal
Payment_sequential	Time-invariant	Structured	Qualitative	Ordinal
Payment_type	Time-invariant	Structured	Qualitative	Ordinal
Payment_installment	Time-invariant	Structured	Quantitative	Discrete
Payment_value	Time-invariant	Structured	Quantitative	Continuous

Table 4: Data Profile: olist_order_payments_dataset

2.2.3 olist_order_reviews_dataset

olist_order_reviews_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about review scores, comments, and time stamps related to the review.
Data Relevance	The data is relevant to calculating customer satisfaction value
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 5: Data Set: olist_order_reviews_dataset

2.2.3.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Review_id	Time-invariant	Structured	Qualitative	Ordinal
Order_id	Time-invariant	Structured	Qualitative	Ordinal
Review_score	Time-variant	Structured	Quantitative	Discrete
Review_comment_title	Time-invariant	Unstructured	Qualitative	Nominal
Review_comment_message	Time-invariant	Unstructured	Qualitative	Nominal
Review_creation_date	Time-invariant	Structured	Qualitative	Ordinal
Review_answer_timestamp	Time-invariant	Structured	Qualitative	Ordinal

Table 6: Data Profile: olist_order_reviews_dataset

2.2.4 olist_customers_dataset

olist_customers_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about the customers, like zip code, city, and state
Data Relevance	The data is relevant to conduct a geospatial analysis and determine the location of High Lifetime Value customers, as well as calculating the churn rate which is important to determine whether the customers are active or have stopped purchasing at Olist.
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 7: Data Set: olist_customers_dataset

2.2.4.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Customer_id	Time-invariant	Structured	Qualitative	Ordinal
Customer_unique_id	Time-invariant	Structured	Qualitative	Ordinal
Customer_zip_code_prefix	Time-invariant	Structured	Quantitative	Discrete
Customer_city	Time-invariant	Structured	Qualitative	Nominal
Customer_state	Time-invariant	Structured	Qualitative	Nominal

Table 8: Data Profile: olist_customers_dataset

2.2.5 olist_order_itmes_dataset

olist_order_itmes_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about the items like seller id, price per item, freight value
Data Relevance	The data is relevant to determine areas with the most profitable sellers after being merged with the seller's dataset
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 9: Data Set: olist_order_itmes_dataset

2.2.5.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Order_id	Time-invariant	Structured	Qualitative	Ordinal
Order_item_id	Time-invariant	Structured	Qualitative	Ordinal
Product_id	Time-invariant	Structured	Qualitative	Ordinal
Seller_id	Time-invariant	Structured	Qualitative	Ordinal
Shipping_limit_date	Time-invariant	Structured	Qualitative	Ordinal
Price	Time-invariant	Structured	Quantitative	Continuous
Freight_value	Time-invariant	Structured	Quantitative	Continuous

Table 10: Data Profile: olist_order_itmes_dataset

2.2.6 olist_sellers_dataset

olist_sellers_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information about the sellers like zip code, city, and state
Data Relevance	The data is relevant to determine areas with the most profitable sellers after being merged with the order item dataset
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 11: Data Set: olist_sellers_dataset

2.2.6.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Seller_id	Time-invariant	Structured	Qualitative	Ordinal
Seller_zip_code_prefix	Time-invariant	Structured	Quantitative	Discrete
Seller_city	Time-invariant	Structured	Qualitative	Nominal
Seller_state	Time-invariant	Structured	Qualitative	Nominal

Table 12: Data Profile: olist_sellers_dataset

2.2.7 olist_geolocation_dataset

olist_geolocation_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides geolocational information about Brazil
Data Relevance	The data is relevant for geospatial analysis as well as for creating visualizations in the later stage of the project
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 13: Data Set: olist_geolocation_dataset

2.2.7.1 Data Profile

Variables	Time-variant/ -invariant	Structured/ Unstructured	Qualitative / Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Geolocation_zip_code_prefix	Time-invariant	Structured	Quantitative	Discrete
Geolocation_lat	Time-invariant	Structured	Quantitative	Discrete
Geolaction_lng	Time-invariant	Structured	Quantitative	Discrete
Geolocation_city	Time-invariant	Structured	Qualitative	Nominal
Geolocation_state	Time-invariant	Structured	Qualitative	Nominal

Table 14: Data Profile: olist_geolocation_dataset

2.2.8 olist_products_dataset

olist_products_dataset	
Source: Olist	
Data Source	The data is an external source, collected by Olist, which is a large department store in Brazilian marketplaces.
Data Collection	The data was collected directly at Olist due to the system
Data Contents	The data set provides information on the product category, name, description, and different benchmark values like length, etc.
Data Relevance	The data is relevant for geospatial analysis as well as for creating Visualizations in the later stage of the project
Data Limitations	The data limitations occur due to system errors when collecting the data at Olist

Table 15: Data Set: olist_products_dataset

2.2.8.1 Data Profile

Variables	Time-variant/-invariant	Structured/Unstructured	Qualitative/Quantitative	Qualitative: Nominal/Ordinal Quantitative: Discrete/Continuous
Product_id	Time-invariant	Structured	Qualitative	Ordinal
Product_category_name	Time-invariant	Structured	Qualitative	Ordinal
Product_name_lenght	Time-invariant	Structured	Quantitative	Discrete
Product_description_lenght	Time-invariant	Structured	Quantitative	Discrete
Product_photos_qty	Time-invariant	Structured	Quantitative	Discrete
Product_weight_g	Time-invariant	Structured	Quantitative	Continuous
Product_length_cm	Time-invariant	Structured	Quantitative	Continuous
Product_height_cm	Time-invariant	Structured	Quantitative	Continuous
Product_width_cm	Time-invariant	Structured	Quantitative	Continuous

Table 16: Data Profile: olist_products_dataset

3 Business Questions

3.1 Revenue Analysis

- What is the average revenue per order, and how does it vary across different product categories?
- Does revenue differ across regions in Brazil?

3.2 Customer Segmentation

- Are there identifiable clusters of customers based on purchase behavior?
- Are customers without writing a review either satisfied or dissatisfied?
- Where are customers with a high lifetime value located?
- What is the customer satisfaction value?
- Do customers who leave a review show higher loyalty or repeat purchase behavior than those who don't?
- What is the churn rate of Olist customers, and what patterns are evident among customers who churn versus those who remain active?

3.3 Optimize Efficiency

- What is the average delivery time by region and product category, and how can it be improved?

3.4 Product Performance

- What are the top 10 products in terms of total sales?

3.5 Time-based Trends

- Are there seasonal trends in sales volume, revenue, or return rates?
- Does customer behavior change during specific times of the year, such as holidays or weekends?