**Brazilian E-Commerce**

**1a. Data Source:**

Brazilian E-Commerce Public Dataset by Olist, [Kaggle](https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce?resource=download).

**1b. Data Collection:**

The dataset was provided by Olist, the largest department store in Brazilian marketplaces. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil.

**1c. Data Contents:**

The dataset includes order status, price, payment and freight performance to customer location, product attributes, geolocation, and reviews written by customers.

**1d. Data Profile:**

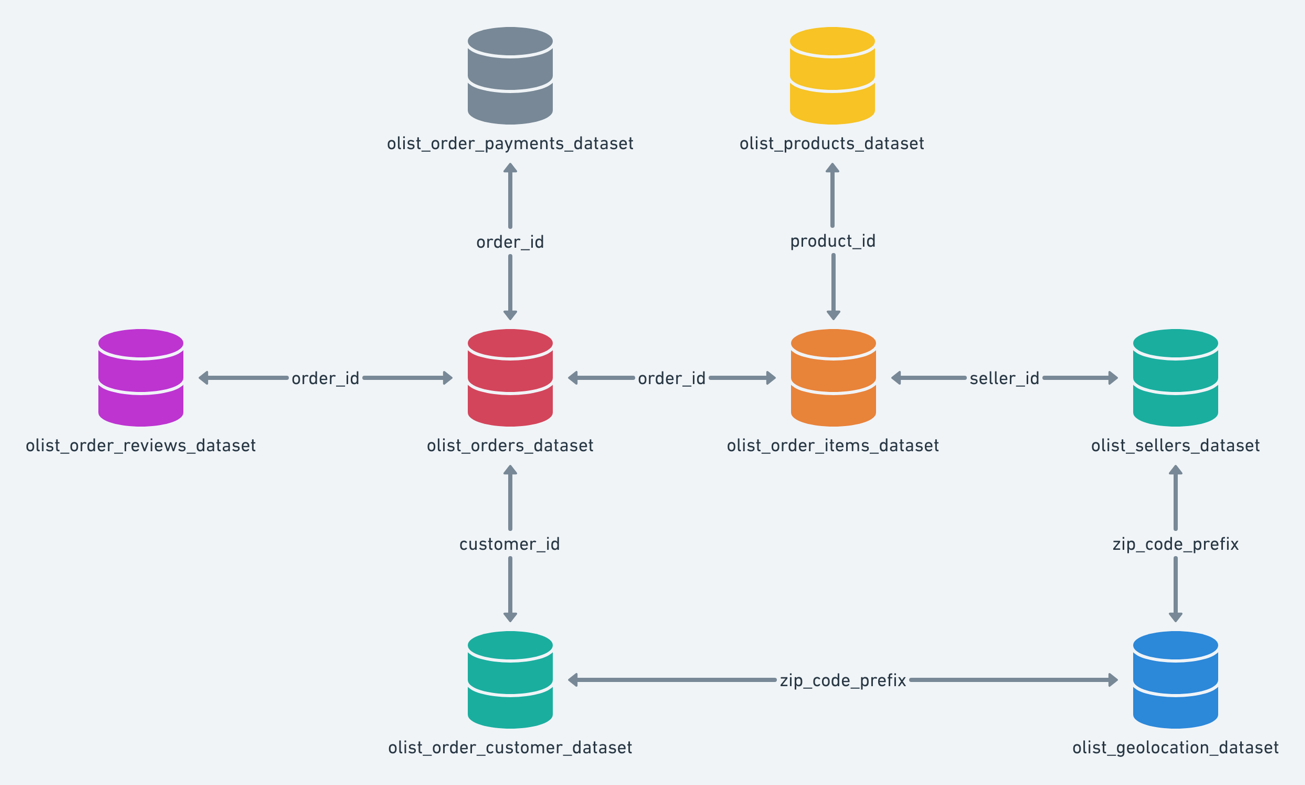
The dataset is split by product category name, sellers, products, orders, reviews, payments, items, geolocation, and customers.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Index | Columns | Description | Time Variant/ Invariant | Data Type |
| 1. | Seller\_id | Seller ID | Invariant | Qualitative |
| 2. | Seller\_zip\_code\_prefix | Seller Zip Code | Invariant | Qualitative |
| 3. | Seller\_city | Seller City | Invariant | Qualitative |
| 4. | Seller\_state | Seller State | Invariant | Qualitative |
| 5. | Product\_category\_name | Product category name in Portuguese | Invariant | Qualitative |
| 6. | Product\_category\_name\_english | Product category name in English | Invariant | Qualitative |
| 7. | Product\_id | Product ID | Invariant | Qualitative |
| 8. | Product\_category\_name | Product Category Name | Invariant | Qualitative |
| 9. | Product\_name\_length | Product name character length in Portuguese | Invariant | Quantitative |
| 10. | Product\_description\_length | Product description character length | Invariant | Quantitative |
| 11. | Product\_photos\_qty | Product photo quantity | Variant | Quantitative |
| 12. | Product\_weight\_g | Product weight in grams | Invariant | Quantitative |
| 13. | Product\_length\_cm | Product length in centimeters | Invariant | Quantitative |
| 14. | Product\_width\_cm | Product width in centimeters | Invariant | Quantitative |
| 15. | Order\_id | Order ID | Invariant | Qualitative |
| 16. | Customer\_id | Customer ID | Invariant | Qualitative |
| 17. | Order\_status | Order Status | Invariant | Quantitative |
| 18. | Order\_purchase\_timestamp | Time of purchase | Variant | Quantitative |
| 19. | Order\_apporved\_at | Time of order approval | Variant | Quantitative |
| 20. | Order\_delivered\_carrier\_date | Date product was delivered to carrier | Variant | Quantitative |
| 21. | Order\_delivered\_customer\_date | Date product was delivered to customer | Variant | Quantitative |
| 22. | Order\_estimated\_delivery\_date | Estimated delivery date | Variant | Quantitative |
| 23. | Review\_id | Review ID | Invariant | Qualitative |
| 24. | Order\_id | Order ID | Invariant | Qualitative |
| 25. | Review\_score | Review Score | Variant | Quantitative |
| 26. | Review\_comment\_title | Review comment title | Invariant | Qualitative |
| 27. | Review\_comment\_message | Review left by costumer | Invariant | Qualitative |
| 28. | Order\_id | Payment Order\_ID | Invariant | Qualitative |
| 29. | Payment\_sequential | Payment Sequential | Invariant | Quantitative |
| 30. | Payment\_type | Payment Type | Invariant | Qualitative |
| 31. | Payment\_installments | Payment Installments | Invariant | Quantitative |
| 32. | Payment\_value | Payment Value | Invariant | Quantitative |
| 33. | Order\_id | Order ID | Invariant | Qualitative |
| 34. | Order\_item | Order Item | Invariant | Qualitative |
| 35. | Product\_id | Product ID | Invariant | Qualitative |
| 36. | Seller\_id | Seller ID | Invariant | Qualitative |
| 37. | Shipping\_limit\_date | Shipping limit date | Variant | Qualitative |
| 38. | Price | Price | Invariant | Quantitative |
| 39. | Freight\_value | Freight Value | Invariant | Quantitative |
| 40. | Geolocation\_zip\_code\_prefix | Zip Code Prefix | Invariant | Qualitative |
| 41. | Geolocation\_lat | Latitude | Invariant | Qualitative |
| 42. | Geolocation\_lng | Longitude | Invariant | Qualitative |
| 43. | Geolocation\_city | City | Invariant | Qualitative |
| 44. | Geolocation\_state | State | Invariant | Qualitative |
| 45. | Customer\_id | Customer ID | Invariant | Qualitative |
| 46. | Customer\_unique\_id | Customer Unique ID | Invariant | Qualitative |
| 47. | Customer\_zip\_code\_prefix | Zip Code Prefix | Invariant | Qualitative |
| 48. | Customer\_city | Customer City | Invariant | Qualitative |
| 49. | Customer\_state | Customer State | Invariant | Qualitative |

**Data Cleaning:**

* In the payments column I changed ‘boletos’ to ‘voucher’ for consistency.
* Deleted the ‘not defined’ payment types in the payments tabs (3 rows deleted).
* Changed ‘s√£o paulo’ to ‘sao paulo’ for consistency.
* Changed ‘\*cidade’ to Curitiba (based on the lat and long).
* Changed ‘¬¥teresopolis’ to Teresopolis.
* Replaced √© with E.
* Replaced √≥ with O.

**Data Schema**



**Limitations and Ethics:**

Since the data was provided by Olist (the owner) there are no ethical issues in terms of use. Some limitations might be that an order might have multiple items, each item might be fulfilled by a distinct seller, and all the identifying stores and partners were replaced with codenames.

**Questions to Explore:**

1. What are the overall sales trends between 2016 and 2018?

* The year of 2017 had the highest revenue.

A picture containing diagram, plot, line, text

Description automatically generated

1. Which products or product categories have the highest sales volume or revenue?

* Housewares have the highest revenue.

A screenshot of a computer program

Description automatically generated with low confidence

1. Are there any seasonal or cyclical patterns in the sales data?

* I was expecting the holiday season to have an increase in revenue to the numbers show otherwise.

A picture containing line, diagram, text, plot

Description automatically generated

1. What is the average order values?
2. How does sales performance vary across different states and cities in Brazil?