Brazilian E-Commerce Public Dataset by Olist

Welcome! This is a Brazilian ecommerce public dataset of orders made at [Olist Store](http://www.olist.com/). The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.

This is real commercial data, it has been anonymized, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

Join it With the Marketing Funnel by Olist

We have also released a [Marketing Funnel Dataset](https://www.kaggle.com/olistbr/marketing-funnel-olist/home). You may join both datasets and see an order from Marketing perspective now!

Instructions on joining are available on this [Kernel](https://www.kaggle.com/andresionek/joining-marketing-funnel-with-brazilian-e-commerce).

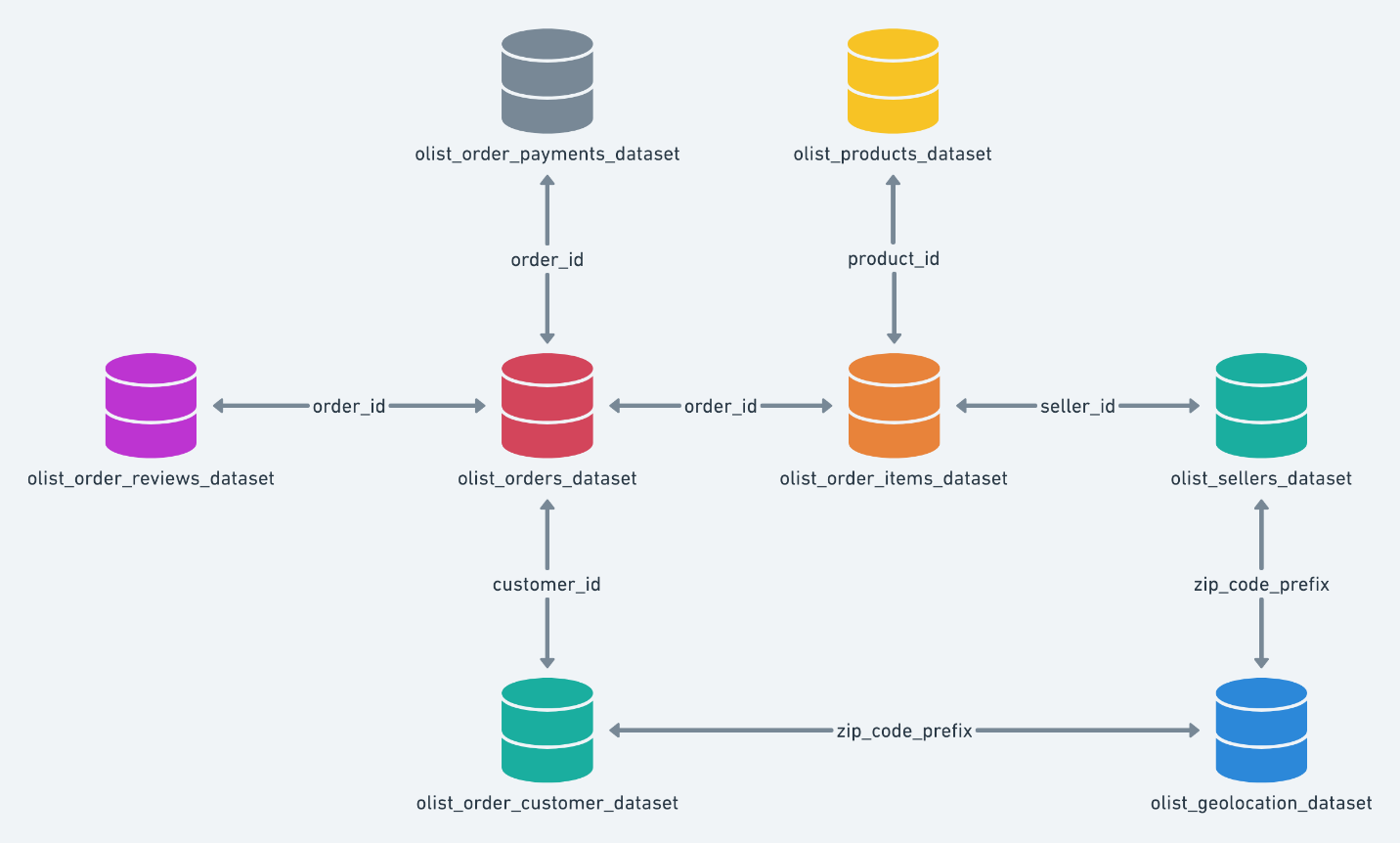
Context

This dataset was generously provided by Olist, the largest department store in Brazilian marketplaces. Olist connects small businesses from all over Brazil to channels without hassle and with a single contract. Those merchants are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners. See more on our website: [www.olist.com](https://www.olist.com/)

After a customer purchases the product from Olist Store a seller gets notified to fulfill that order. Once the customer receives the product, or the estimated delivery date is due, the customer gets a satisfaction survey by email where he can give a note for the purchase experience and write down some comments.

Attention

1. An order might have multiple items.
2. Each item might be fulfilled by a distinct seller.
3. All text identifying stores and partners were replaced by the names of Game of Thrones great houses.



products

RangeIndex: 32951 entries, 0 to 32950

# Column Non-Null Count Dtype

0 product\_id 32951 non-null object

1 product\_category\_name 32341 non-null object

2 product\_name\_lenght 32341 non-null float64

3 product\_description\_lenght 32341 non-null float64

4 product\_photos\_qty 32341 non-null float64

5 product\_weight\_g 32949 non-null float64

6 product\_length\_cm 32949 non-null float64

7 product\_height\_cm 32949 non-null float64

8 product\_width\_cm 32949 non-null float64

customers

RangeIndex: 99441 entries, 0 to 99440

# Column Non-Null Count Dtype

0 customer\_id 99441 non-null object

1 customer\_unique\_id 99441 non-null object

2 customer\_zip\_code\_prefix 99441 non-null int64

3 customer\_city 99441 non-null object

4 customer\_state 99441 non-null object

orders

RangeIndex: 99441 entries, 0 to 99440

# Column Non-Null Count Dtype

0 order\_id 99441 non-null object

1 customer\_id 99441 non-null object

2 order\_status 99441 non-null object

3 order\_purchase\_timestamp 99441 non-null object

4 order\_approved\_at 99281 non-null object

5 order\_delivered\_carrier\_date 97658 non-null object

6 order\_delivered\_customer\_date 96476 non-null object

7 order\_estimated\_delivery\_date 99441 non-null object

orders\_items

RangeIndex: 112650 entries, 0 to 112649

# Column Non-Null Count Dtype

0 order\_id 112650 non-null object

1 order\_item\_id 112650 non-null int64

2 product\_id 112650 non-null object

3 seller\_id 112650 non-null object

4 shipping\_limit\_date 112650 non-null object

5 price 112650 non-null float64

6 freight\_value 112650 non-null float64

memory usage: 6.0+ MB

geo\_location

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 1000163 entries, 0 to 1000162

Data columns (total 5 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 geolocation\_zip\_code\_prefix 1000163 non-null int64

1 geolocation\_lat 1000163 non-null float64

2 geolocation\_lng 1000163 non-null float64

3 geolocation\_city 1000163 non-null object

4 geolocation\_state 1000163 non-null object

dtypes: float64(2), int64(1), object(2)

memory usage: 38.2+ MB

product\_name\_eng

RangeIndex: 71 entries, 0 to 70

# Column Non-Null Count Dtype

0 product\_category\_name 71 non-null object

1 product\_category\_name\_english 71 non-null object

dtypes: object(2)

memory usage: 1.2+ KB

Olist is the largest department store in Brazilian marketplaces. It connects small businesses from all over Brazil to channels without hassle and with a single contract. Those merchants are able to sell their products through the Olist Store and ship them directly to the customers using Olist logistics partners.

Olist is experiencing good growth. Order counts, customer counts and revenue are all increasing. However, there are a few areas that Olist needs to keep and eye on to ensure that this continues.

Any advertising campaign should focus on the southeastern states: Sao Paulo, Rio de Janeiro, Minas Gerais, Rio Grande do Sul, and Parana. These states accounted for approximately 75% of sales in 2018. Olist should also concentrate on the top five categories: health\_beauty, watches\_gifts, bed\_bath\_table, sports\_leisure, and computers\_accessories. These categories accounted for approximately 42% of sales. If the next five categories; housewares, furniture\_decor, auto, baby, and cool\_stuff were added, then approximately 60% of sales would be covered.

satisfied.