

# Executive Summary

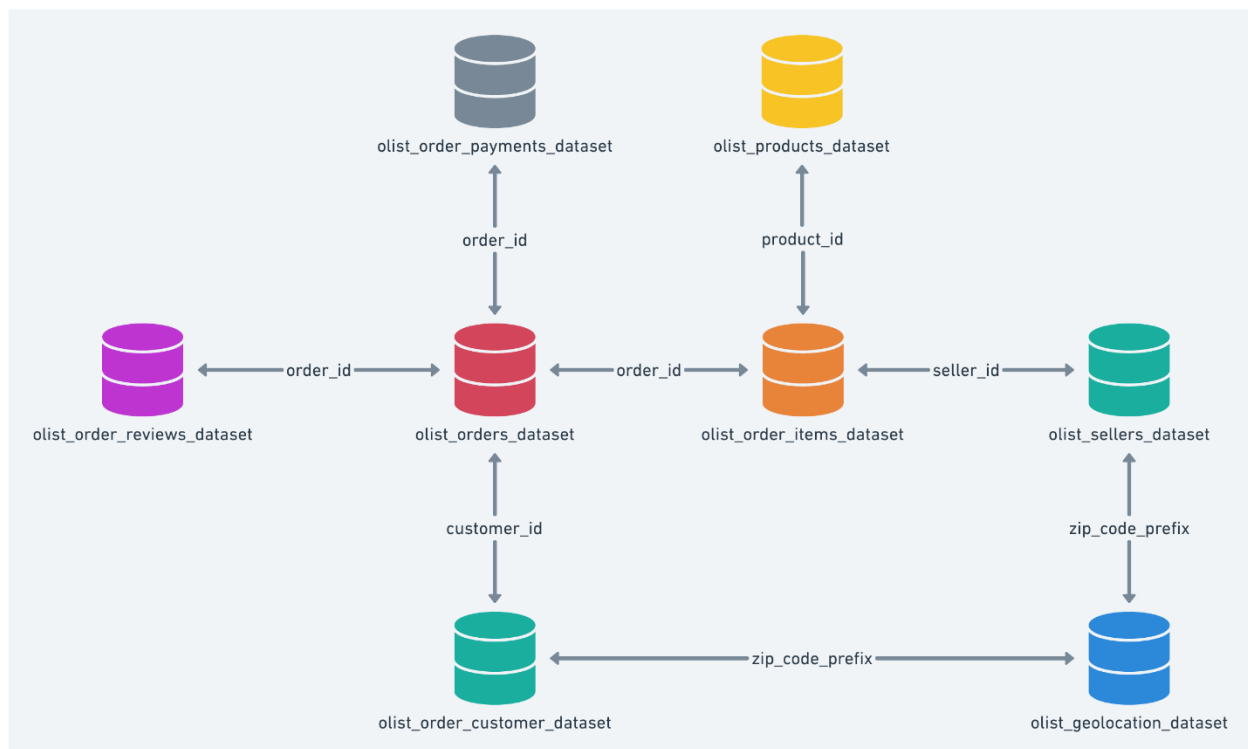
Olist is a Brazilian e-commerce platform that connects buyers and sellers throughout the country. We were asked to conduct an analysis on the factors behind on-time and late delivery trends.

## Data Source

**Data Source:** The data comes from Olist Store, a Brazilian e-commerce platform that connects small businesses with multiple sales channels. The data is publicly-available on [Kaggle](#).

**Data Collection:** The data was collected by Olist, through its website, which tracks customer reviews through surveys that are automatically sent out after the order is fulfilled.

**Data Content:** The dataset includes customer location information, geographic information for Brazilian cities, order item data, payment data, customer review data, order fulfillment data, product information data, seller location data, and product category information. Key information includes: buyer and seller locations; order, payment approval, and delivery dates; product dimensions and photos; and product categories. The schema below outlines the connections between the various datasets.



**Data Profile:**

1.) The new csv files has the following dimensions:

<u>Variables</u>	<u>Time Variant/ Time Invariant</u>	<u>Quantitative/ Qualitative</u>	<u>Nominal/Ordinal Discrete/Continuous</u>	<u>Python Data Type</u>
customer_zip_code	Invariant	Quantitative	Discrete	int64
customer_city	Invariant	Qualitative	Nominal	object
customer_state	Invariant	Qualitative	Nominal	object
order_purchase_timestamp	Variant	Quantitative	Continuous	object
order_approved_at	Variant	Quantitative	Continuous	object
order_delivered_carrier_date	Variant	Quantitative	Continuous	object
order_delivered_customer_date	Variant	Quantitative	Continuous	object
order_estimated_delivery_date	Variant	Quantitative	Discrete	object
shipping_limit_date	Variant	Quantitative	Continuous	object
price	Invariant	Quantitative	Continuous	float64
freight_value	Invariant	Quantitative	Continuous	float64
seller_zip_code	Invariant	Quantitative	Discrete	int64
seller_lat	Invariant	Quantitative	Continuous	float64
seller_lng	Invariant	Quantitative	Continuous	float64
seller_city	Invariant	Qualitative	Nominal	object
seller_state	Invariant	Qualitative	Nominal	object
Delivery-Purchase_Time_Difference_Minutes	Variant	Quantitative	Continuous	float64
Delivery-Purchase_Time_Difference_Min	Variant	Quantitative	Continuous	float64
Estimated-Purchase_Time_Difference_Min	Variant	Quantitative	Continuous	float64

Approval-Purchase_Time_Difference_Min	Variant	Quantitative	Continuous	float64
Carrier_Delivered-Order_Time_Difference_Min	Variant	Quantitative	Continuous	float64

**Limitations and Ethics:** The data is genuine commercial data that has been anonymized for privacy and ethical reasons. The datasets do contain somewhat sensitive information, including uniquely-identifying customer IDs, but there is no personally-identifying information present in these datasets.

## Questions

### Clarifying Questions

- 1.) **How often are customer orders delivered early, on-time, or late, based on the estimated arrival time provided at the time of purchase?**

The vast majority of orders are delivered on-time relative to their estimated arrival date. I ended up not testing for this as it became clear that the numbers of orders fulfilled after their estimated delivery time were an infinitesimally small portion of the overall dataset.

- 2.) **In which zip codes are buyers getting the most on-time deliveries?**

I ended up revising this question to look at the zip codes by order fulfillment time length.

- 3.) **From which zip codes are sellers providing the most on-time deliveries?**

There's no strong relationship between a seller's zip code and the time it takes for delivery of an order.

- 4.) **Is there a connection between freight price and on-time delivery?**

- 5.) **What's the typical time lag between handing over the order to the logistics partner and customer receipt of the order? How does the time lag vary for on-time versus late deliveries?**

### Adjoining Questions

- 6.) **What's the impact of longer seller-to-carrier times on overall fulfillment times (i.e. purchase to delivery)?**

7.) How accurate is the estimated delivery time with actual fulfillment time?

**Funneling Questions**

8.) In which buyer Zip Codes, States, and/or cities are we projecting the most accurate estimated delivery times? Where are we most incorrect?

**Elevating Questions**

9.) What portions of the ordering-fulfillment cycle would create the greatest impact in ensuring on-time delivery?

## **Hypotheses**

- 1.) The effect of Customer Zip Code on fulfillment and estimated fulfillment times is a function of their distance from the main southern coastal regions of the nation.
- 2.) Of all the stages in the fulfillment process, the time lag between ordering and delivery to the carrier has the greatest impact on overall fulfillment time.