

# **Project Brief: Olist Brazilian E-Commerce Sales Analytics**

1.0 Executive Summary & Business Objectives

2.0 Data Sources & Engineering (ETL)

3.0 Data Model & DAX Measures

4.0 Analysis & Key Insights

5.0 Recommendations

6.0 Appendix

# 1.0 Executive Summary & Business Objectives

## Business Context

Olist is a Brazilian online marketplace that connects small businesses to major e-commerce platforms. It operates similarly to Amazon's marketplace model, enabling sellers across the country to reach customers through Olist's platform. With thousands of sellers, millions of orders, and diverse product categories, Olist generates a rich dataset that provides insights into customer behavior, sales performance, logistics, and satisfaction.

## Project Objective

The objective of this project is to provide leadership with a clear understanding of Olist's recent performance between 2016 and 2018 and customer experience by:

- **Analyzing sales trends** to identify which product categories and sellers are driving the most revenue.
- **Evaluating delivery performance** to measure on-time vs delayed orders and determine how delays impact customer satisfaction.
- **Assessing customer behavior** by distinguishing repeat customers from one-time buyers and quantifying their contribution to sales.

## 2.0 Data Sources & Engineering (ETL)

### Data Sources

The analysis utilizes the public Olist dataset, consisting of the following key tables:

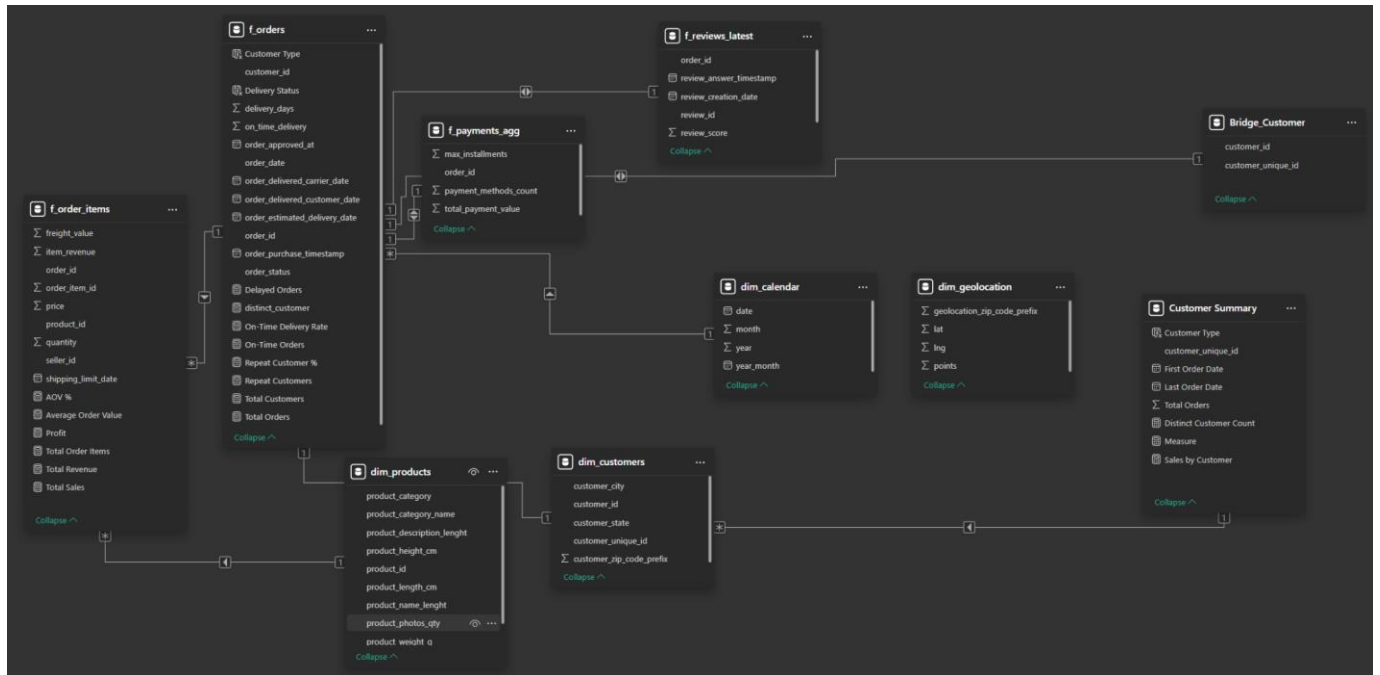
- Olist\_customers\_dataset.csv
  - Purpose: Master list of all customers.
  - Key Columns: customer\_id, customer\_unique\_id (the true identifier for a customer), customer\_zip\_code\_prefix, customer\_city, customer\_state.
  - ETL Steps: Loaded as a dimension table (dim\_customers). Removed duplicates on customer\_id to ensure one record per customer.
- Olist\_geolocation\_dataset.csv
  - Purpose: Mapping of Brazilian zip codes to latitude and longitude coordinates.
  - Key Columns: geolocation\_zip\_code\_prefix, geolocation\_lat, geolocation\_lng, geolocation\_city, geolocation\_state.
  - ETL Steps: To optimize performance, the table was grouped by geolocation\_zip\_code\_prefix and the average latitude and longitude were calculated, significantly reducing the row count. This cleaned table was loaded as dim\_geolocation.
- Olist\_order\_items\_dataset.csv
  - Purpose: Contains each item purchased within every order.
  - Key Columns: order\_id, order\_item\_id, product\_id, seller\_id, price, freight\_value.
  - ETL Steps: Loaded as a fact table (f\_order\_items). A new column, item\_revenue, was created from the price column to clearly represent the sales value for each line item.
- Olist\_order\_payments\_dataset.csv
  - Purpose: Details of all payment transactions for orders.
  - Key Columns: order\_id, payment\_sequential, payment\_type, payment\_installments, payment\_value.
  - ETL Steps: Aggregated to the order level (f\_payments\_agg) by summing payment\_value and calculating the maximum number of installments per order\_id, creating a one-to-one relationship with the orders table.
- Olist\_order\_reviews\_dataset.csv
  - Purpose: Customer reviews and scores for orders.
  - Key Columns: order\_id, review\_score, review\_creation\_date, review\_answer\_timestamp.
  - ETL Steps: Loaded as a fact table (f\_reviews\_latest). The table was sorted by review\_answer\_timestamp and duplicates were removed on order\_id to keep only the most recent review for each order.
- Olist\_orders\_dataset.csv
  - Purpose: The core fact table with metadata for every order.

- Key Columns: order\_id, customer\_id, order\_status, order\_purchase\_timestamp, order\_approved\_at, order\_delivered\_carrier\_date, order\_delivered\_customer\_date, order\_estimated\_delivery\_date.
- ETL Steps: Loaded as a fact table (f\_orders). Critical calculated columns were added:
  - order\_date: Extracted the date from order\_purchase\_timestamp.
  - delivery\_days: Calculated the number of days between order and delivery.
  - on\_time\_delivery: This curated star schema ensures optimal performance for reporting and analysis in Power BI.
- Olist\_products\_dataset.csv
  - Purpose: Information about each unique product.
  - Key Columns: product\_id, product\_category\_name, product\_name\_length, product\_description\_length, product\_photos\_qty, product\_weight\_g, product\_length\_cm, product\_height\_cm, product\_width\_cm.
  - ETL Steps: Loaded as a dimension table (dim\_products). The Portuguese category names were translated to English by merging with the translation table. Duplicates were removed on product\_id.
- Olist\_sellers\_dataset.csv
  - Purpose: Master list of all sellers.
  - Key Columns: seller\_id, seller\_zip\_code\_prefix, seller\_city, seller\_state.
  - ETL Steps: Loaded as a dimension table (dim\_sellers). Removed duplicates on seller\_id to ensure one record per seller.
- Product\_category\_name\_translation.csv
  - Purpose: Maps Portuguese product category names to English.
  - Key Columns: product\_category\_name, product\_category\_name\_english.
  - ETL Steps: Used as a reference table to enrich the dim\_products table during its ETL process.

ETL Architecture: A strict "raw" and "clean" separation was maintained in Power Query. All original CSV files were loaded as connection-only "raw" queries. Transformations were applied to create new "clean" fact (f\_\*) and dimension (dim\_\*) tables, which were then loaded into the Power BI data model. This ensures reproducibility and easy refreshes of the data.

## 3.0 Data Model & DAX Measures

Data Model Relationships: A star schema was implemented in Power BI to ensure efficient and accurate analysis. The model centers around the `f_orders` and `f_order_items` fact tables, which are connected to various dimension tables for slicing and filtering.



Key Performance Indicators (DAX Measures): The following core measures were created to power the dashboard analytics:

- Total Sales:  $\text{Total Sales} = \text{SUM}(\text{f\_order\_items}[\text{item\_revenue}])$ 
  - The sum of revenue from all items sold.
- Total Orders:  $\text{Total Orders} = \text{DISTINCTCOUNT}(\text{f\_orders}[\text{order\_id}])$ 
  - The distinct count of all orders placed.
- Total Customers:  $\text{Total Customers} = \text{DISTINCTCOUNT}(\text{dim\_customers}[\text{customer\_unique\_id}])$ 
  - The distinct count of all unique customers.
- Average Order Value (AOV):  $\text{AOV} = \text{DIVIDE}([\text{Total Sales}], [\text{Total Orders}])$ 
  - The average revenue generated per order.
- On-Time Delivery %:  $\text{On-Time Delivery \%} = \text{DIVIDE}([\text{On-Time Deliveries}], [\text{Delivered Orders}], 0)$ 
  - The percentage of delivered orders that were on time or early.
- Delivered Orders:  $\text{Delivered Orders} = \text{CALCULATE}([\text{Total Orders}], \text{f\_orders}[\text{order\_status}] = \text{"delivered"})$ 
  - The count of orders with a status of "delivered".

- On-Time Deliveries: On-Time Deliveries = CALCULATE([Delivered Orders],  
f\_orders[on\_time\_delivery] = 1)
  - The count of delivered orders that met the estimated delivery date.
- Repeat Customers: Repeat Customers = CALCULATE([Total Customers],  
FILTER(VALUES(dim\_customers[customer\_unique\_id]), [Total Orders] >= 2))
  - The count of unique customers who have placed two or more orders.
- Repeat Customer %: Repeat Customer % = DIVIDE([Repeat Customers], [Total Customers], 0)
  - The percentage of all customers who are repeat customers.
- Sales by Customer: Sales by Customer = CALCULATE([Total Sales],  
TREATAS(VALUES('Customer Summary'[customer\_unique\_id]),  
dim\_customers[customer\_unique\_id]))
  - A measure to correctly filter total sales by customer type (New vs. Repeat) using a bridging technique.

## 4.0 Analysis & Key Insights

### 4.1 Delivery & Customer Analysis

Put dashboard pic here

Customer Behavior:

- The vast majority (97%) of Olist's customers are one-time buyers. This indicates a significant challenge with customer retention and loyalty.
- Analysis reveals a significant challenge with customer retention: only 3% of customers are repeat buyers.

Delivery Performance:

Put dashboard pic here

- The overall on-time delivery rate is a strong 89%. This means 11% of orders experienced delays.
- Delayed orders have a direct and negative impact on customer satisfaction. The data shows a clear correlation: the average review score for on-time orders is 4.2, compared to 3.8 for delayed orders. This significant drop can impact the platform's reputation.
- Geographic analysis reveals that delivery delays are not evenly distributed. The state of São Paulo (SP) in Brazil as a whole is a significant hotspot for delays, indicating a potential regional logistics or carrier issue that requires targeted intervention.

### 4.2 Sales Performance

Product Trends:

- The top two product categories by revenue are health\_beauty and watches\_gifts. These categories should be prioritized for marketing, inventory, and seller onboarding.

Temporal Trends:

- Sales peak consistently between March and August each year, indicating a strong seasonal pattern that can be leveraged for inventory planning and promotional campaigns.

Customer Revenue Contribution:

- New customers are responsible for the vast majority (94%) of all sales revenue. This aligns with the finding that they represent 97% of the customer base, but highlights a

critical dependency on constantly acquiring new customers rather than generating repeat business.



## 5.0 Recommendations

1. Boost Customer Retention & Lifetime Value:
  - a. Launch a targeted loyalty program to convert the vast segment of one-time buyers into repeat customers, increasing customer lifetime value and creating a more stable revenue base.
2. Optimize Logistics & Customer Satisfaction:
  - a. Improve delivery performance in priority regions like São Paulo in Brazil by partnering with additional carriers or investing in local warehouse capacity to address the root cause of delays.
  - b. Implement a proactive customer recovery process for delayed orders, such as offering discounts or expedited shipping, to mitigate negative satisfaction and protect review scores.
3. Capitalize on Product & Seasonal Trends:
  - a. Align inventory and marketing strategies with seasonal demand, focusing budgets and stock on the high-growth period from March to August.
  - b. Double down on top-performing categories like health\_beauty and watches\_gifts through featured promotions and strategic seller onboarding to maximize revenue.

## 6.0 Appendix

### A: Full List of DAX Measures:

f\_orders[customer\_id] → dim\_customers[customer\_id]

f\_order\_items[order\_id] → f\_orders[order\_id]

f\_order\_items[product\_id] → dim\_products[product\_id]

f\_order\_items[seller\_id] → dim\_sellers[seller\_id]

f\_orders[order\_date] → dim\_calendar[date]

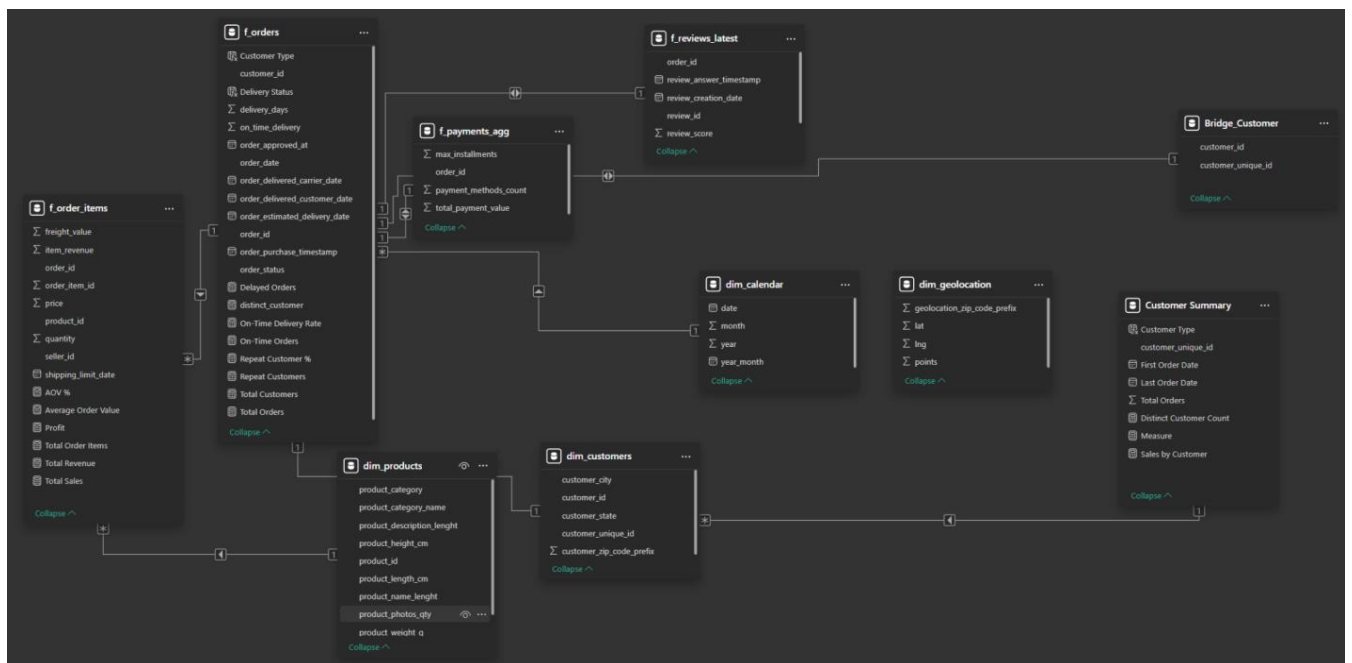
f\_payments\_agg[order\_id] → f\_orders[order\_id]

f\_reviews\_latest[order\_id] → f\_orders[order\_id]

dim\_customers[customer\_zip\_code\_prefix] → dim\_geolocation[geolocation\_zip\_code\_prefix]

dim\_sellers[seller\_zip\_code\_prefix] → dim\_geolocation[geolocation\_zip\_code\_prefix]

### B: Data Model Schema Diagram:



## C: Key Power Query M Code Snippets

### Creating the Date Dimension Table (dim\_calendar):

```
= let
    Start = #date(2016,1,1),
    End   = #date(2018,12,31),
    List  = List.Dates(Start, Duration.Days(End-Start)+1, #duration(1,0,0,0)),
    Tbl   = Table.FromList(List, Splitter.SplitByNothing(), {"date"})
in
    Table.TransformColumnTypes(
        Table.AddColumn(
            Table.AddColumn(
                Table.AddColumn(Tbl, "year", each Date.Year([date])),
                "month", each Date.Month([date])),
                "year_month", each Date.ToText([date], "yyyy-MM")),
        {"date", type date}))
```

### Optimizing Geolocation Data (dim\_geolocation):

```
= let
    Source = raw_geolocation,
    #"Grouped Rows" = Table.Group(Source, {"geolocation_zip_code_prefix"}, {
        {"lat", each List.Average([geolocation_lat], type number),
        {"lng", each List.Average([geolocation_lng], type number)
    })
in
    #"Grouped Rows"
```

## D: GitHub Repository

<https://github.com/moses1252/Olist-Insights-Driving-Growth-in-Brazilian-E-Commerce>