E-Commerce Revenue & Customer Behavior Analysis

Brazilian Olist Dataset | SQL + Tableau | Portfolio Project

Giovanni Valdes

2026

Executive Summary

This project analyzes the publicly available Brazilian Olist e-commerce dataset to surface actionable insights about revenue generation, customer value distribution, and retention behavior. Using SQL in Google BigQuery to query and aggregate data across multiple relational tables, and Tableau to communicate findings visually, the analysis identifies four core business findings — each with direct strategic implications.

$15.4M	$159.83	38.05%	3.00%
Total Revenue	Avg. Order Value	Top 10% Revenue Share	Repeat Customer Rate

Project Overview

Dataset

The **Brazilian E-Commerce Public Dataset (Olist)**, sourced from Kaggle, simulates real-world marketplace operations across multiple relational tables. The three tables central to this analysis are: **orders** (purchase status and timestamps), **order_items** (product-level pricing and freight), and **customers** (unique buyer identifiers). Tables are joined via shared keys — *order_id* and *customer_id*.

Objectives

- Calculate total revenue from completed transactions
- Identify how revenue is distributed across the customer base
- Determine the split between one-time and repeat buyers
- Identify the top-performing product categories
- Measure average order value and explore levers to increase it

Tools & Methodology

All data processing and metric calculation was performed using **SQL in Google BigQuery**. Key techniques include multi-table joins, CTEs (Common Table Expressions), window functions (*NTILE* for percentile ranking), and conditional aggregation. Analysis was restricted to orders with a status of *'delivered'* to ensure only completed transactions were counted. Findings were then visualized in **Tableau**.

Dashboard Overview

The Tableau dashboard below summarises all four key findings in a single view, providing a snapshot of revenue performance, category mix, customer concentration, and retention behavior.

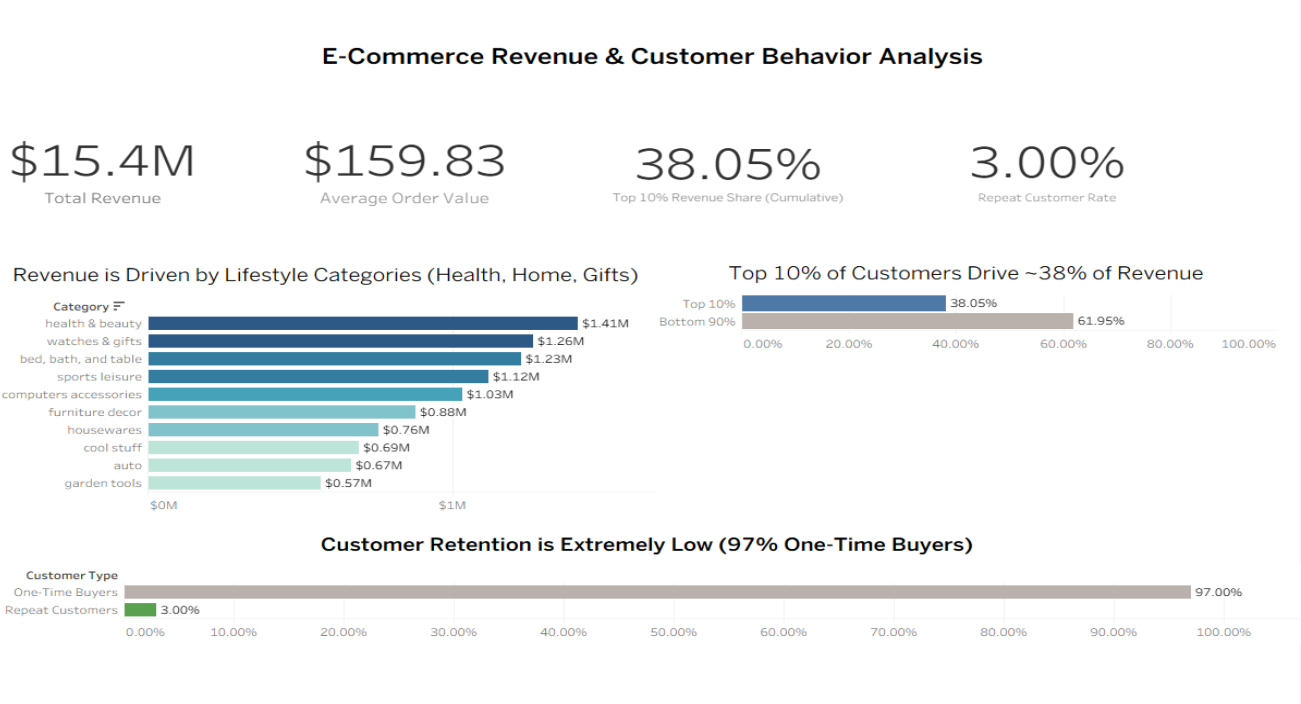

Figure 1 — E-Commerce Revenue & Customer Behavior Dashboard (Tableau)

Findings & Analysis

Finding 1 — Total Revenue & Average Order Value

The business generated approximately **$15.4M in total revenue** from delivered orders, with an **average order value (AOV) of $159.83**. Revenue was calculated by summing *price + freight_value* at the item level and aggregating across all completed orders. The AOV positions the platform at a mid-range price point — high enough to suggest considered purchases, but with meaningful room to grow through upselling and bundling.

Key SQL Query — Total Revenue & AOV

```
SELECT
  SUM(oi.price + oi.freight_value)                        AS total_revenue,
  SUM(oi.price + oi.freight_value) / COUNT(DISTINCT o.order_id)  AS avg_order_value
FROM `ecommerce_data.orders` o
JOIN `ecommerce_data.order_items` oi ON o.order_id = oi.order_id
WHERE o.order_status = 'delivered';
```

Why it matters: Including *freight_value* captures the true economic value of each transaction. Filtering on *order_status = 'delivered'* removes cancelled and pending orders, ensuring revenue reflects cash actually collected.

Finding 2 — Revenue Concentration Across Customer Segments

Using *NTILE(100)* to rank customers into percentile buckets, the analysis reveals that the **top 10% of customers (cumulative) contribute ~38% of total revenue**, while the bottom 90% account for the remaining ~62%. Breaking this down further into non-overlapping buckets shows a clear tiering effect:

Segment	Revenue	Share of Total
Top 1%	~$1.60M	10.4%
Top 2–5%	~$2.51M	16.3%
Top 6–10%	~$1.74M	11.3%
Bottom 90%	~$9.55M	61.9%

Important note on methodology: the segmentation buckets above are non-overlapping (e.g. Top 2–5% excludes Top 1%). The headline figure of 38% shown in the dashboard uses a separate cumulative query — it represents all customers in the top decile combined. Both approaches are intentional and answer different analytical questions.

Key SQL Query — Cumulative Top 10%

```
WITH customer_revenue AS (
  SELECT o.customer_id,
         SUM(oi.price + oi.freight_value) AS revenue
  FROM `ecommerce_data.orders` o
  JOIN `ecommerce_data.order_items` oi ON o.order_id = oi.order_id
  WHERE o.order_status = 'delivered'
  GROUP BY o.customer_id
),
ranked AS (
  SELECT customer_id, revenue,
         NTILE(100) OVER (ORDER BY revenue DESC) AS percentile
  FROM customer_revenue
),
total AS (SELECT SUM(revenue) AS total FROM customer_revenue)
SELECT SUM(revenue) AS top_10_revenue,
       SUM(revenue) / (SELECT total FROM total) AS pct
FROM ranked WHERE percentile <= 10;
```

Finding 3 — Top Product Categories by Revenue

Revenue is driven by lifestyle and consumer-oriented categories rather than any single dominant segment. The top five categories each surpass **$1M in revenue**, indicating a healthy, diversified product mix:

Rank	Category	Revenue
1	Health & Beauty	$1.41M
2	Watches & Gifts	$1.26M
3	Bed, Bath & Table	$1.23M
4	Sports & Leisure	$1.12M
5	Computers & Accessories	$1.03M

Category names were originally stored in Portuguese and translated to English using a mapping table (*category_translation*) joined via *product_category_name*. This multi-table join across four tables — orders, order_items, products, and category_translation — demonstrates the value of a well-structured relational schema.

Key SQL Query — Category Revenue

```
SELECT ct.string_field_1 AS category,
       SUM(oi.price + oi.freight_value) AS category_revenue
FROM `ecommerce_data.orders` o
JOIN `ecommerce_data.order_items` oi  ON o.order_id   = oi.order_id
JOIN `ecommerce_data.products` p       ON oi.product_id = p.product_id
JOIN `ecommerce_data.category_translation` ct
  ON p.product_category_name = ct.string_field_0
WHERE o.order_status = 'delivered'
GROUP BY category
ORDER BY category_revenue DESC
LIMIT 10;
```

Finding 4 — Customer Retention: A Critical Risk

The most significant finding of this analysis is the **extremely low repeat purchase rate**. Of approximately 93,000 unique customers, **~97% made only a single purchase**, with just ~3% returning for a second order. This means the business is almost entirely dependent on new customer acquisition to sustain revenue — a costly and fragile growth model.

Customer Type	Count	Share
One-Time Buyers	~90,557	97.0%
Repeat Customers	~2,801	3.0%

Key SQL Query — Repeat vs One-Time Buyers

```
WITH customer_orders AS (
  SELECT c.customer_unique_id,
         COUNT(DISTINCT o.order_id) AS order_count
  FROM `ecommerce_data.orders` o
  JOIN `ecommerce_data.customers` c ON o.customer_id = c.customer_id
  WHERE o.order_status = 'delivered'
  GROUP BY c.customer_unique_id
)
SELECT
  CASE WHEN order_count = 1 THEN 'One-Time Buyers'
       ELSE 'Repeat Customers' END AS customer_type,
  COUNT(*) AS customer_count,
  COUNT(*) * 1.0 / (SELECT COUNT(*) FROM customer_orders) AS percentage
FROM customer_orders
GROUP BY customer_type;
```

Note on unique IDs: This query joins on *customer_unique_id* rather than *customer_id*. In the Olist schema, a single customer can have multiple *customer_id* values across orders; *customer_unique_id* is the true person-level identifier. Using the wrong key would overstate the one-time buyer rate.

Strategic Recommendations

Each recommendation below is tied directly to a finding. The business has strong revenue fundamentals — the opportunity lies in converting that volume into durable, compounding growth.

Improve Customer Retention → Finding 4

With 97% of customers purchasing only once, even a modest improvement in retention would have an outsized impact on revenue. Priority actions include launching a loyalty program, building automated post-purchase email sequences, and deploying personalized product recommendations based on prior purchase history.

Increase Average Order Value → Finding 1

At $159.83, the AOV has clear upside. Product bundling, 'frequently bought together' modules, and a free-shipping threshold (e.g., orders over $200) are proven levers. Even a 10% AOV lift across the existing order volume would add over $1.5M in revenue.

Double Down on Top Categories → Finding 3

Health & Beauty, Watches & Gifts, and Bed, Bath & Table collectively drive nearly $4M in revenue. These categories warrant priority inventory investment, targeted digital marketing spend, and expanded product assortment to capture more wallet share.

Develop High-Value Customer Programs → Finding 2

The top 10% of customers account for 38% of revenue. These customers should be identified and enrolled in a VIP program with early access to new products, exclusive discounts, and dedicated support — increasing both retention and lifetime value for the segment that matters most.

Conclusion

This analysis demonstrates that the Olist platform is a revenue-generating business with a diversified product mix and a broad customer base — strong foundations for growth. However, the 97% one-time buyer rate is the defining challenge: without addressing retention, the business is essentially refilling a leaking bucket, spending on acquisition to replace customers it has already lost.

The path to scalable, sustainable revenue runs through customer lifetime value. Improving retention by even a few percentage points, combined with incremental AOV gains in the top-performing categories, would compound meaningfully over time. The data is clear on where the opportunity lies — the next step is execution.

Giovanni Valdes | Data Analytics Portfolio Project | 2026
Dataset: Brazilian E-Commerce Public Dataset (Olist) via Kaggle | Tools: Google BigQuery (SQL), Tableau