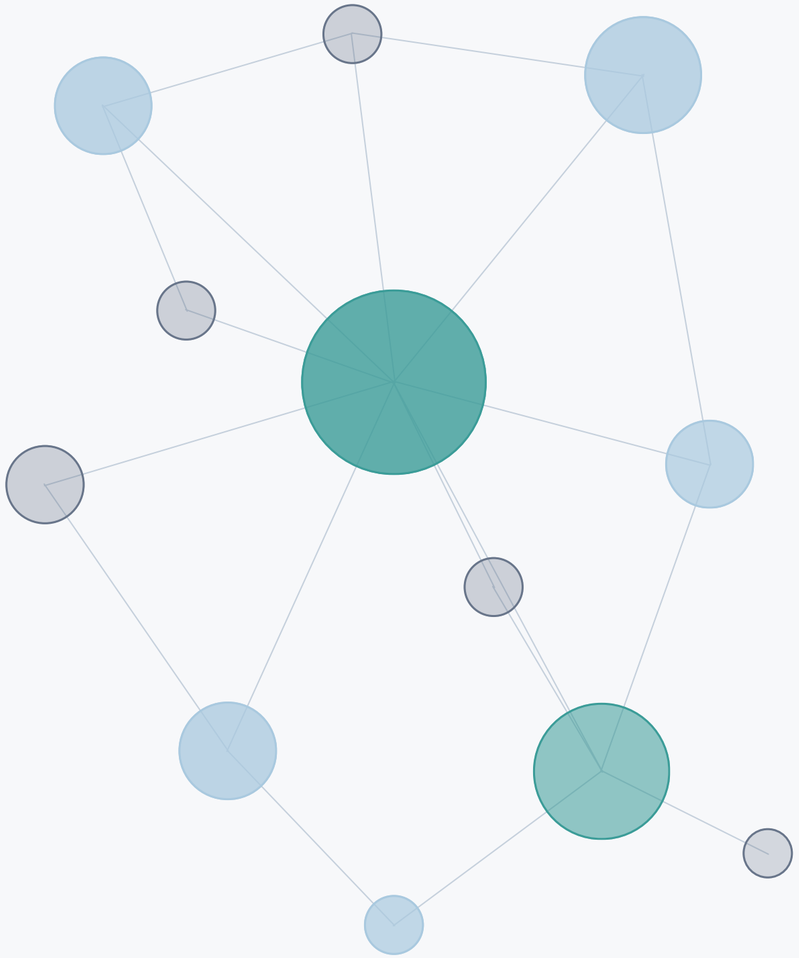
BRAZILIAN OLIST E-COMMERCE ANALYTICS

E-Commerce Analytics & Business Intelligence

A Data-Driven Analysis of the Brazilian Olist Marketplace



PRESENTED BY

Basit Ibrahim

TOOLS

Python | SQL | Power BI

PERIOD

Historical dataset covering 2016–2018

Turning Marketplace Data into Business Insights

1. Analysed a real-world Brazilian e-commerce marketplace dataset
2. Investigated sales, customers, products, sellers, deliveries and customer satisfaction
3. Used Python, MySQL and Power BI across the analytical workflow



A Marketplace Dataset with Over 1.2 Million Records

9

datasets

1,289,091

combined cleaned rows across the nine files

99,441

orders

96,478

delivered orders

93,358

unique customers with delivered purchases

3,095

registered sellers

Data covers approximately September 2016 to August 2018. The 1,289,091 figure is the combined row count across nine datasets, not the number of unique marketplace transactions.

Reliable Insights Start with Reliable Data

Source files

Original files preserved unchanged

Cleaning and validation

- Audited missing values
- Investigated duplicate records
- Validated primary and foreign keys
- Corrected date and numeric types
- Standardised postal code formats
- Translated categories into English
- Investigated unusual financial values
- Exported nine cleaned datasets

9 cleaned datasets

Ready for SQL and Power BI

261,831

exact duplicate geolocation rows removed from the processed copy

Preserved

original source files kept as downloaded

Retained

missing values kept where appropriate, never invented

The Marketplace Generated R$13.22M in Delivered Product Revenue



R$13.22M Product revenue

R$2.20M Freight value

R$15.42M Delivered GMV

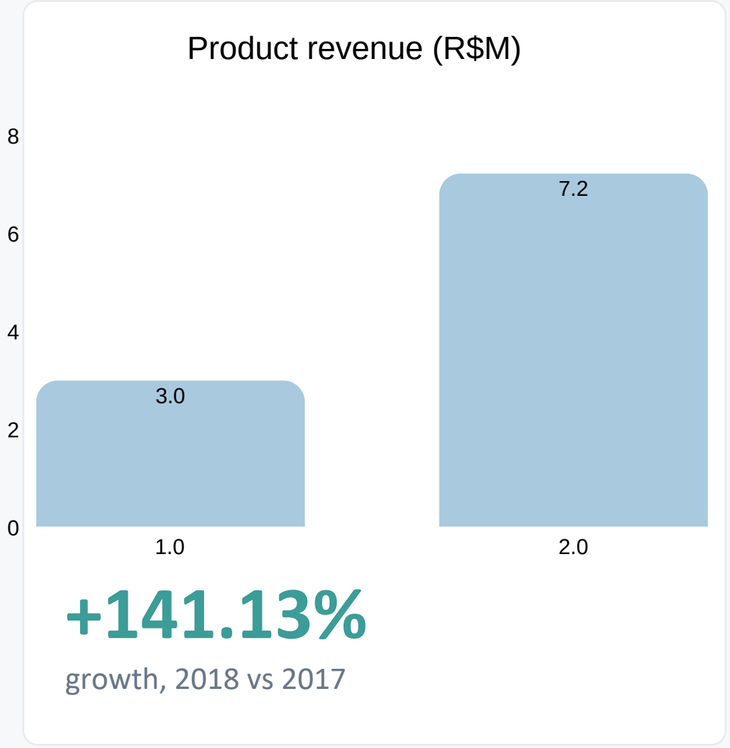
R$137.04 Average order value (excl. freight)

96,478 Delivered orders

Product revenue is the sum of item prices on delivered orders. Delivered GMV adds freight. Neither is Olist's net company revenue.

Order Growth, Rather Than Higher Spending, Drove Revenue Expansion

SAME-PERIOD COMPARISON: JANUARY–AUGUST 2017 VS JANUARY–AUGUST 2018



+141.13%

growth, 2018 vs 2017



+139.94%

growth, 2018 vs 2017

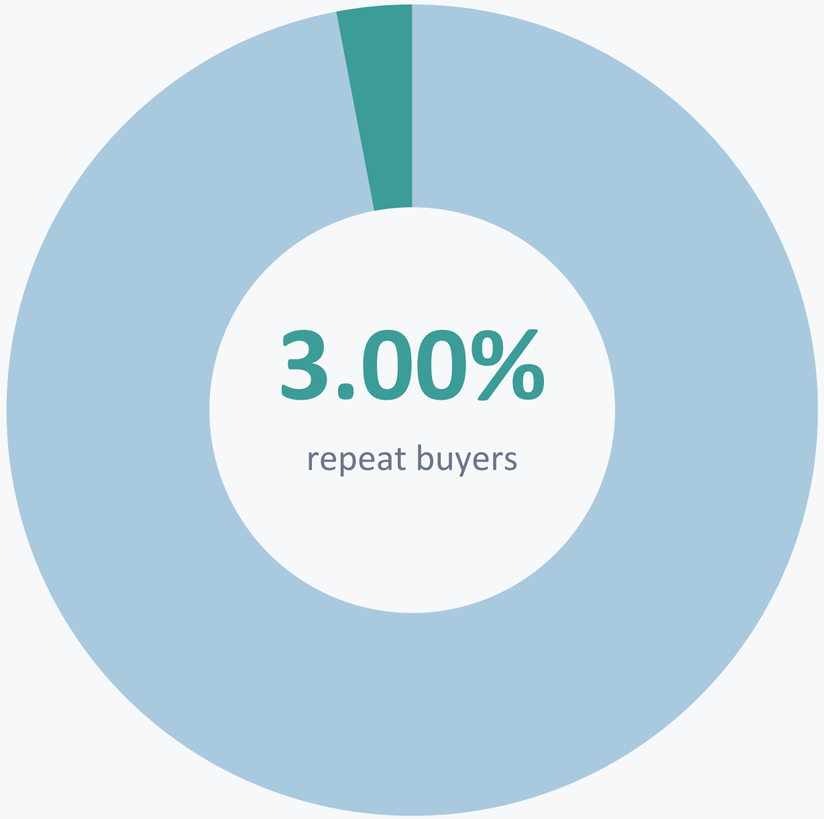


+0.49%

growth, 2018 vs 2017

Revenue growth was overwhelmingly associated with more delivered orders, not a meaningful rise in average order value.

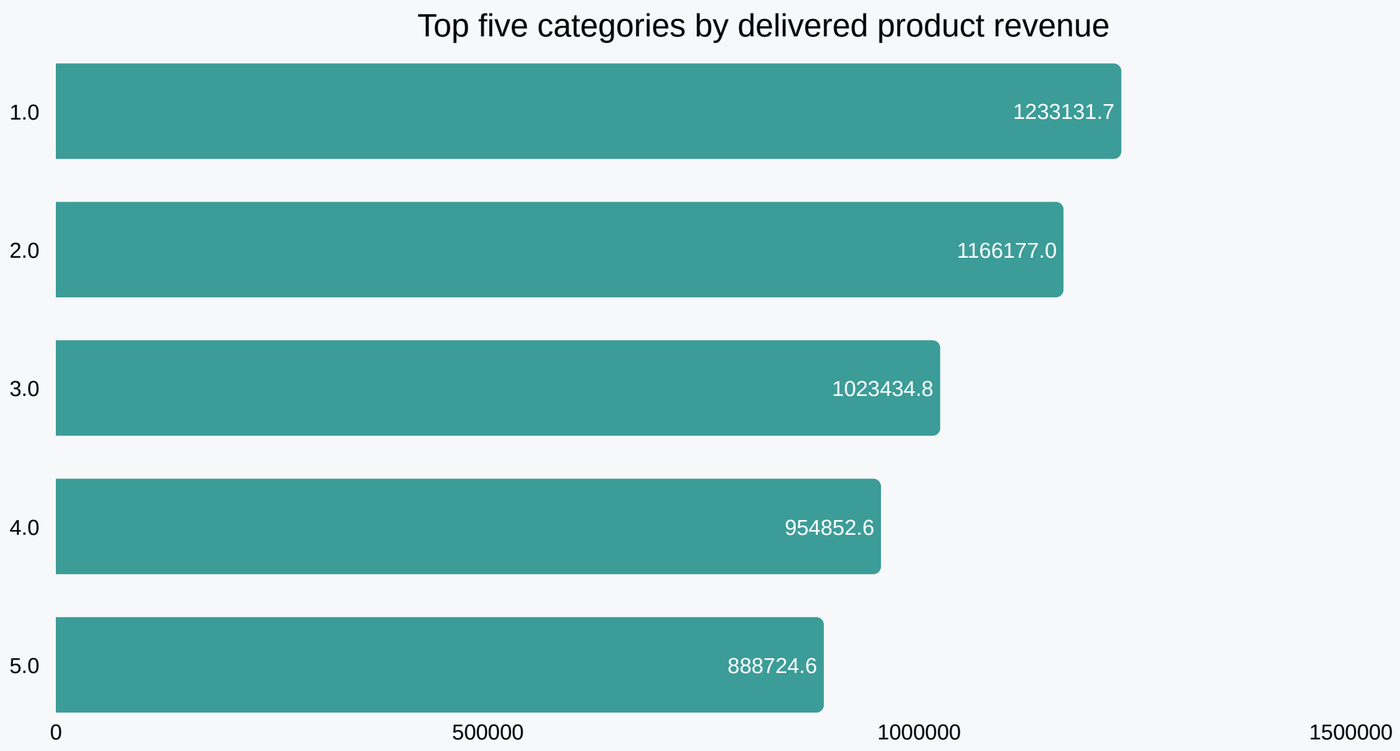
Only 3% of Observed Buyers Made Repeat Purchases



93,358	Unique delivered buyers
90,557	One-time buyers
2,801	Repeat buyers (more than one delivered order)
3.00%	Repeat customer share

A very high share of one-time buyers suggests retention is worth investigating. The data spans Sept 2016 to Aug 2018, so this is not lifetime behaviour.

Health & Beauty Led Product Revenue



25.89%

of delivered product revenue came from the top three categories

62.43%

came from the top ten categories

Revenue is spread across several categories rather than dominated by a single one.

A Broad Seller Base Contributes to Marketplace Revenue

3,095

registered sellers

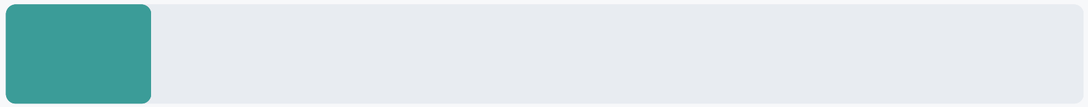
2,970

sellers with delivered sales

13.27%

of product revenue from the top 10 sellers

Seller revenue concentration



Top 10 sellers: 13.27% All other sellers: 86.73%

Geographic concentration



São Paulo sellers: 64.36% Other states: 35.64%

Top seller by delivered product revenue: R$226,987.93. Sellers without delivered sales are not necessarily inactive.

Revenue is spread across a broad group of sellers, although São Paulo is a major geographic concentration.

More Than 91% of Delivered Orders Arrived Before the Estimated Date



12.56 days

average time from purchase to delivery

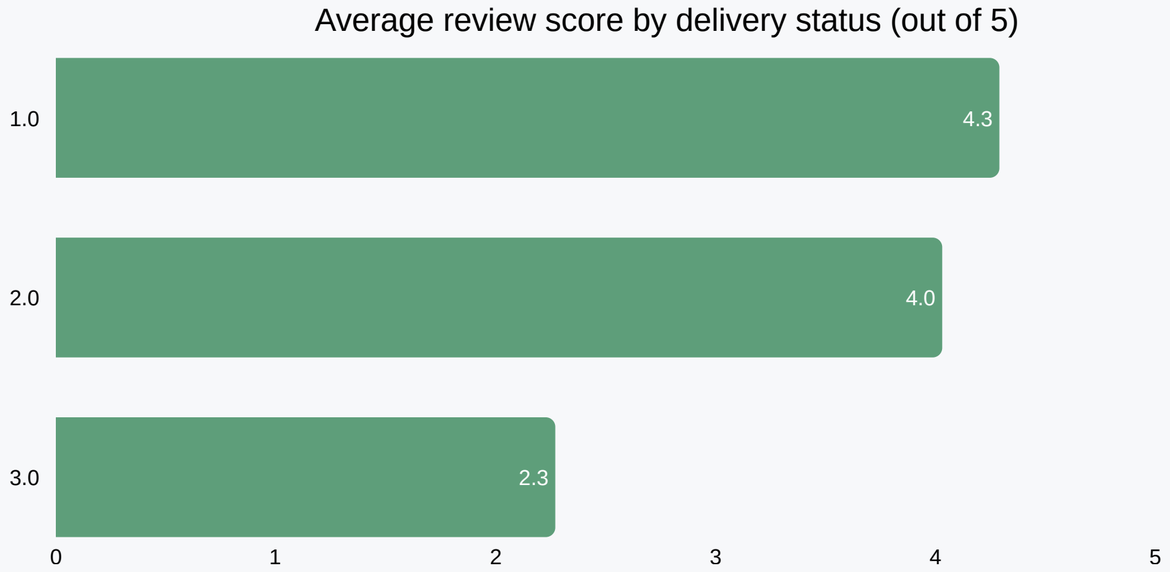
6.77%

of delivered orders arrived late (6,534 orders)

Denominator: 96,470 delivered orders with valid delivery-date comparisons.

Delivery was generally strong, but a group of late deliveries remains and is worth examining by region.

Late Deliveries Were Strongly Associated with Lower Review Scores



2.02 points

lower average score for late vs. early deliveries (4.29 vs. 2.27)

4.09 / 5

average review score across 99,224 review records



Positive (4–5)	**77.07%**
Neutral (3)	**8.24%**
Negative (1–2)	**14.69%**

This is an observed association, not proof that late delivery caused the lower scores.

Four Opportunities Emerge from the Analysis



Customer retention

Investigate why repeat purchases are low. Test post-purchase engagement, loyalty and targeted retention campaigns.

Evidence: 3.00% repeat buyers



Delivery performance

Investigate regions with higher late-delivery rates and assess logistics performance.

Evidence: 6.77% late deliveries



Category strategy

Protect strong-performing categories while investigating categories with lower customer satisfaction.

Evidence: top ten categories = 62.43% of revenue



Seller performance

Identify high-performing sellers and investigate practices that could be replicated across the marketplace.

Evidence: top 10 sellers = 13.27% of revenue

An End-to-End Analytics Workflow

Google Colab

Notebook-based analysis

Python / Pandas

Data cleaning and exploratory analysis

MySQL

Relational queries, joins, CTEs, window functions and business analysis

Power BI

Data modelling, DAX measures and interactive dashboards

From Raw Marketplace Data to Actionable Business Insights

01

Revenue growth was primarily driven by order volume

02

Repeat purchasing was limited in the observed period

03

Delivery performance was closely associated with customer satisfaction

---

Thank You

Basit Ibrahim

E-Commerce Analytics & Business Intelligence