Return-Driven Revenue Loss: A Brazilian E-Commerce Olist Store Analysis

BY DUSTIN SHERRATT

TOTAL ORDERS
95,697

AVERAGE REVIEW SCORE
2.56 (LATE)

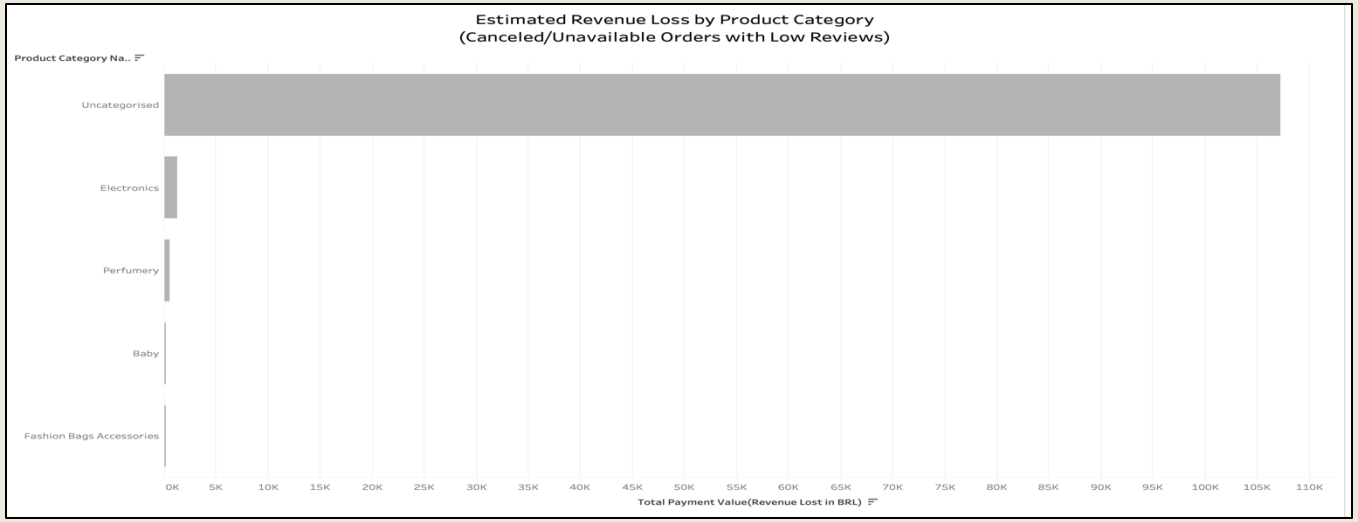
HIGH RISK SELLERS
76

BUSINESS CONTEXT

Olist, a large Brazilian e-commerce marketplace, partners with thousands of third-party sellers to deliver products to customers across the country. Despite strong order volume, the platform faces a recurring challenge: a significant portion of orders are either cancelled, unavailable, or lead to customer dissatisfaction due to delayed delivery or poor seller behaviour.

These fulfilment failures not only result in revenue loss, but also erode customer trust and lead to increased returns, support costs, and negative reviews.

This analysis investigates the root causes of these failed transactions-highlighting key risk areas across delivery performance, product categories, seller reliability, and geographic distribution-so that Olist can optimize platform operations, improve customer experience, and reduce return-driven losses.



METHODS

- **Tools Used**: Python (Pandas, Seaborn), Tableau, Excel, Google Colab
- **Data Source**: Olist Brazilian E-commerce Dataset (Kaggle) – 9 linked CSV files
- **Merges**: Joined on order_id, customer_id, seller_id, zip_code_prefix

Data Cleaning & Preparation:

- Removed duplicate orders and reviews
- Standardized inconsistent city/state names using fuzzy matching
- Mapped ZIP codes to most frequent cities and full state names
- Translated product categories to English and filled missing values
- Filtered orders with valid chronological timestamps and key fulfilment statuses

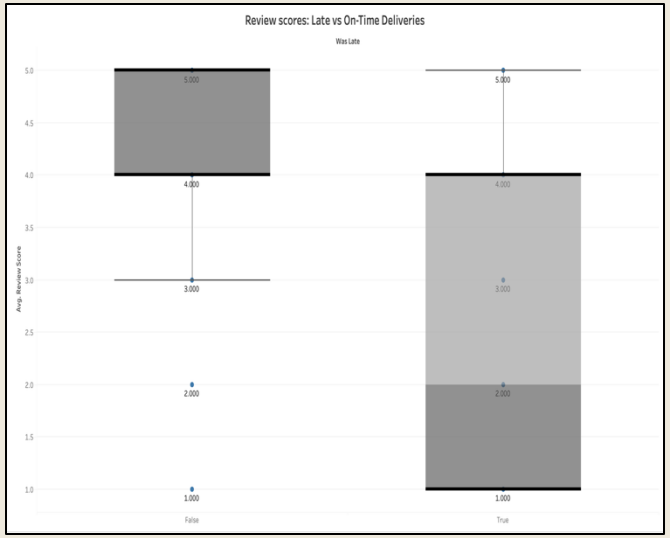
ANALYSIS

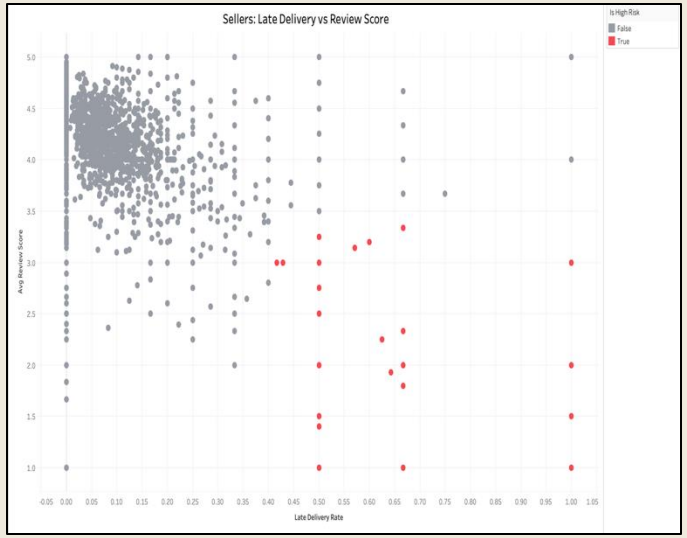
Approximately 8% of all curated orders were marked as cancelled or unavailable, directly contributing to lost revenue. These failures were concentrated in categories like Uncategorised, Electronics and Perfumery, highlighting operational friction.

Using a binary flag 'was_late', it was found that late orders received 1.5 fewer stars on average compared to on-time ones. This confirms that delivery delays are a strong predictor of negative customer sentiment.

Some sellers had Late delivery rates >40% and Average review scores below 3.5. By combining these two metrics, we identified a group of 76 high-risk sellers responsible for a disproportionate share of low satisfaction orders.

States like **São Paulo** and **Rio de Janeiro** had higher concentrations of problematic orders, likely due to **volume and logistics complexity**.





CONCLUSION

This project revealed the key operational breakdowns driving revenue loss on the Olist platform. Through curated and merged datasets, we found that:

- Late deliveries significantly reduce customer satisfaction, often resulting in poor reviews.
- A small group of high-risk sellers consistently contribute to delayed, low-rated, and uncategorised orders.
- Certain product categories are more prone to fulfilment issues.
- Poor product listing practices, like using "uncategorised" labels, hurt discoverability and trust.

By targeting these risk points, Olist can take data-driven action to:

- Prioritize logistics performance
- Flag or retrain low-performing sellers
- Enforce better product categorization

Explore the interactive visualizations: Tableau