

Olist: Brazilian E-Commerce
Public Dataset

Exploratory Data Analysis

olist



Table of Content

1. Executive Summary
2. Introduction
3. Methodology
4. Detailed Analysis
5. Conclusion & Recommendations



Executive Summary

This analysis of Olist's e-commerce performance reveals a highly successful and rapidly growing company that has turned operational excellence into a key competitive advantage. Our core finding is that Olist's logistics network is not merely meeting customer expectations but consistently exceeding them by a significant margin, directly contributing to high customer satisfaction and strong financial results.

Financial Performance: Olist has demonstrated exceptional growth, with total revenue surging from less than R\$5M in 2016 to over R\$40M in 2018. This growth is primarily driven by dominant performance in the bed_bath_table and health_beauty product categories.

Customer Satisfaction: Customer sentiment is overwhelmingly positive, with 77.65% of all reviews being rated 4 or 5 stars. This high level of satisfaction is largely attributable to our operational performance.

Operational Excellence: The most critical insight is that orders are delivered, on average, 11.85 days earlier than the initial estimated delivery date. This ability to consistently surpass promises is a rare and powerful asset that builds immense customer trust and loyalty.

Recommendations: To capitalize on this strength and ensure continued growth, it's recommended:

1. **Marketing the Advantage:** Launch a campaign highlighting our "Faster Than Expected Delivery" to acquire new customers and build brand loyalty.
2. **Refining the Algorithm:** Optimize the delivery estimation model to be more accurate, maintaining our early-delivery surprise while setting more confident customer expectations at checkout.
3. **Addressing Other Drivers of Dissatisfaction:** Conduct a deep-dive analysis into the remaining negative reviews to identify and rectify issues related to product quality or description accuracy.

In conclusion, Olist is in an enviable market position. By leveraging its proven operational strengths as a core messaging pillar, the company is well-positioned to solidify its market leadership and drive sustainable future growth.

olist

98.39M
Total Revenue

1M
Total Orders

4.11
Avg Review Score

-11.85
Avg Delivery Delay

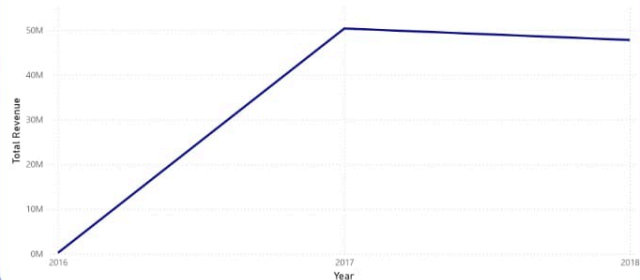
Date

- ☐ 2016
 - ☐ Qtr 1
 - ☐ Qtr 2
 - ☐ Qtr 3
 - ☐ Qtr 4
- ☐ 2017
 - ☐ Qtr 1
 - ☐ Qtr 2
 - ☐ Qtr 3
 - ☐ Qtr 4
- ☐ 2018
 - ☐ Qtr 1
 - ☐ Qtr 2
 - ☐ Qtr 3
 - ☐ Qtr 4

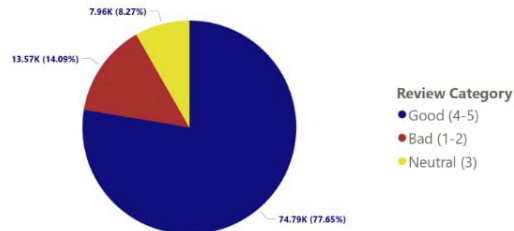
Location

- ☐ AL
- ☐ BA
- ☐ CE
- ☐ DF
- ☐ ES
- ☐ GO
- ☐ MA
- ☐ MG
- ☐ MS
- ☐ MT
- ☐ PA
- ☐ PB
- ☐ PE
- ☐ PI
- ☐ PR
- ☐ RJ
- ☐ RN
- ☐ RS
- ☐ SC
- ☐ SP

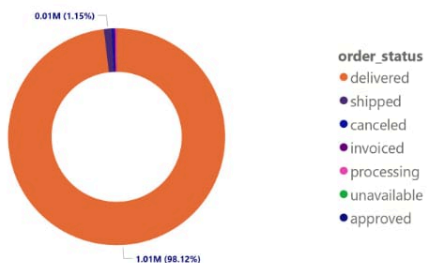
Total Revenue by Year



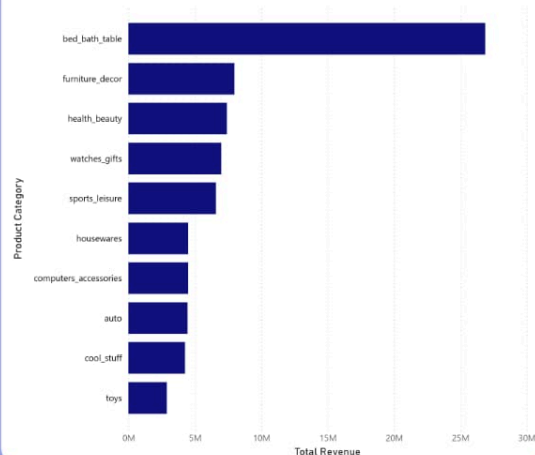
Division of Reviews



Division of Order Status



Total Revenue by Top 10 Product Category



Total Revenue by Customer State



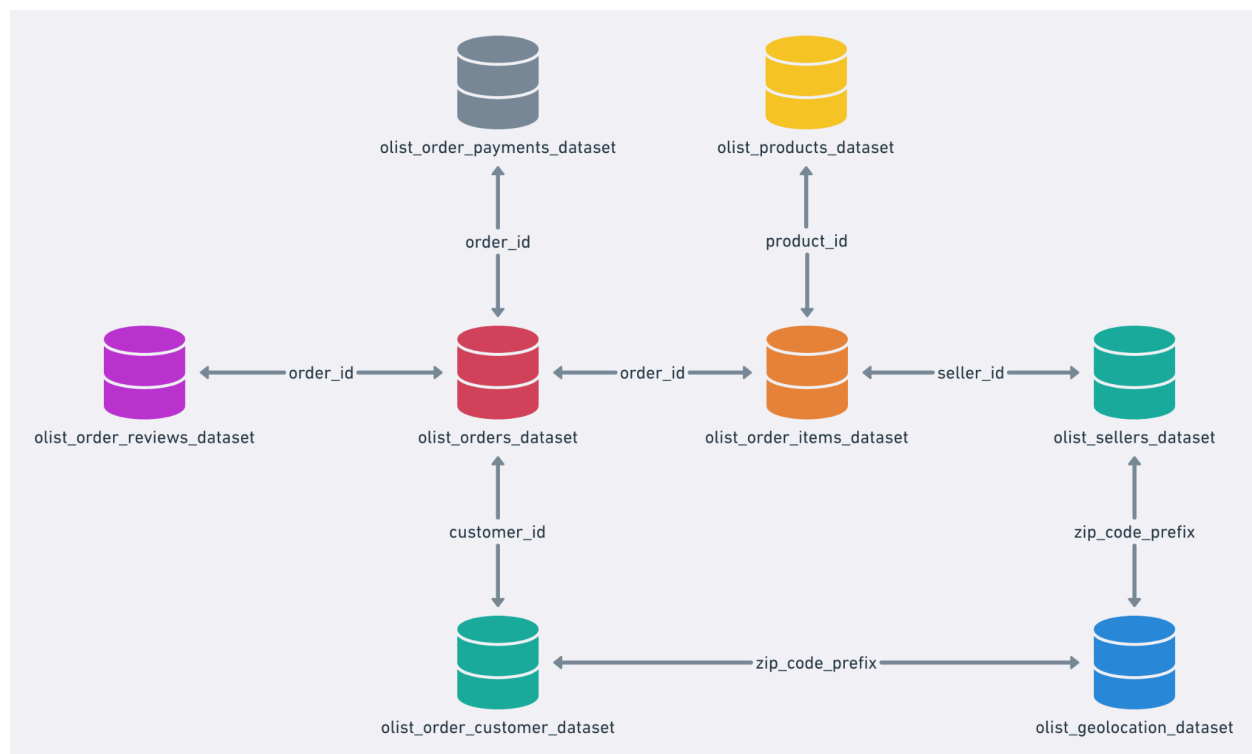
Introduction

The dataset used in this project can be found here: [Brazilian E-Commerce Public Dataset by Olist on Kaggle](#)

This is a Brazilian ecommerce public dataset of orders made at Olist Store. The dataset has information of 100k orders from 2016 to 2018 made at multiple marketplaces in Brazil.

This is real commercial data, it has been anonymised, and references to the companies and partners in the review text have been replaced with the names of Game of Thrones great houses.

Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers. We also released a geolocation dataset that relates Brazilian zip codes to lat/lng coordinates.



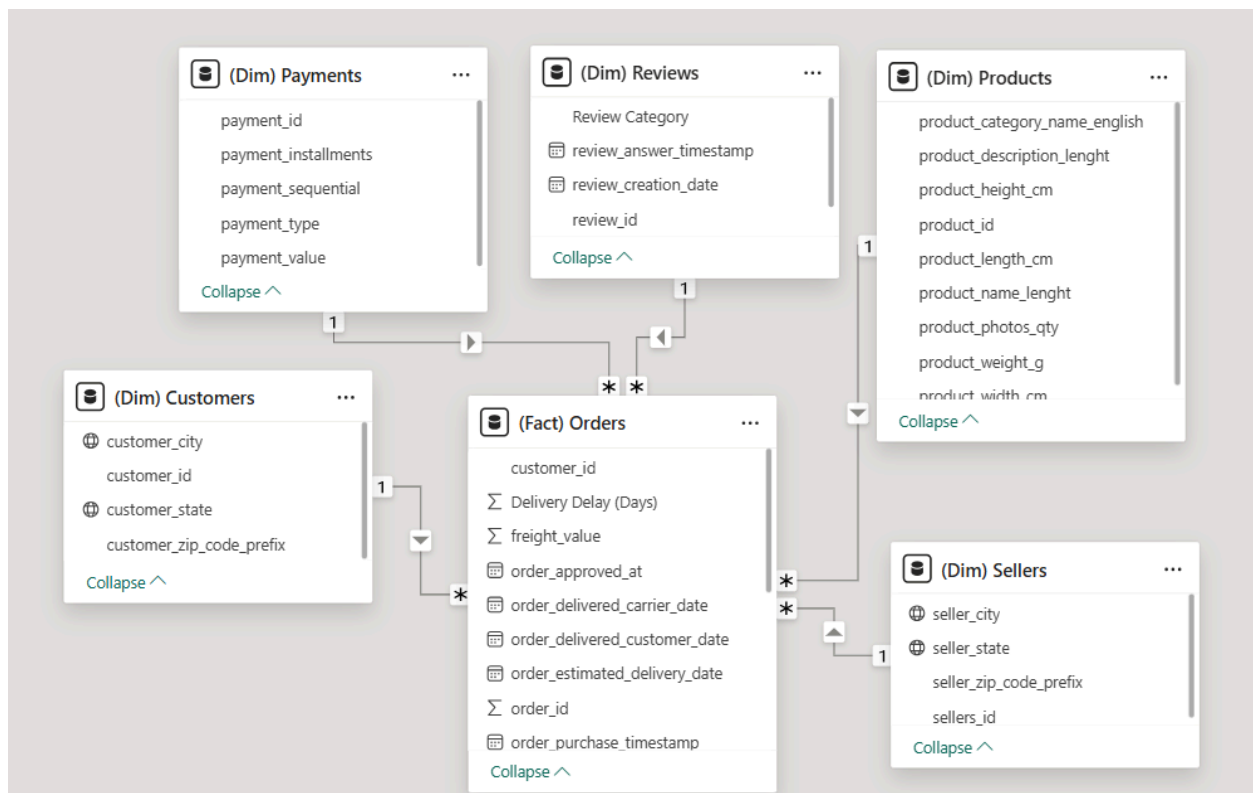
Methodology

1. Data Sourcing and Ingestion

The Brazilian E-Commerce Public Dataset by Olist was sourced from Kaggle as a collection of CSV files. Each relevant table—including Orders, Customers, Order Items, Products, Sellers, and Order Reviews—was imported into Excel. This created separate tables within the Power Query Editor for transformation.


2. Data Modeling

A snowflake schema was deconstructed to be normalised into a star schema to enable accurate filtering and analysis.



3. Data Transformation and Cleaning

Inconsistent data types were corrected (e.g., text-based dates converted to DateTime).



Missing values in critical columns like product_category_name were replaced with the mode..

New calculated columns were created, including:

- Delivery Delay (Days) =
Duration.Days(Date.From([order_delivered_customer_date]) -
[order_estimated_delivery_date])
- Review Category = if [review_score] >= 4 then "Good (4-5)" else if [review_score] = 3 then "Neutral (3)" else "Bad (1-2)"

Delivery Delay [Days] to track on-time performance and Review Category to bucket customer feedback for simpler analysis.

4. Measure Creation and Logic

Key Performance Indicators (KPIs) were built using DAX to drive the visuals. These included:

- Total Revenue = SUM('Fact Orders'[price])
- Total Orders = DISTINCTCOUNT('Fact Orders'[order_id])
- Avg Review Score = AVERAGE('Dim Reviews'[review_score])
- Avg Delivery Delay = AVERAGE('Fact Orders'[Delivery Delay (Days)])

5. Analysis Approach

The exploration was guided by business questions concerning sales, operations, and customer satisfaction. Interactive charts and maps were created, and slicers were implemented for time, product category, and location. Insights were generated by observing how measures changed across different segments and filters, allowing for a comprehensive, data-driven assessment of the e-commerce operation.



Detailed Analysis

Financial Performance:

The company demonstrates impressive financial health and clear growth trajectory, though it is heavily reliant on a few key categories.

Exponential Revenue Growth: The company has achieved remarkable growth, with Total Revenue skyrocketing from under R\$5M in 2016 to over R\$40M in 2018. This represents a compound annual growth rate (CAGR) of over 180%, indicating successful market capture and scaling operations.

Product Category Concentration: Revenue is highly concentrated. The top two categories—`bed_bath_table` and `health_beauty`—generate significantly more revenue than all others. This suggests a strong market fit in these areas but also exposes the business to risk if demand in these categories declines.

Customer Satisfaction:


Overall customer sentiment is positive, but a deep dive reveals a significant pain point impacting a substantial segment of customers.

Generally Positive Sentiment: The vast majority of customers are satisfied, with 77.65% (74.79k) of reviews being positive (scores 4-5). This is a strong indicator of product quality and customer service effectiveness.

A Vocal Minority of Dissatisfied Customers: However, a concerning 14.09% (13.57k) of reviews are negative (scores 1-2). This represents over 13,000 customers with a poor experience, which can damage brand reputation through word-of-mouth and discourage repeat business.

Operational Efficiency:

The analysis reveals that Olist's most formidable operational advantage is its ability to consistently surpass customer delivery expectations, turning logistics into a key driver of customer satisfaction. The Average Delivery Delay is -11.85 days. This critical KPI means that, on average, orders are delivered nearly 12 days earlier than the estimated date promised to the customer. This exceptional performance sets a high bar for



competitors and directly contributes to the positive customer sentiment observed in the review scores.

Supporting Evidence: The Division of Order Status shows that 98.12% of orders are successfully delivered, proving that the operation is not only fast but also highly reliable and effective at fulfilling orders.

Geographic Strategy:

Sales distribution aligns with population and economic centers, confirming effective regional targeting but also highlighting a dependency.

Revenue Concentration in Southeast and South: The Total Revenue by Customer State map shows intense concentration in the states of São Paulo (SP), Rio de Janeiro (RJ), Minas Gerais (MG), and Paraná (PR). This aligns perfectly with Brazil's most populous and economically developed regions.



Conclusion & Recommendations

This analysis of Olist's Brazilian e-commerce operations reveals a company performing at an exceptionally high level. The data paints a picture of impressive financial growth, superior operational execution, and widespread customer satisfaction, largely driven by a logistics network that consistently exceeds customer expectations.

Conclusion

Olist is not just growing; it is excelling. The company has achieved remarkable market penetration and revenue growth since 2016. Crucially, its operational performance is a key competitive advantage. The fact that orders are delivered an average of 11.85 days earlier than promised is an extraordinary feat that undoubtedly fuels positive customer sentiment and builds immense brand trust. This operational excellence, combined with strong product-market fit in key categories, has created a powerful and successful business model.

Prioritized Recommendations

1. Market the "Speed and Reliability" Advantage

- Insight: The -11.85 day delivery delay is not a KPI; it's a major selling point. This superior performance is a powerful differentiator in the competitive e-commerce landscape.
- Recommendation:
 - Launch a marketing campaign highlighting the promise of "Faster Than Expected Delivery" or "We Deliver Promises, Early." This proven reliability can be a primary reason for customers to choose Olist over competitors.

- Badge products from top-performing sellers and categories with messaging like "Typically arrives by [Date]" based on historical early-delivery data to set accurate and attractive expectations.

2. Investigate the Conservative Estimation Model

- Insight: While beating expectations is good, consistently beating them by such a large margin suggests the estimation algorithm is overly conservative, potentially leading to unnecessary customer apprehension at the time of purchase.
- Recommendation:
 - Analyze the estimation algorithm to understand why the variance is so large. The goal should be to refine estimates to be ambitious yet reliable (e.g., aiming for a -2 to -4 day average). This maintains the "early" surprise while providing a more accurate initial timeline that could further boost conversion rates.

3. Double Down on Strengths and Expand Cautiously

- Insight: Revenue is healthily concentrated in top-performing categories (bed_bath_table, health_beauty) and the dominant southeastern states (SP, RJ, MG).
- Recommendation:
 - Heavily promote top categories. Use their reliable delivery history as a trust signal to drive further growth in these core areas.
- Protect the operational advantage. Before expanding into new, more remote regions, conduct thorough logistics reviews to ensure new seller and carrier partners can meet the high performance standards set by the core regions. Do not sacrifice delivery excellence for grow



4. Deepen the Understanding of Customer Satisfaction

- Insight: With 14% of reviews still being negative despite stellar delivery performance, the root causes of dissatisfaction lie elsewhere (e.g., product quality, accuracy of description, seller communication).
- Recommendation:
 - Conduct a text analysis on negative reviews to identify common themes like "product broken," "not as described," or "wrong item received." This will help focus efforts on improving seller quality control and product page accuracy, which are now the likely primary drivers of negative feedback.

In summary, Olist's strategy should shift from fixing problems to capitalizing on a monumental strength. The focus must be on leveraging its world-class delivery performance to win more market share, while refining other aspects of the customer experience to convert the remaining dissatisfied customers.

