

# ***Olist Documentation***

## ***Business Performance: The Executive Overview***

### ***Workflow:***

- ***KPI Construction:** Established core business health metrics: Total Orders (99K), Average Order Value (\$161.13), and Total Revenue (\$15.92M).*
- ***Financial Segmentation:** Integrated a Payment Type Tree Map to visualize revenue distribution across credit cards, boleto, and vouchers.*
- ***Trend Mapping:** Developed a Revenue by Month time-series to identify peak sales periods and seasonal dips.*

### ***Findings & Insights:***

- ***Revenue Engines:** "Health & Beauty" and "Watches/Gifts" are the primary drivers, contributing significantly to the \$15.84M total.*

***Market Concentration:** Geospatial analysis confirms a heavy concentration of sales within South America, specifically major Brazilian hubs.*

## ***Customer Analysis: Understanding the Buyer***

### ***Workflow:***

- ***Loyalty Modeling:** Calculated the split between Repeat vs. One-Time Customers, finding a massive 93.93% reliance on new acquisitions.*
- ***Regional Spend Profiling:** Aggregated spend data to identify that regions like PB and AC have the highest average spend per customer (~\$250+).*

### ***Findings & Insights:***

- ***Retention Opportunity:** With only 6.07% repeat customers, the business is in a "growth-by-acquisition" phase rather than a "loyalty" phase.*
- ***Price Sensitivity:** The Price Distribution Histogram shows most orders are under \$100, despite an average spend per customer of \$166.*

## *Product Analysis: Inventory & Sales*

### *Workflow:*

- *Volume vs. Value: Compared Top Selling Products (by count) against Top Selling Categories (by revenue).*
- *Inventory Health: Tracked 113K total items sold, identifying Garden Tools as the volume leader.*

### *Findings & Insights:*

- *Efficiency Gap: Only 45.85% of products are currently meeting optimal delivery efficiency targets.*
- *Category Dominance: While Garden Tools lead in item count, Health & Beauty dominates in total revenue contribution.*

## *Seller Analysis: Supply Chain Health*

### *Workflow:*

- *Activity Audit: Identified 3,095 Active Sellers out of a total pool of 3.10K.*
- *Performance Ranking: Ranked sellers by both Revenue and Order Count to identify high-value partners.*

### *Findings & Insights:*

- *Category Concentration: The supply side is heavily skewed toward Sports/Leisure and Toys, suggesting a specialized seller base.*
- *Stability: Retention and performance over time (2016-2018) remain remarkably flat, indicating a stable but non-growing seller ecosystem.*

## *Operational Metrics: Logistics & Cost*

### *Workflow:*

- *Tier Categorization: Segregated logistics into Rapid (8.1K), Standard (42.72K), and Late (60.2K) deliveries.*
- *Cost-Weight Correlation: Analyzed the relationship between product weight and freight value to find shipping inefficiencies.*

## ***Findings & Insights:***

- ***The Late Crisis:*** 54.11% of all orders are categorized as "Late," creating a major operational bottleneck.
- ***Freight Outliers:*** The scatter plot shows significant freight cost variation for items between 20K and 30K grams, suggesting inconsistent shipping partner pricing.

## ***Final Recommendations for the Presentation***

1. ***Prioritize Retention:*** Launch a loyalty program targeting the 93.93% one-time buyers to increase repeat purchase rates.
2. ***Fix the Logistics Bottleneck:*** Address the 54.11% late delivery rate by incentivizing sellers in high-volume categories like Health & Beauty to ship faster.
3. ***Optimize High-Weight Freight:*** Audit shipping contracts for items over 20kg to standardize freight costs and improve margins.
4. ***Capitalize on "Rapid" Delight:*** Since the review scores plateau at 4.1 for most deliveries, focus on growing the 7.32% "Rapid" tier to push the brand toward 5-star excellence.

## ***Cleaning***

### ***Translation Table***

***Header Promotion:*** Promoted first rows to headers for mapping Portuguese to English.

***Data Typing:*** Standardized category columns as Text.

### ***2. Location & Geolocation***

***Zip Codes:*** Set to Text to preserve leading zeros; applied Trim to remove whitespace.

***Text Normalization:*** Standardized cities (Capitalize Each Word) and states (Uppercase) for consistent grouping.

***Coordinates:*** Set Lat/Lng to Decimal Number for accurate geographic plotting.

### ***3. Products Table (Null Management)***

***Strategic Replacement:*** Handled 3% missing values by replacing null categories with "unknown\_category" and dimensions/weights with 0.

***Integrity:*** Ensured all dimensions were set to Whole Number to prevent calculation errors.

### ***4. Financial & Item Tables***

*Currency Precision: Set Price, Freight, and Payment values to Decimal Number (Total Revenue: \$15.84M).*

*Payment Type: Applied Trim and Lowercase to standardize methods (e.g., "credit\_card").*

*Sequencing: Set installments and item IDs to Whole Number.*

## **5. Date & Time Standardization**

*Temporal Logic: Converted all timestamp and shipping limit columns to Date/Time format to enable delivery-tier and monthly trend analysis.*

## **Business Performance**

### **Interpretations:**

1. *Revenue Concentration: The marketplace has achieved a massive \$15.84M in Total Revenue, but it is heavily dependent on the Health & Beauty and Watches/Gifts categories as primary drivers.*
2. *Spending Power: The Average Order Value (AOV) of \$161.31 indicates a mid-to-high-tier consumer base, suggesting customers are willing to spend more per transaction than on a typical discount platform.*
3. *Payment Strategy: The Credit Card dominance in the Treemap suggests that installment-based purchasing is the "engine" of the marketplace, making credit availability a key factor for sales growth.*

*Recommendation: Diversify the product portfolio. While Health & Beauty is strong, the business should run promotions in underperforming categories (like "Home Appliances") to reduce financial risk if the top categories fluctuate. **Customer Analysis***

### **Interpretations:**

1. *Retention Gap: A staggering 93.93% of customers are one-time buyers, indicating that Olist is currently a "discovery" platform rather than a "loyalty" platform.*
2. *Regional Spend Paradox: Customers in regions like Paraíba (PB) and Acre (AC) show the highest average spend per customer (~\$250+), despite having lower total order volumes than São Paulo.*
3. *Order Volume vs. Value: Most orders cluster under \$100, meaning high-volume, low-cost items drive the platform's daily traffic.*

***Recommendation: Launch a "Second Purchase" discount campaign specifically targeting the 93% of one-time buyers to increase the Lifetime Value (LTV) of the existing customer base.***

### ***Product Analysis***

#### ***Interpretations:***

- 1. Volume Leader: Garden Tools lead in item count (over 4K units), proving that household utility items are the most consistent movers in the inventory.***
- 2. Revenue vs. Quantity: There is a mismatch between volume and value; categories like Watches/Gifts move fewer units than Garden Tools but contribute more significantly to the total \$15.84M revenue.***
- 3. Logistics Health: Only 45% of products are meeting optimal shipping targets, suggesting that product physical characteristics (size/weight) may be slowing down fulfillment.***

***Recommendation: Bundle high-volume "Garden Tools" with higher-margin accessories to increase the profit-per-box and improve the efficiency of the shipping process.***

### ***Seller Analysis***

#### ***Interpretations:***

- 1. Supply Stability: There are 3,095 Active Sellers, showing a robust and mature vendor ecosystem with very little "ghost" account activity.***
- 2. Niche Dominance: Sellers are highly concentrated in the Sports/Leisure and Toys sectors, showing that the platform attracts specialized merchants.***
- 3. Geographic Hub: The majority of top-performing sellers are concentrated in the South/Southeast (SP/RJ), which explains why delivery to the North is often slower.***

***Recommendation: Recruit and onboard sellers in the North and Northeast regions to create local fulfillment hubs and reduce the cross-country shipping burden.***

### ***Operational Metrics***

#### ***Interpretations:***

- 1. The "Late" Crisis: 54.15% of all deliveries are "Late" (taking 10+ days), which is the single biggest threat to the platform's reputation.***
- 2. Shipping Rariness: Only 7.32% of orders are "Rapid," meaning fast shipping is an exception rather than the standard for Olist.***
- 3. Freight Inconsistency: The scatter plot shows that high-weight items (20kg+) have unpredictable freight costs, suggesting a lack of standardized shipping contracts for heavy goods.***

***Recommendation: Standardize freight rates with carriers for items over 20kg to provide customers with more predictable checkout costs and protect seller margins.***

