

## Analysis Report: Olist E-Commerce Performance

### Project Overview

- **Project Name:** E-COMMERCE SALES PERFORMANCE DASHBOARD – OLIST BRAZIL
- **Objective:** To design an interactive 4-page Power BI dashboard that analyzes Olist's e-commerce sales, profitability, delivery efficiency, and customer satisfaction to support data-driven business decisions.
- **Dataset:** Olist Brazilian E-Commerce Public Dataset (9 CSV files, ~100,000 orders).

### Dataset Summary

The dataset consisted of 9 CSV files, providing a complete picture of the e-commerce process, including orders, payments, reviews, customers, products, and sellers.

Key cleaning steps performed in Power Query included:

- Checking data types for all columns, especially date/time fields.
- Assessing key columns (like order\_purchase\_timestamp) for null values, which was critical for building accurate delivery metrics.

### Data Model Overview

Section	Details
Tables Used	orders, order_items, customers, products, sellers, order_payments, order_reviews, geolocation, product_category_name_translation
Relationships	A star-schema-like model was created. orders and order_items serve as the central fact tables. This model is a <b>Star Schema with minor Snowflake extensions</b> — often called a <b>“hybrid schema”</b>
Key Calculations / Measures	<b>Total Sales</b> = SUM(order_items[price]) <b>Total Freight Cost</b> = SUM(order_items[freight_value]) <b>Total Profit</b> = [Total Sales] - [Total Freight Cost] <b>Profit Margin %</b> = DIVIDE([Total Profit], [Total Sales]) <b>Total Customers</b> = DISTINCTCOUNT(customers[customer_unique_id]) <b>Average Review Score</b> = AVERAGE(order_reviews[review_score]) <b>Total Validated Deliveries</b> = CALCULATE(DISTINCTCOUNT(orders[order_id]), orders[order_status] = "delivered", NOT ISBLANK(orders[order_delivered_customer_date]))  <b>On-Time Deliveries</b> = CALCULATE(COUNTROWS(orders), FILTER(orders, orders[order_delivered_customer_date] <= orders[order_estimated_delivery_date] && orders[order_status] = "delivered" && NOT ISBLANK(orders[order_delivered_customer_date])))

Section	Details
	<p><b>On-Time Delivery %</b> = <math>\text{DIVIDE}([\text{On-Time Deliveries}], [\text{Total Validated Deliveries}])</math></p> <p><b>Average Delivery Days</b> = <math>\text{AVERAGEX}(\text{FILTER}(\text{orders}, \text{orders}[\text{order\_status}] = \text{"delivered"} \&amp;\&amp; \text{NOT ISBLANK}(\text{orders}[\text{order\_delivered\_customer\_date}]))</math>,  <math>\text{DATEDIFF}(\text{orders}[\text{order\_purchase\_timestamp}], \text{orders}[\text{order\_delivered\_customer\_date}], \text{DAY})</math></p>
Notes	<p>The model's accuracy hinges on the Total Validated Deliveries measure, which correctly filters out orders without a valid delivery date. Cross-filter direction was set to "Both" for orders &lt;=&gt; payments and orders &lt;=&gt; order_items to allow filtering from the dimension side.</p>
Diagram	<p>The diagram illustrates a data model with the following tables and their attributes:</p> <ul style="list-style-type: none"> <li><b>customers</b>: customer_city, customer_id, customer_state, customer_unique_id, customer_zip_code_prefix</li> <li><b>geolocation</b>: geolocation_city, geolocation_lat, geolocation_lng, geolocation_state, geolocation_zip_code_prefix</li> <li><b>orders</b>: customer_id, order_approved_at, order_delivered_carrier_date, order_delivered_customer_date, order_estimated_delivery_date, order_id, order_purchase_timestamp, order_status</li> <li><b>order_payments</b>: order_id, payment_installments, payment_sequential, payment_type, payment_value</li> <li><b>sellers</b>: seller_city, seller_id, seller_state, seller_zip_code_prefix</li> <li><b>order_items</b>: order_id, orders.order_approved_at, price, product_id, seller_id, shipping_limit_date, Valid Order</li> <li><b>order_reviews</b>: order_id, review_answer_timestamp, review_creation_date, review_id, review_score</li> <li><b>products</b>: product_category_name, product_height_cm, product_id, product_length_cm, product_weight_g</li> </ul> <p>Key relationships include:</p> <ul style="list-style-type: none"> <li>customers (1) to orders (M)</li> <li>geolocation (1) to orders (M)</li> <li>orders (1) to order_payments (M)</li> <li>orders (1) to order_items (M)</li> <li>orders (1) to order_reviews (M)</li> <li>products (1) to order_items (M)</li> <li>sellers (1) to order_items (M)</li> </ul>

## Analysis & Insights

The 4-page dashboard provided key insights for each target audience.

### Page 1: Executive Overview

- **Insight:** The business shows strong sales growth, with sales climbing from 2016 through 2018.
- **Insight:** Sales are highly concentrated geographically. The state of **São Paulo (SP)** is responsible for a significantly larger portion of sales than any other state.

### Page 2: Sales Performance

- **Insight:** Product performance is split into two clear types:
  - **High-Value Products:** watches\_gifts has the highest *sales* (revenue) but not the highest number of orders.

- **High-Volume Products:** sports\_leisure has the highest *number of orders* (volume) but lower total sales.
- **Insight:** The dashboard's interactivity allows managers to click a state (e.g., "SP") to instantly see the top-selling products and sellers *in that specific state*.

### Page 3: Marketing & Customer Insights

- **Insight:** The customer base grew consistently from 2016 to 2018.
- **Insight:** credit\_card is the dominant payment method, used in over 76% of all orders. This suggests customers are comfortable with credit, but also presents an opportunity to convert the 24% who use other methods.
- **Insight:** The "**Average Review Score by Category**" chart is a direct guide for marketing, showing clear "winners" (e.g., books, cds\_dvds\_mus...) to promote and "losers" (e.g., food) to investigate.

### Page 4: Operations & Delivery

- **THE KEY INSIGHT:** The dashboard revealed a critical "faster but later" paradox.
  - **Avg. Delivery Days (Speed):** This has **improved**. Average delivery time dropped from ~20 days in 2016 to ~12 days in 2018.
  - **On-Time Delivery % (Promise):** This has **worsened**. The on-time rate fell from ~98% to ~90%.
- **Conclusion:** The problem is **not** the delivery speed; the problem is the **delivery estimate**. The company is promising delivery too fast, failing to meet its promise, and creating unhappy customers (as seen in the Avg. Review Score).
- **Insight:** The "**Seller Report Card**" table clearly identifies the specific sellers who are consistently late, have low review scores, and are hurting the platform's reputation.

### Conclusions

1. The Olist platform is growing, but this growth is creating operational stress that is damaging customer satisfaction.
2. The single most critical business problem identified is the **inaccuracy of delivery estimates**, which is driving down the on-time delivery rate despite faster actual shipping.
3. A small number of low-performing sellers are responsible for a large share of delivery delays.

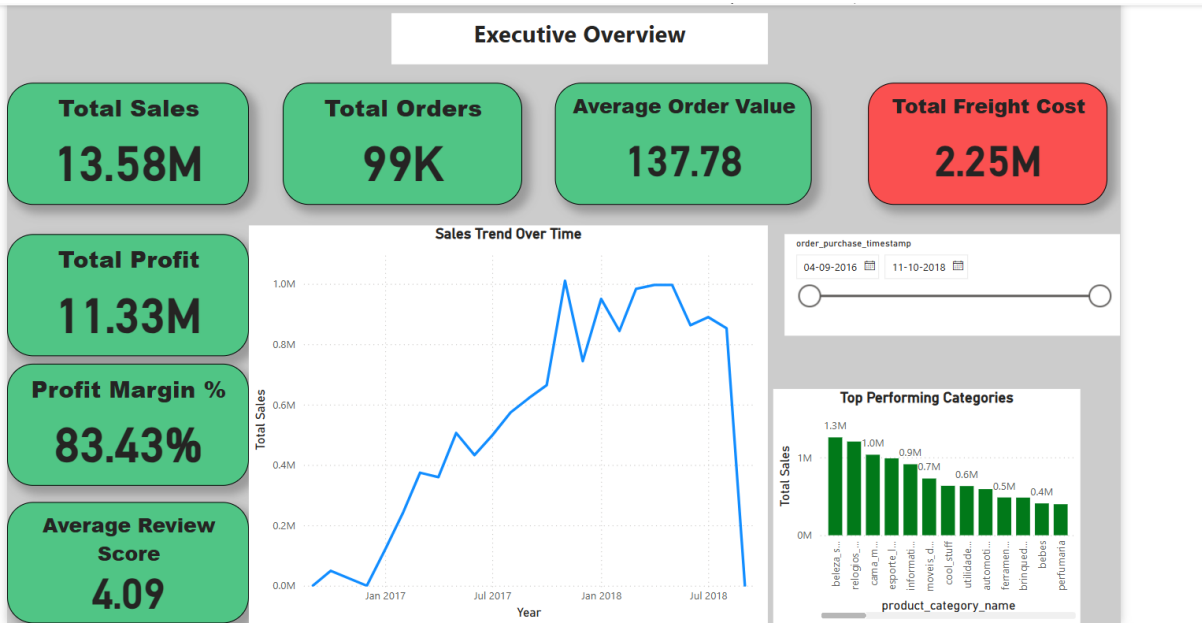
### Recommendations

1. **Operations: Immediately review and adjust the delivery estimation logic.** It is better to promise delivery in 15 days and deliver in 12 (high satisfaction) than to promise in 10 and deliver in 12 (low satisfaction).
2. **Seller Management:** Use the "**Seller Report Card**" on Page 4 to **review the bottom 10% of sellers**. Provide warnings, retraining, or delist sellers who consistently fail to meet on-time targets.

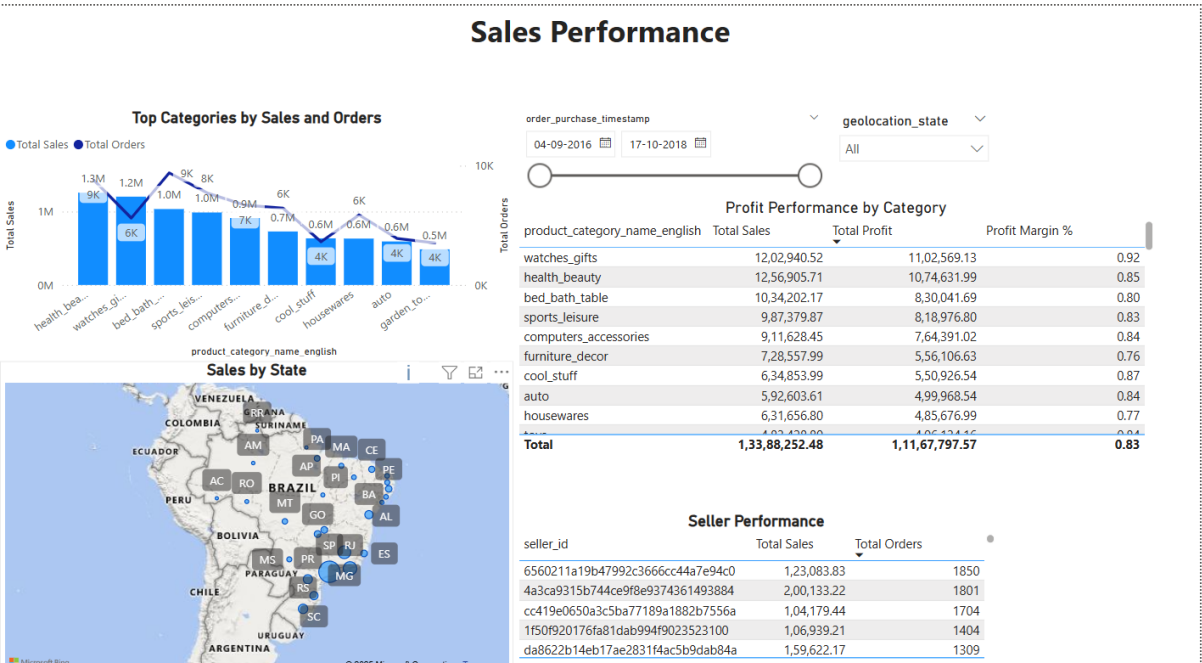
3. **Marketing:** Use the "Average Review Score by Category" chart on Page 3 to guide marketing spend. **Double down on promoting high-rated categories** like books and watches\_gifts and stop promoting low-rated categories until their quality/fulfillment issues are resolved.

Dashboard Overview

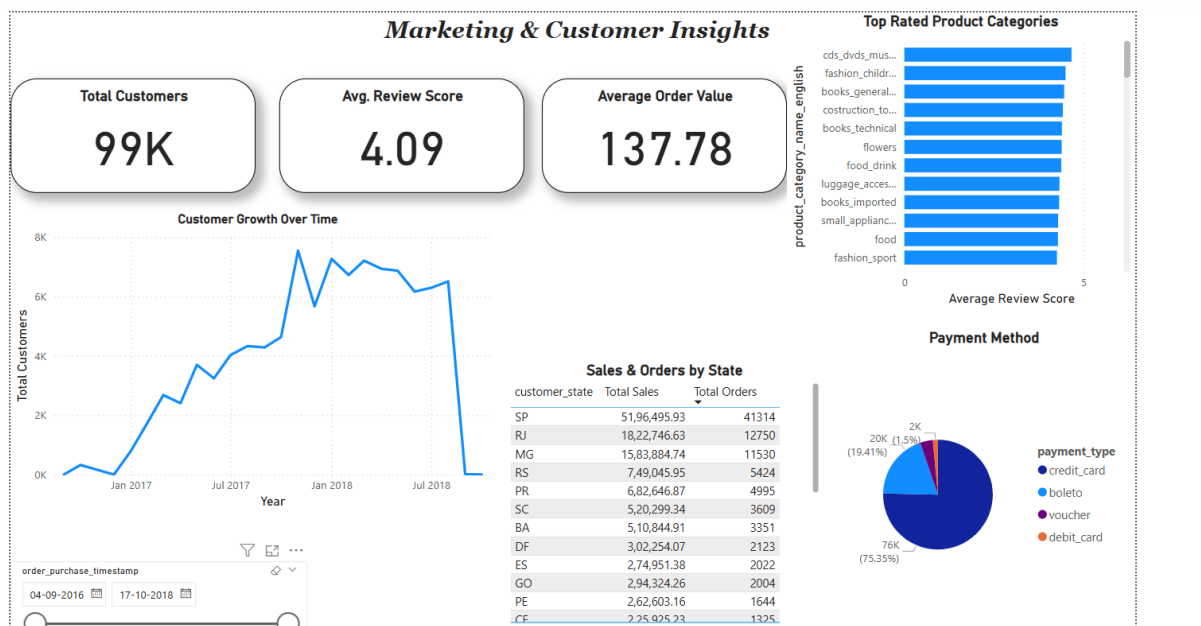
- Page 1: Executive Overview



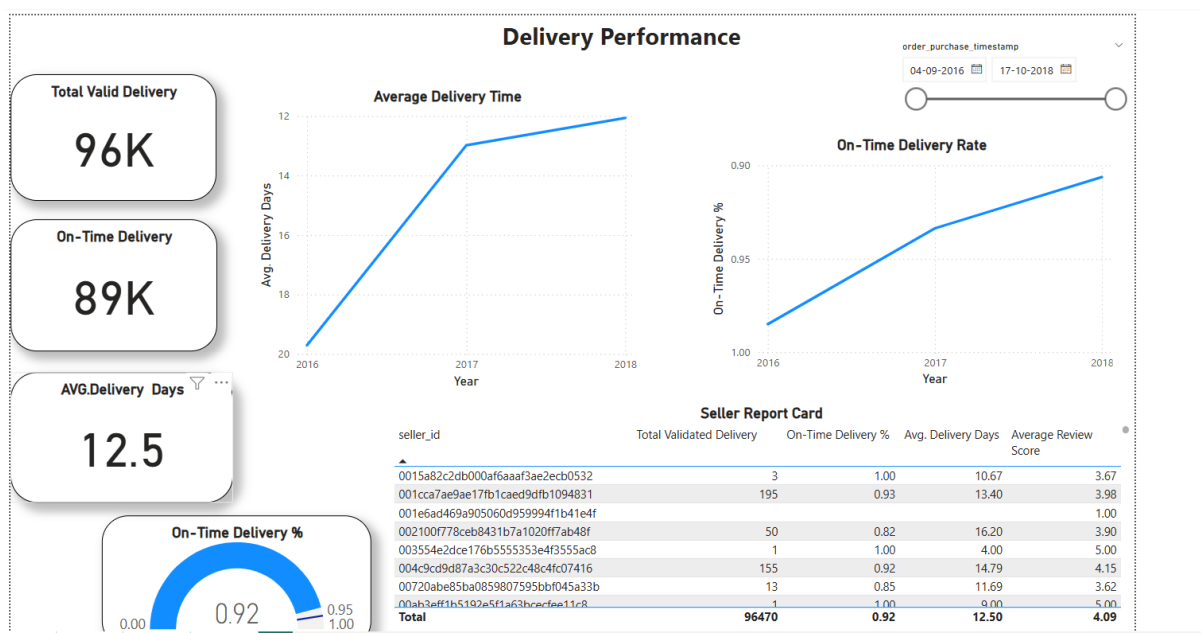
- Page 2: Sales Performance



- Page 3: Marketing & Customer Insights



- Page 4: Operations & Delivery



### Notes / Limitations

- **Profit Definition:** "Profit" is calculated as Total Sales - Total Freight Cost Since the dataset does **not include product cost** (the cost of making or buying each product), I can't calculate *true profit*.  
So instead, a **simplified version**: Profit = Total Sales — Total Freight Cost
- **Data Accuracy:** All delivery metrics are based only on confirmed deliveries (with actual delivery dates). This ensures accuracy and avoids counting incomplete or missing data.