AR

E-Commerce Sales & Logistics Analysis

OLIST BRAZILIAN MARKETPLACE

PREPARED BY : FADLIANSYAH RAHMAN

DATE : JAN 2026

About Task

BUSINESS PROBLEM

Olist is a Brazilian e-commerce marketplace that connects customers with thousands of independent sellers across multiple product categories. As transaction volume increases, operational complexity emerges in areas such as product performance monitoring, customer behavior analysis, seller contribution evaluation, and delivery reliability. Raw transactional data stored across multiple tables makes it difficult to gain a unified view of marketplace performance.

This project addresses the need for a structured analytics pipeline that consolidates fragmented data into a single analytical view. The goal is to enable data-driven insights that support strategic decisions related to sales optimization, seller management, and logistics performance.

OBJECTIVE

The main objectives of this analysis are to build an analytics-ready master dataset, evaluate overall marketplace performance, identify top-performing product categories, analyze customer purchasing patterns, assess seller contribution distribution, and measure delivery reliability using real transaction data.

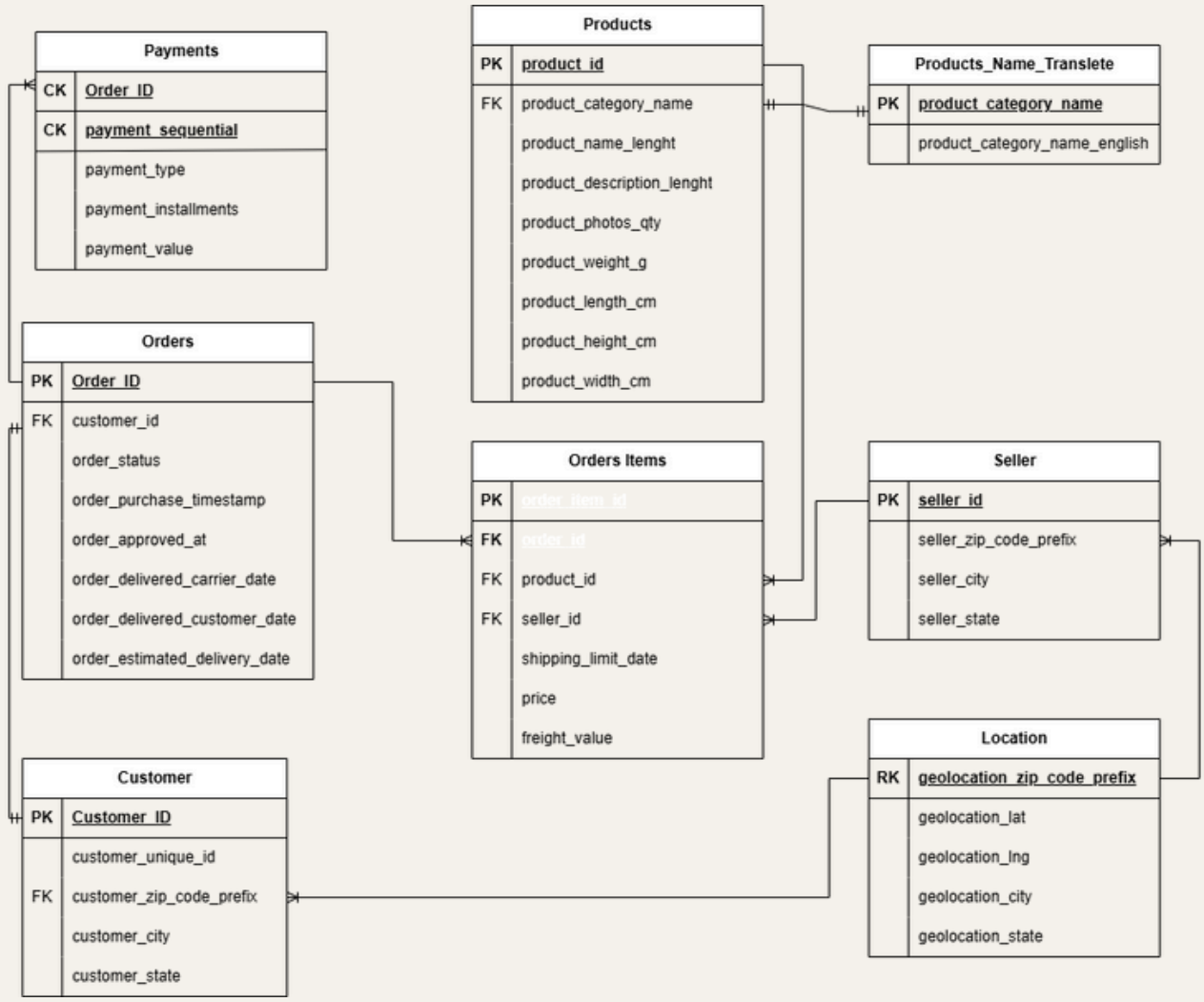
TOOLS

Google BigQuery is used for data validation, cleaning, transformation, feature engineering, and master table creation at order-item level.

Looker

Google Looker Studio is used to visualize metrics and insights, allowing stakeholders to explore marketplace performance interactively based on the prepared dataset.

Data Information



THE DATA MODEL IS DESIGNED WITH ORDERS AS THE CENTRAL ENTITY, REPRESENTING THE FULL TRANSACTION LIFECYCLE. CUSTOMERS, PAYMENTS, AND DELIVERY TIMESTAMPS ARE DIRECTLY LINKED TO ORDERS, WHILE PRODUCTS AND SELLERS ARE CONNECTED THROUGH THE ORDER ITEMS TABLE. ORDER ITEMS ACT AS A BRIDGE TABLE THAT RESOLVES THE MANY-TO-MANY RELATIONSHIP BETWEEN PRODUCTS AND SELLERS. THIS STRUCTURE ENSURES ANALYTICAL FLEXIBILITY WHILE PRESERVING DATA INTEGRITY AND ALIGNS WITH STANDARD E-COMMERCE DATA MODELING PRACTICES.

NULL CHEK

Data preparation begins with data quality checks, including null validation on key fields, duplicate detection on primary keys, and validation of delivery timestamps. Orders with incomplete or invalid delivery information are excluded from logistics analysis to prevent biased results.

```sql
-- check duplicates

-- Customer
SELECT customer_id,COUNT(*) AS cnt FROM `ecomercebrazil.olist_customers_dataset`
GROUP BY customer_id
HAVING COUNT(*) > 1;


-- Order
SELECT order_id,COUNT(*) AS count_order FROM `ecomercebrazil.olist_orders_dataset`
GROUP BY order_id HAVING COUNT(*) > 1;


-- Order Itemms
SELECT order_id,order_item_id,COUNT(*) AS count_orde_items FROM `ecomercebrazil.olist_order_items_dataset`
GROUP BY order_id, order_item_id HAVING COUNT(*) > 1;


--product
SELECT product_id,COUNT(*) AS count_product FROM `ecomercebrazil.olist_products_dataset`
GROUP BY product_id HAVING COUNT(*) > 1;


-- product translete
SELECT product_category_name,COUNT(*) AS count_translete FROM `ecomercebrazil.product_category_name_translation`
GROUP BY product_category_name HAVING COUNT(*) > 1;


-- seller
SELECT seller_id,COUNT(*) AS count_seller FROM `ecomercebrazil.olist_sellers_dataset`
GROUP BY seller_id HAVING COUNT(*) > 1;


-- payment
SELECT order_id,payment_sequential,COUNT(*) AS count_payments FROM `ecomercebrazil.olist_order_payments_dataset`
GROUP BY order_id, payment_sequential HAVING COUNT(*) > 1;


--location
SELECT geolocation_zip_code_prefix,COUNT(*) AS count_location FROM `ecomercebrazil.olist_geolocation_dataset`
GROUP BY geolocation_zip_code_prefix HAVING COUNT(*) > 1;
```

VIEW TABLE

After validation, all relevant tables are integrated into a single master analytical view (v_master). Orders serve as the base table, joined with order items for granularity, products and sellers for dimensional context, payments for revenue calculation, and category translation for reporting clarity.

```sql
-- orders #for relevant order business

CREATE OR REPLACE VIEW `ecomercebrazil.v_orders_clean` AS
SELECT
  order_id,
  customer_id,
  order_status,
  order_purchase_timestamp,
  order_approved_at,
  order_delivered_carrier_date,
  order_delivered_customer_date,
  order_estimated_delivery_date
FROM `ecomercebrazil.olist_orders_dataset`
WHERE customer_id IS NOT NULL
AND order_purchase_timestamp IS NOT NULL
AND order_status NOT IN ('canceled', 'unavailable');
```

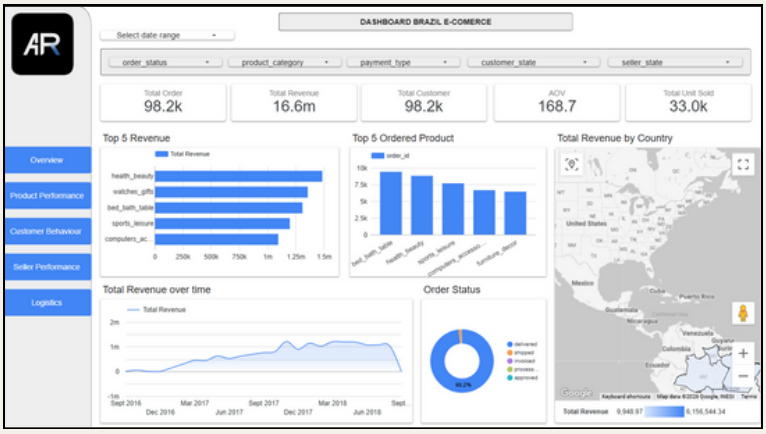
```sql
-- Master Table ( FINAL FACT TABLE )
CREATE OR REPLACE VIEW `ecomercebrazil.v_master` AS
SELECT
-- order
  o.order_id,
  o.order_status,
  o.order_purchase_timestamp,
  o.order_delivered_customer_date,
  o.order_estimated_delivery_date,

-- custommer
  c.customer_id,
  c.customer_city,
  c.customer_state,

-- order item
  oi.order_item_id,
  oi.product_id,
  oi.seller_id,
  oi.price,
  oi.freight_value,
  (oi.price + oi.freight_value) AS item_total_value,
```

Analysis

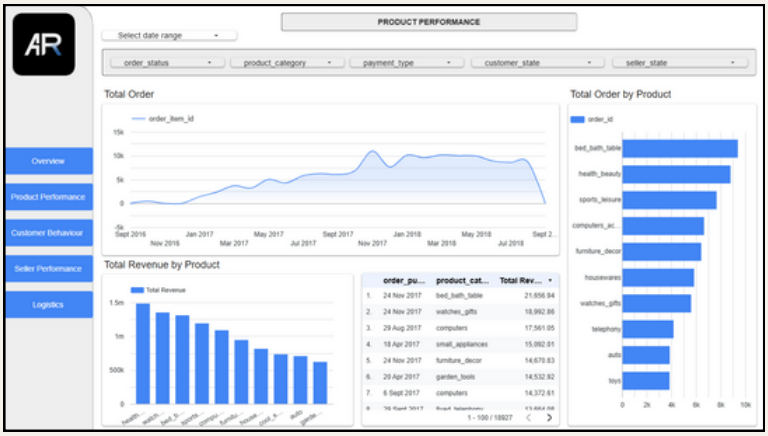
Overview



Overall, the e-commerce business demonstrates a mature operational scale with 98.2K total orders and total revenue of 16.6M. The Average Order Value (AOV) of 168.7 indicates that most transactions fall within the mid-value range, reflecting a stable and consistent purchasing pattern rather than impulsive or high-end buying behavior.

Revenue trends show steady growth from 2016 to early 2018, followed by a flattening pattern. This suggests that the business has entered a mature stage, where growth is no longer driven primarily by volume expansion but by efficiency and optimization. Operationally, performance remains strong, with 98.2% of orders successfully delivered, highlighting a generally reliable fulfillment process.

Product Performance

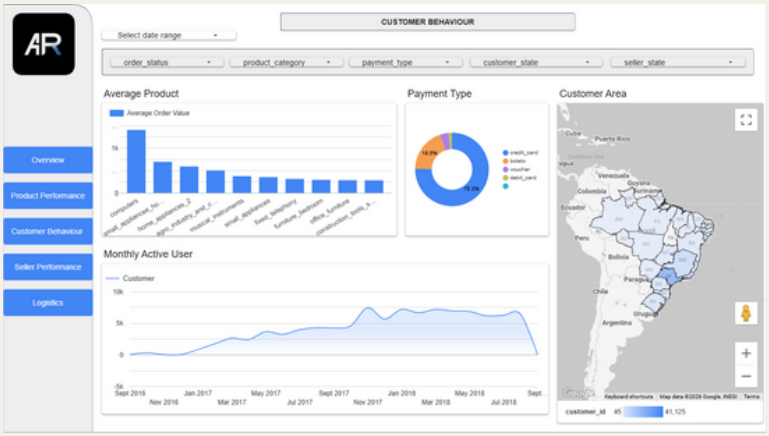


Product performance analysis reveals that revenue is highly concentrated in a few key categories, particularly Health & Beauty, Watches & Gifts, and Bed Bath Table. Among these, Bed Bath Table stands out as a core category, contributing significantly to both order volume and total revenue.

However, a clear distinction exists between high-volume products and high-revenue products. Some categories generate large numbers of orders but contribute less to revenue, while others deliver high revenue with relatively fewer transactions. This indicates a dual revenue structure driven by volume-based categories and margin-driven categories, each requiring a differentiated commercial strategy.

Analysis

Customer Behavior



From a customer behavior perspective, credit card payments dominate (75.3%), reflecting strong customer trust in digital transactions and a well-established payment infrastructure. This dominance also presents opportunities for installment-based promotions and partnerships with financial institutions.

The Monthly Active User (MAU) trend shows consistent growth until mid-2018, indicating healthy customer retention. In addition, categories such as electronics and home appliances record the highest Average Order Value, suggesting that customers are willing to make higher-value purchases in categories perceived as more reliable or essential.

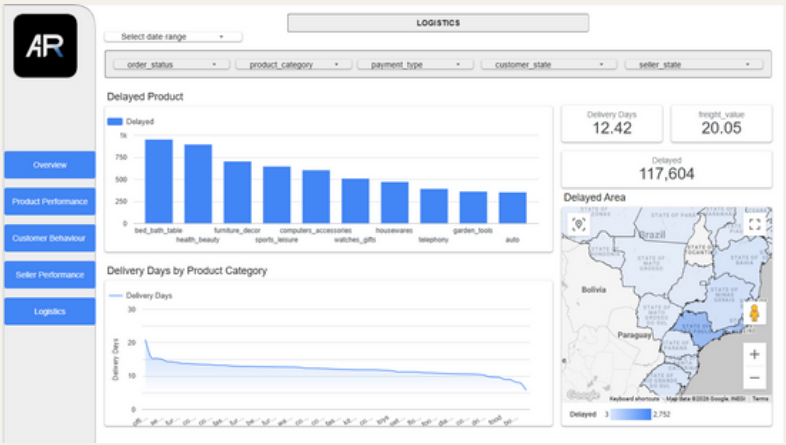
Seller Performance

The platform is supported by 3,095 sellers, yet revenue and order contributions are heavily concentrated among a small number of top-performing sellers. The top 5 sellers account for a disproportionately large share of total revenue and orders, while the majority of sellers remain in the long-tail segment with limited contribution.

This concentration creates a dependency risk on top sellers but also highlights an opportunity to strengthen mid-tier sellers through performance incentives, capability development, and service quality improvements. A more balanced seller ecosystem would improve resilience and long-term sustainability.

Analysis

Delivery Performance



From a logistics standpoint, the average delivery time is 12.42 days, with a total of 117,604 delayed orders. Delays are most prevalent in categories such as Bed Bath Table, Furniture Decor, and Sports Leisure, which typically involve larger or bulkier items and higher delivery complexity.

While overall delivery success rates remain high, logistics performance particularly for bulky products represents a critical operational bottleneck. Persistent delays may negatively impact customer satisfaction and reduce repeat purchase rates if not addressed strategically.

Conclusions

KEY INSIGHTS

1. The e-commerce business has reached a stable and mature phase, with revenue growth beginning to plateau.
2. High order volume does not always translate into high revenue, emphasizing the need for differentiated product strategies.
3. Customers exhibit mature digital behavior, with strong reliance on credit card payments and stable purchasing patterns.
4. Seller performance is highly concentrated, presenting both dependency risks and opportunities for seller ecosystem development.
5. Logistics, especially for large and bulky products, remains the primary challenge affecting customer experience.

FINAL TAKEAWAYS

The analysis indicates that future business strategy should shift away from pure volume growth toward profitability optimization, logistics efficiency, and seller performance balancing. Strengthening high-performing product categories, developing mid-tier sellers, and improving delivery performance for high-risk product categories will be critical to sustaining long-term growth.
This dashboard serves as a strategic decision-support tool for commercial, operations, and supply chain teams, enabling data-driven actions to enhance business resilience and operational excellence.

WHAT I LEARN

Through this project, I learned how to build a structured view model to unify multiple tables into a single analytical layer, enabling consistent and scalable analysis. I also strengthened my SQL skills by writing more complex analytical queries, handling joins, derived metrics, and data validation across large transactional datasets. In addition, I improved my ability to design effective dashboards by focusing on clear structure, appropriate visual choices, and business-oriented storytelling rather than visuals alone.

CONNECT WITH ME

fadliansyahfadel01@gmail.com

linkedin.com/fadliansyahrahman

github.com/FasaH-1