

DIGITAL EGYPT PIONEERS INITIATIVE – DEPI

Data Analyst Specialist Course

(Technical Path)

Fully Automated Olist E-commerce Analysis

Submitted By:

Tech Vanguard

Asmaa

Moustafa

Riham Salah Ali Ali

Youmna

Yassein

Group: ALX3_DAT1_G3

Supervised By:

Dr. Waleed Mohamed

November 2025

1. Contents

2. Abstract.....	3
3. Introduction	4
4. Problem Statement.....	4
5. Project Objectives	5
6. Tools and Technologies Used.....	5
7. Dataset Description.....	6
8.The Methodology “The Olist Data Warehouse Journey: From Zero to Data Hero”	7
Part 1: The Master Plan (Where We Draw the Map)	7
Part 2: The Cleanup & Staging Zone (ETL In Action)	7
Part 3: Building the Star Schema (The Transformation).....	10
Part 4: The Operational Layer (Making it Easy!)	13
8.5 Reporting and Strategic Delivery (The Payoff) (Power BI & Next Steps)	15
9. Project Pipeline Architecture.....	16
9.1 Who Does What in Our Pipeline? Automation vs Manual:.....	16
10. Visualizations.....	18
10.1 Executive Summary Dashboard.....	18
10.1.1 Key Performance Indicators (KPI Cards)	18
10.1.2 Chart:Revenue Over Time.....	19
10.1.3 Monthly Revenue & Monthly Orders (KPIs Cards)	20
10.1.4 Performance Breakdown & Segmentation	20
10.1.5 Top 5 Products Sold (Bar Chart).....	21
10.2 Financial Performance Dashboard	21
10.2.1 The KPI :Core Financial Health Metrics	22
10.2.2 Monthly Sales by Top 2 Products (Bar Chart).....	23
10.2.3 Orders in Weekday vs. Weekend (Donut Chart)	23
10.3 Sellers’ Performance Dashboard.....	24
10.3.1 Seller KPIs	24
10.3.2 Operational Risk & Quality Control	25
10.3.3 Sellers with Highest Processing Time (Bar Chart)	26
10.3.4 Stacked Bar Chart (Top 5 seller).....	26
10.3.5 Sellers by State (Map).....	26
10.4 Customers’ Analysis.....	27

10.4.1 Customer Scale and Value Metrics	27
10.4.2 Geographic & High-Value Segmentation	28
10.4.3 Behavioral Analysis & Risk Management.....	29
10.5 Operational Efficiency Dashboard.....	30
10.5.1 End-to-End KPIs	31
10.5.2 second rows KPIs	32
10.5.3 Geographic & Historical Analysis	33
10.5.4 Order Status Distribution (Donut Chart).....	35
10.6 Product Analysis Dashboard.....	35
10.6.1 Core Product Financials (KPIs).....	36
10.6.2 Profitability and Contribution.....	36
10.6.3 Return Risk and Loss Analysis.....	37
10.6.4 Most Ordered vs. Most Returned Product (KPI Cards).....	37
10.6.5 deep analysis in fashion.....	38
10.7 Reviews Dashboard Documentation	39
10.7.1 Overall Satisfaction & Risk Metrics (KPIs).....	40
10.7.2 Failure Analysis & Risk Focus	40
11.Q&A	42
12. Results and Findings.....	42
13. Overall Recommendations	44

2. Abstract

This project -fully automated system- analyzes the Olist E-Commerce dataset (2016-2018) to uncover specific insights into **Return Rate, customer behavior, and end-to-end logistics performance**, addressing the business challenge of unstable growth and compromised profitability. The **Methodology** employed a structured **ELT (Extract, Load, Transform)** pipeline: **Python (Pandas/SQLAlchemy)** handled data acquisition and pre-processing, loading clean data into **MySQL Staging Tables**. The heavy lifting was performed by an atomic **SQL Stored Procedure** that constructed the **OLAP Data Warehouse** and calculated derived metrics, including a dynamic **Date Dimension**. Final analysis was enabled through interactive dashboards in **Power BI**, the project identifies critical failures.

The findings confirm that the platform is fundamentally constrained by 12.48 day average delivery time and that its main revenue source, the **Home & Comfort** category, simultaneously generates the highest financial risk through returns and poor reviews. The **Solution** delivers targeted, data-backed recommendations focused on **reducing delivery time fragmentation**, mitigating product-specific risks, and implementing a strategy to boost customer retention. The entire solution is operationalized via a **Streamlit UI** to facilitate one-click, repeatable data updates.

3. Introduction

Olist is a large Brazilian e-commerce marketplace that connects small businesses to customers nationwide. The company provides a rich, multi-table-9 tables- raw dataset used for e-commerce analytics. Understanding this data enables valuable insights into customer behavior, sales trends, delivery performance, and product distribution.

In this project, data is analyzed and visualized using Python, MySQL, PowerBI. The analysis aims to provide actionable conclusions based on real-world e-commerce dynamics.

4. Problem Statement

E-commerce businesses frequently face critical challenges such as late deliveries, inconsistent product performance, low customer retention, and difficulty understanding sales patterns. Without proper analysis, these challenges remain hidden and actively erode profitability. **This project aims to address these problems directly by analyzing Olist dataset from all eight analytical dashboards and identifying specific, actionable patterns that influence customer satisfaction, sales, logistics performance, and identifying the key risks and strategic priorities for the committee thereby establishing a data-driven path for strategic business optimization.**

5. Project Objectives

- **Dimensional Model Implementation:** Implement a robust OLAP Star Schema in MySQL, including Fact, Dimension, and dynamic DimDate tables, based on informed EDA.
- **Data Pipeline & Quality:** Establish an ELT pipeline (Python ETL + SQL Stored Procedure) to load Olist data, ensuring quality and calculating derived metrics (e.g., delivery_delay_days).
- **Interactive Reporting & Operationalization:** Deploy interactive Power BI dashboards and enable one-click execution of the pipeline via a Streamlit UI.
- **Strategic Business Recommendations:** Provide data-backed strategies to reduce the 12.48-day average delivery time and mitigate financial risk from the Home & Comfort category.

6. Tools and Technologies Used

Tools and Technologies Used: Our Data Stack

- **Python (Pandas, SQLAlchemy, python-dotenv)**
 - Role: ETL Orchestration (Extract, Transform, Load) and Data Staging.
 - Key Use: Leveraging Pandas for fast, in-memory data cleaning and transformation (removing duplicates, handling nulls). SQLAlchemy was used for secure database connectivity, and python-dotenv for loading sensitive credentials.
- **MySQL**
 - Role: Data Warehouse (DW) Repository and Transformation Engine.
 - Key Use: Hosting the staging layer and the final Star Schema (Fact and Dimension tables). MySQL executed the complex, high-performance Stored Procedures which handled all the SCD Type 1 updates, lookups, and calculated metric generation.
- **Power BI**
 - Role: Business Intelligence (BI) and Data Visualization.
 - Key Use: Consuming the finalized Star Schema, developing interactive, high-value dashboards, and performing business analysis.
- **Figma**
 - Role: Modeling the whole process of our project

- **draw.io**
 - Role: Conceptual Modeling.
 - Key Use: Designing the Entity-Relationship Diagram (ERD) and creating the Star Schema Blueprint

to guide the data modeling phase.

- **Jupyter Notebook & Visual Studio Code (VS Code)**
 - Role: Development Environment.
 - Key Use: Jupyter Notebook was essential for initial Exploratory Data Analysis (EDA) and prototyping Python functions. VS Code served as the primary Integrated Development Environment (IDE) for writing and orchestrating the core Python and SQL scripts.

7. Dataset Description

The Olist dataset 2016-2018 implements a **star schema design** optimized for historical analysis; analytical queries, comprising eight interconnected tables and the ninth table for translating the product category column from pt>en:

- **Customers** (5 fields): Location-based identification with zip code, city, state, ID and unique ID
- **Products** (9 fields): Product categories and specifications, and Product ID
- **Sellers** (4 fields): Location-based identification with zip code, city, and state
- **Orders** (8 fields): Order details including timestamps “Comprehensive order lifecycle tracking from purchase through delivery”, Order status and Order ID.
- **Order_items** (7 fields): Individual items in each order details with pricing and freight values
- **Order_payments** (5 fields): Payment types, its value, installment and order
- **Order_reviews** (7 fields): Customer ratings and feedback with timestamps
- **Geolocation** (5 fields): Geographic reference data linking zip codes to coordinates

8.The Methodology “The Olist Data Warehouse Journey: From Zero to Data Hero”

Forget the complex jargon. This is the real story of how we built the ultimate data brain for Olist. Our goal: turn messy files into a lightning-fast Star Schema ready for Power BI glory

Part 1: The Master Plan (Where We Draw the Map)

Before we touch any code, we strategize. This phase is all about understanding the market and designing the perfect house for our data.

1. The Detective Work (Planning & Design)

- **Market Deep Dive:** We started by exploring the e-commerce world. What are the hot trends? What makes Olist succeed? This research helped us define our **Key Business Questions (KBQs)** the non-negotiable questions the business needs answers to.
- **The Blueprint:**
 - **The Data Cheat Sheet (Metadata):** We created a list of every single column in the raw data, detailing its name and what it means. This sheet was our team's guiding light!
 - **The ERD Sketch:** We designed the initial **Entity-Relationship Diagram (ERD) using draw.io** our architectural blueprint. We positioned the main **Fact** tables (the numbers) and the surrounding **Dimension** tables (the context) and mapped out all the connections (**PKs** and **FKs**).

2. Security and Setup (The Foundation)

We built a fortress!

- **Virtual Environment:** We used a **Virtual Environment** to lock down all our Python tools (Pandas, Streamlit, etc.). No more "it works on my machine" issues!
- **Secure Credentials:** We tucked all sensitive login details into a private .env file, loaded securely by python-dotenv. Security first!
- **Team Sync:** Everyone agreed on one single location for the raw files (D:\data\raw). Simple, but genius for avoiding file errors!

Part 2: The Cleanup & Staging Zone (ETL In Action)

This is where the raw, dirty data lands safely and gets its first scrub-down before the big transformation.

1. The MySQL Staging Layer (Our Safety Net)

- **The Magic Strategy:** We set *every single column* in our staging tables to the flexible type: **VARCHAR(255)**.
- **Why it's Crucial:** Raw data is often wild, with mixed types or weird characters. By using VARCHAR, we guarantee that the data **loads successfully** 100% of the time. We handle the complex type conversions later when we have total control!

Code Example (SQL DDL):

```
-- 2.1. staging_customers

DROP TABLE IF EXISTS staging_customers;

CREATE TABLE staging_customers (
    customer_id VARCHAR(255),
    customer_unique_id VARCHAR(255),
    customer_zip_code_prefix VARCHAR(255),
    -- ... and so on!
);
```

The Python Bridge (The Automated Engine)

Our main Python script is the workhorse. It extracts the raw data, cleans it in memory using Pandas, and loads it into staging.

- **Secure Connection:** We safely read passwords from .env and use urllib.parse.quote to make sure even passwords with crazy special characters don't break the connection URL!
- **Generic Cleaning Function:** This is where the initial Exploratory Data Analysis (EDA) happens instantly:
 - We trim whitespace from column names and text data.
 - We hunt down "**fake nulls**" (like the string 'nan' or 'NULL') and replace them with proper missing values (pd.NA).
 - We instantly remove any fully duplicated rows (df.drop_duplicates()).
- **The Atomic Load:** This is a powerhouse function! We use a transactional block that says: "Either **TRUNCATE** (delete) the old data and **load** the new data completely, or fail trying." This ensures our staging data is always fresh, fast, and complete!
Idempotency for the win!

Code Example (Staging Load - Snippet 3):

```
def load_staging(df: pd.DataFrame, table_name: str) -> None:  
  
    try:  
  
        with engine.begin() as conn:  
  
            # Temporarily disable foreign key checks for a clean  
            TRUNCATE  
  
            conn.execute(text("SET FOREIGN_KEY_CHECKS = 0;"))  
  
            conn.execute(text(f"TRUNCATE TABLE {table_name};"))  
  
            conn.execute(text("SET FOREIGN_KEY_CHECKS = 1;"))  
  
  
        # Load the cleaned Pandas DataFrame  
  
        df.to_sql(table_name, con=conn, if_exists="append",  
index=False)  
  
  
    except Exception as e:  
  
        # ... error handling and raising the exception ...  
  
        raise
```

- **The Grand Finale of Python:** After all files are safely staged, the Python script's final job is simple: make one single call to our powerful **SQL Stored Procedure**.

Code Example (Main Orchestration - Snippet 4):

```
def run_olist_etl(raw_folder: str) -> tuple[bool, str]:  
  
    # ... (code to loop through files and call load_staging) ...  
  
  
    # Run the incremental stored procedure once all staging tables  
    # are loaded  
  
    print("\n[DW LOAD] Calling stored procedure  
sp_LoadOlist_DW_Incremental() ...")  
  
  
    with engine.begin() as conn:
```

```
conn.execute(text("CALL sp_LoadOlist_DW_Incremental();"))
```

```
print("    -> Data warehouse loaded successfully!")
```

```
# ... return success ...
```

Part 3: Building the Star Schema (The Transformation)

The data is clean, and now we model it for maximum speed and analytic power!

Dimensional Modeling: The Final Blueprint

- **Surrogate Keys (SKs):** We gave every Dimension table (Customer, Product, etc.) a new, simple, auto-incrementing integer called a **Surrogate Key** (CustomerSK). These integers replace the complex original IDs, making our reporting lookups **lightning fast!**
- **The Date Dimension:** This is an analyst's best friend! We built a dedicated DimDate table by pulling *all* date fields from, We built it in MySQL by pulling all necessary dates from ('Fact Order'.'Purchase_date', 'estimated_delivery_date', 'delivery_customer_date'), marking the starting/ending dates and adding granular columns (Month Name, Quarter, is_weekend for Brazil context). This means Power BI never has to guess about time-based analysis!

Code Example (SQL DDL - DimCustomer):

```
-- 3.2 DimCustomer

CREATE TABLE DimCustomer (
    CustomerSK INT AUTO_INCREMENT PRIMARY KEY, -- The new, fast key
    customer_id CHAR(32) NOT NULL, -- The original ID
    (business key)
    customer_unique_id VARCHAR(255) NULL,
    zip_code_prefix INT NULL,
    city VARCHAR(255) NOT NULL,
    geo_sk INT NULL, -- Foreign Key to
    DimGeolocation
```

```

CONSTRAINT u_customer UNIQUE (customer_id),
CONSTRAINT fk_dimcustomer_geo
    FOREIGN KEY (geo_sk) REFERENCES DimGeolocation(geo_sk)
) ENGINE=InnoDB;

```

Fact Tables: The Measures

We split our transactions into four focused Fact tables, connecting them to the Dimensions using the Surrogate Keys.

- **The Unknown Member Rule:** This is a Data Warehousing secret weapon! Every Fact table has a **DEFAULT -1** on its Foreign Keys. We also manually insert a record with SK = -1 into every Dimension.
 - **Result:** If we have a sale missing a Customer ID, it links to the safe 'Unknown Customer' member instead of crashing the load. **We preserve every transaction!**

Code Example (SQL DDL - Example Fact Table):

```

/* 5.1 FactOrders */

CREATE TABLE FactOrders (
    order_id CHAR(32) NOT NULL,
    customer_sk INT NOT NULL DEFAULT -1, -- FK to DimCustomer,
    defaulting to -1 (UNKNOWN)

    order_status VARCHAR(50) NOT NULL,
    purchase_date_key VARCHAR(20) NULL,
    delivery_delay_days INT NULL,           -- A calculated metric!
    is_returned TINYINT NOT NULL DEFAULT 0,
    CONSTRAINT pk_fact_orders PRIMARY KEY (order_id),
    CONSTRAINT fk_fact_orders_dim_customer
        FOREIGN KEY (customer_sk) REFERENCES DimCustomer(CustomerSK)
) ENGINE=InnoDB;

```

The ETL Engine: sp_LoadOlist_DW_Incremental (The SQL Power)

All the real heavy lifting happens inside this single, powerful SQL Stored Procedure.

- **Dimensional Loading (SCD Type 1):** For Dimensions, we use the super-efficient **ON DUPLICATE KEY UPDATE** pattern. If we load a customer who already exists, we simply **update** their city or state. The data is always current!

Code Example (SQL DML - DimCustomer ETL):

```
/* 6.2 DimCustomer */

INSERT INTO DimCustomer (...)

SELECT

    -- ... Cleaning and Casting logic here ...

    COALESCE(g.geo_sk, -1) -- Get the SK or default to -1

FROM staging_customers s

LEFT JOIN DimGeolocation g -- Join to get the Surrogate Key (geo_sk)

    ON g.zip_code_prefix = s.z

    AND g.city = UPPER(s.city_txt)

ON DUPLICATE KEY UPDATE -- SCD Type 1: Update if the record exists

    customer_unique_id = VALUES(customer_unique_id),

    zip_code_prefix = VALUES(zip_code_prefix),

    geo_sk = VALUES(geo_sk);
```

- **Fact Loading & Calculations:** We calculate key metrics like delivery_delay_days right here, during the load. This offloads work from Power BI and makes reports scream! We also perform all our lookups to swap original IDs for the fast CustomerSK values.

Code Example (SQL DML - FactOrders ETL with Calculations):

```
/* 6.5 FactOrders */

INSERT INTO FactOrders (...)

SELECT

    TRIM(o.order_id) AS order_id,

    COALESCE(dc.CustomerSK, -1) AS customer_sk, -- The lookup,
    defaulting to -1

    -- ... Date formatting and keys ...
```

```

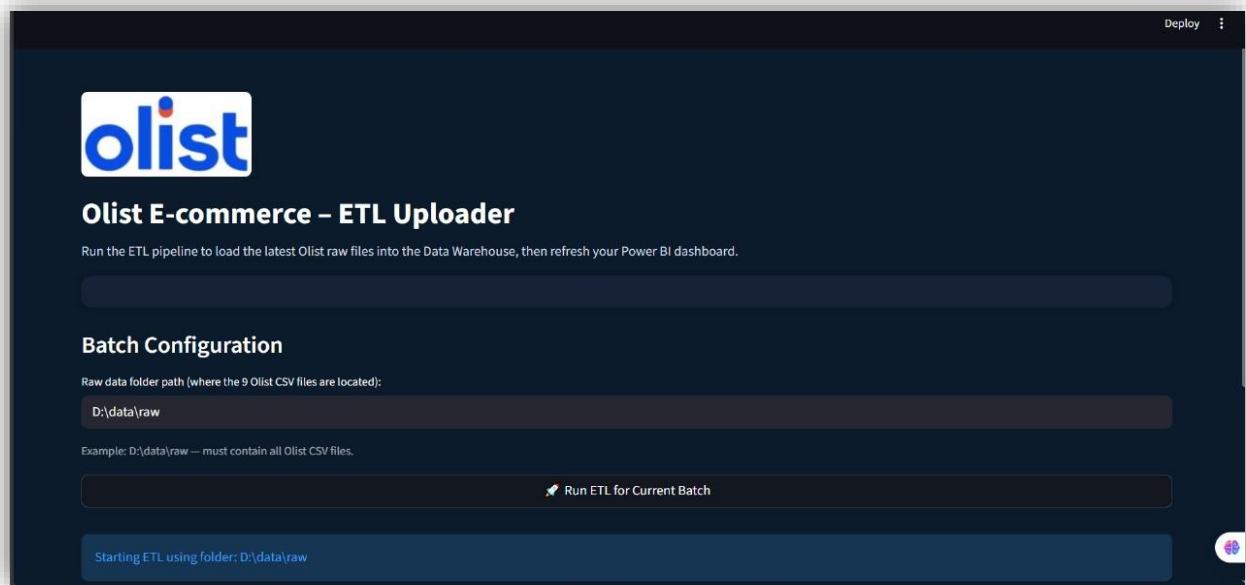
CASE WHEN dcust_dt IS NULL OR est_dt IS NULL
      THEN NULL
      ELSE DATEDIFF(dcust_dt, est_dt) END AS delivery_delay_days, --
Calculated Metric!

-- ... other calculated metrics ...

FROM staging_orders o
LEFT JOIN DimCustomer dc
ON dc.customer_id = o.customer_id -- Lookup to get CustomerSK
-- ...

```

Part 4: The Operational Layer (Making it Easy!)



The final step: building a beautiful, user-friendly interface using **Streamlit** so anyone can run this complex pipeline with one click.

- **The Dashboard Interface:** We used custom CSS to give our ETL tool a professional, dark aesthetic. Good design boosts user confidence!
- **One-Click Execution:** The user enters the file path, hits the "**Run ETL**" button, and that's it!

- **Instant Feedback:** We use Streamlit's loading spinner (st.spinner) to show the process is running. When it finishes, we give a huge success message and **balloons!** (Because a successful data load deserves a celebration.)

Code Example (Execution and Feedback):

```
if run_button:  
  
    # ... folder path validation ...  
  
    else:  
  
        with st.spinner("ETL process in progress... please wait ⏳"):  
  
            try:  
  
                run_olist_etl(raw_folder)  
  
                success = True  
  
                message = "ETL process finished successfully."  
  
            except Exception as e:  
  
                success = False  
  
                message = f"ETL process FAILED with unexpected error:  
{e}"  
  
  
        if success:  
  
            st.success(message)  
  
            st.balloons() # Success reward!  
  
        else:  
  
            st.error(message)
```

The Streamlit interface successfully abstracts all the code complexity into a simple tool, allowing analysts to load the freshest data whenever they need it!

All The Steps Above completes the documentation of The entire technical stack of our Project:

1. **Conceptual/Logical Design (The Methodology)**
2. **DDL/DML (The Data Warehouse Structure and Transformation Logic)**

3. Python ETL (The Data Staging Engine)
4. Streamlit UI (The Operational Interface)

8.5 Reporting and Strategic Delivery (The Payoff) (Power BI & Next Steps)

Now for the grand finale getting the insights out and providing actionable business advice.

Visualization & Recommendations (Showing Off the Results):

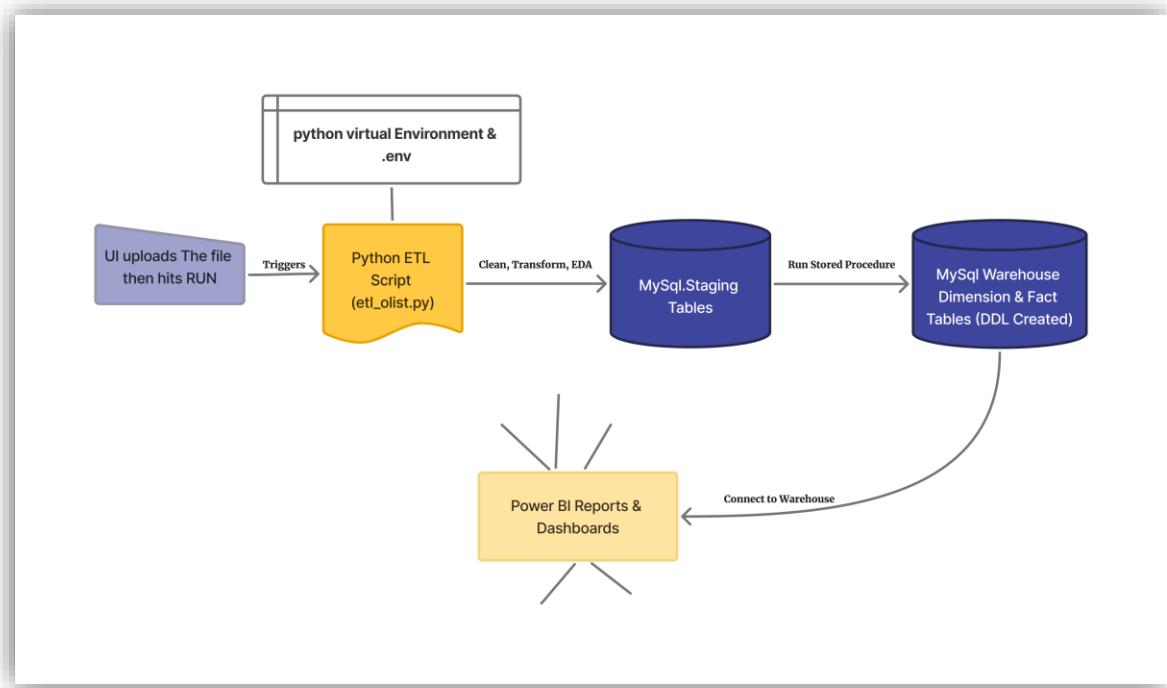
- BI Connection: We established a **Direct Connection** from our completed MySQL data warehouse into Power BI.
- Interactive Dashboard: We built professional and interactive dashboards that clearly presented our complex findings to any user.

Finally, we used the clear patterns and insights from the dashboards to formulate solid, data-driven recommendations for stakeholders on how to improve things like sales and delivery.

6. Interactive dashboards were created for:

- Executive Summary
- Finance Performance
- Sellers' Performance
- Customers' Analysis
- Operational Efficiency
- Reviews
- Product Analysis
- Q&A

9. Project Pipeline Architecture



9.1 Who Does What in Our Pipeline? Automation vs Manual:

It's awesome that we have a fully automated system, but it's crucial to remember that the machine can only execute the brilliant plan we designed. Here's a clear, simple breakdown of the machine's job versus the essential human strategy behind the Olist Data Warehouse:

The Machine's Job: Full Automation on Demand

When the user hits "Run ETL" in Streamlit, our pipeline takes over completely. This is the realm of consistent, lightning-fast execution.

1. Data In-and-Out (E & L):

- The Python script handles the boring but necessary work: securely reading the raw CSV files and connecting to the database.
- It uses a crucial technique to **atomically reload** the staging tables meaning the old data is wiped and the new batch is loaded in one clean transaction.

2. First-Pass Cleanup (The Fast Scrub):

- Python's Pandas library instantly cleans the data *in memory*: stripping extra spaces, spotting and replacing those sneaky '**fake nulls**' (like the string 'nan'), and dropping duplicated rows. No messy data gets past the first gate!

3. Heavy Transformation (The Power-Up):

- Control is handed to the high-performance MySQL **Stored Procedure** (our data factory).
- **Type Casting:** It converts all that generic staging data (remember the safety net VARCHAR) into the correct types (INT, DATE, etc.).
- **Modeling Magic:** It automatically looks up and inserts all necessary **Surrogate Keys (SKs)**, connecting transactions to context.
- **SCD Type 1:** It uses smart **INSERT ... ON DUPLICATE KEY UPDATE** logic to update dimension records instantly keeping our dimensions current without human intervention.

4. Metric Calculation (The Brain):

- We make the machine do the math! It calculates complex business metrics like **delivery_delay_days** and **ship_leadtime_days** directly in the pipeline, ensuring Power BI reports are simple and screamingly fast.

5. Final Load & Structure:

- The transformed and enriched data is inserted into its final, perfect home: the **Star Schema**

Our Job: Strategy, Design, and Wisdom

The machine can only execute. The most crucial work the **strategic thinking** that gives the data meaning remains in the hands of the analyst and developer.

1. Conceptual Design & Strategy (The Master Plan):

- We conducted the **Market Research** to understand trends.
- We defined the **Key Business Questions (KBQs)** this is the **most valuable human analysis!** We decided *what* needs to be measured.
- We designed the entire **Star Schema** structure, choosing the Fact and Dimension tables and defining their relationships.

2. Data Quality & Robustness (The Safety Measures):

- We made the **strategic decision** to use VARCHAR(255) for staging a choice that guarantees successful imports.

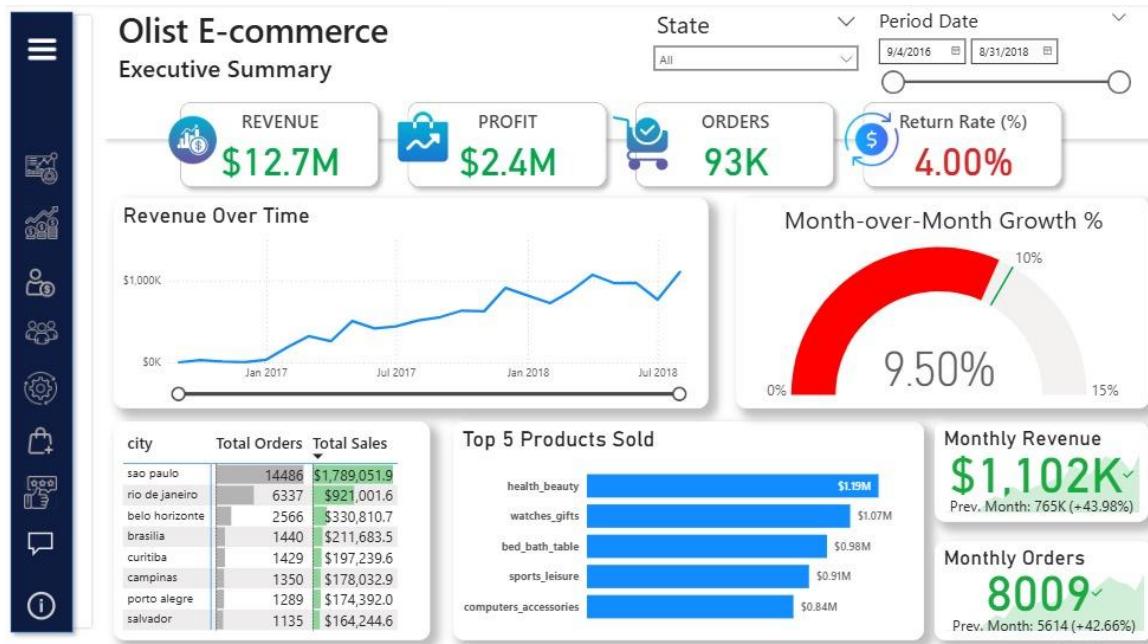
- We manually enforced a data quality standard by inserting the '**Unknown Member**' (-1 SK) into every dimension, ensuring we never lose a transaction due to missing data.

3. Defining the Transformation Logic (Writing the Code):

- The execution of the SQL Stored Procedure is automated, but **writing the code itself is our job**. We manually defined every rule, every join, and every formula needed to make the data useful.

10. Visualizations

10.1 Executive Summary Dashboard



10.1.1 Key Performance Indicators (KPI Cards)

The top cards use formulas that strictly filter sales and orders where `order_status = "delivered"`, ensuring the metrics reflect successful transactions.

Total Revenue	\$12.7M	Target: \$12.0M	Insight: Strong revenue generation, exceeding the target.	Recommend: Maintain current sales/marketing spend.
----------------------	---------	-----------------	---	---

Net Profit	\$2.4M	Target: \$2.0M	Insight: 18.9% Net Profit Margin. Significantly above the industry average of 10%	Recommend: Maintain pricing/cost strategy. Preserve current Net Profit Margin of 18.9%. Action: Implement a quarterly audit of Product Cost (COGS) data quality, reporting any discrepancies > 2% to the Finance team by the 15th of the following month.
Total Orders	93,000	Target: 85,000	Insight :High volume indicates strong market presence and customer demand.	Recommend: Operational/Negotiation Focus Reduce Average Shipping Cost per Order by 8% without impacting delivery speed. Action: By Q2 2019 , renegotiate contracts with the top three shipping partners using the 93K order volume as leverage.
Failure Rate (%)	3.94%	Target: < 5%	Insight: This suggests that customers are highly satisfied with the products they receive and that our delivery process is smooth.	Product/Logistics Focus, Reduce the overall Failure Rate to 3.5% or below. Action: By March 31, 2019 , conduct a root cause analysis (RCA) on the 3.94% rate, specifically isolating returns (RMA) from cancellations. Based on the RCA, launch a pilot program (e.g., improving product descriptions or strengthening packaging) targeting the leading failure cause and the time of cancellation.

10.1.2 Chart:Revenue Over Time

Insight

Recommendation

Strong, Consistent Growth: The revenue chart shows a clear, **upward growth trend** over the entire period. The recent **9.48% Month-over-Month (MoM) growth** confirms the business is accelerating rapidly.

Recommendation: The growth is aggressive. Operations must check if we have enough **inventory and shipping capacity** to handle a potential 10% increase every single month without causing delays or customer dissatisfaction.

10.1.3 Monthly Revenue & Monthly Orders (KPIs Cards)

Metric	Key Insight	Recommendation
Monthly Revenue (42.75% increase) & Monthly Orders (42.01% increase)	The massive approx 42% surge in both orders and revenue confirms that recent growth is driven by actual customer demand and higher sales volume , not just higher prices.	Recommendation: Immediately identify which products and cities drove this 42 %spike. Focus marketing and seller support on those specific areas to capitalize on the momentum.

10.1.4 Performance Breakdown & Segmentation

City Performance (Table) Top 8 Cities

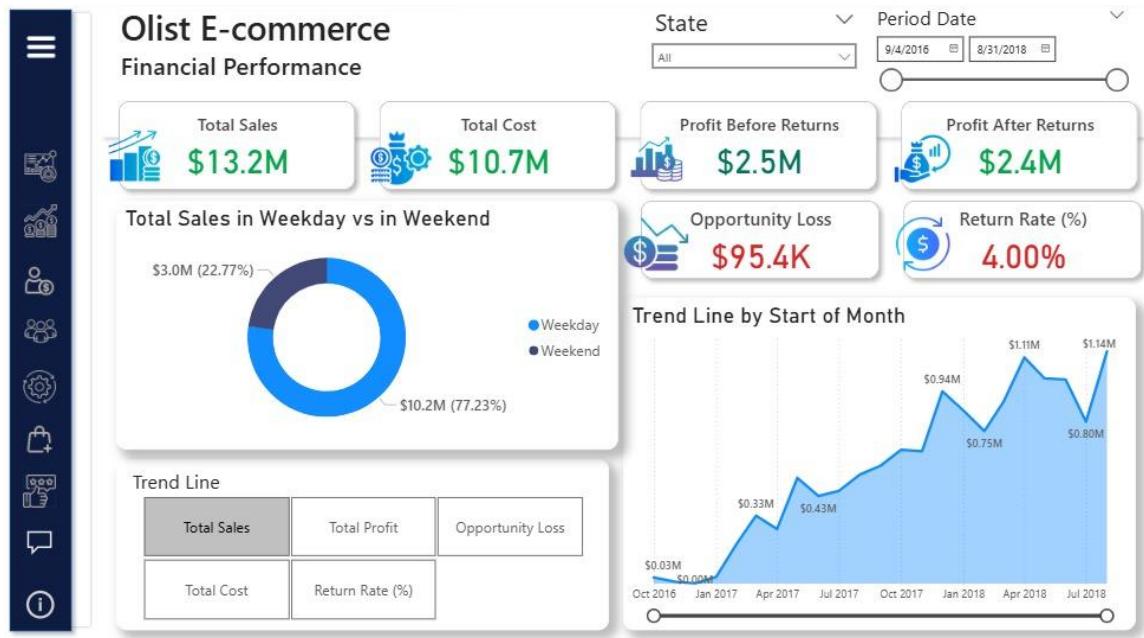
Insight	Recommendation
Market Concentration: São Paulo completely dominates the market, bringing in more than double the sales volume of the second city (Rio de Janeiro). We are highly dependent on this single city.	Recommendation: Allocate resources to attract customers in mid-tier cities (like Curitiba and Porto Alegre) to lessen our reliance on São Paulo and balance our geographic risk.

10.1.5 Top 5 Products Sold (Bar Chart)

Insight	Recommendation
<p>Top Sellers Drive Revenue: Health & Beauty and Watches & Gifts are the clear top-performing categories. Our customers prefer these lifestyle and gift-related products.</p>	<p>Recommendation: Prioritize finding new suppliers and expanding product selection within these two dominant categories to maintain our competitive edge and meet existing high demand.</p>

10.2 Financial Performance Dashboard

This dashboard moves beyond the Executive Summary by quantifying **costs, risks, and profitability drivers**. It clearly separates gross profit from net profit after accounting for product returns and cancellations.



10.2.1 The KPI :Core Financial Health Metrics

Metric	Value	Measure Logic	Key Insight
Total Sales	\$13.2M	SUM('Fact Sales'[price])	This is the Gross Revenue from <i>all</i> orders placed, including those that were canceled or returned. This is the maximum potential income.
Total Cost	\$10.7M	SUM('Fact Sales'[product_cost])	This is the Total Cost of Goods Sold (COGS) for <i>all</i> products ordered. This high cost is the reason profitability must be watched closely, Explore strategies to reduce supplier costs, negotiate better deals, or optimize inventory to maintain healthy margins.
Profit Before Returns	\$2.5M	[Total Sales] - [Total Cost]	This represents the theoretical Gross Profit margin (approx 19%), assuming every single order was delivered and kept.
Opportunity Loss	\$95.1K	Profit lost on items where is_returned = 1.	This \$95.1K is the direct profit lost due to failures (returns/cancellations). This is a quantifiable target for operational improvement; improving packaging, delivery reliability, and product quality checks, Implement proactive customer communication to clarify product details and expectations before purchase. Analyze return patterns to identify high-risk products or sellers and take corrective actions. Consider return policies or incentives that encourage customers to keep products when feasible.
Profit After Returns	\$2.4M	[Profit Before Returns] - [Opportunity Loss]	This is the True Net Profit from the successful, non-returned transactions. The difference between Before and After (\$100K) confirms the profitability loss due to returns is relatively low.

Metric	Value	Measure Logic	Key Insight
Return Rate (%)	3.94%	Total Canceled/Returned Orders divided by Total Delivered Orders.	This low rate is consistent with the Executive Summary, confirming high product quality and fulfillment reliability.

10.2.2 Monthly Sales by Top 2 Products (Bar Chart)

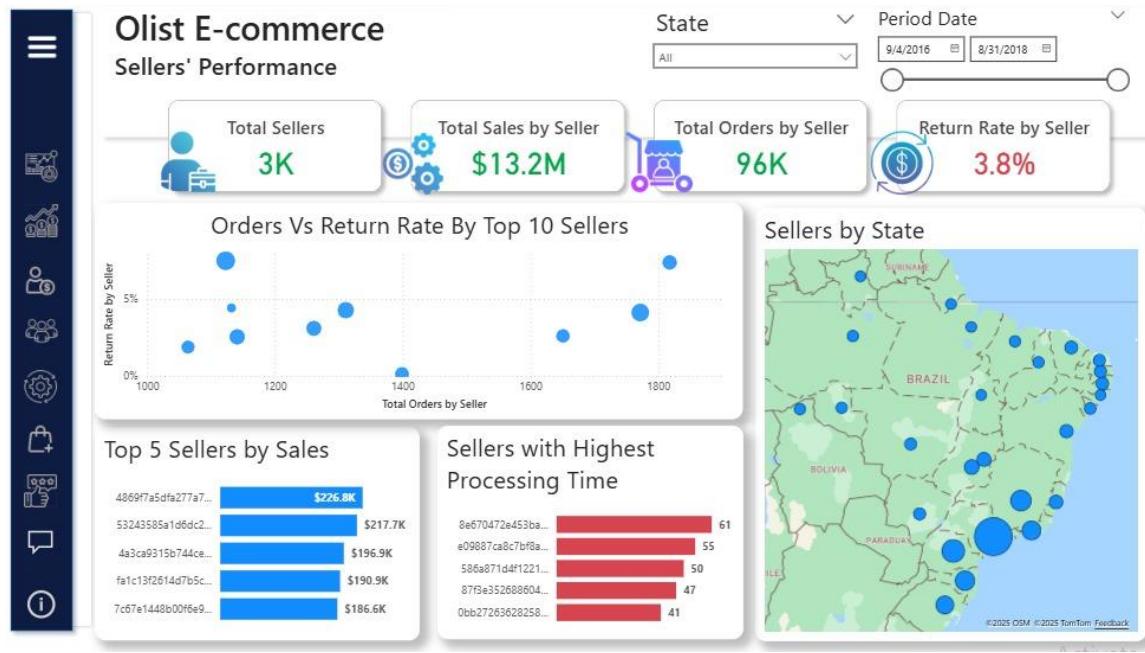
Description	Key Insight	Recommendation
Tracks the monthly Total Sales Revenue for the two top-performing product categories: Health & Beauty and Watches & Gifts .	Both categories show a clear upward trend , with sales peaking in 2018. This visual proves these categories are not only top sellers overall but are also the primary drivers of the platform's recent month-over-month growth .	Recommendation: Since these products are the growth engine, analyze their specific profit margins ($[price] - [product_cost]$) compared to other categories. If margins are strong, double down on marketing investment for these two product lines.

10.2.3 Orders in Weekday vs. Weekend (Donut Chart)

Description	Key Insight	Recommendation
Compares the total number of orders placed on Weekdays (Monday-Friday) versus Weekends (Saturday-Sunday) .	Orders are significantly higher during the Weekdays (72.8% of volume) . This suggests customers use the platform more for planned or work-related purchases rather than impulsive weekend shopping.	Recommendation: Shift advertising spend to focus on Weekday morning and evening commuter hours . Furthermore, investigate running weekend-specific flash sales to test if that strategy can boost the currently lower weekend volume.

10.3 Sellers' Performance Dashboard

This dashboard evaluates the efficiency, risk, and contribution of our supplier base, enabling targeted actions to improve the supply chain and customer satisfaction.



10.3.1 Seller KPIs

Metric	Key Insight	Recommendation
Total Sellers (3K) & Total Sales (\$13.2M)	The platform has a large, diverse seller base, which minimizes risk if any single seller leaves. The total sales confirm that these sellers are driving the overall business volume.	Recommendation: Use the large seller pool to continuously introduce competition. For example, introduce a tiered commission structure to reward top-performing sellers.
Total Orders (96K)	The high order volume shows that the platform provides significant business opportunity for its sellers.	Recommendation: Leverage this high order volume in negotiations with logistics partners to secure better shipping rates, which can then be passed on as a benefit to the sellers and encourage them to sell more.

Metric	Key Insight	Recommendation
Return Rate by Seller (3.8%)	This low rate is excellent and confirms that the <i>average</i> seller maintains a high standard of quality for products and fulfillment.	Recommendation: Use this low rate as the benchmark. Focus quality checks on sellers whose individual return rates exceed the 3.8% average.

10.3.2 Operational Risk & Quality Control

10.3.2.1 Orders Vs Return Rate by Top 10 Sellers (Scatter Plot)

Insight	Recommendation
Crucial Risk Identification: This plot clearly identifies " outlier " sellers that generate high order volume (X-axis) but also have a high Return Rate (Y-axis). These sellers pose the biggest threat to profitability and customer experience.	Recommendation: Immediate Action must be taken on any seller located in the top-right quadrant. Launch a Seller Improvement Program (SIP) for these specific high-risk accounts, providing performance warnings and mandatory training.

Quadrant Interpretation:

Quadrant	Volume (X-axis)	Return Rate (Y-axis)	Interpretation / Risk Level
Top Left	Low	High	Low Priority Risk: These are smaller sellers with quality control issues. They don't generate high volume, so their impact on overall revenue is small.
Top Right	High	High	CRITICAL Risk: These are high-volume sellers that cause high failure rates. They represent the biggest threat to customer experience and profitability.
Bottom Left	Low	Low	Potential Growth: These are reliable small sellers. They are good candidates for mentorship and scaling programs.

Quadrant	Volume (X-axis)	Return Rate (Y- axis)	Interpretation / Risk Level
Bottom Right	High	Low	IDEAL: These are our most valuable, high-volume, reliable sellers. They should be rewarded and protected.

10.3.3 Sellers with Highest Processing Time (Bar Chart)

Insight	Recommendation
Logistical Bottlenecks: The sellers shown here have the longest average time to prepare and ship an order (e.g., 50+ days). This is a severe weakness in our supply chain that directly leads to slow delivery for the customer.	Recommendation: Establish a strict Maximum Processing Time Service Level Agreement (SLA) (e.g., 7 days). Issue warnings to the top 10 slowest sellers, and reduce their listing visibility if they fail to meet the SLA.

10.3.4 Stacked Bar Chart (Top 5 seller)

Insight	Recommendation
Value Concentration (80/20 Rule): A very small number of sellers (the Top 5) are responsible for a large portion of the platform's revenue. This creates a risk if one of these top sellers leaves.	Recommendation: Create a VIP Seller Status for the Top 5. Offer them priority support, faster payment cycles, and higher visibility to ensure their loyalty and continued partnership.

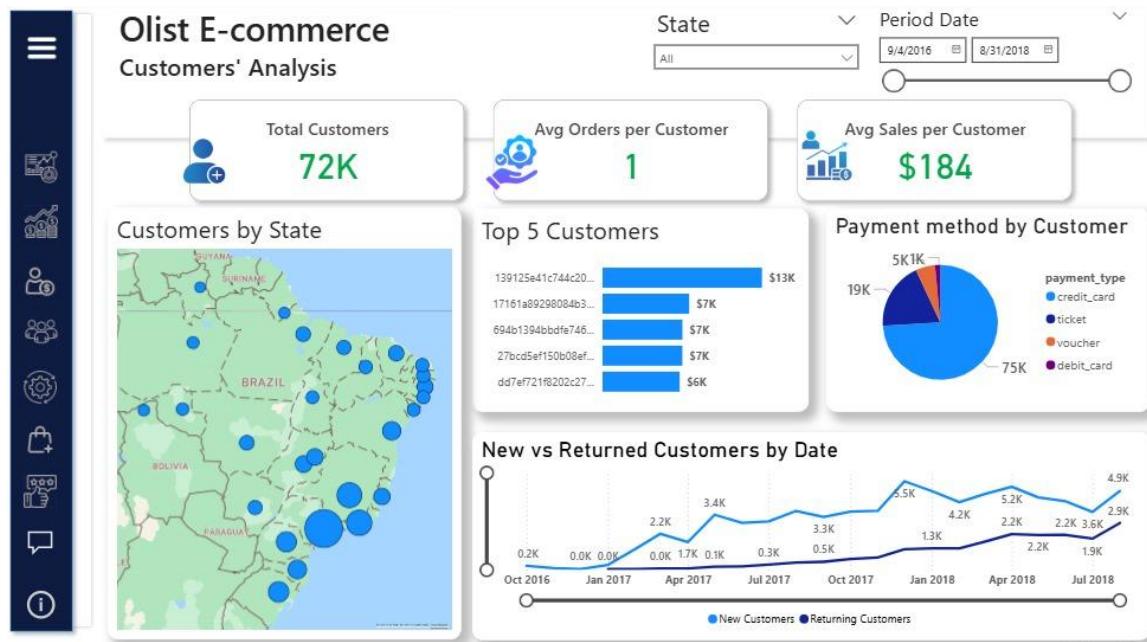
10.3.5 Sellers by State (Map)

Insight	Recommendation
Geographic Alignment: Seller concentration is highest in the Southeast (São Paulo, Rio de Janeiro) , matching where most of our customers are. This is efficient for logistics within that region.	Recommendation: To achieve better national coverage and reduce shipping costs to other regions, offer incentives (like reduced commissions) to recruit high-quality sellers in underserved areas (e.g., North and Northeast Brazil).

3. Integrated Analysis (Customer & Seller) : Market Concentration (Demand vs. Supply)

Insight	Recommendation
<p>Market and Supply Concentration: São Paulo completely dominates both customer demand (orders/sales) and seller supply (as seen on the Seller Map). This heavy concentration in the Southeast is efficient for logistics <i>within</i> that region.</p>	<p>Recommendation: While demand is high in São Paulo, the heavy concentration of sellers there creates logistical bottlenecks and risk. Offer incentives (e.g., free shipping credits) to recruit high-quality sellers in underserved states outside the Southeast to improve delivery times and lower costs for customers nationwide.</p>

10.4 Customers' Analysis



10.4.1 Customer Scale and Value Metrics

The top row KPIs quantify the size and average value of your customer base.

Metric	Value	Measure Logic	Key Insight
Total Customers	90K	DISTINCTCOUNT(customer_id)	A large base of 90K customers confirms strong market reach, but it implies a need for effective segmentation to maximize marketing ROI.
Avg Orders per Customer	1	Orders Count / Total Customers	An average order count of exactly 1 shows that Olist is primarily an acquisition platform (one-time buyers). Customer retention is not yet a strength So

Metric	Value	Measure Logic	Key Insight
			Focus campaigns specifically on driving the second purchase (e.g., "Welcome back" discounts)
Avg Sales per Customer	\$147	Total Sales / Total Customers	An average lifetime value of \$147 is a solid base. The strategy should focus on moving this value upward through repeat purchases .

10.4.2 Geographic & High-Value Segmentation

10.4.2.1 Customers by State (Map)

Description	Key Insight	Recommendation
Maps the Total Customers by their geographic state (using Avg Lat/Long).	The map confirms that the customer base is heavily concentrated in the Southeast region (especially São Paulo and Rio de Janeiro), matching both seller location and order volume.	Recommendation: Launch an experimental retention campaign targeting customers in lower-density states (e.g., North/Northeast) to test if geographic loyalty can be increased with localized incentives or faster, region-specific shipping offers.

10.4.2.2 Top 5 Customers (Bar Chart)

Description	Key Insight	Recommendation
Ranks customers by Total Sales (X-axis) and includes Orders Count and Total Profit After Returns in the tooltip.	The top 5 customers generate significant sales approx. \$47K total and high profit, confirming they are true VIPs .	Recommendation: Immediately create a CRM (Customer Relationship Management) profile for these top 5 customers. Assign them a dedicated account manager and offer exclusive first-look access to new products to secure their continued loyalty.

10.4.3 Behavioral Analysis & Risk Management

10.4.3.1 Payment Method by Customer (Pie Chart)

Description	Key Insight	Recommendation
Shows the volume of orders by payment type (Credit Card, Boleto, Voucher, Debit Card).	Credit Card is the dominant payment method approx 75%, indicating trust in the online process. Ticket is the second-largest payment type (approx. 18-20%).	Recommendation: Since Ticket payments often have a higher cancellation rate (due to non-payment), review the current Ticket processing window (Period before approving the Request). Consider offering a small discount (e.g., 2%) for immediate Ticket payment confirmation to increase conversion and reduce order loss risk.

10.4.3.2 New vs Returned Customers by date

Metric	Trend/Value	Key Insight Derived from Measure	Enhanced Recommendation (Simplified)
New Customers	High volume 5.5K peak; much higher than Returning Customers.	Acquisition Dependency: We rely heavily on finding new buyers. This is costly and risky for long-term health.	Shift Focus & Budget. Allocate more marketing money to keeping existing customers.
Returning Customers	Low volume 2.5K peak); trending up slowly.	Retention Opportunity: Many first-time buyers are not coming back. We have a big chance to earn more money from customers we already have.	Improve Customer Experience (CX) and launch simple loyalty rewards.
Ratio (New : Returning)	Approximately 2:1	Cost Risk: It costs much more to find a new customer than	Goal: Get more repeat buyers to make the

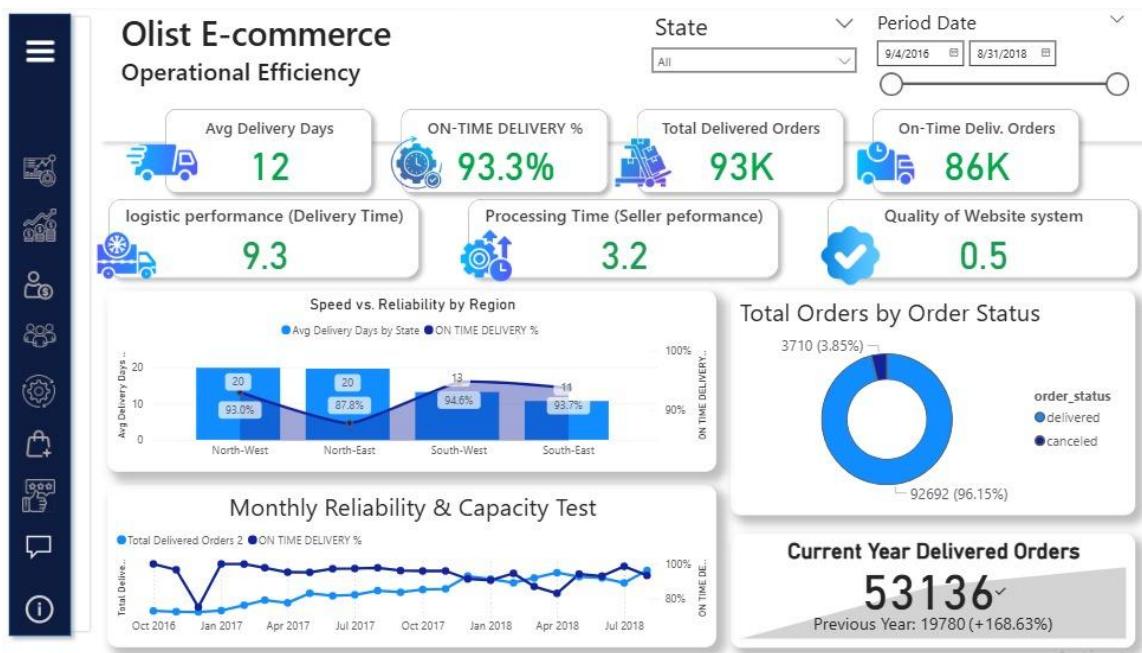
Metric	Trend/Value	Key Insight Derived from Measure	Enhanced Recommendation (Simplified)
		to sell to an old one. This ratio is inefficient.	business more profitable and stable.

Where To invest from the City,Seller, and Customer Details and analysis:

Opportunity	Data Rationale	Actionable Recommendation
Diversify Supply	The gap between supply and demand in the North/Northeast states is large. Seller concentration is too low to effectively serve these regions.	Offer financial incentives (e.g., discounted commissions or freight subsidies) to recruit high-quality sellers in underserved states . This will lower delivery costs and cut shipping times for local customers, boosting sales in those regions.
Targeted Acquisition	The Customer Map shows sparse but existing customers in the North. Sales per customer in these areas might be high, but acquisition is low.	Launch localized marketing campaigns in low-density states. Focus messaging on the improved regional shipping speeds resulting from the new local sellers to drive new customer acquisition.
Protect Key Markets	São Paulo is the engine. Its high order volume and sales must be maintained.	Establish VIP Support for the highest-volume sellers in São Paulo to prevent them from moving to competitors. Use them as a stable base while investing in expanding supply elsewhere.

10.5 Operational Efficiency Dashboard

This dashboard provides a detailed, component-level breakdown of the platform's delivery performance, focusing on the speed, reliability, and causes of logistical delays.



10.5.1 End-to-End KPIs

Metric	Value	Measure Logic	Key Insight
Avg Delivery Days	12 days	{Customer Delivery Date} - {Purchase Date}	A 12-day average delivery time is the total customer waiting period. This benchmark must be continually challenged to maintain competitiveness.
On-Time Delivery %	93.3%	{Orders Delivered On-Time} \ {Total Delivered Orders}	A reliability rate of over 93% is excellent, meaning the platform meets its delivery promise most of the time. This is a crucial metric for customer trust.
Total Delivered Orders	93K	Distinct count of orders where status is "delivered".	This volume matches the 93K orders from the Executive Summary, confirming that the vast majority of orders placed are successfully delivered.
On-Time Deliv Orders	87K	Total delivered orders that met the estimated delivery date.	The 87K figure is the absolute number of successfully fulfilled promises. (6.45% aren't on time)

10.5.2 second rows KPIs

KPI (Time Component)	Value (Days)	Measure Logic (Key Dates)	Key Insight	Recommendation
Payment Processing	0.51	{Approved Date} - {Purchase Date}	Excellent Efficiency (Website System): Processing the payment and approving the order takes less than half a day. This indicates the website and payment systems are performing optimally.	Recommendation: None needed for this metric. The system is highly efficient.
Processing Time	3.21	{Carrier Date} - {Purchase Date}	Seller Performance Bottleneck: This represents the time the seller takes to prepare the product and hand it over to the carrier. This is a controllable, internal factor that accounts for a large portion of the overall time.	Recommendation: This 3.21 days must be reduced. Incentivize or penalize sellers based on this metric (as discussed in the Seller Dashboard) to drive it closer to 1-2 days.
Delivery Time	9.27	{Customer Date} - {Carrier Date}	Logistics Performance Dominance: The actual shipping time by the carrier accounts for the largest part (9.27 days) of the 12-day customer waiting period.	Recommendation: Re-negotiate contracts with carriers in high-volume regions. Explore offering a slightly higher price tier to the customer for a guaranteed 5-7 day service level to reduce this average (Implement a tiered shipping model with Standard, Express, and Priority options . This gives customers flexibility based on

KPI (Time Component)	Value (Days)	Measure Logic (Key Dates)	Key Insight	Recommendation
				urgency and allows the platform to generate additional revenue. Express and Priority tiers can reduce average delivery time significantly while enhancing customer satisfaction).
Avg Delivery Days (End-to-End)	12.48	{Customer Date} - Purchase Date}	Total Customer Waiting Time.	Insight: Seller Time (3.21) + Logistics Time (9.27) \approx 12.48 days. The logistics phase is the priority target for improvement.

10.5.3 Geographic & Historical Analysis

10.5.3.1 Avg Delivery Days by State/Direction (Bar Chart)

Visual	Key Insight	Recommendation
Ranks delivery time by region (e.g., North, Northeast, South East).	Geographic Inequality: Delivery times are significantly worse in the North and Northeast regions (e.g., 19.6 days) compared to the Southeast (10.7 days). This creates an unfair customer experience and hinders national growth.	Recommendation: This confirms the need for Seller Supply Diversification . Recruit or create regional fulfillment centers in the North/Northeast to bypass the long cross-country shipping times and reduce delivery time inequality.

10.5.3.4 Combined Area/Line Chart (bottom-Left)

Visual	Key Insight	Strategic Recommendation
Combined Area/Line Chart (On-Time Delivered Orders vs. Total Delivered Orders by Month)	Absolute Capacity Failure: The chart provides clear visual proof that during peak periods (late 2017 and late 2018), the Total Delivered Orders (system load) outpaces the On-Time Delivered Orders (system limit) . The growing gap between the two lines	Recommendation: Fund a Strategic Logistics Investment aimed at breaking the capacity ceiling . This requires either negotiating guaranteed surge capacity with new or existing carriers for Q3/Q4 or investing in

Visual	Key Insight	Strategic Recommendation
	in high-volume months confirms that the logistics network does not have the surge capacity needed to handle peak demand reliably.	regional fulfillment hubs to handle the high volume without sacrificing the customer promise.

10.5.3.5 (Combined BAR Chart/LineChart)

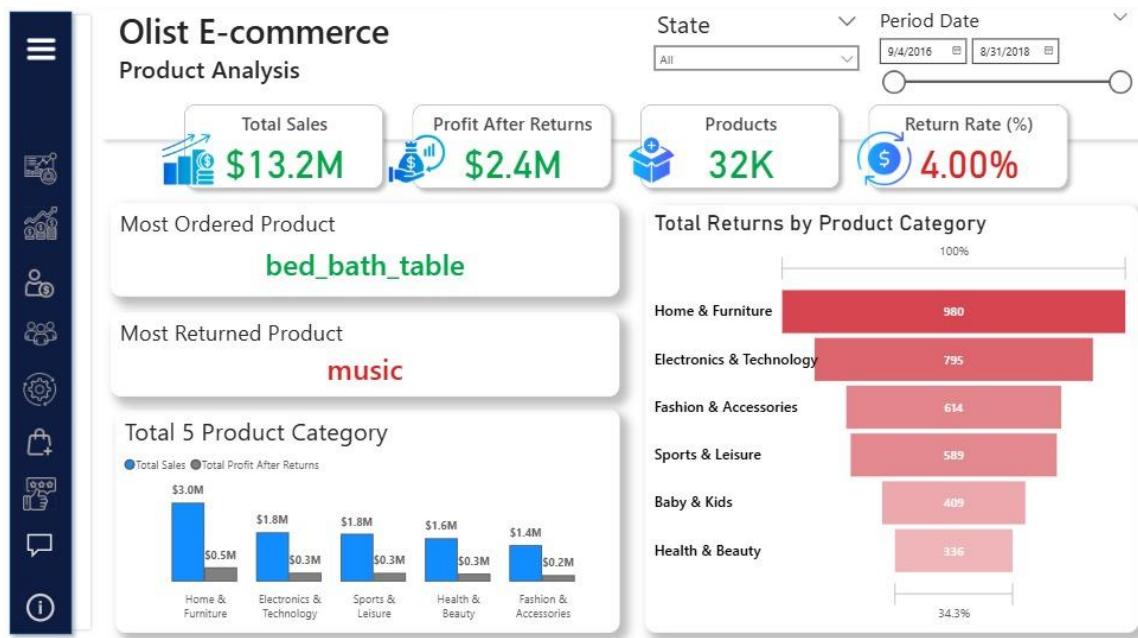
Priority Area	Data Justification	Recommendation
Highest Priority: North-East	Reliability is only approx 68% and speed is 19.6 days . This region is actively damaging the brand's reputation.	Recommendation 1: Immediate Carrier Audit. This reliability failure must be addressed first. Assign a regional logistics manager to determine if the primary carrier in the North-East is the bottleneck. Replace or augment carriers if needed.
Target for Supply Investment: North-West	Speed is the absolute slowest approx 21 days . This confirms the critical need for local supply to cut cross-country shipping time.	Recommendation 2: Fund Regional Fulfillment Hubs. Begin pilot programs to recruit sellers or establish a micro-fulfillment center in the North-West to drastically reduce the 21-day average. This is the only way to solve the speed problem.
Protect and Leverage: South-East	This region delivers the fastest (10.7 days) and is the most reliable (approx 95%).	Recommendation 3: Study the Southeast Model. Document the logistics and seller hand-off processes used in the Southeast and try to replicate them in the South-West and North-West to leverage best practices across the platform.

10.5.4 Order Status Distribution (Donut Chart)

Visual	Key Insight	Recommendation
Shows total orders broken down by status: delivered, shipped, canceled.	High Success Rate: The vast majority of orders (over 93%) are delivered . The percentage of canceled orders is extremely low.	Recommendation: Maintain focus on the canceled segment. Since the volume is low, it should be possible to perform root-cause analysis on every single canceled order to drive the cancellation rate even lower.

10.6 Product Analysis Dashboard

This dashboard evaluates the performance of the product catalog by linking sales, profitability, and return risk, enabling strategic decisions on inventory management and product focus



10.6.1 Core Product Financials (KPIs)

Metric	Value	Measure Logic	Key Insight
Total Sales	\$13.2M	Sum of sales price.	Confirms the gross revenue driven by the entire product catalog.
Total Profit After Returns	\$2.4M	Profit before returns minus opportunity loss.	This is the crucial net profit figure, showing that \$2.4M is the real value generated after accounting for all failed orders.
Products	32K	Distinct count of product IDs.	The large size of the catalog (32K products) suggests a long tail distribution where most sales come from a small fraction of products.
Return Rate (%)	3.94%	Total Canceled/Returned Orders / Total Delivered Orders.	The overall return risk is low, which is a good indicator of general product quality and delivery success.

10.6.2 Profitability and Contribution

Top 5 Product Category (Clustered Column Chart)

Visual Focus	Key Insight	Recommendation
Total Sales vs. Total Profit After Returns by category.	"Home & Comfort" (Bed, Bath, Table) is the absolute dominant category, driving the highest sales and highest net profit approx 0.8M This category is the platform's core strength.	Recommendation: Prioritize and Protect. Allocate the largest marketing budget to the Home & Comfort category. Dedicate a specific product management team to ensuring seller quality and identifying new, high-margin products within this vertical.
Secondary Contributors: "Sports & Health & Beauty" and "Computers &		Recommendation: For categories with strong profit potential, like Sports & Health & Beauty , launch a targeted seller acquisition campaign to bring in more specialized suppliers and capture greater market share.

Visual Focus	Key Insight	Recommendation
Accessories" are key secondary growth areas.		

10.6.3 Return Risk and Loss Analysis

Total Returns by Product Category (Funnel Chart)

Visual Focus	Key Insight	Recommendation
Ranks categories by Total Returns (absolute count) .	<p>Return Risk is Concentrated: Even though "Home & Comfort" is the biggest seller, it also has the highest absolute number of returns (999). Electronics & Tools is the second riskiest category (704 returns), despite having lower total sales than some others.</p>	<p>Recommendation: Conduct a Root-Cause Analysis on the top two returning categories: Home & Comfort and Electronics & Tools. This should investigate whether returns are due to product damage (logistics) or poor quality/misrepresentation (seller/product page).</p>

10.6.4 Most Ordered vs. Most Returned Product (KPI Cards)

Metric	Value	Key Insight	Recommendation
Most Ordered	bed_bath_table	Confirms that this is the highest demand category.	Pricing & Profit Optimization, Inventory Strategy and safety stock to reach zero out of stock days for best to 10 selling products in bed_bath_table for next quarter. Cross-Sell/Up-Sell Strategy to track and report the AOV lift attributable to the cross-sell

Metric	Value	Key Insight	Recommendation
			feature, aiming for the 10%AOV increase target.
Most Returned	Music	High Operational Cost: Even if music products (e.g., vinyl, CDs) don't sell a lot, high returns mean high costs for shipping, processing, and handling. This item is hurting profit margins badly.	Immediate Audit and Mitigation. Stop the financial bleeding by fixing the problem or removing the product temporarily.

10.6.5 deep analysis in fashion

Visual / Metric	Value	Insight	Strategic Implication
Top 5 Product Category (Profit/Sales)	Total Sales: 1.4M	This category is a significant revenue generator , ranking fifth in Total Sales (just behind Sports & Leisure).	Keep it: The category is too large to ignore and must be maintained as a core part of the catalog.
Top 5 Product Category (Profit After Returns)	Total Profit: 0.2M (estimated based on chart size)	It generates the lowest net profit among the Top 5 categories. This suggests that high operational costs (potentially high return processing) are eroding the gross sales value.	Improve Profitability: Actions must be taken to convert more of the gross sales into net profit.
Total Returns by Product Category	Total Returns: \$627	This category ranks fourth in absolute returns (627). While not the highest, this large return volume confirms that the low net profit is likely due to the cost of processing these returns.	Address Root Cause: Investigate if returns are due to poor quality, sizing issues, or delayed shipping impacting seasonal relevance.

Visual / Metric	Value	Insight	Strategic Implication
Most Returned Product Card	Product Name: fashion_children_clothe	The single " Most Returned Product " on the entire platform falls under the fashion umbrella. This specific product has a disproportionately high risk.	Immediate Audit: This single product's poor performance is likely skewing the entire category's profitability downward.

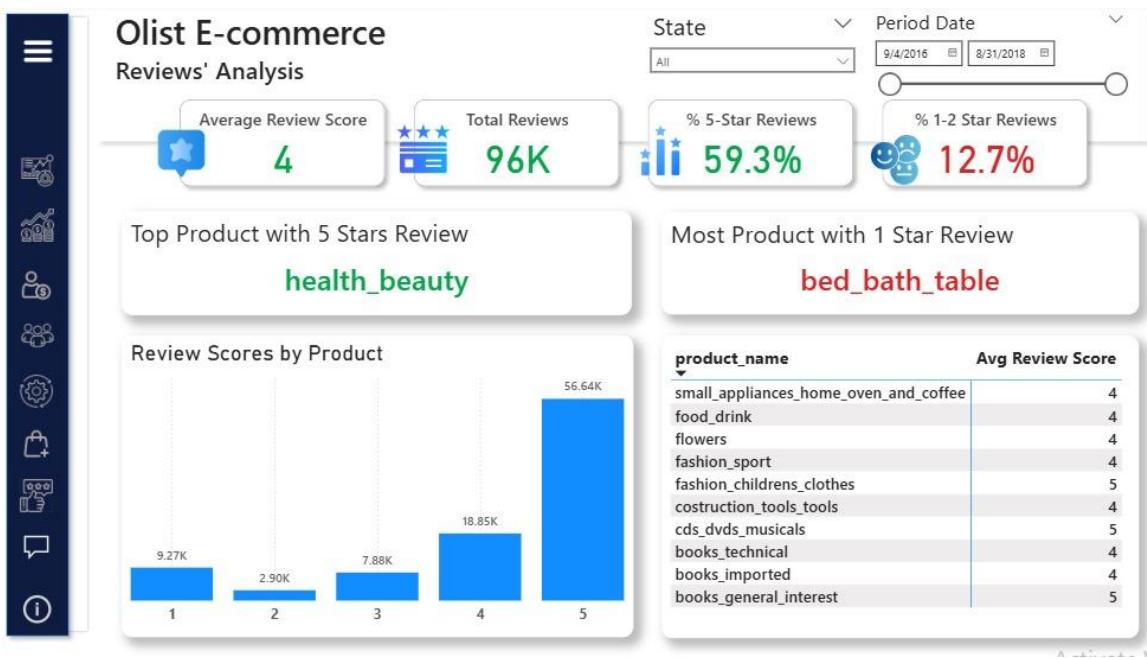
Recommendation for Fashion & Accessories

The strategy for this category must be **Risk Mitigation** and **Profit Optimization**.

- **Primary Recommendation: Mandatory Data Quality Audit.** Immediately review the top-returning products in this category. For clothing, enforce stricter quality controls on seller listings, focusing on standardized sizing charts and high-resolution product photography to reduce customer expectations mismatch.
- **Secondary Recommendation: Re-evaluate Logistics.** Due to the high number of returns (627), consider if faster delivery times are needed for fashion items to meet immediate customer demand (e.g., event shopping), or if the long average delivery time (12.48 days) is leading to higher customer cancellations/returns.

10.7 Reviews Dashboard Documentation

This dashboard measures customer satisfaction by analyzing feedback scores. It links review quality to specific products and overall operational performance to prioritize corrective action.



10.7.1 Overall Satisfaction & Risk Metrics (KPIs)

Metric	Value	Measure Logic	Key Insight
Average Review Score	4.29	Average of all review scores (1-5).	A high average score of 4.29 indicates strong overall customer satisfaction. The platform generally delivers a positive experience.
Total Reviews	96K	Total count of all reviews.	The high volume of reviews (96K) provides a statistically robust sample, confirming the average score is highly reliable.
% 5-Star Reviews	59.3%	5-Star reviews / Total Reviews.	Nearly 60% of all customer feedback is top-tier, which is an excellent marketing asset and proof of quality for most sellers.
% 1-2 Star Reviews	12.7%	1-2 Star reviews / Total Reviews.	This 12.7 failure rate represents the core risk area. This percentage must be targeted and reduced to secure brand reputation.

10.7.2 Failure Analysis & Risk Focus

10.7.2.1 Reviews Score by Product (Stacked Column Chart)

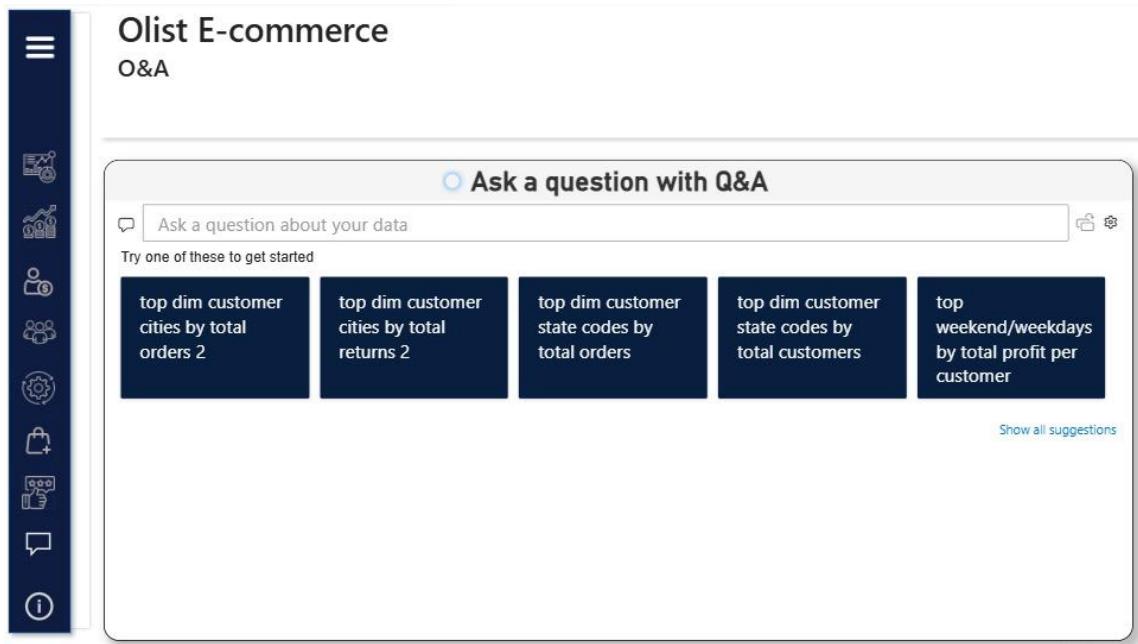
Visual Focus	Key Insight	Recommendation
Shows the distribution of review scores (1 to 5) by order count.	The chart clearly shows that the majority of orders	Recommendation: Prioritize 1-Star Review Investigation. Use the volume of 1-star reviews as a

Visual Focus	Key Insight	Recommendation
	result in 5-star reviews, confirming the overall success rate. However, the 1-star and 2-star columns still represent a substantial number of failed transactions.	leading indicator of severe operational issues (e.g., product never shipped, severe damage, or extreme lateness). Route all 1-star reviews to a dedicated investigative team for immediate seller/carrier action.

10.7.2.2 Top/Most Reviewed Products (KPI Cards & Table)

Metric(Visual)	Value	Key Insight	Recommendation
Top Product with 5 Stars	health_beauty	The Health & Beauty category consistently excels, providing the most reliable 5-star experience. This success should be leveraged.	Recommendation: Promote Success. Highlight sellers in the Health & Beauty category on the platform's homepage and in marketing materials to associate the brand with quality and success.
Most Product with 1 Star	bed_bath_table	This is a critical operational conflict. Home & Comfort is the platform's highest sales and profit category, but it also generates the highest volume of 1-star reviews.	Recommendation: Connect Reviews to Returns. Cross-reference 1-star reviews in the Home & Comfort category with the high absolute returns observed in the same category. This dual failure suggests major quality control issues that directly threaten the platform's main revenue stream.
Avg Review Score by Product Name (Table)	fashion_childrens_clothes has low scores; others are high.	The table shows that specific products, like fashion_childrens_clothes (identified earlier as the most returned product), have low average scores, confirming their role as reputation liabilities.	Recommendation: Product-Level Intervention. The intervention must be specific. Place the bottom 10 products/sellers with the worst average review scores under immediate operational review.

11.Q&A



The Q&A page is a crucial dashboard feature that transforms reporting into a **self-service analytical tool**. It utilizes **Natural Language Query (NLQ)** to let any stakeholder ask ad-hoc questions about the **Star Schema Data Warehouse** in plain English, receiving instant visual answers. Featuring pre-populated query suggestions tied to key dimensions (like customer cities and state codes), the page successfully **abstracts technical complexity**, empowering teams (Logistics, Marketing, Executive) to retrieve data-driven insights immediately without relying on custom BI reports.

12. Results and Findings

The Olist E-commerce platform is experiencing rapid growth, with \$12.7M in revenue and 93K orders. The operational integrity of the business is generally strong (Return Rate of 3.94 and Avg Review Score of 4.29). However, three critical and interconnected challenges threaten future profitability and national expansion.

1. Logistical Fragility & Service Inequality

The Problem

The platform's logistics network lacks surge capacity and is unable to handle peak volume, leading to service failure.

- **Priority Target:** The **Carrier Transit Time** (actual shipping) accounts for 74.19 (9.27 days) of the 12.48 day total customer wait time.

- **Proof of Failure:** The **On-Time Delivery %** fails cyclically, with the volume of late orders surging during high-sales months (Q3/Q4), proving the system hits an absolute capacity ceiling.
- **Geographic Risk:** Delivery speed is severely unequal, with the **North-East** region experiencing 19.6 day average delivery times and approx 68 reliability, actively damaging the brand's national reputation.

Strategic Recommendation

Invest in Regional Surge Capacity. Allocate a dedicated budget to establish **regional logistics hubs** (or highly incentivized local sellers) in the **North and Northeast**. This dual action will reduce the \$19.6 day shipping bottleneck and create the necessary surge capacity to handle peak sales volume reliably.

2. Profitability Risk in Core Categories

The Problem

The platform's most profitable category is also the biggest source of dissatisfaction, creating a high-risk scenario.

- **Financial Driver: Home & Comfort** (Bed, Bath, Table) is the top category by **Total Sales** and generates the highest **Total Profit After Returns**.
- **Reputation Liability:** This same **Home & Comfort** category generates the highest volume of absolute returns 999 and is the category with the **Most Product with 1 Star** reviews.
- **Seller Accountability:** **Seller Fulfillment Time** (3.21 days) is too slow and directly contributes to the total delivery time.

Strategic Recommendation

Protect the Core Asset. Launch a mandatory **Quality Assurance Program** targeting sellers in the **Home & Comfort** category. Enforce stricter **Seller Fulfillment Time** SLAs (e.g., target 1-2 days instead of 3.21 days) and conduct audits to reduce the returns and 1-star reviews that are currently eroding net profit.

3. Customer Value & Retention

The Problem

The customer base is large 90K\$ Total Customers, but their engagement is shallow.

- **Shallow Engagement:** The **Average Orders per Customer** is only **1**, meaning the platform is excellent at acquisition but poor at retention.

- **Profit Leakage:** The **Low-Margin Frequent Buyers** segment (high order count, low total sales) consumes excessive operational resources (shipping, handling) relative to the revenue they generate, creating profit leakage.

Strategic Recommendation

Shift Focus to Lifetime Value (LTV). Stop giving broad discounts. Instead, launch a **Tiered Loyalty Program** that incentivizes customers to complete their **second purchase** and encourages low-margin buyers to increase their average order value (AOV) by setting **minimum spend thresholds** for benefits (like free shipping).

13. Overall Recommendations

Priority	Action	Target Goal
1. Logistics & Expansion	Fund Strategic Logistics Investment to break the capacity ceiling. Invest in regional fulfillment hubs in the North/Northeast to cut the -day shipping time and provide guaranteed surge capacity for peak seasons.	Reduce Avg Delivery Days to below days and increase On-Time Delivery % to consistently.
2. Core Asset Protection	Launch a mandatory Quality Assurance Program targeting sellers in the Home & Comfort category. Enforce stricter Seller Fulfillment Time SLAs (target 1-2 days) and conduct audits to reduce the 1-star reviews and returns that threaten the main profit stream.	Reduce returns and 1-star reviews in the Home & Comfort category by .
3. Customer Value	Shift marketing focus from acquisition to Lifetime Value . Implement a Tiered Loyalty Program that requires minimum spend thresholds for benefits, thereby discouraging low-margin frequent buyers and driving the average order count above .	Increase Avg Orders per Customer to .