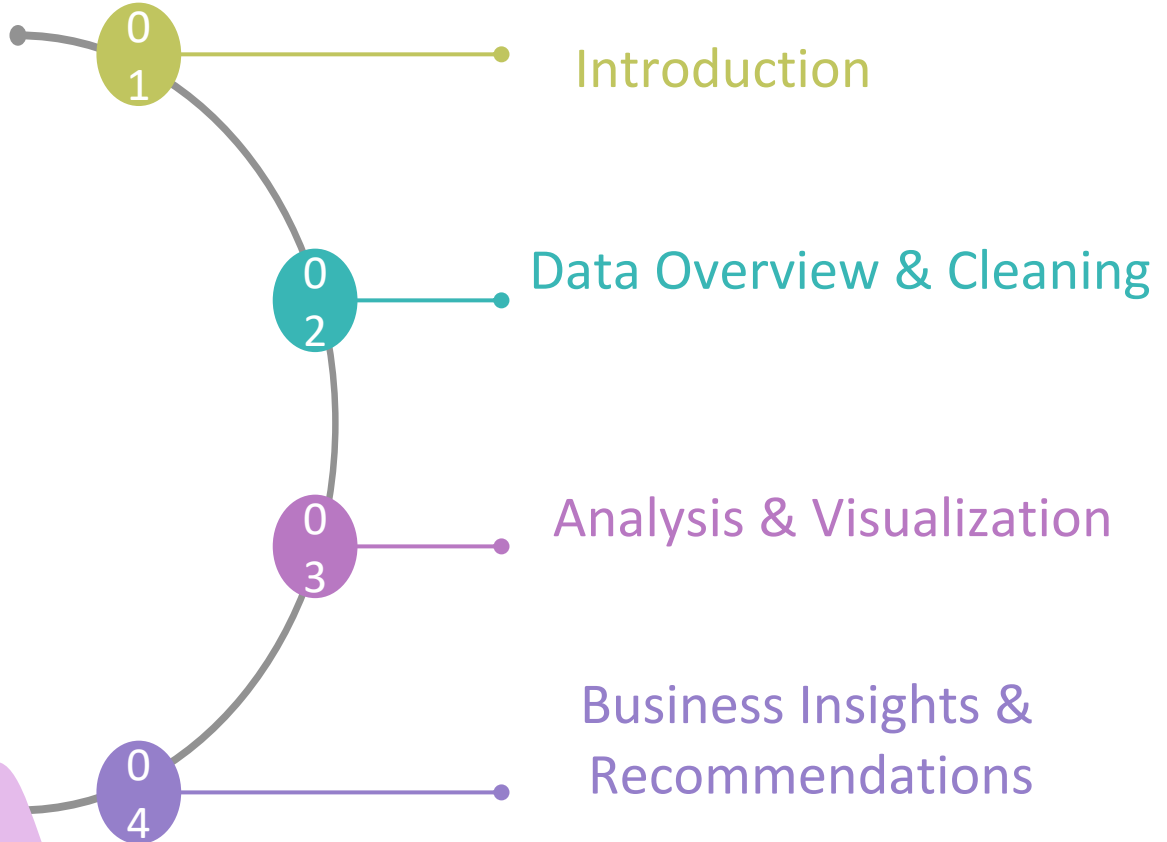
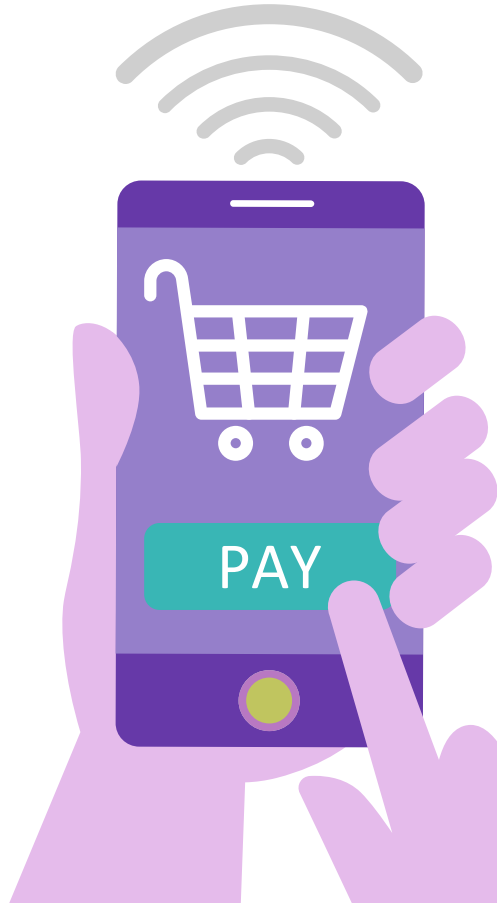


Brazilian E-Commerce

Hagar Abdelsalam



Agenda



Introduction

- The Brazilian E-Commerce Public Dataset by Olist offers a rich and comprehensive collection of data that captures the dynamics of the online retail environment in Brazil. This dataset encompasses detailed information about orders, products, customers, and payments. It includes key attributes such as order statuses, purchase and delivery timestamps, product pricing and shipping costs, customer locations, and various payment details like installment information and payment types.
- By leveraging this dataset, one can uncover trends in revenue, order fulfillment, and customer satisfaction, ultimately leading to actionable business strategies that enhance customer experience and drive revenue growth.

Data Overview

Dataset Source:

Brazilian E-Commerce Public Dataset by Olist

Key Data Attributes:

Order Date:

- *order_purchase_timestamp*

Represents when an order was placed, providing a timeline for sales analysis.

Payment Details:

- *payment_value*

Indicates the monetary value of the payment.

- *payment_type*

Specifies the method of payment (e.g., credit card, boleto, etc.).

Order Status:

- *order_status*

Shows the current state of the order (e.g., approved, canceled, delivered), which is crucial for performance and reconciliation.

Data Overview

Key Data Attributes

Product Prices:

- price*

The cost of the product.

- freight_value*

The shipping cost associated with the order.

Data Integration:

All information is merged using a unique identifier (`order_id`), enabling a comprehensive view of each transaction.

Analysis Opportunities:

- Evaluate sales performance over time.
- Analyze payment behavior and accuracy.
- Compare product pricing and shipping costs.
- Assess order fulfillment and customer satisfaction based on order status.

Tools Utilized:

Excel for data extraction and merging, Python (Pandas, Matplotlib, Seaborn) for analysis, and Power BI for interactive visualization.

First step before Cleaning is Integrate Data : I used Excel for data extraction and merging :

2-Merge the sheets in one sheet

4-Remove Errors

[illegible]

Cleaning

Data Cleaning :is the process of identifying and correcting errors in a dataset to improve its quality and reliability.

Data Issues :

- 1- Missing Data columns (20 Columns)
- 2- Duplicate Data : Total duplicate rows (27)
- 3- Consistency : Order_status have not valid values
- 4- Data Type & Format : there are some columns need to change it's type and format

Cleaning Steps

1. Fix Data Types

2. Data Type Conversion

3. Handling Missing Values

4. Duplicate Removal

5. Outlier Detection and Treatment

6. Date Consistency and Formatting

Cleaning Steps

Fix Data Types:

- 1-Replacing 'unavailable' with the most common value (Mode).
- 2-Replacing 'not_defined' with the most common value (Mode).
- 3-Replacing 'boleto' with it's English Translation (ticket).
- 4-Replacing 'customer_state'[cut] with it's Full-name .
- 5-Replacing 'seller_state'[cut] with it's Full-name .

Data Type Conversion :

- 1.Payment_installments (From Float to Int.)
- 2.Split columns Like [order_purchase_timestamp, order_delivered_customer_date, order_estimated_delivery_date , review_answer_timestamp] to separate columns date and time so we can convert it
- 3.Drop the columns like 'order_purchase_timestamp' column because we didn't need it anymore

Cleaning Steps

Data Type Conversion :

4. order_purchase_date (From Object to datetime.)
5. order_delivered_customer_date (From Object to datetime.)
6. Order_estimated_delivery_date (From Object to datetime.)
7. Shipping_limit_date (From Object to datetime.)
8. Review_creation_date (From Object to datetime.)
9. Review_answer_date (From Object to datetime.)

Cleaning Steps

Handling Missing Values :

1-Dropping Rows with Missing Critical Data like:

Rows with missing `product_weight_g` were removed, as weight is crucial for shipping and logistics analysis.

Similarly, rows missing `payment_value` or other critical numeric columns were dropped.

2-Handling Missing Categorical Data like :

In the `Product_category_name_english` column, missing values were replaced with a default value ("Unknown") to ensure that every record had a defined category.

For `order_delivered_customer_date` and `order_delivered_customer_time`, missing values were either dropped or filled with a placeholder (e.g., "Not Delivered"), based on the relevance to the analysis.

Cleaning Steps

Duplicate Removal:

- 1-Remove duplicate rows from the DataFrame
- 2-Ensure that we remove Duplicate rows

Outlier Detection and Treatment : Nothing to do

Date Consistency and Formatting :

- 1.Order_delivered_customer_date & Order_delivered_customer_time
- 2.Order_estimated_delivery_date & Order_estimated_delivery_time
- 3.Shipping_limit_date & Shipping_limit_time
- 4.Review_creation_date & Review_creation_time
- 5.Review_answer_date & Review_answer_time

Cleaning Steps

Conclusion

The cleaning process addressed key data quality issues, including:

- 1-Splitting and reformatting date/time fields.
- 2-Converting data types and handling missing values appropriately.
- 3-Detecting and removing duplicates and outliers.
- 4-Standardizing categorical data for consistency.

Data Analysis

Duplicate Removal:

- 1-Remove duplicate rows from the DataFrame
- 2-Ensure that we remove Duplicate rows

Outlier Detection and Treatment : Nothing to do

Date Consistency and Formatting :

- 1.Order_delivered_customer_date & Order_delivered_customer_time
- 2.Order_estimated_delivery_date & Order_estimated_delivery_time
- 3.Shipping_limit_date & Shipping_limit_time
- 4.Review_creation_date & Review_creation_time
- 5.Review_answer_date & Review_answer_time

Data Analysis

The primary goals were to evaluate **sales performance**, **payment accuracy**, **profitability**, **growth metrics**, **product-level insights**, **seller performance**, **cities analysis**, and potential **forecasting** opportunities. By examining key metrics such as total revenue, top-selling products, top-performing cities, and delivery times, we gain insights into overall business performance and areas for optimization.

This report outlines key analyses performed on an e-commerce dataset, focusing on:

1-Sales Performance

2-Payment Accuracy (previously covered)

3-Proftability Analysis

4-Revenue Forecasting

5-Payment Analysis

6-Cities Analysis

7-Seller Performance Analysis

8-Product Level Analysis

9-Time Analysis (days with highest sales)

Data Analysis

Sales Performance

1-Total Revenue :

- Result:** represents the sum of product price plus freight cost for all delivered orders.
- Insight:** Reflects actual income from fulfilled orders.

2-Expected Revenue:

- Result:** Potential or recognized income from orders likely to be fulfilled.
- Insight:** Useful for short-term revenue forecasting.

3-Canceled Orders :

- Result:** Total count of canceled orders.
- Insight:** A high cancellation rate may indicate issues in payment, stock, or customer satisfaction.

Data Analysis

Payment Accuracy

1-Expected Installment Amount:

- Result:** [expected installment amount] shows the theoretical cost per installment.
- Insight:** Large discrepancies may indicate data entry errors or billing issues.

2-Payment Success Rate:

Goal: Evaluate how different payment types perform in terms of order completion or “success” (e.g., delivered, shipped, invoiced).

- Insight :**Shows which payment methods have the highest completion rate.
- A lower success rate may require deeper investigation into payment gateway issues or customer preferences.

Data Analysis

Payment Analysis

1-Average Payment:

Insight: Provides a simple overview of how much customers pay on average (including price + freight).

2-Payment Type Distribution :

- Insight:** Shows how many orders each payment method handles.

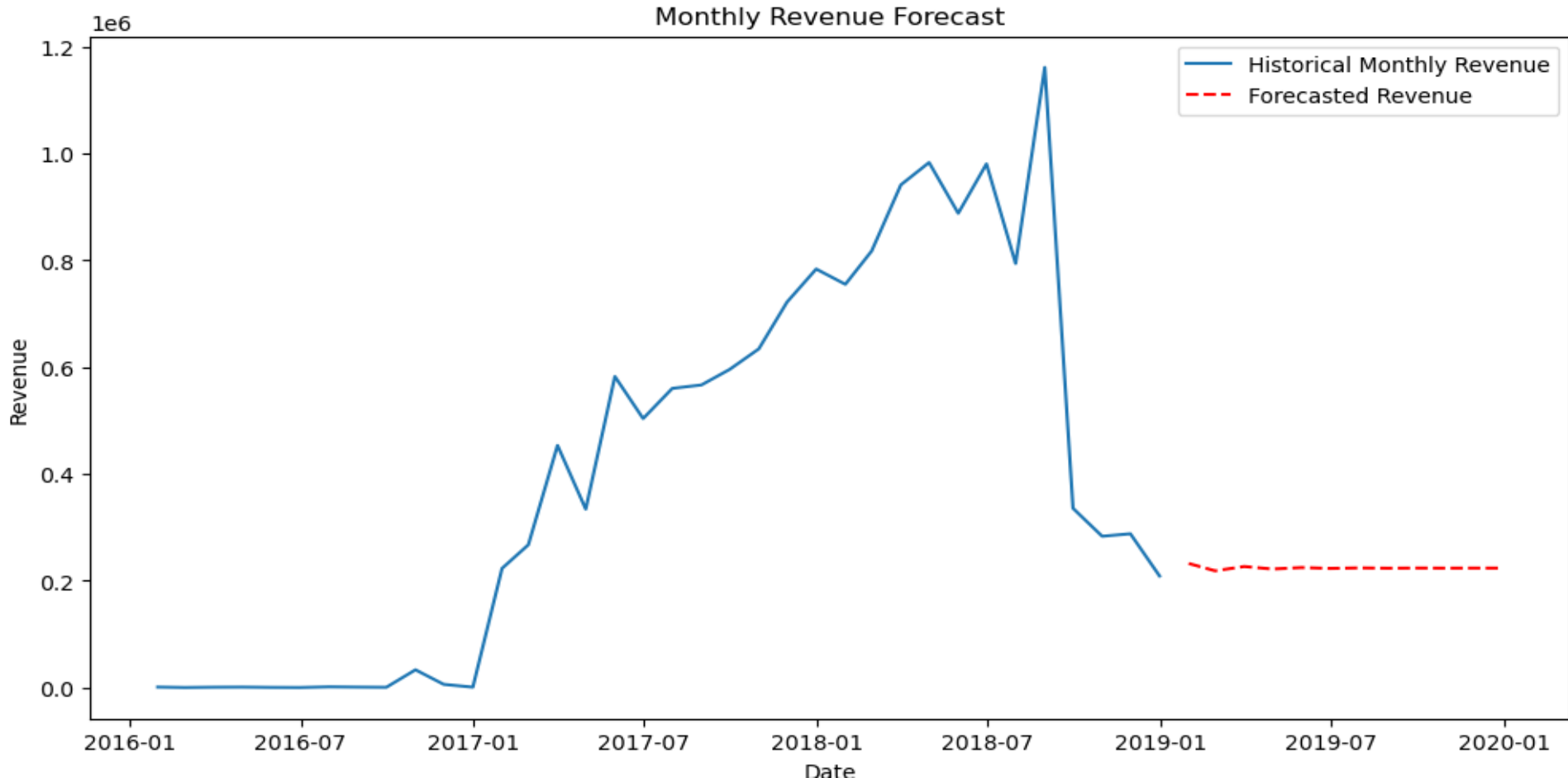
- Use Case:** Identify popular payment methods and underused ones that may need improvement or better promotion.

3-Installment Value Analysis :

Insight: Ensures billing matches the number of installments. Large discrepancies could indicate data-entry errors.

Data Analysis

Revenue Forecasting



Data Analysis

Revenue Forecasting

Goal of the Monthly Revenue Forecast :

1. Predict Future Revenue
2. Identify Trends and Seasonality

Insights :

1. Historical Peak and Decline:

1. The historical line shows a clear rise in revenue up to a **peak** at around mid/late 2017, followed by a **notable drop**.
2. This shift might indicate a one-time event (e.g., a major promotion or market expansion) or the end of a seasonal trend.

2. Forecasted Downward/Stable Trend:

1. The red dotted line suggests **flat or slightly decreasing** revenue in the near term.
2. If the forecast is accurate, the business should **prepare for lower sales** or plan interventions (e.g., promotions, product launches) to boost revenue.

Data Analysis

Revenue Forecasting

Insights :

1.Planning Implications:

- 1.Inventory: Avoid overstocking if demand is projected to remain low or decline.
- 2.Marketing & Promotions: Consider more aggressive campaigns or new partnerships to counteract a downward trend.
- 3.Cost Management: If revenue is likely to stay low, focusing on cost efficiencies and profit margins becomes even more critical.

2.Validation & Adjustment:

- 1.Forecasts are projections, not guarantees. External factors (economic shifts, competitor moves, etc.) can alter real outcomes.
- 2.Revisit the forecast regularly and incorporate new data to refine predictions.

Data Analysis

Geographical Insights

1-Order Distribution by City :

- Insight:** Shows how many orders come from each city.
- Use Case:** Focus marketing or shipping improvements on high-volume regions.

2-Top Cities by Sales :

- Insight:** Identifies which cities generate the most revenue.
- Use Case:** Allocate resources, advertising, or inventory based on high-sales cities.

Data Analysis

Geographical Insights

1-Order Distribution by City :

- Insight:** Shows how many orders come from each city.
- Use Case:** Focus marketing or shipping improvements on high-volume regions.

2-Top Cities by Sales :

- Insight:** Identifies which cities generate the most revenue.
- Use Case:** Allocate resources, advertising, or inventory based on high-sales cities.

Data Analysis

Seller Performance Analysis

Average Order Value by Seller :

Goal: Determine the average revenue per order for each seller to identify top-performing sellers or those needing support.

Insight:

Sellers with a higher AOV might be selling premium products or effectively upselling. Those with low AOV might benefit from bundling strategies or marketing support.

Data Analysis

Product-Level Analysis

1-Revenue by Product :

Goal: Identify which products generate the most total revenue.

Insight : Reveals top-grossing products. Helps with inventory planning and targeted marketing of high-revenue SKUs.

2-Average Order Value (AOV) per Product :

Goal: Determine the average amount customers spend per product. This can differ from total revenue if some products sell fewer units at a higher price.

Insight : Helps identify premium-priced items (high AOV) vs. lower-priced or high-volume items.

Data Analysis

Product-Level Analysis

3-Return or Cancellation Rate per Product :

Goal: Calculate how often each product gets canceled or returned, indicating potential quality or stock issues.

•**Insight :**

1-High cancellation/return rates could mean poor product descriptions, quality issues, or mismatched expectations.

2-Products with low cancellation rates are generally well-received and reliable.

Data Analysis

Time Analysis

Which Days Have the Highest Sales :

Goal: Identify daily sales patterns (e.g., weekends vs. weekdays) to optimize promotions, staffing, and inventory.

Insight :

- 1-Identifies peak days (e.g., Saturdays might have the highest sales).
- 2-Allows you to align marketing campaigns and ensure adequate staffing on peak days.

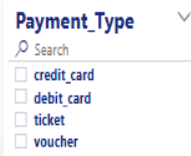
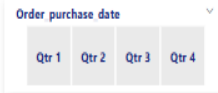
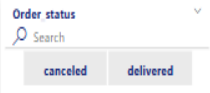
Visualization

olist

Monthly Financial

Customer

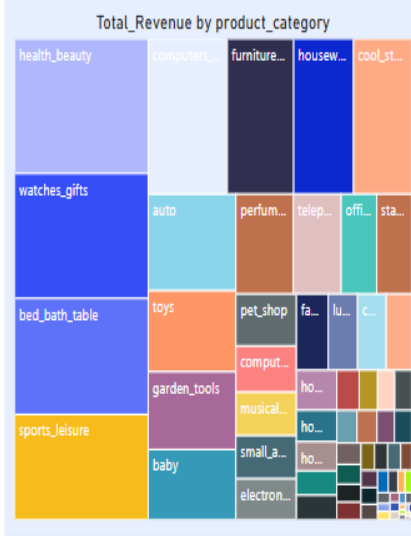
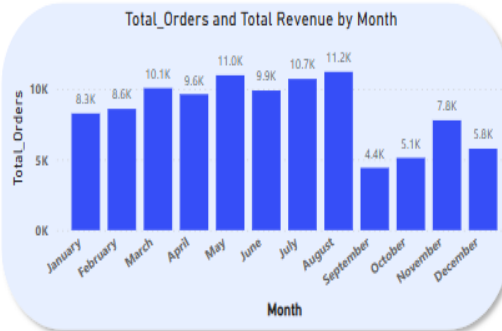
Seller Performance



1bn
Total Revenue

102K
Total Orders

14.36K
Avg_Order_Value

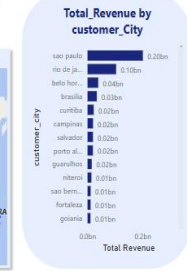
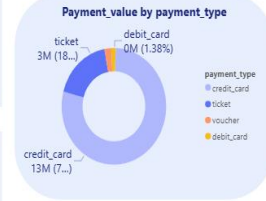


olist

Monthly Financial

Customer

Seller Performance

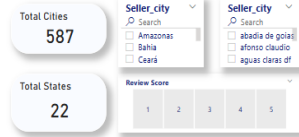


olist

Monthly Financial

Customer

Seller Performance



Visualization



KPI Cards (1bn Total Revenue, 102k Orders, 143.6k ...)

Key Analytical Findings & Trends

- **1bn Total Revenue:** Indicates a significant overall sales volume, suggesting the business has reached a substantial market presence.
- **102k Orders:** Reflects the total count of transactions. A large number of orders can mean a broad customer base or a high purchase frequency.
- **14.36k (e.g., SKUs or Unique Customers):** If this number represents **unique customers**, it highlights your reach and potential for loyalty programs. If it represents **SKUs**, it indicates product variety.

Visualization



KPI Cards (1bn Total Revenue, 102k Orders, 143.6k ...)

Business/Strategic Insights

- The scale of revenue and orders underscores the **importance of efficient logistics, inventory management, and customer service** to handle volume at this level.
- If 14.36k is the **number of unique customers**, consider **customer retention strategies** (e.g., loyalty programs, personalized offers). If it's **SKUs**, evaluate whether the product assortment is optimal or if there's an opportunity to **streamline or expand** certain categories.

Visualization

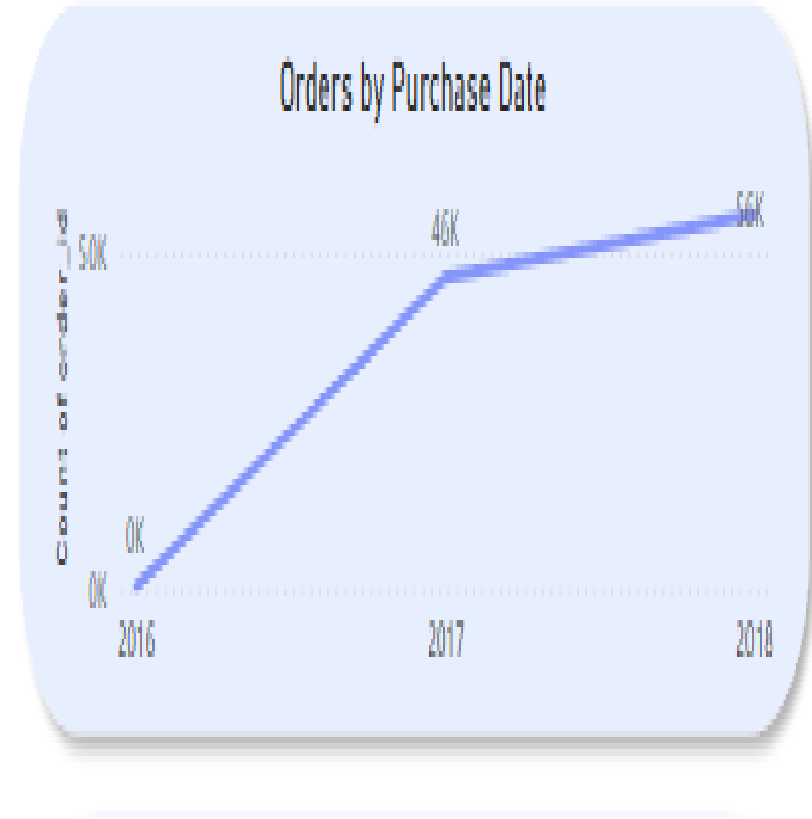
Orders by Purchase Date (Line Chart)

Key Analytical Findings & Trends

- Shows how **order volume changes over time** (daily, weekly, or monthly).
- You may see **seasonal peaks** (e.g., holidays) or **steady growth/decline** periods.

Business/Strategic Insights

- **Seasonality:** If there's a clear spike (e.g., around November–December), you can **plan promotions and inventory** well in advance.
- **Trend Analysis:** A sustained upward trend suggests strong brand growth, whereas a plateau or dip may signal **market saturation** or the need for new marketing initiatives.



Visualization

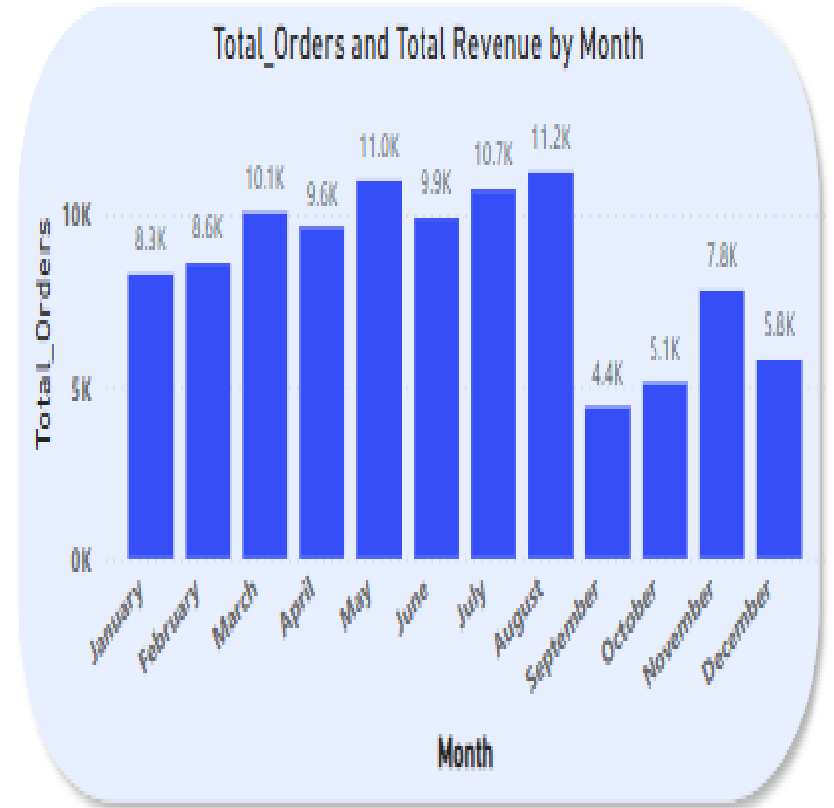
Total Orders and Total Revenue by Month (Bar Chart)

Key Analytical Findings & Trends

- Compares **monthly order counts** to **monthly revenue**.
- Highlights **average order value** differences if revenue and orders don't rise at the same pace.

Business/Strategic Insights

- **Identify High-Revenue Months:** Pinpoint months with high revenue to **replicate successful campaigns** or product launches in future periods.
- **Spot AOV Changes:** If a month has relatively fewer orders but higher revenue, investigate **premium product sales**, successful upsells, or special promotions.



Visualization

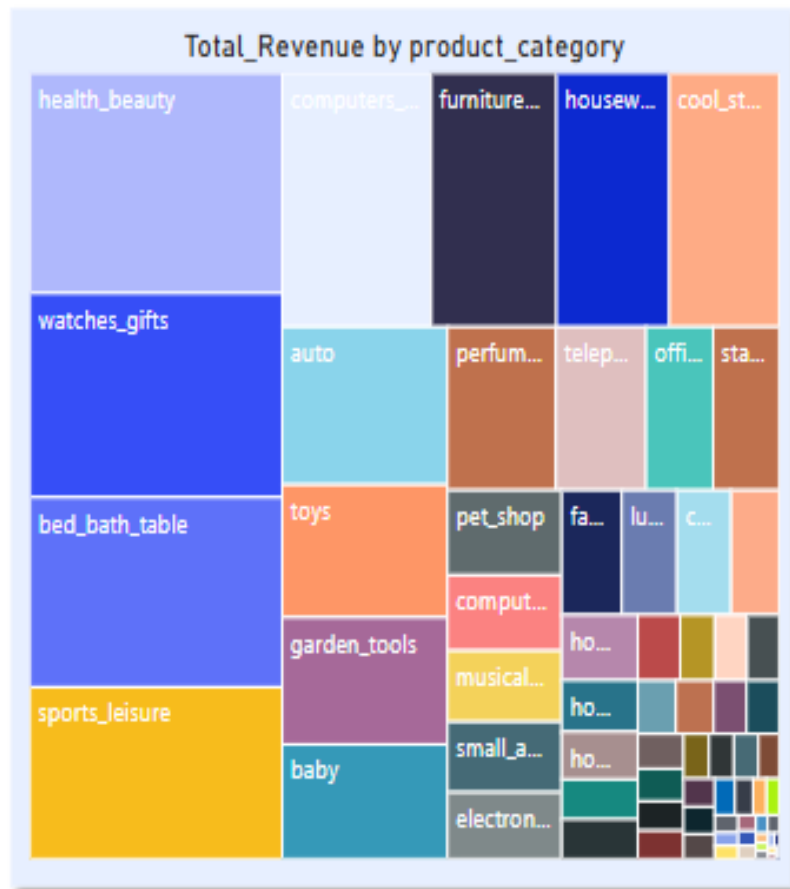
Product Category (Treemap)

Key Analytical Findings & Trends

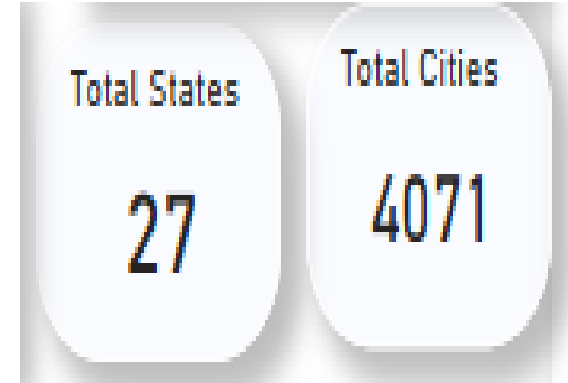
- Displays **revenue contribution** by category.
- Reveals **top-performing categories** (largest boxes) vs. niche categories (smaller boxes).

Business/Strategic Insights

- **Focus on Best Sellers:** Invest more in top categories through **marketing, inventory**, or product development.
- **Diversify:** If one category dominates, there's a risk concentration. Consider expanding or strengthening other categories to **reduce reliance** on a single segment.



Visualization



KPI Cards (e.g., “No. of States: 27”, “No. of Sellers”, etc.)

Key Analytical Findings & Trends

- **No. of States:** Reflects the geographic coverage of your business.
- **No. of Sellers:** Shows the breadth of your seller network on the platform.
- Other KPI cards (e.g., “No. of Products,” “No. of Categories,” etc.) indicate **platform scale** and **variety**.

Business/Strategic Insights

- **Geographic Reach:** If you cover 27 states, you have near-national coverage. Evaluate logistics and shipping partners to ensure consistent service.
- **Seller Management:** A high number of sellers requires robust onboarding, quality control, and dispute resolution processes.
- **Variety vs. Complexity:** More categories or products can attract diverse customers but may increase operational complexity (inventory management, marketing segmentation).

Visualization

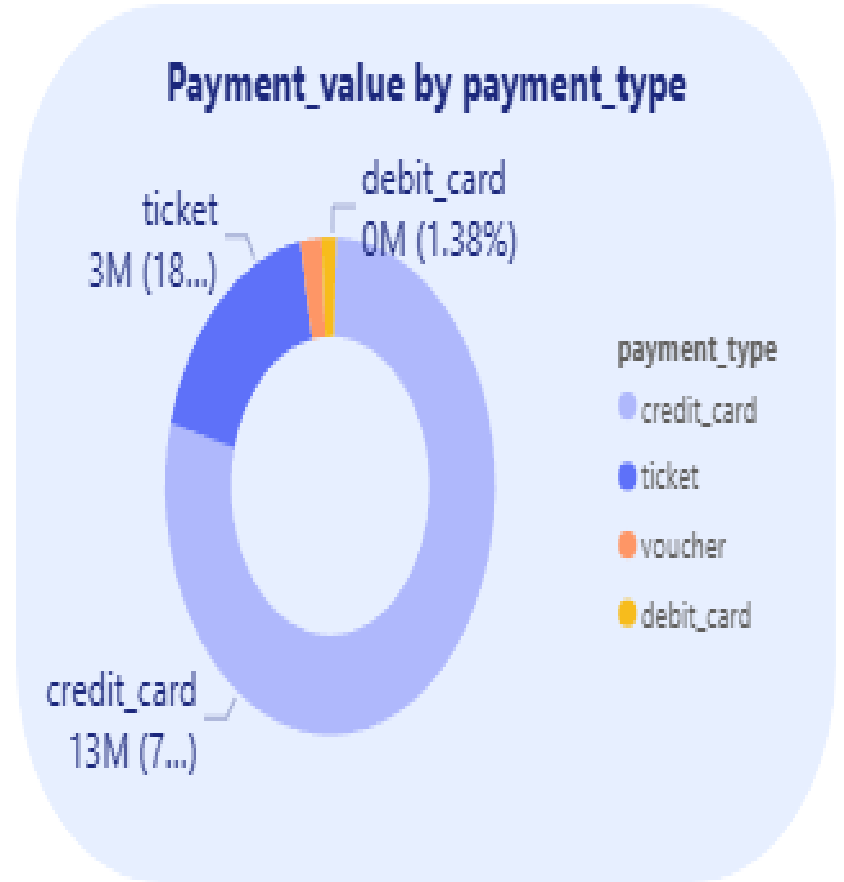
Payment Value by Payment Type (Donut Chart)

Key Analytical Findings & Trends

- Shows the **distribution of total payment value** across different payment methods (e.g., credit_card, boleto, voucher, etc.).
- Identifies which payment types **contribute most to revenue** (largest slice) vs. those that are less used (smaller slices).

Business/Strategic Insights

- Focus on Top Payment Methods:** Ensure the most popular method(s) offer a seamless checkout. A poor user experience here risks lost sales.
- Opportunity for Growth:** If an alternative payment type is growing, you could promote it to capture a niche market or reduce payment friction.
- Risk Diversification:** Over-reliance on a single payment method can be risky if that gateway has issues. Consider maintaining multiple options.



Visualization

Total Orders by Customer City (Map Visualization)

Key Analytical Findings & Trends

- Highlights where orders originate, showing **city-level distribution** on a map.
- Large circles or darker regions indicate **high order volumes**.

Business/Strategic Insights

- **Regional Demand:** Pinpoint high-demand cities to allocate resources (marketing, local inventory, shipping partnerships).
- **Expansion Strategy:** If certain regions are underrepresented, you might run targeted promotions or improve shipping options to boost orders there.
- **Logistics Optimization:** High-density areas may benefit from local fulfillment centers or same-day delivery options.



Visualization

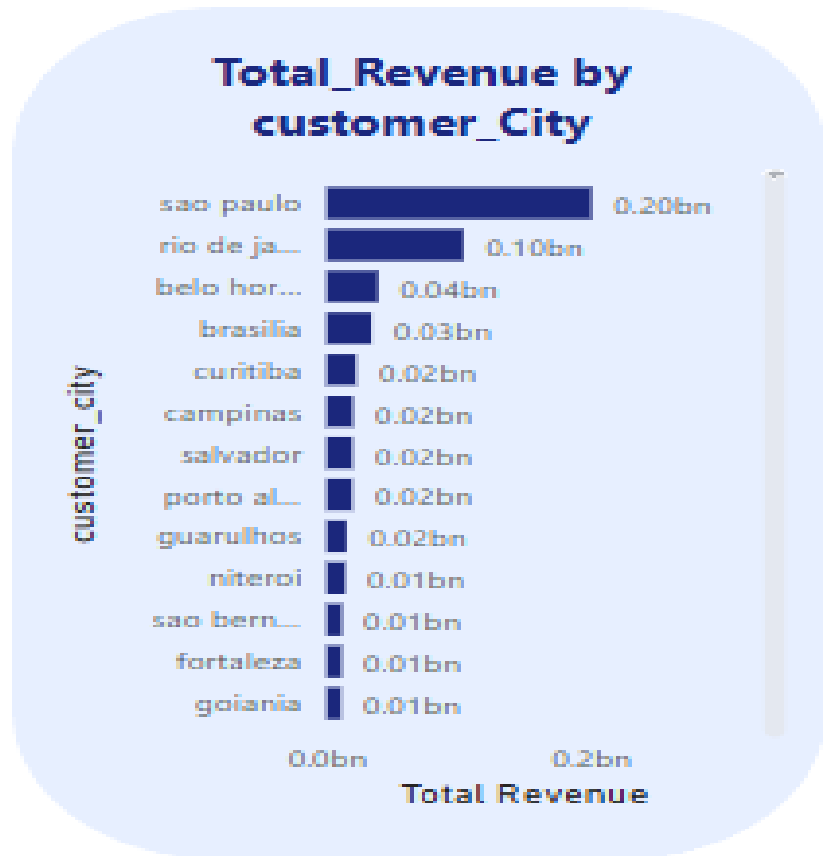
Total Revenue by Customer City (Bar Chart)

Key Analytical Findings & Trends

- Ranks cities by **total revenue** rather than order count.
- Highlights potential differences in **average order value** across cities.

Business/Strategic Insights

- **High-Value Markets:** Cities generating significant revenue might justify more marketing spend or premium services.
- **Compare Order Count vs. Revenue:** Some cities may have fewer orders but higher revenue (premium customers). Tailor your approach accordingly (e.g., specialized offerings or loyalty programs).
- **Regional Promotions:** If certain cities have high revenue potential, local events or influencer partnerships can further boost sales.



Visualization

Total Order Amount by Customer State (Bar Chart)

Key Analytical Findings & Trends

- Shows **which states** contribute the most to your total order amount.
- Identifies any **regional concentration** in sales.

Business/Strategic Insights

- **Focus on Top States:** Invest in state-specific campaigns or partnerships where you already see strong sales.
- **Explore Underserved States:** If a state has high population but low sales, investigate potential barriers (shipping, payment methods, marketing reach).
- **Scalability:** High sales in multiple states require robust logistics to ensure timely delivery and customer satisfaction.



Visualization

KPI Cards: Total Cities (587) & Total States (22) Key

Analytical Findings & Trends

Geographical Reach: Indicates the breadth of the platform's coverage (number of unique cities and states).

- Shows a **significant footprint**, suggesting the business operates nationwide (or across multiple regions).

Business or Strategic Insights

- **Logistics and Delivery:** With 22 states involved, ensure shipping partnerships and fulfillment centers are positioned for efficient coverage.
- **Localized Marketing:** High city/state coverage may warrant **regional campaigns** or targeted promotions (e.g., city-specific deals).
- **Growth Potential:** If some states or cities are underrepresented, there's an opportunity to expand seller recruitment or marketing efforts there.

Total Cities
587

Total States
22

Visualization

Freight Value by Seller_ID (Horizontal Bar Chart)

Key Analytical Findings & Trends

- Displays how **freight costs** vary across different sellers.
- Identifies which sellers incur **higher or lower freight charges**, possibly indicating differences in shipping distance, product size/weight, or negotiated courier rates.

Business or Strategic Insights

- **Shipping Cost Optimization:** Sellers with **consistently high freight costs** may need better logistics partners or cost-sharing strategies to remain competitive.
- **Competitive Positioning:** If certain sellers offer lower freight costs, they can attract more price-sensitive customers. Encourage best practices or shared shipping solutions across sellers.



Visualization

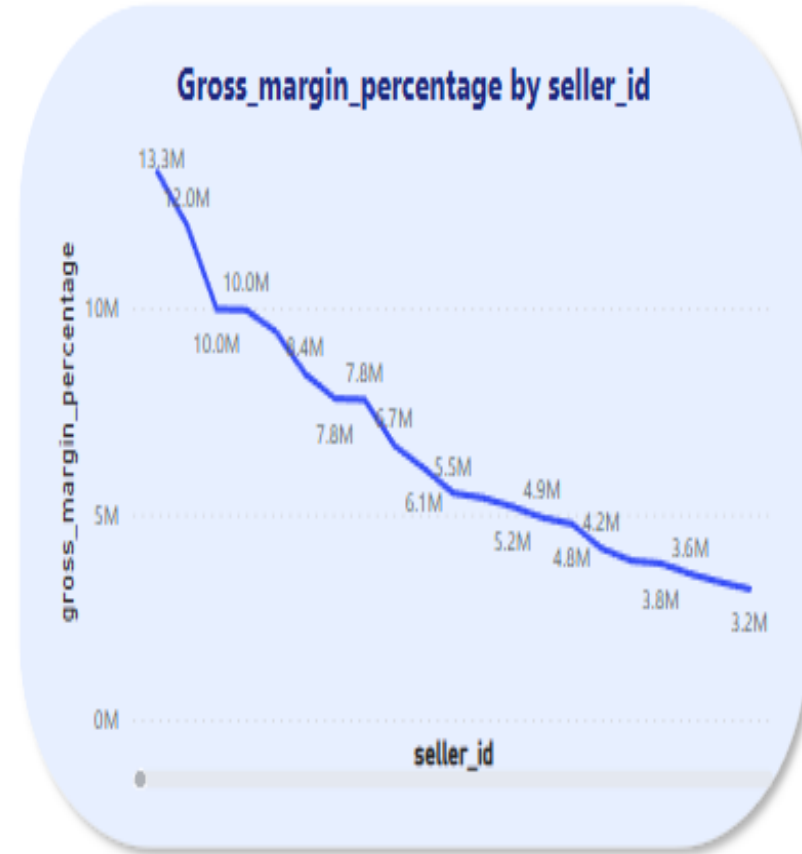
Gross_Margin_Percentage by Seller_ID (Line Chart)

Key Analytical Findings & Trends

- Illustrates **profitability** differences among sellers by showing each seller's gross margin percentage.
- Some sellers may have **high margins**, indicating efficient cost structures or premium pricing, while others have **lower margins**, suggesting narrower profits.

Business or Strategic Insights

- **Profitability Focus:** High-margin sellers might be role models—study their cost management or pricing strategies to share insights with lower-margin sellers.
- **Contract or Commission Adjustments:** If certain sellers have unsustainably low margins, renegotiate terms or provide resources (e.g., marketing support, bulk shipping rates) to improve viability.
- **Product Mix Strategy:** Encourage underperforming sellers to diversify into higher-margin product categories if feasible.



Visualization

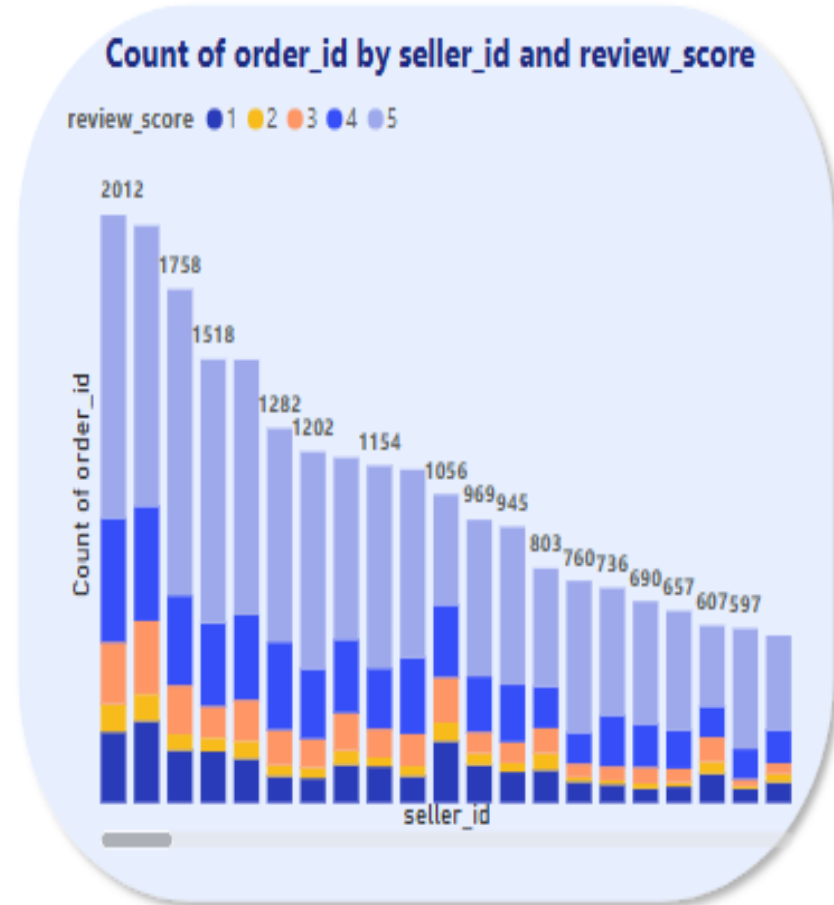
Count of Orders by Seller_ID and Review_Score (Stacked Bar Chart)

Key Analytical Findings & Trends

- Shows **order volume** per seller, broken down by **review scores** (e.g., 1–5 stars).
- Reveals which sellers generate **high order counts** and how satisfied customers are with those orders.

Business or Strategic Insights

- Quality vs. Volume:** A seller with large order volumes but predominantly low review scores needs immediate quality improvements or better customer service.
- Seller Benchmarking:** Sellers consistently getting high scores can be studied for best practices (packaging, product quality, shipping speed).
- Targeted Interventions:** Offer training or stricter guidelines to sellers with subpar reviews, potentially raising overall platform reputation.



Business Insights & Recommendations

Business Insights

1. Strong Overall Sales Volume

1. High total revenue and large order counts indicate a robust market presence.
2. Many sellers and products suggest a diverse marketplace but also operational complexity.

2. Payment Method Dominance & Diversity

1. One or two payment types (e.g., credit card) often dominate.
2. Alternative payment methods may be gaining traction but remain smaller in share.

3. Product Category Concentration

1. Certain categories consistently generate higher revenue.
2. Niche or emerging categories might show growth potential, even if current volumes are smaller.

Business Insights & Recommendations

Business Insights

4 .Seller Performance Variability

1. Some sellers exhibit high freight costs or low margins, while others maintain healthier margins and lower shipping costs.
2. High-volume sellers can have subpar customer reviews, risking brand reputation.
3. Best-performing sellers (in margin and review scores) provide valuable benchmarks

5 .Customer Satisfaction Indicators

1. Stacked bar charts of order volume vs. review scores show that a subset of sellers or products drives negative feedback.
2. Return or cancellation rates reveal quality, shipping, or product-description issues.

6. Geographical Reach & Trends

1. Large coverage across multiple states/cities suggests broad demand but also challenges in logistics and localized marketing.
2. Certain cities/states yield higher revenue or order counts, indicating potential for deeper market penetration.

Business Insights & Recommendations

Business Insights

7. Time-Based Patterns

1. Sales often spike during certain months or days (e.g., weekends, holidays).
2. Seasonality can be leveraged for targeted campaigns and inventory planning.

8. Freight Cost & Gross Margin

1. Freight costs can significantly affect final margins, varying widely by seller or region.
2. Gross margin disparities suggest differences in cost control, product mix, or shipping efficiency.

9. Revenue Forecasting

1. Forecast models may show flat or slightly declining trends, indicating the need for interventions to sustain or increase revenue.
2. Historical spikes might correlate with major campaigns or holiday seasons.