Olist E-Commerce Performance

Purpose of the Study

This report distils more than 100 000 orders, 50 000 customers, and 112 product categories from the Olist Brazilian e-commerce dataset into a concise, decision-ready narrative. The goal is to show where the marketplace is winning, where it is leaking value, and how those insights should inform merchandising, logistics, finance, and customer-experience strategies for the coming year.

Data and Methodology

Kaggle's Olist relational database, January 2017 – August 2018 sales history.

Normalised duplicate category labels and built a time-series calendar.

Created analytical measures in DAX (e.g., On-Time Delivery, Review Buckets).

Proposed by

Zahra Noury – August 2025

Product and Sales

Key Business Questions to Answer

- Which product categories generate the highest total revenue?
- Which categories deliver the highest average order value (AOV)?
- Where do we see the strongest repeat-purchase behaviour?
- How is demand distributed across the days of the week?

Findings and Insights

Most money comes from Home-related goods.

- Furniture, garden, and home decor bring in about R$ 4.3 million, more than any other group.
- If demand here drops, overall revenue takes a big hit.

Some smaller categories have very big baskets.

- Shoppers who buy Arts & Hobbies or Office & Business items spend R$ 260–290 per order, nearly double the site average.
- Even though these groups sell fewer units, each sale is highly profitable.

Electronics & Office items are repeat-purchase magnets.

- These groups generate lots of orders, meaning customers come back often or buy in bulk.
- Repeat buyers are easier to keep than to first time buyer. Nurturing them raises long time value.

Sales dip on weekends; especially Saturday.

- Orders fall 35 % from mid-week highs (Tue–Wed) to Saturday.
- If we raise AOV (average order value) or weekend orders just a little, it moves the top line quickly.

Snapshot

- Totals: R$ 15.42 M revenue, AOV = R$ 153, = 96 K orders

Action plans

Prioritise ad spend, search placement, and inventory depth for Home, Health & Beauty, Electronics.

Introduce premium bundles, "complete-the-set" cross-sells and Buy-Now-Pay-Later in Arts & Hobbies, Office & Business.

Launch loyalty points or "subscribe & save" programmes for Electronics and Office.

Run Friday-evening push notifications plus Saturday free-shipping codes.

Analytical Workflow

The original product category column was inconsistent, with several duplicates and unclear groupings. To resolve this, I created a new cleaned and standardised category column (groupings) to enable accurate product-level analysis.

Additionally, to explore consumer behaviour patterns, I created a new DAX column:

Purchase Weekday = FORMAT(orders[order_purchase_timestamp], "dddd")

This helps identify peak order days across the week.
I also created a state table to correctly link each state to its full name. Initially, the map was confusing the Brazilian state abbreviations with U.S. states. Additionally, since the geolocation table couldn't be directly connected to the customer table due to many-to-many relationships, a secondary table was needed to bridge the geolocation and customer tables.
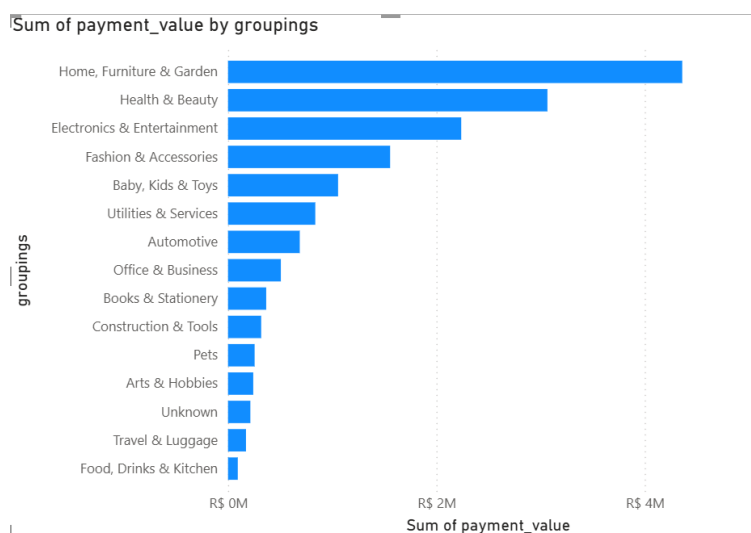

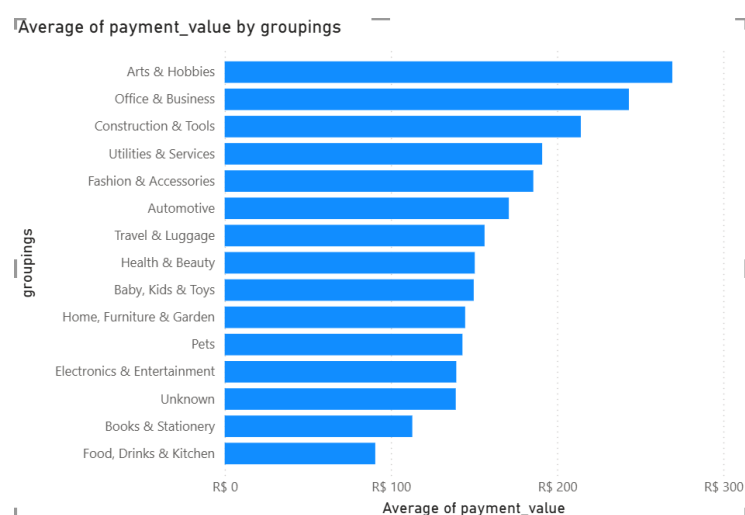*Figure 1 - Total Revenue (Sum of payment_value) by Product Category*

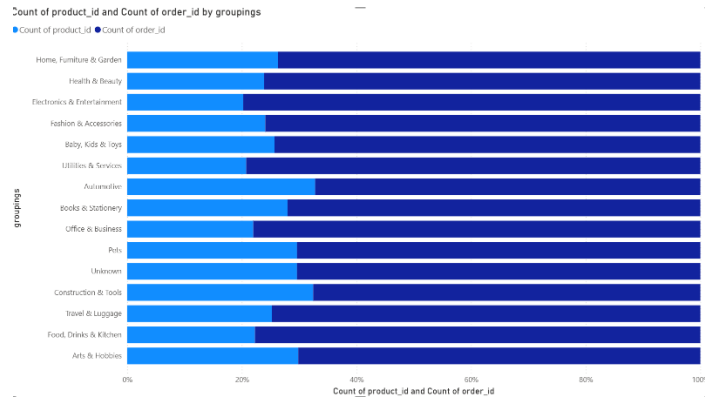*Figure 2 - Average Order Value (Mean payment_value) by Product Category*

Count of product_id and Count of order_id by groupings
● Count of product_id ● Count of order_id

Figure 3 - Relative Share of Order Count vs Count by Product Items

Count of order_id by Purchase Weekday
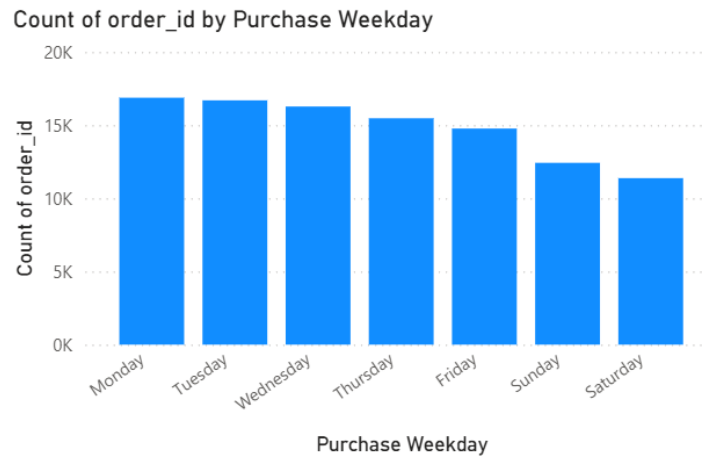
Figure 4 - Order Volume by Day of Week

groupings	Count of product_id	Count of order_id
Home, Furniture & Garden	10156	34556
Health & Beauty	6179	21730
Electronics & Entertainment	3927	17387
Fashion & Accessories	2550	8633
Baby, Kids & Toys	2342	7221
Utilities & Services	1099	4490
Automotive	1900	4235
Books & Stationery	1219	3397
Office & Business	555	2482
Pets	719	1947
Construction & Tools	696	1768
Unknown	610	1603
Food, Drinks & Kitchen	287	1196
Travel & Luggage	349	1092
Arts & Hobbies	363	913
Total	**32951**	**112650**

Figure 5 - Summary Table - Orders and SKUs per Category

Payment

Key Business Questions to Answer

Which payment methods drive the largest share of revenue?

Which methods deliver the highest average ticket?

How is revenue distributed across instillment plans?

What actionable risks or opportunities emerge when we combine the findings above?

Findings and Insights

Total Revenue by Category

- Home, Furniture & Garden is the clear revenue leader (R$ 4.3 M), followed by Health & Beauty and Electronics & Entertainment (R$ 3 M each).
- Confirms that lifestyle-oriented categories dominate top-line sales.

AOV by Category

- Arts & Hobbies commands the highest AOV (R$ 290), with Office & Business and Construction & Tools also above R$ 200.

Snapshot

- Total paid value: R$ 16.01 M, Average instillments on card payments: 3.8

Action plans

Negotiate blended rates with a second acquirer or enable Brazil's instant payment system as a low-fee instant option.

Keep interest-free up to 6; apply graduated interest 7–12

Auto-send QR codes and two reminder emails; cancel unpaid Boletos after 48 h.

Offer 1 % instant cash-back for debit and meal-voucher payments during checkout.

Analytical Workflow
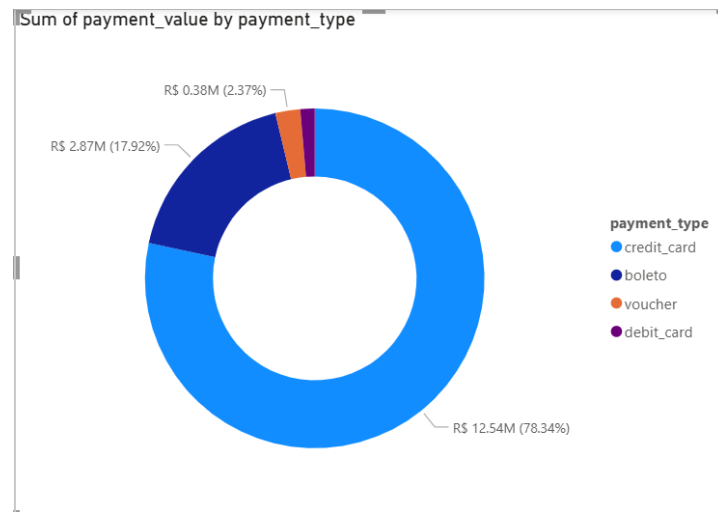

Figure 6 - Total Revenue (Sum payment_value) by Payment Type

payment_type	Sum of payment_value	%GT Sum of payment_value ▼
credit_card	R$ 12,542,084.19	78.34%
boleto	R$ 2,869,361.27	17.92%
voucher	R$ 379,436.87	2.37%
debit_card	R$ 217,989.79	1.36%
Total	**R$ 16,008,872.12**	**100.00%**

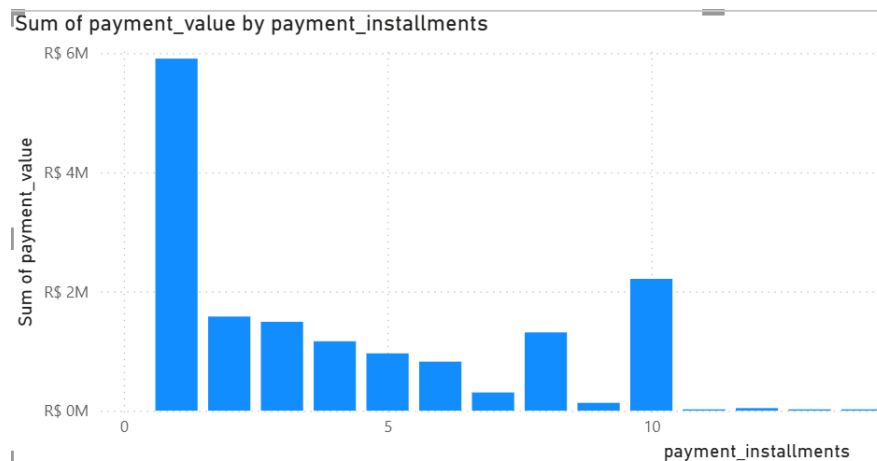*Figure 7 - KPI Table – Payment Value & Method Share*

Figure 8 - Revenue by Number of Instalments

Customers and Geolocation

Key Business Questions to Answer

Which Brazilian states bring in the most revenue and house the largest customer bases?

Where are the highest-spending customers (average order value) located?

How do volume leaders (many customers) differ from value leaders (high AOV) across the map?

What regional opportunities or risks emerge when we combine volume, value, and geography?

Findings and Insights

Revenue and Customer Count by State

- Sao Paulo alone drives 37 % of revenue (R$ 6 M) and 42 K customers; far ahead of any other state. Rio de Janeiro (13 %) and Minas Gerais (12 %) round out the top three.
- High population plus strong logistics make the Southeast the commercial core.

Revenue Share Table

- The top five states account for 73 % of national revenue but only 63 % of customers, indicating slightly higher spend per customer in these hubs.
- Confirms revenue concentration in a handful of urban states.

Average Order Value by State

- Paraiba, Acre, and Rondonia post the highest AOV (R$ 240–250) despite modest order counts. Sao Paulo ranks in the bottom quartile on AOV (R$ 150).
- Smaller, often remote states generate fewer orders but higher-value baskets; possibly due to higher shipping thresholds or product rarity.

Customer Density Map

- Bubbles cluster along the Southeast coast; the North and interior show sparse but widespread demand pockets.

Action plans

Add a micro-fulfilment hub near Sao Paulo to speed delivery and protect share.

Launch targeted digital ads and free-shipping thresholds in Paraiba, Acre, Rondonia.

Expand seller onboarding in the North/Northeast to dilute revenue concentration.

Use geotargeted emails offering regional best-sellers and localised promotions for Minas Gerais & Parana.

Analytical Workflow

state.state_name	Count of customer_id	%GT Sum of payment_value
São Paulo	41746	37.47%
Rio de Janeiro	12852	13.39%
Minas Gerais	11635	11.70%
Rio Grande do Sul	5466	5.57%
Paraná	5045	5.07%
Santa Catarina	3637	3.89%
Bahia	3380	3.85%
Distrito Federal	2140	2.22%
Goiás	2020	2.19%
Espírito Santo	2033	2.04%
Pernambuco	1652	2.03%
Ceará	1336	1.75%
Pará	975	1.36%
Mato Grosso	907	1.17%
Maranhão	747	0.95%
Paraíba	536	0.88%
Mato Grosso do Sul	715	0.86%
Piauí	495	0.68%
Rio Grande do Norte	485	0.64%
Alagoas	413	0.61%
Sergipe	350	0.47%
Tocantins	280	0.38%
Rondônia	253	0.38%
Amazonas	148	0.17%
Acre	81	0.12%
Amapá	68	0.10%
Roraima	46	0.06%
Total	**99441**	**100.00%**

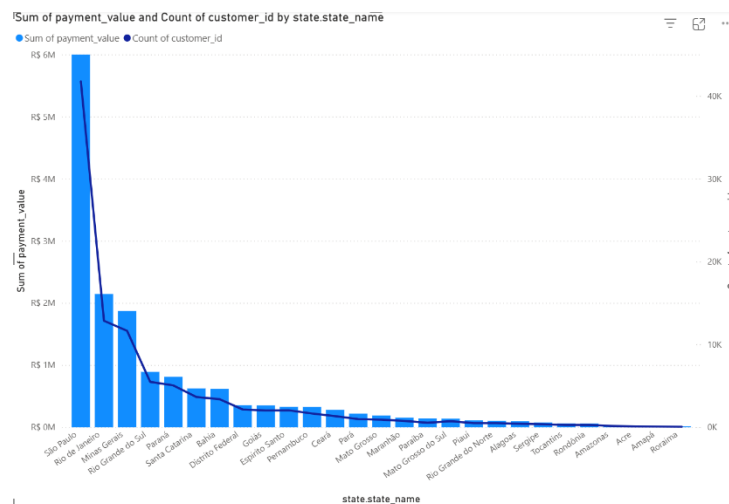*Figure 9 - Customer Count & Revenue Share Table by State*

Figure 10 - Total Revenue (bars) & Customer Count (line) by Brazilian State
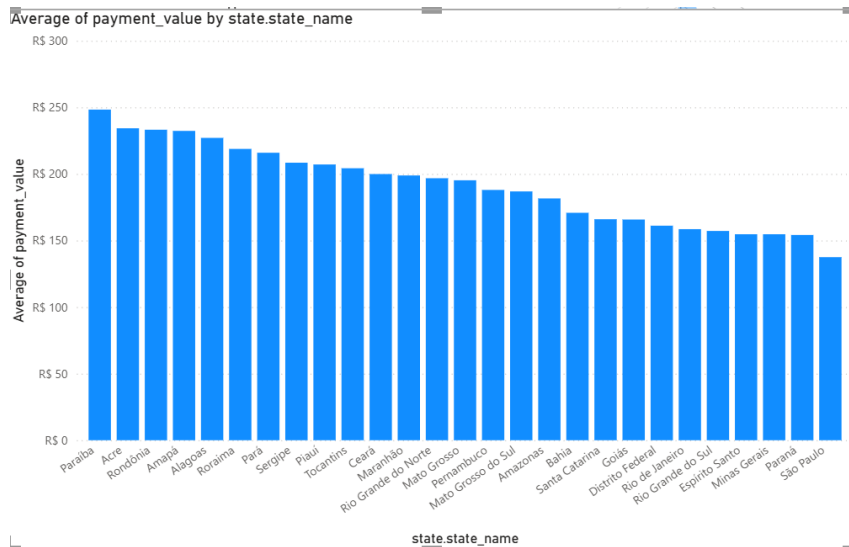
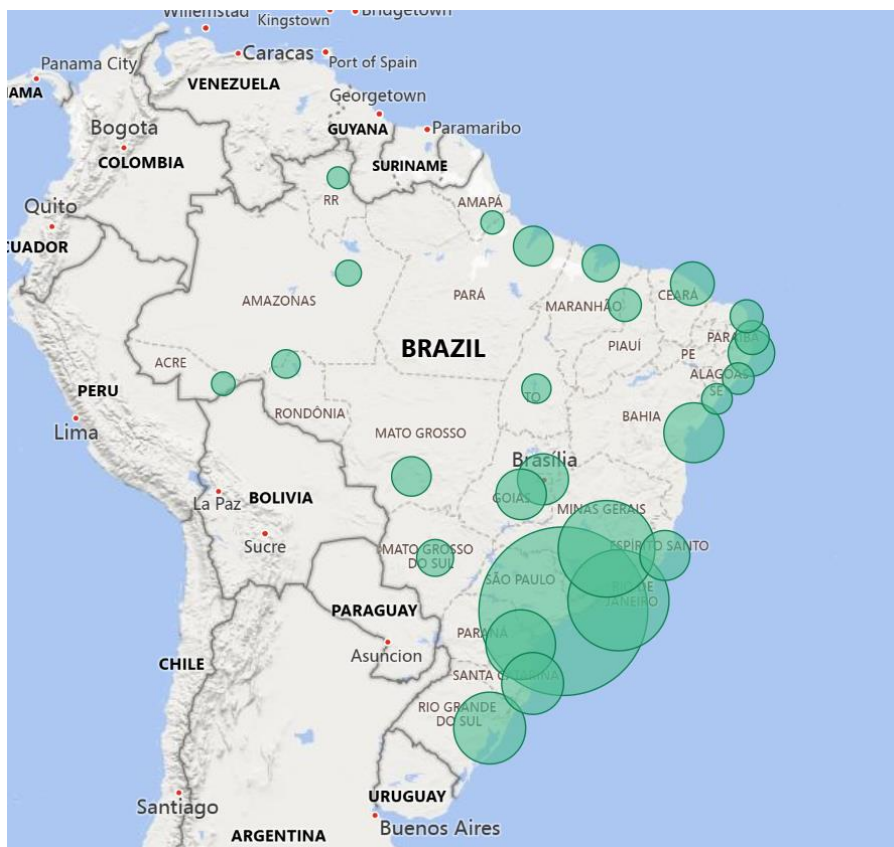*Figure 11 - Average Order Value by State*

Figure 12 - Customer Density Map – Bubble Size Indicates Order Volume

Delivery Performance

Key Business Questions to Answer

What portion of total order value is product price vs. freight cost by category?

Does a higher freight cost deter customers?

What is our overall on-time-delivery (OTD) rate and how is revenue distributed across order statuses?

Which states and product categories experience the largest delivery-time overruns?

Across all orders, how long does the "purchase to delivery" journey take and where are most orders concentrated on that timeline?

Findings and Insights

Price vs Freight by Category

- Freight accounts for 17–20 % of value in top categories (Home, Health & Beauty, Electronics).
- Freight cost is material yet consistent among leaders.

Average Price vs Average Freight

- Arts & Hobbies has the highest freight-to-price ratio, while Food/Drinks absorbs the lowest.
- Bulky or fragile items inflate shipping costs in creative-goods segment.

Freight Cost vs Order Count

- No visible negative slope: order volume is not suppressed by higher freight spend.
- Customers appear inelastic to shipping fees within observed range.

Revenue by Order Status

- 96.3 % of revenue comes from orders in "delivered" status.
- Confirms strong fulfilment execution and low payment leakage.

Delivery Gap by State

- Acre, Rondonia, Amapa show > 18-day overruns; Alagoas beats estimates by 11 days.
- Geography and logistics infrastructure drive large variance.

Delivery Gap by Category

- Travel & Luggage overruns estimate by 12.6 days, versus 10.6 for Food/Drinks.
- Weight, size, and handling complexity delay travel gear most.

Delivery Gap Histogram

- 50.5 % of orders arrive within 6-15 days of the estimate

- Most delays cluster in a manageable 10-day window.

Late-Order Drill-down

- Late deliveries are concentrated in Home items shipped to Sao Paulo metro by a handful of sellers (e.g., seller ID 4a3ca… shows 27 late orders).
- Pinpoints vendor-level issues for targeted remediation.

Snapshot

- Average Days to Approve: 0.52 · To Carrier: 2.71 · Purchase to Delivery: 12.50

Action plans

Partner with regional 3PLs or set cross-dock points in Acre/Rondonia/Amapa to cut transit lag.

Issue late-shipment scorecards and penalties/bonuses for sellers flagged in Sankey.

Offer premium "Same-Week Delivery" (paid upgrade) on Travel & Luggage and other +12-day categories.

Display freight-to-item cost ratio at checkout; highlight when it's below category average.

Analytical Workflow

To quantify delivery performance, I first compared the promised date with the actual arrival date for every order. That single calculation: Days Estimate Gap, tells us whether a parcel was early (negative value) or late (positive value) and powers the state.

On-Time Delivery Rate – the share of orders where the gap is zero or negative.

Late Orders – a flag I reuse in drill-downs to isolate problem sellers, cities or SKUs.

Days between Estimate and actual delivery

```
Days Estimate Gap =
VAR d = orders[order_delivered_customer_date]
VAR e = orders[order_estimated_delivery_date]
RETURN IF (
    NOT ISBLANK(d) && NOT ISBLANK(e),
    DATEDIFF ( e, d, DAY ) * -1)
```

On-time delivery rate

```
On Time Delivery Rate =
DIVIDE (
    CALCULATE (
        COUNTROWS ( orders ),
        orders[order_delivered_customer_date] <= orders[order_estimated_delivery_date]),
    COUNTROWS ( orders ))
```

End-to-end customer wait time

```
Avg Days Purchase to Delivery =
AVERAGEX (
    FILTER (
        orders,
        NOT ISBLANK ( orders[order_delivered_customer_date] )
            && NOT ISBLANK ( orders[order_purchase_timestamp] )),
    DATEDIFF (
        orders[order_purchase_timestamp],
        orders[order_delivered_customer_date],
        DAY))
```

Internal approval latency

```
Avg Days to Approve =
AVERAGEX (
    FILTER (
        orders,
        NOT ISBLANK ( orders[order_approved_at] )
            && NOT ISBLANK ( orders[order_purchase_timestamp] )),
    DATEDIFF (
        orders[order_purchase_timestamp],
        orders[order_approved_at],
        DAY))
```

Warehouse hand-off latency

```
Avg Days to Carrier =
AVERAGEX (
    FILTER (
        orders,
```

```
    NOT ISBLANK ( orders[order_delivered_carrier_date] )
        && NOT ISBLANK ( orders[order_approved_at] )),
  DATEDIFF (
    orders[order_approved_at],
    orders[order_delivered_carrier_date],
    DAY))
```

Unique orders (for scatter & Sankey)

Orders Count = DISTINCTCOUNT (order_items[order_id])

Total freight spend

Total Freight Cost = SUM (order_items[freight_value])

Late-order flag for root-cause analysis

```
Late Orders =
CALCULATE (
  [Orders Count],
  orders[Delivery Status] = "Late")
```
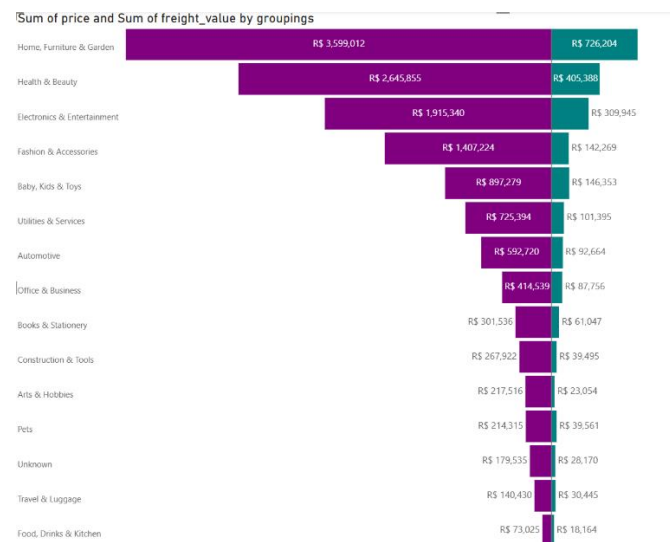

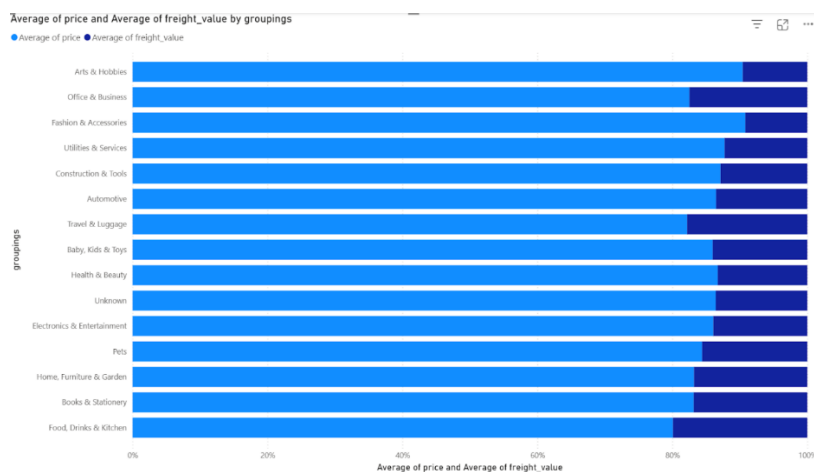*Figure 13 - Total Price vs Freight Value by Category*

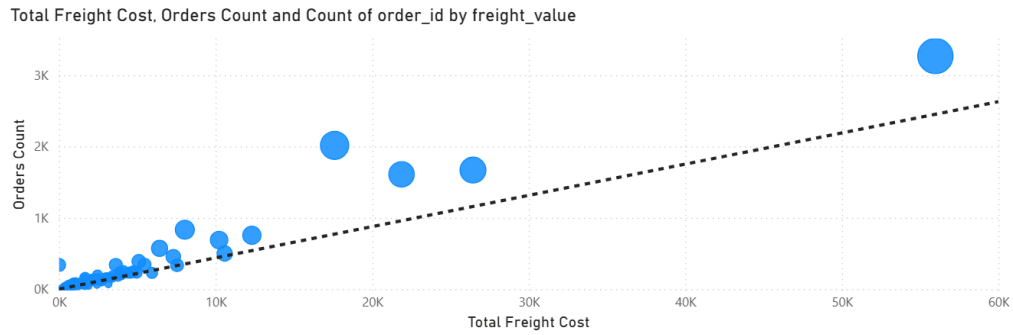*Figure 14 - Average Price vs Average Freight Ratio by Category*
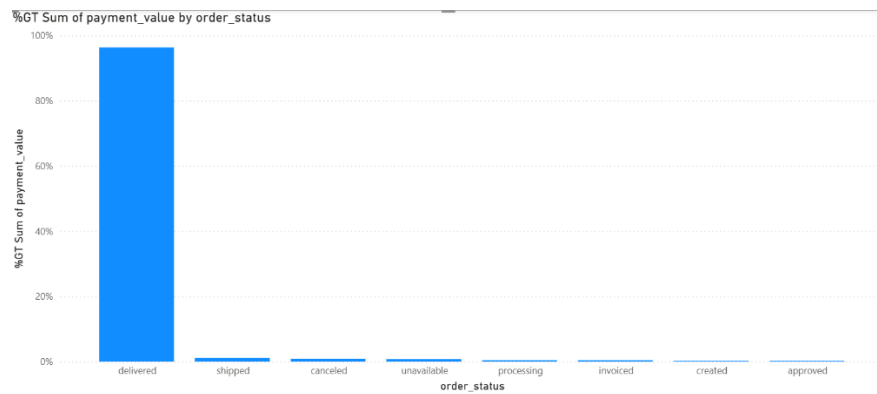
Total Freight Cost, Orders Count and Count of order_id by freight_value

Figure 15 - Total Freight Cost vs Order Count

%GT Sum of payment_value by order_status

Figure 16 - Revenue Share by Order Status
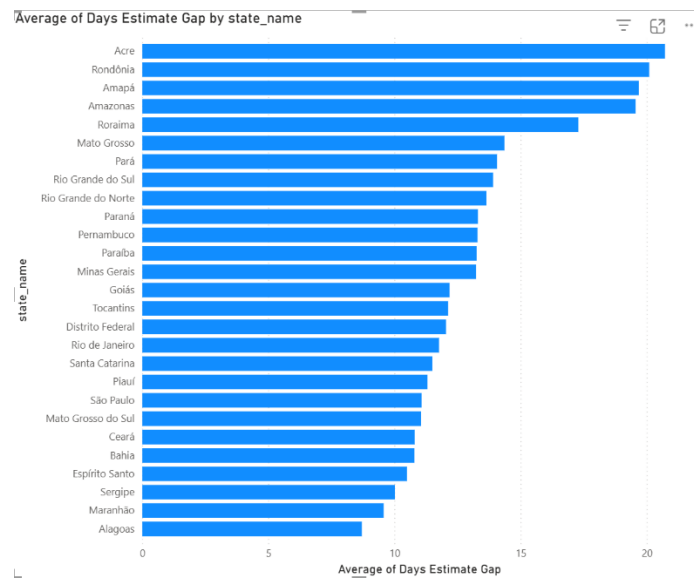
Average of Days Estimate Gap by state_name

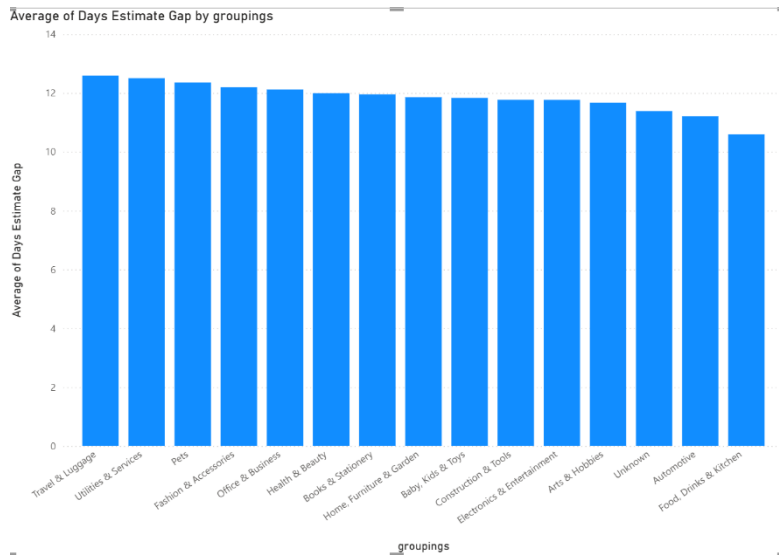*Figure 17 - Average Delivery-Time Gap by State*

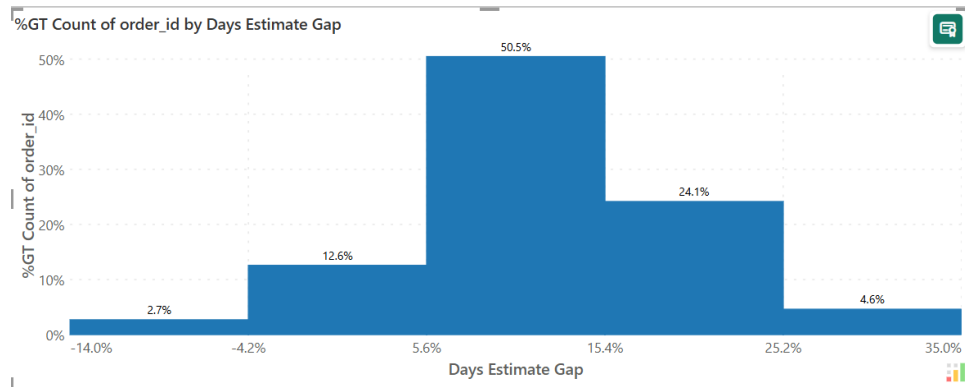*Figure 18 - Average Delivery-Time Gap by Product Category*

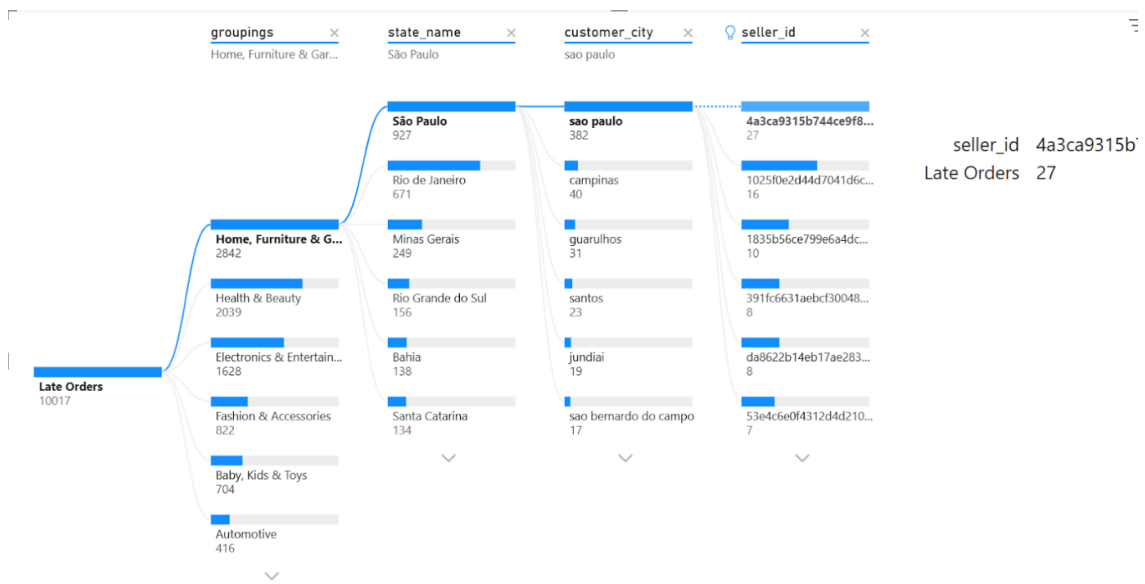*Figure 19 - Distribution of Orders by Delivery Gap Buckets*

Figure 20 - Late Orders by Category → State → City → Seller

Reviews and Sentiment

Key Business Questions to Answer

How is revenue distributed across review-score tiers?

Which order statuses attract the best and worst feedback?

Where do states and product categories rank on average review score?

Do delivery-time gaps or instillment plans influence customer ratings?

Are there seasonal shifts in the mix of positive / neutral / negative reviews?

What key themes emerge in positive vs. negative comment text?

Findings and Insights

Revenue by Review Score

- Five-star orders generate 70 % of revenue; less than 3-star orders contribute less than 10 %.
- High satisfaction clearly links to revenue concentration.

Score by Order Status

- Delivered orders average 4.1; all other statuses fall below 2.6.
- Fulfilment success is the main driver of strong ratings.

State Leaderboard

- Northern states Amapa & Amazonas top the chart (4.18) while Rio de Janeiro lags (3.87).
- Local logistics or seller mix affect customer perception.

Category Leaderboard

- Travel & Luggage (4.33) and Books & Stationery (4.29) lead; Office & Business trails (3.78).
- Product type influences expectations.

Monthly Review Mix

- Positive reviews stay ≥ 75 % all year; March and December see slight upticks in negatives.

Word Clouds

- Negative comments-focus: "not product", "delay", "missing", "wrong". Positive focus: "recommend", "delivery", "excellent".
- Quality and fulfilment drive sentiment polarity.

Delay vs Rating

- Mild positive correlation (R = 0.45) between longer gaps & higher rating for instillment-heavy orders.
- Multi-instillment purchases tolerate delays better.

Review Bucket by Category

- Office & Business shows the largest neutral/negative share of revenue, Travel & Luggage the smallest.

Snapshot

- Overall Average Review = 4.16 across 99 k orders.

Action plans

Let customers send back the 5 best-selling items for any reason.

Create a helpdesk just for top region with faster reply to times and messages in local language.

Add extra support staff and turn on FAQ chatbots for the busy months of March and December.

Analytical Workflow

To make sentiment analysis readable I collapsed the 1-to-5 star ratings into three buckets:

- Positive (4–5): satisfied customers
- Neutral (3): lukewarm feedback
- Negative (1–2): dissatisfaction

```
Review Bucket =
VAR s = order_reviews[review_score]
RETURN
   SWITCH (
     TRUE (),
     s <= 2, "Negative",
     s = 3, "Neutral",
     s >= 4, "Positive",
"No Review")
```
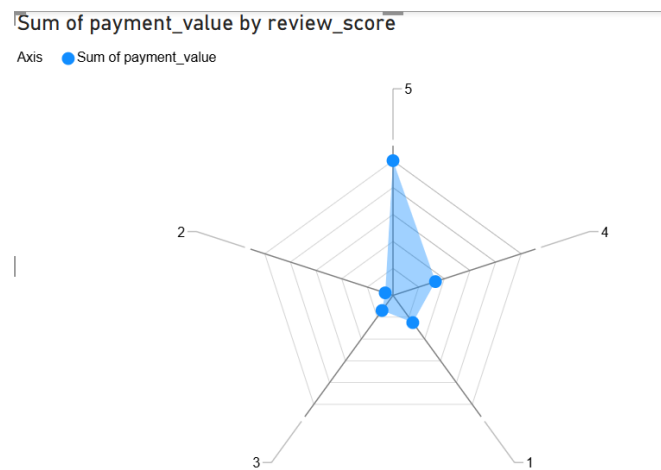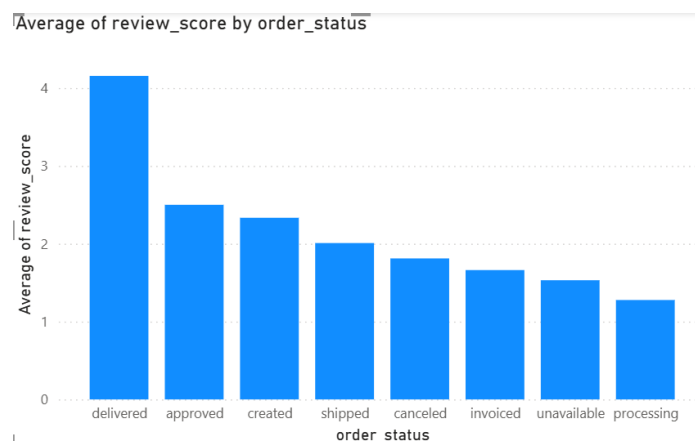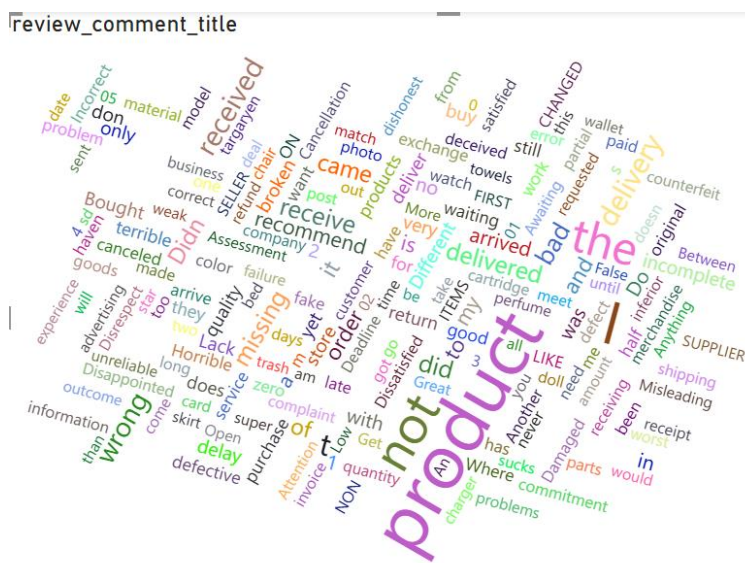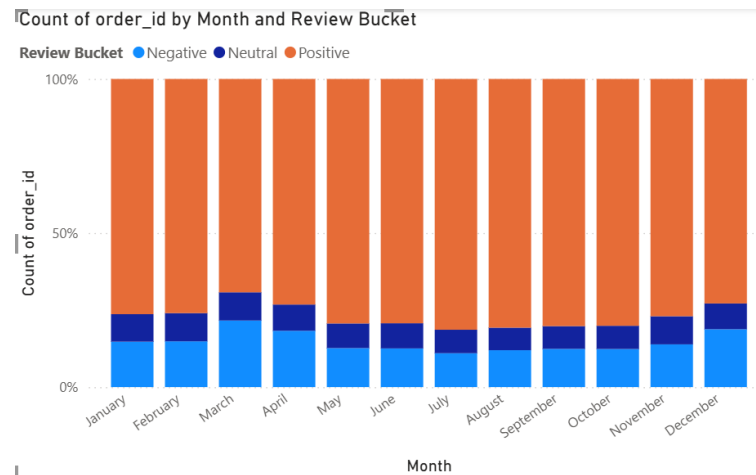

Figure 21 - Revenue Share by Review-Score Tier

Figure 22 - Average Review Score by Order Status

state.state_name	Count of customer_id	Average of review_score
Amapá	68	4.19
Amazonas	148	4.18
Paraná	5045	4.18
São Paulo	41746	4.17
Minas Gerais	11635	4.14
Rio Grande do Sul	5466	4.13
Mato Grosso do Sul	715	4.12
Rio Grande do Norte	485	4.11
Mato Grosso	907	4.10
Tocantins	280	4.10
Santa Catarina	3637	4.07
Distrito Federal	2140	4.06
Rondônia	253	4.05
Acre	81	4.05
Goiás	2020	4.04
Espírito Santo	2033	4.04
Paraíba	536	4.02
Pernambuco	1652	4.01
Piauí	495	3.92
Rio de Janeiro	12852	3.87
Bahia	3380	3.86
Ceará	1336	3.85
Pará	975	3.85
Sergipe	350	3.81
Maranhão	747	3.76
Alagoas	413	3.75
Roraima	46	3.61
Total	**99441**	**4.09**

Figure 23 - State-Level Average Review Score

groupings	Average of review_score
Travel & Luggage	4.33
Books & Stationery	4.29
Food, Drinks & Kitchen	4.26
Pets	4.24
Health & Beauty	4.18
Utilities & Services	4.15
Arts & Hobbies	4.15
Baby, Kids & Toys	4.12
Construction & Tools	4.10
Fashion & Accessories	4.09
Automotive	4.09
Home, Furniture & Garden	4.05
Electronics & Entertainment	4.03
Unknown	3.91
Office & Business	3.78
Total	**4.09**

Figure 24 - Category-Level Average Review Score

Count of order_id by Month and Review Bucket

Review Bucket ● Negative ● Neutral ● Positive

Figure 25 - Monthly Mix of Positive / Neutral / Negative Reviews

review_comment_title

Figure 26 - Negative Review Comment Word Cloud

review_comment_title

Figure 27 - Positive Review Comment Word Cloud
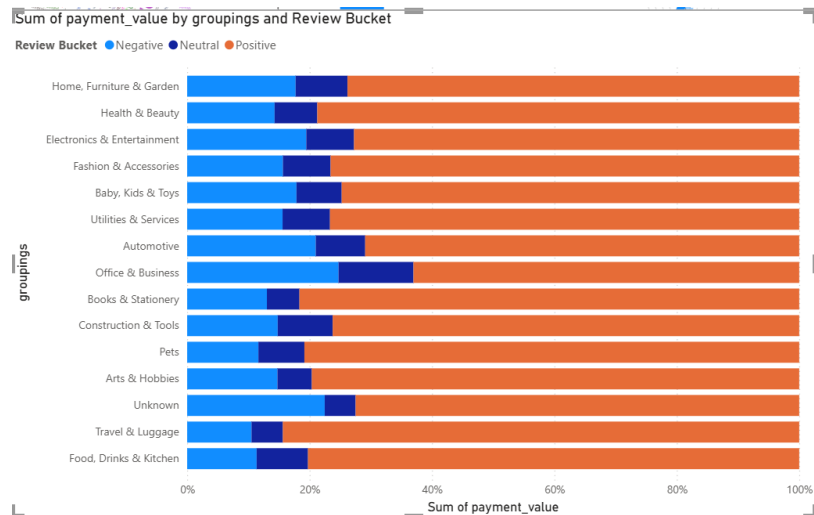
Figure 28 - Value-Weighted Review Bucket by Product Category
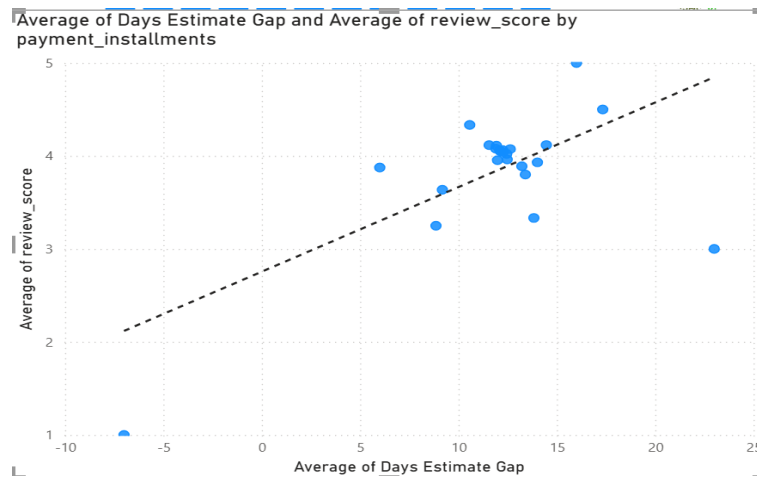

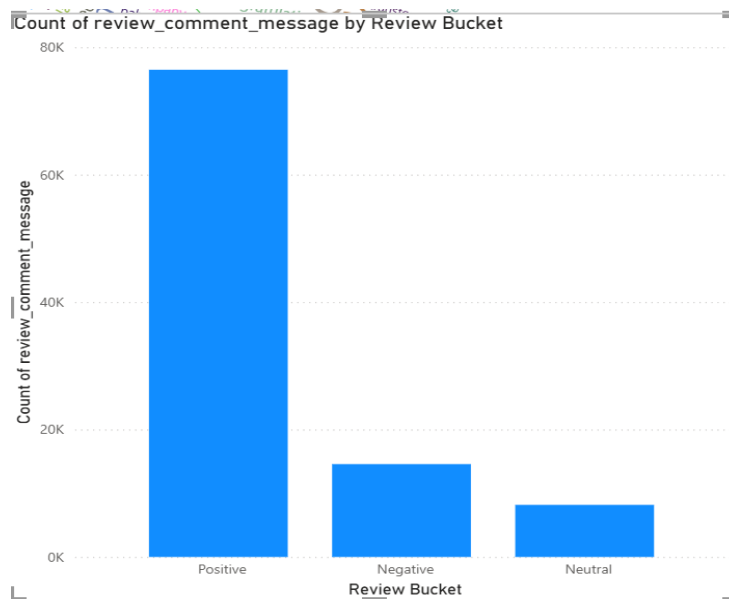*Figure 29 - Delivery Gap vs Average Review Score*

Figure 30 - Number of review messages vs Average Review Bucket

Forecasting

Key Business Questions to Answer

What is the projected revenue trajectory for the next six months?

Which product categories fall into growth, stable, or decline clusters based on the forecast?

How big is each cluster's current revenue footprint and how should inventory, marketing, and pricing adapt?

Findings and Insights

Monthly Revenue Forecast

- Total marketplace revenue is projected to climb from R$ 1.0 M (Aug 2018) to R$ 1.5 M (Mar 2019): a 45 % lift, with the 95 % confidence band spanning R$ 1.1 M – 2.0 M.
- Continues the upward trend observed since early-2017 but with widening uncertainty.

Category-level Forecast

- Growth Cluster: Home, Furniture & Garden; Health & Beauty; Fashion & Accessories; Automotive; Construction & Tools; Pets; Arts & Hobbies; Food, Drinks & Kitchen show growths by 15 %.
- Stable Cluster: Electronics & Entertainment; Baby, Kids & Toys; Office & Business; Books & Stationery hover within ± 5 % of current run-rate.
- Decline Cluster: Utilities & Services; Travel & Luggage forecast negative growth (-6 % to -10 %).

Cluster Footprint

- Growth cluster already represents 58 % of 2018 revenue, stable 34 %, decline 8 %.
- Resource allocation should mirror forward-looking potential, not just historical share.

Action plans

Re-order popular items more often and keep extra inventory so they never run out.

Sell accessories together with main products to lift the average basket size without big discounts.

Run short markdown tests to see how price-sensitive products are before deciding to drop them from the range.
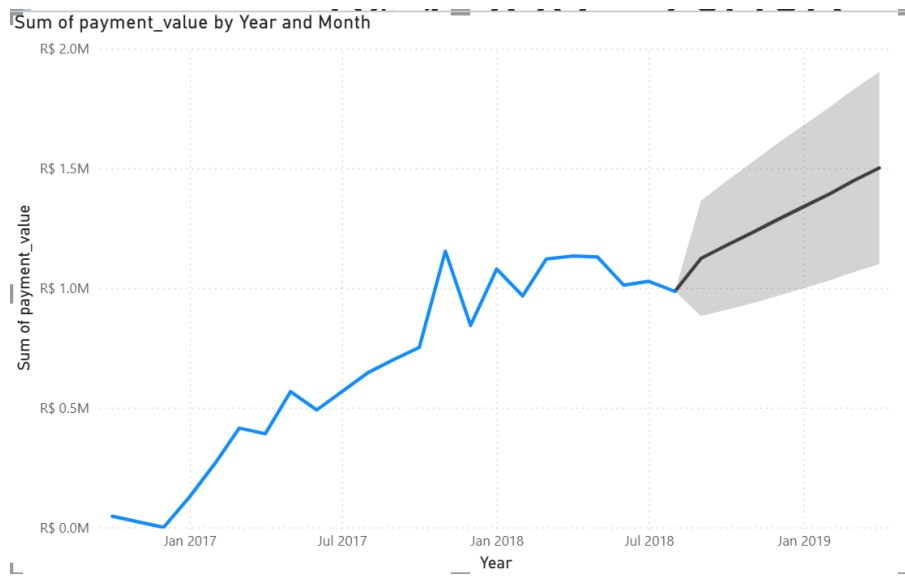
Sum of payment_value by Year and Month

Figure 31- Monthly Revenue (Actual vs 6-Month Forecast with 95 % Confidence Band)