



WE
CHALLANGE

Presented by: Mai Atef

followed process

- *Understand the Data .*
- *Understand the relationships between different tables .*
- *Data Cleaning .*
- *Exploratory Data Analysis (EDA) .*
- *Specific Analyses*
- *Customer Analysis:*
 - *Analyze customer demographics, purchase behavior, and lifetime value.*
- *Sales Analysis:*
 - *Analyze overall sales trends, seasonality, and best-selling products.*
- *Product Analysis:*
 - *Analyze product categories, pricing strategies, and inventory turnover.*
- *Seller Analysis:*
 - *Evaluate seller performance and geographical distribution.*

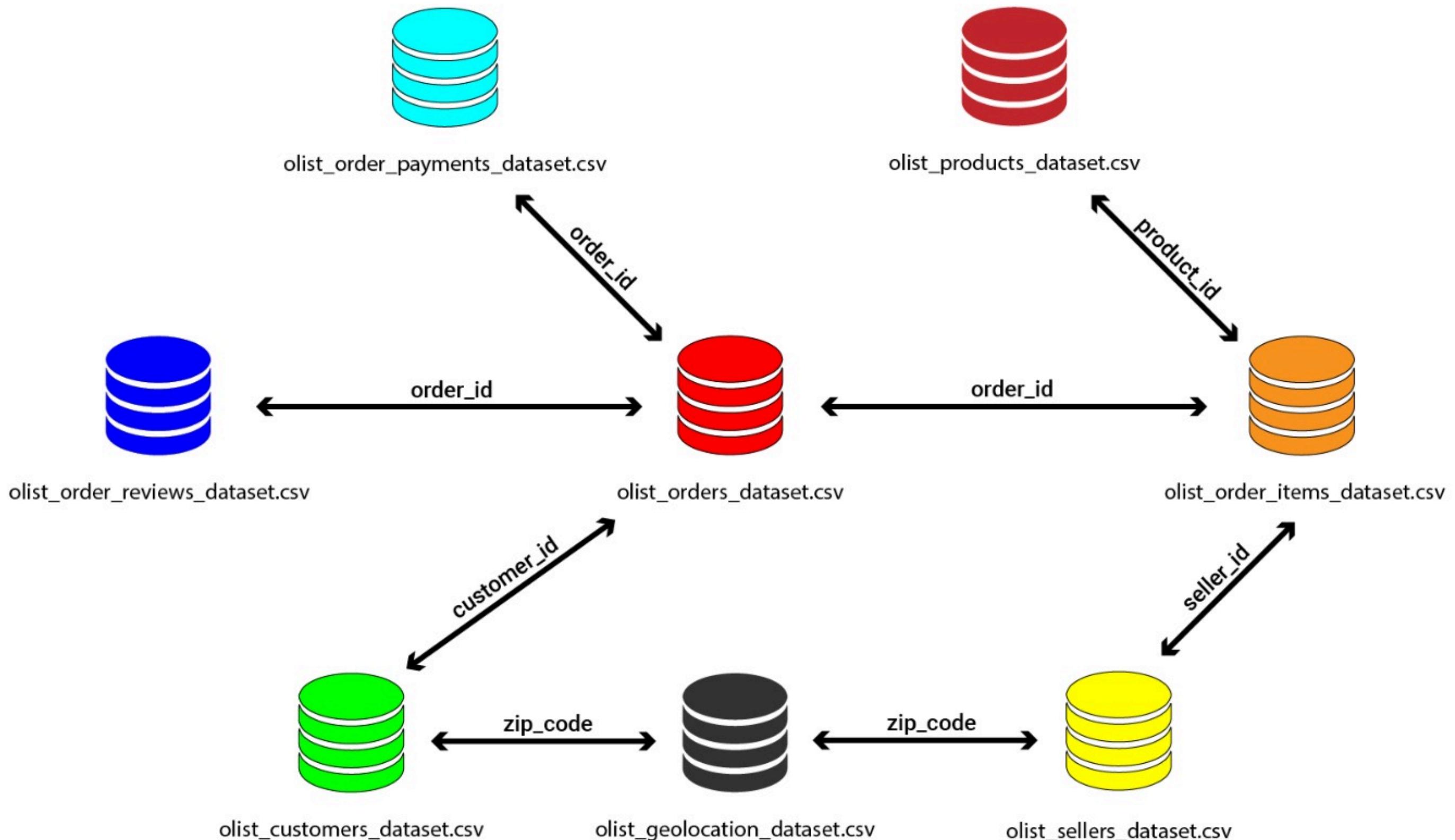
Brazilian E-Commerce Public Dataset by Olist

- ***olist_orders_dataset*** : Contains information about orders, including order status, purchase timestamp, and customer ID.
- ***olist_order_items_dataset*** : Contains detailed information about the items within each order, including product ID, seller ID, price, and freight value.
- ***olist_order_reviews_dataset*** : Contains customer reviews of orders, including review scores and comments.
- ***olist_products_dataset*** : Contains information about products, including product category, name, and dimensions.

Understand dataset

- ***olist_order_payments_dataset***: Contains payment details for orders, including payment type and installments.
- ***olist_customers_dataset***: Contains information about customers, including customer ID and geolocation.
- ***olist_geolocation_dataset***: Contains geolocation data, including zip codes and coordinates.
- ***olist_sellers_dataset***: Contains information about sellers, including seller ID and geolocation

Understand the relationships between different tables



Sales Analysis

(Best-Selling Products based on revenue)

Revenue by Product Category : determine which products generate the most revenue.

```
34 WITH TotalOrders AS (
35     SELECT sum(price) AS totalRevenue
36     FROM order_items
37 )
38 SELECT products.product_category_name ,
39 SUM(order_items.price)*100.0/totalRevenue "Revenue by Product Category"
40 FROM products
41 INNER JOIN order_items ON products.product_id=order_items.product_id
42 CROSS JOIN TotalOrders
43 GROUP BY products.product_category_name
44 ORDER BY "Revenue by Product Category" DESC
45 LIMIT 30;
```

Insights

- This can help you understand which products are most valuable to the business.
- guide marketing and inventory strategies to maximize revenue.
- Inventory Management: Ensure that these products are always in stock to avoid missed sales opportunities.
- Dependency Risk: Identify the risk of revenue dependency on a few products and diversify the product range.

(Best-Selling Products based on revenue)



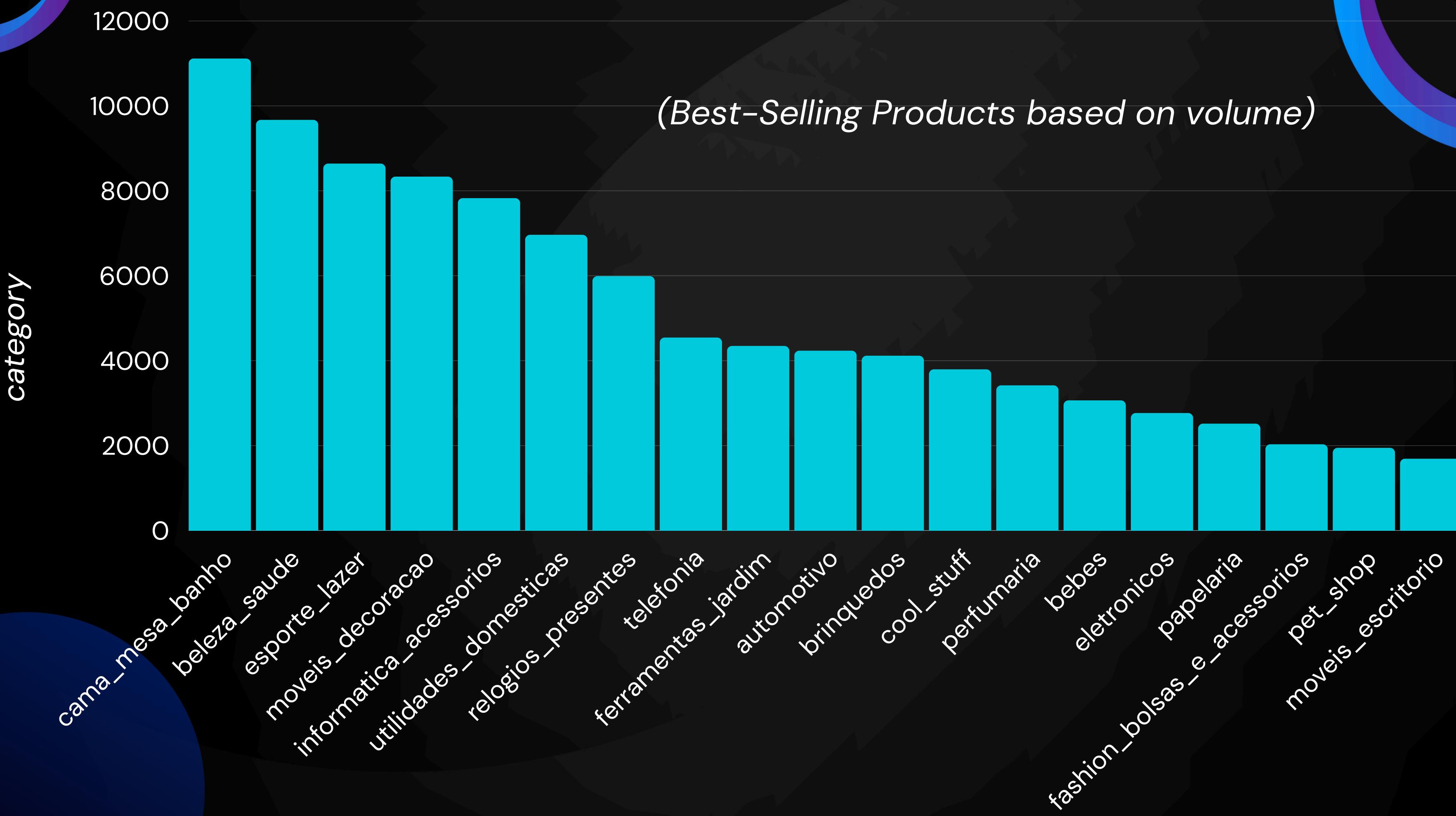
(Best-Selling Products based on volume)

Insights

- **Customer Preferences**
- *This indicates their high popularity and demand.*
- **Inventory Management:** Ensure that these products are always in stock to avoid missed sales
- **Marketing Focus:** Marketing efforts could be concentrated on promoting these top-selling products further, as they already have a strong customer preference
- **Product Bundling:** Consider bundling top-selling products with less popular items to boost the sales of slower-moving inventory.

```
1 SELECT product_category_name ,  
2 COUNT(order_items.order_id) "number of orders"  
3 FROM oproducts  
4 INNER JOIN order_items ON oproducts.product_id=order_items.product_id  
5 GROUP BY oproducts.product_category_name  
6 ORDER BY "number of orders" DESC  
7 LIMIT 30;
```

top 10 solid order per product category



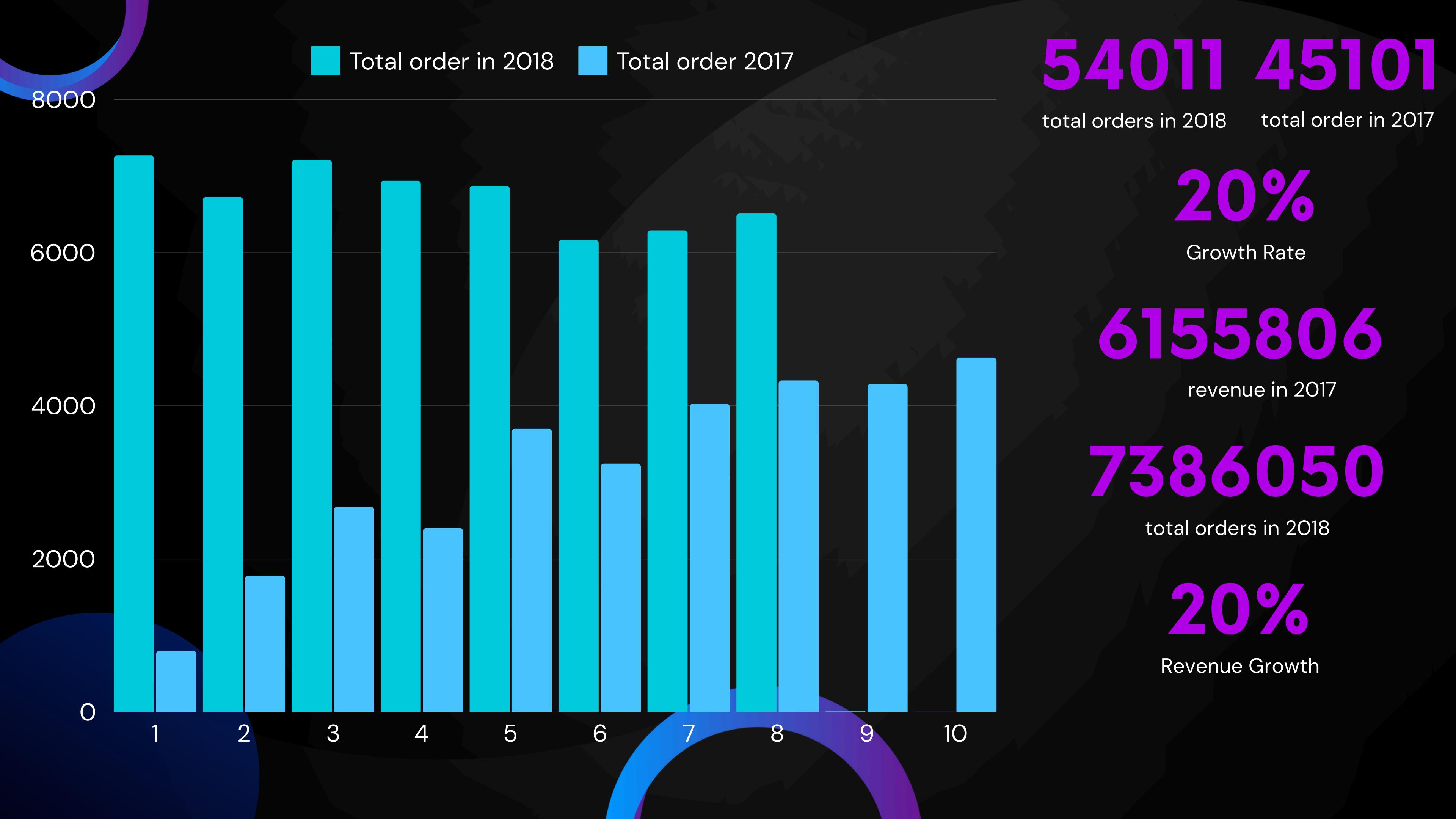
number of orders

Sales Analysis

Total Orders Comparison Between 2017 and 2018

```
1 SELECT
2     COUNT(orders.order_id) AS "Total order",
3     strftime('%Y-%m', orders.order_purchase_timestamp) AS month
4 FROM orders
5 WHERE strftime('%Y', orders.order_purchase_timestamp) IN ('2017')
6 GROUP BY month
7 ORDER BY month DESC;
8
9 SELECT
10    COUNT(orders.order_id) AS "Total order",
11    strftime('%Y-%m', orders.order_purchase_timestamp) AS month
12 FROM orders
13 WHERE strftime('%Y', orders.order_purchase_timestamp) IN ('2018')
14 GROUP BY month
15 ORDER BY month DESC;
```

- *Identify Monthly Trends*
- **Seasonal Patterns** :to meet expected demand during high-traffic periods.
- **Overall Growth Trend** : indicates
- *Understanding the factors behind the growth can help replicate success in future years.*



Sales Analysis

Total Orders By day of week

```
SELECT
    strftime('%w', orders.order_purchase_timestamp) AS day_of_week,
    COUNT(orders.order_id) AS "Total orders"
FROM orders
INNER JOIN customers ON customers.customer_id=orders.customer_id
GROUP BY day_of_week
ORDER BY day_of_week ASC
```

Peak Order Days:

Monday and Tuesday are the highest order days, contributing the most significant percentage of total orders (16.29% and 16.05%, respectively). This suggests that customers are most active at the start of the workweek.

Sunday has the lowest order percentage at 12.03%, indicating that it is a less active day for orders compared to weekdays.

Weekend Orders:

Saturday shows a lower order percentage (10.95%), which is the lowest among all days. This might indicate a drop in activity during the weekend, possibly due to customers being occupied with personal activities or shopping less.



Customer Analysis

Businesses have always tried to keep their customer base engaged and satisfied with the services provided by them

customer Reviews

```
1 SELECT
2   SUM(CASE WHEN order_reviews.review_score > 3 THEN 1 ELSE 0 END)
3   AS positive,
4   SUM(CASE WHEN order_reviews.review_score <= 3 THEN 1 ELSE 0 END)
5   AS negative
6 FROM
7   order_reviews;
```

Analyzing customer reviews provides valuable insights into customer satisfaction, product quality, service effectiveness, and areas needing improvement.

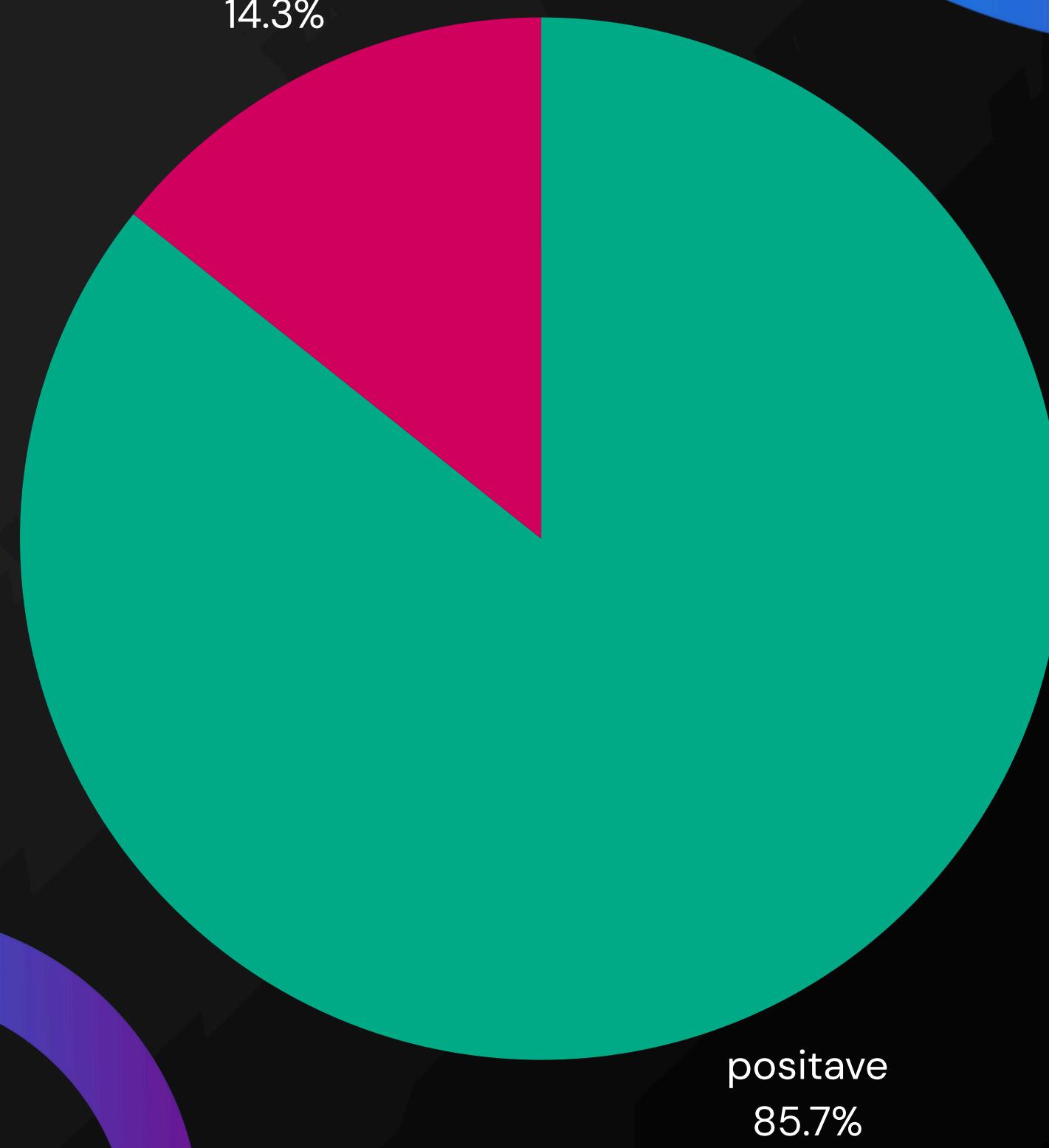
positive negative

negative

14.3%

positive

85.7%



Customer Analysis

compare between review on product who delivered late and order most expensive

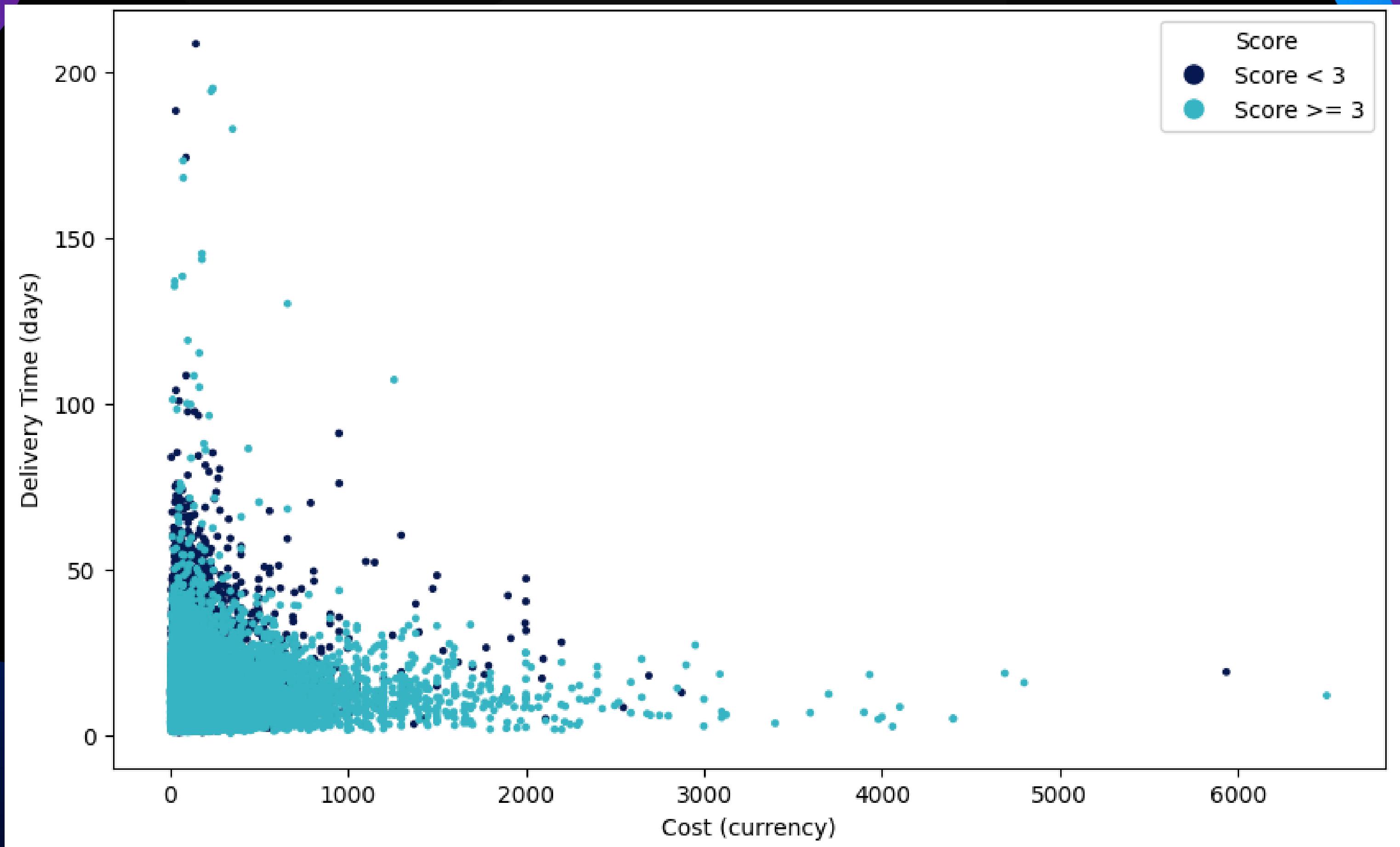
```
SELECT
    julianday(orders.order_delivered_customer_date) - julianday(orders.order_purchase_timestamp)
    SUM(order_items.price) AS cost,
CASE
    WHEN order_reviews.review_score > 3 THEN 1
    ELSE 0
END AS score
FROM orders
JOIN order_items ON order_items.order_id = orders.order_id
JOIN order_reviews ON order_reviews.order_id = orders.order_id
GROUP BY orders.order_id
```

time taken for delivery increases, the chances of being dissatisfied with the product increases significantly.

The increase in the price of an item does not cause too much dissatisfaction if it is delivered on time

- compare between review on product who deleviered late and order most expensive

```
✓ 3s   ▶  
import pandas as pd  
import matplotlib.pyplot as plt  
  
# Load data from CSV  
df = pd.read_csv('/content/zeroScore.csv')  
print(df.head())  
  
# Define colors in hex format  
df['color'] = df['score'].apply(lambda x: '#071952' if x == 0 else '#37B7C3')  
  
# Create the dot plot with swapped columns  
plt.figure(figsize=(10, 6))  
plt.scatter(df['cost'], df['timeee'], c=df['color'], s=6)  
  
# Adding titles and labels  
plt.title('Cost vs Delivery Time with Score as Color')  
plt.xlabel('Cost (currency)')  
plt.ylabel('Delivery Time (days)')  
  
# Update legend to match the colors used  
plt.legend(handles=[  
    plt.Line2D([0], [0], marker='o', color='w', markerfacecolor='#071952', markersize=10, label='Score < 3'),  
    plt.Line2D([0], [0], marker='o', color='w', markerfacecolor='#37B7C3', markersize=10, label='Score >= 3')  
], title='Score')  
  
# Show plot  
plt.show()
```

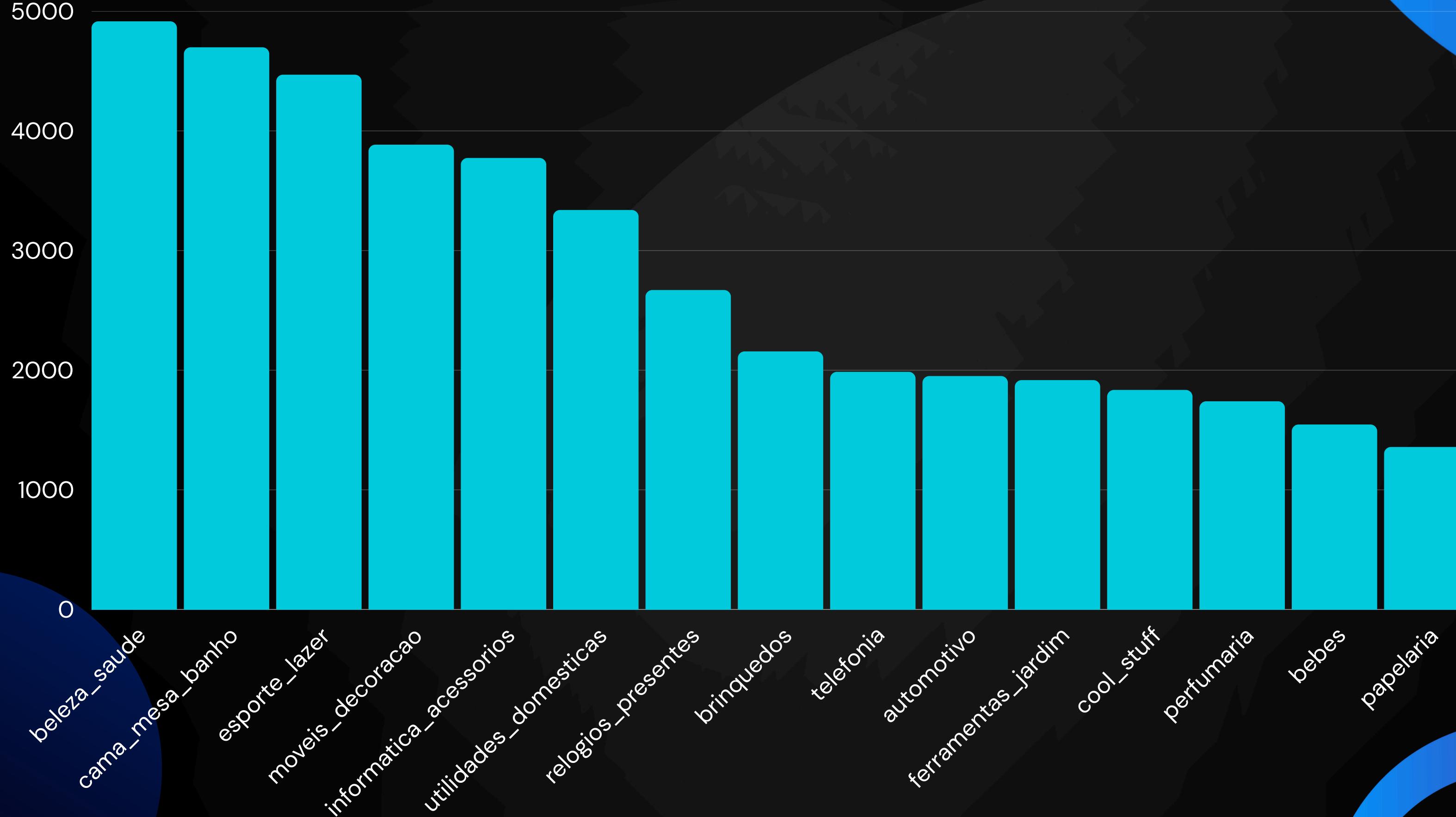


Customer Analysis

- Top 15 reviewed Product : most product satisficed by customer

```
1 SELECT products.product_category_name,  
2     SUM(CASE WHEN order_reviews.review_score > 3 THEN 1 ELSE 0 END)  
3     AS positave  
4 FROM  
5     order_reviews  
6 JOIN order_items ON order_reviews.order_id=order_items.order_id  
7 JOIN products ON order_items.product_id=products.product_id  
8 GROUP BY products.product_category_name  
9 ORDER BY positave DESC  
10 LIMIT 15;
```

- *Top 15 reviewed Product*



order delivery status

Most Orders Delivered:

96.5% of orders have been delivered, showing efficient fulfillment.

Few Orders Still in Transit:

Only 1.1% are still in the "Shipped" status.

Low Cancellation Rate:

0.6% of orders are canceled, indicating minimal cancellation issues.

Stock Issues Noted:

0.6% of orders are marked as "Unavailable," suggesting some stock or supply issues.

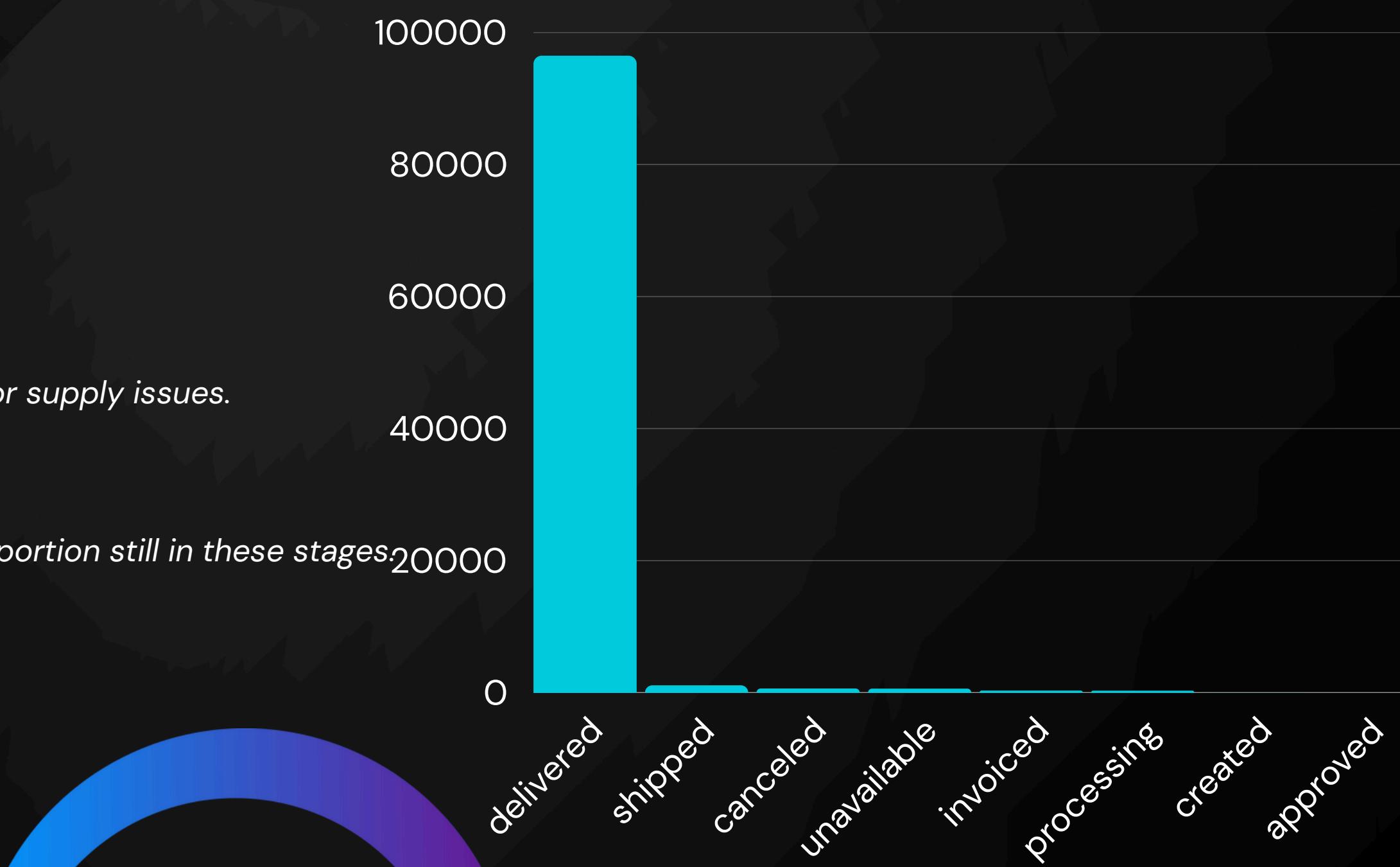
Minor Processing and Invoicing Delays:

0.3% of orders are either "Processing" or "Invoiced," showing a small portion still in these stages.

Very Few Orders at Initial or Approved Stage:

Only 0.01% of orders are "Created," and 0.02% are "Approved."

```
1 SELECT COUNT(orders.order_id)
2   orders.order_status
3   FROM orders
4   GROUP BY orders.order_status
5   ORDER BY totalOrder DESC
```



order delivery status

```
SELECT  
    AVG(julianday(orders.order_delivered_customer_date)  
        -julianday(orders.order_purchase_timestamp))  
    AS average_delivery_time_days  
FROM  
    orders  
WHERE  
    orders.order_status = 'delivered';
```



```
SELECT  
    AVG(julianday(orders.order_estimated_delivery_date)  
        -julianday(orders.order_purchase_timestamp))  
    AS average_delivery_time_days  
FROM  
    orders  
WHERE  
    orders.order_status = 'delivered';
```

12.5 day
actuall Delivery Time

23.7 day
Estimated Delivery Time

47.2%

Percentage Difference

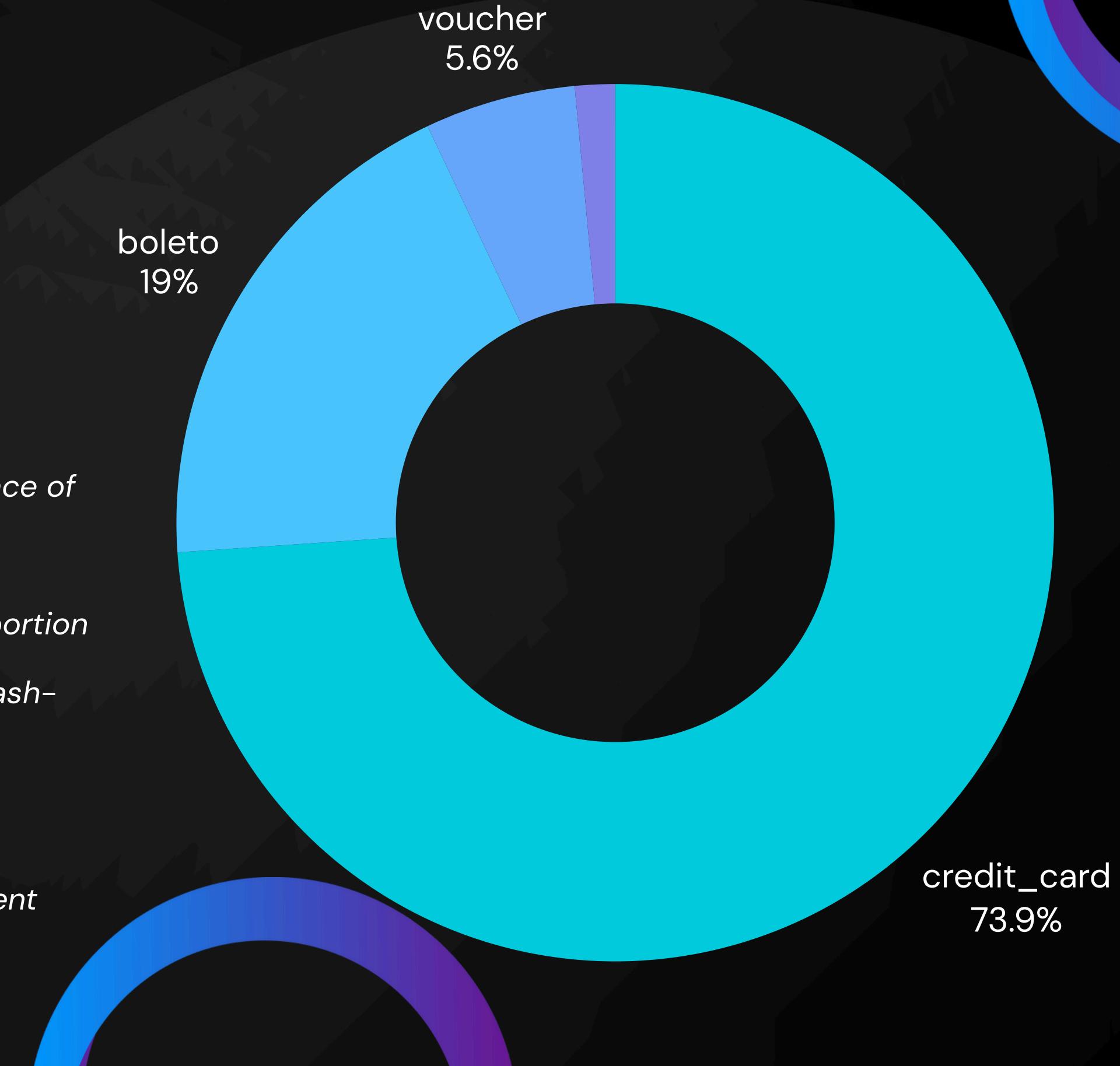
- *payment Method*

```
1 SELECT  
2 payment_type,  
3 COUNT(payment_type) AS total  
4 FROM order_payments  
5 GROUP BY payment_type  
6 ORDER BY total DESC;
```

Insights:

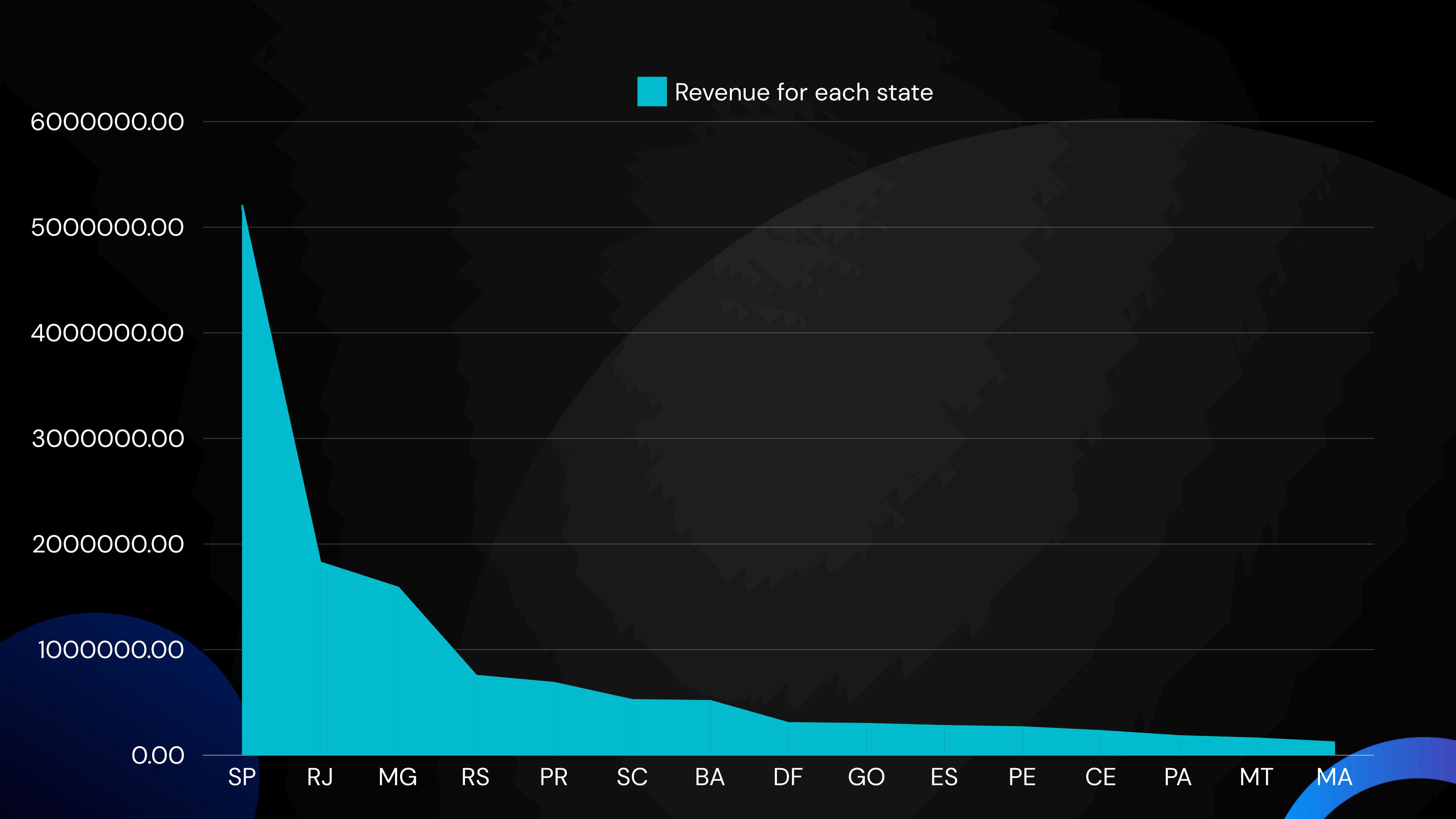
- **Credit Card:** it indicates a strong preference for card payments among customers. This may reflect the convenience and widespread acceptance of credit cards.
- **Boleto:** A significant number of boleto payments might suggest that a portion of the customer base prefers this method due to its convenience for cash-based transactions.
- **Payment Preferences:** The distribution of payment methods can help understand customer preferences and guide decisions on which payment options to prioritize or promote

credit_card boleto voucher
debit_card not_defined



- now finding the Top 15 state which generates the highest revenue. The state named 'SP' generates is at the top

```
2 SELECT customer.customer_city,  
3 sum(order_items.price) AS "Revenue for each city"  
4 FROM customer  
5 INNER JOIN orders ON customer.customer_id=orders.customer_id  
6 INNER JOIN order_items ON orders.order_id=order_items.order_id  
7 GROUP BY customer.customer_city  
8 ORDER BY "Revenue for each city" DESC  
9 LIMIT 15;
```



- now finding the city which generates the highest orders. The state named 'sao paulo' generates is at the top

```
SELECT customers.customer_city,  
COUNT(order_items.order_id) AS "total order for each city"  
FROM customers  
INNER JOIN orders ON customers.customer_id=orders.customer_id  
INNER JOIN order_items ON orders.order_id=order_items.order_id  
GROUP BY customers.customer_city  
ORDER BY "total order for each city" DESC  
LIMIT 25;
```

top 15 city which generates the highest orders

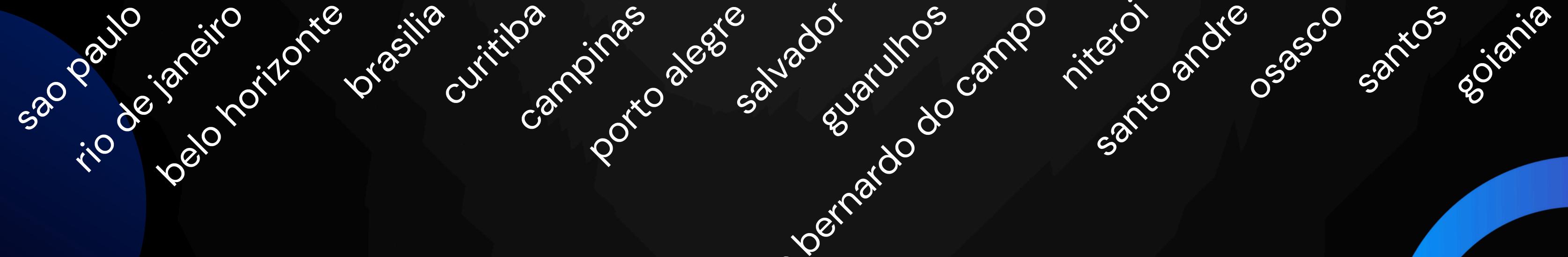
20000.00

15000.00

10000.00

5000.00

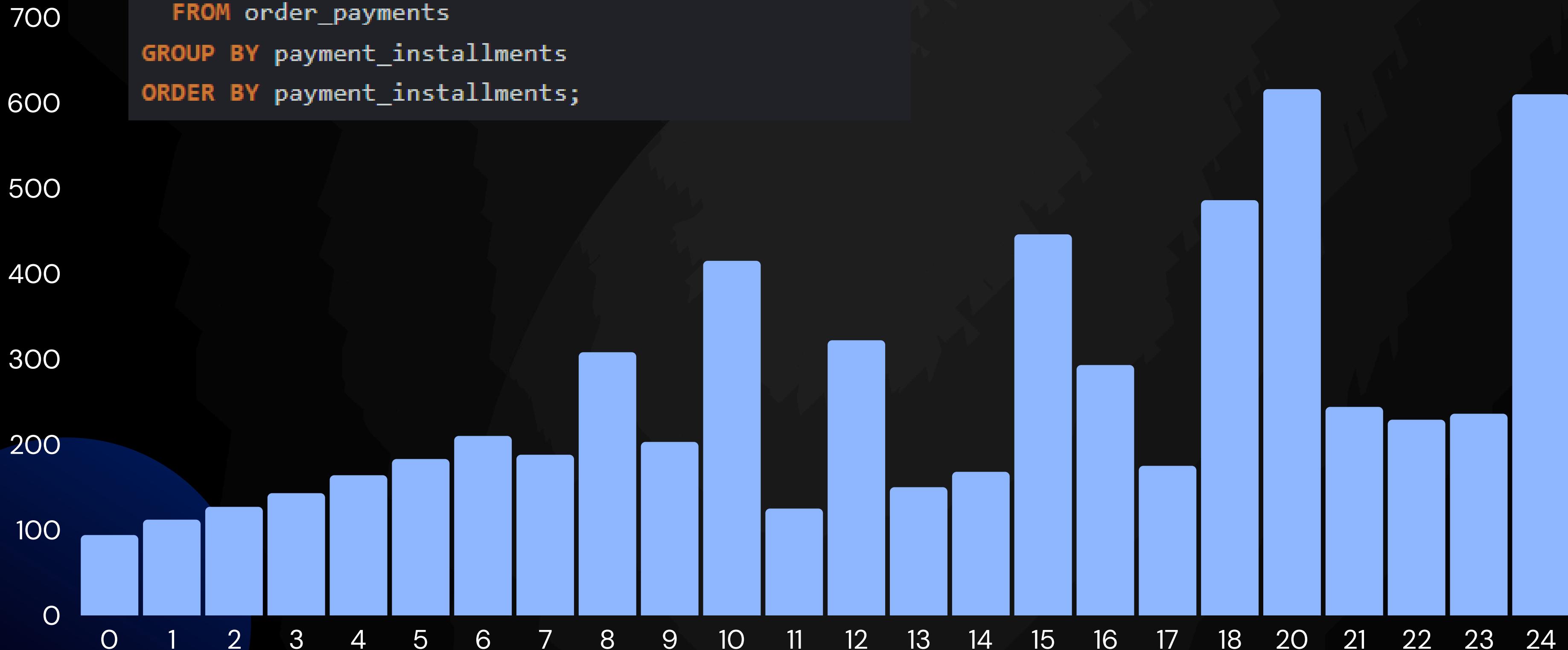
0.00



- how customers are choosing to pay for their purchases over time.

Based on the analysis, draw insights that can inform business decisions

```
SELECT order_payments.payment_installments,  
       round(AVG(payment_value)) AS total_payment_value  
  FROM order_payments  
 GROUP BY payment_installments  
 ORDER BY payment_installments;
```



- *insights*

High Installment Customers (20 or 24 installments)

Tend to make higher-value purchases.

Likely more dependent on credit.

Prefer longer repayment periods.

Low Installment Customers (0-5 installments)

Tend to make lower-value purchases.

Likely have a higher capacity for upfront payments.

Prefer to minimize their debt period.

Identifying why there are significant jumps at specific installment counts (e.g., 10 and 20 installments) can help in designing similar promotions for other installment ranges.

Based on the analysis, draw insights that can inform business decisions

THANK YOU

For watching this presentation

any question ?