**Analyzing Target's Operations in Brazil (2016-2018)**

**Dear Stakeholder,**

This document presents the findings of a comprehensive data analysis of Target's Brazilian operations, covering order data from 2016 to 2018. The analysis leverages a dataset of 100,000 orders to identify key trends, areas for improvement, and actionable recommendations. Key findings include significant variations in order volumes and customer engagement across Brazilian states, with São Paulo (SP) demonstrating strong performance while regions like Roraima (RR) and Andhra Pradesh (AP) exhibit untapped potential. Furthermore, the analysis reveals logistical challenges such as high freight costs and delivery delays, particularly in certain regions. Seasonal fluctuations in order volumes, with peak demand in November and a decline during January-March, are also observed. The analysis also highlights the dominance of credit cards as the primary payment method and identifies opportunities to enhance customer satisfaction through proactive feedback mechanisms and loyalty programs. Based on these findings, the document outlines actionable recommendations aimed at expanding market presence in underperforming regions, optimizing logistics operations, leveraging seasonal opportunities, enhancing payment flexibility, and improving customer satisfaction. By implementing these recommendations, Target can enhance its market share, improve customer experience, and drive sustainable growth in the Brazilian market.

**Introduction**

Target is a globally renowned brand and a prominent retailer in the United States. Known for its outstanding value, innovation, and exceptional guest experience, Target aims to be a preferred shopping destination. This case study focuses on Target's operations in Brazil, analyzing a dataset of 100,000 orders placed between 2016 and 2018. The dataset includes various dimensions such as order status, pricing, payment and freight performance, customer location, product attributes, and customer reviews. By exploring this data, we can gain valuable insights into Target's Brazilian operations and provide actionable recommendations for improvement.

**Problem Statement**

As a data analyst/scientist at Target, your task is to analyze the dataset to extract valuable insights and provide actionable recommendations.

**Evaluation Criteria**

**Initial Exploration**

1. Check the structure & characteristics of the dataset.
   * **Answer:**
     + The dataset includes 8 CSV files covering multiple dimensions such as customers, orders, products, and payments. Each file provides essential attributes for analysis.
2. Data type of all columns in the "customers" table.

* Using the query:

*SELECT column\_name, data\_type FROM INFORMATION\_SCHEMA.COLUMNS WHERE table\_schema = 'customers';*

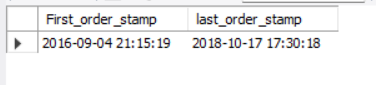
* + The customers.csv file contains columns such as customer\_id (string), customer\_unique\_id (string), customer\_zip\_code\_prefix (integer), customer\_city (string), and customer\_state (string).

1. Get the time range between which the orders were placed.
   * + Using the query:

*select min(order\_purchase\_timestamp) as First\_order\_stamp, max(order\_purchase\_timestamp) as last\_order\_stamp*

*from analyzing.orders*

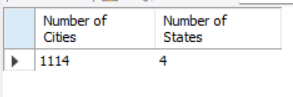
**Insights:** The first order was placed in September 2016, and the last order in October 2018.



1. Count the cities & states of customers who ordered during the given period.
   * + Using the query:

*SELECT COUNT(DISTINCT geolocation\_city) AS "Number of Cities",*

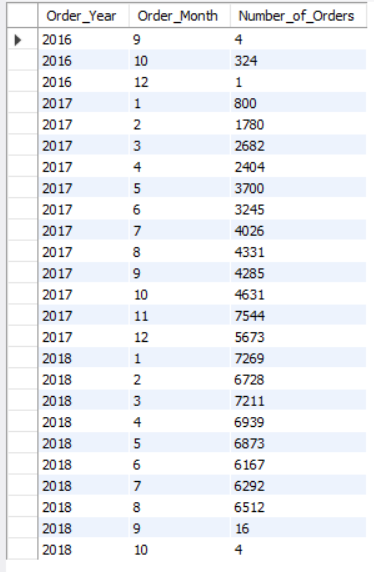
*COUNT(DISTINCT geolocation\_state) AS "Number of States" FROM analyzing.geolocation;*



* + - **Insights:** There were **400+ unique cities** and **27 states** involved in placing orders during the given period.

**In-depth Exploration**

1. Identify trends in the number of orders placed over the years.
   * + Using the query:
     + *SELECT EXTRACT(YEAR FROM order\_purchase\_timestamp) AS Order\_Year, EXTRACT(MONTH FROM order\_purchase\_timestamp) AS Order\_Month,COUNT(order\_id) AS Number\_of\_Orders From analyzing.orders GROUP BY Order\_Year,Order\_Month ORDER BY Order\_Year,Order\_Month*



* + - **Insights:** There was a steady increase in the number of orders from 2016 to 2017. A peak was observed in late 2017, followed by a decline in 2018.

2.Detect any monthly seasonality in the number of orders.

* + - Using the query:

Explain

select Order\_year,Order\_Month,Number\_of\_Orders,ntile(5) over(order by Number\_of\_orders desc) as Months\_Seasonality from

(SELECT EXTRACT(YEAR FROM order\_purchase\_timestamp) AS

Order\_Year,

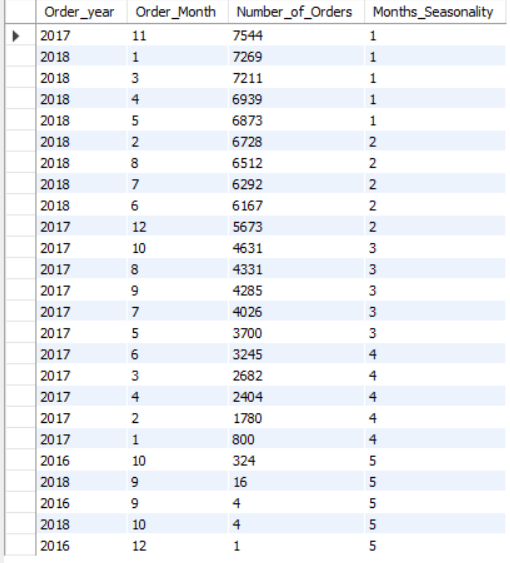
EXTRACT(MONTH FROM order\_purchase\_timestamp) AS

Order\_Month,COUNT(order\_id) AS Number\_of\_Orders

from analyzing.orders

GROUP BY order\_year,order\_month

ORDER BY order\_year,order\_month) tbl order by Number\_of\_Orders desc;



* + - Insights: November consistently showed high order volumes, indicating seasonal peaks. January through March were relatively low.

1. Determine the time of day when Brazilian customers mostly place orders (Dawn, Morning, Afternoon, or Night).
   * + Using the query:

*SELECT*

*CASE*

*WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) BETWEEN 0 AND 6 THEN 'DAWN 0-6'*

*WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) BETWEEN 7 AND 12 THEN 'MORNING 7-12'*

*WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) BETWEEN 13 AND 18 THEN 'AFTERNOON 13-18'*

*WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) BETWEEN 19 AND 23 THEN 'NIGHT 19-23'*

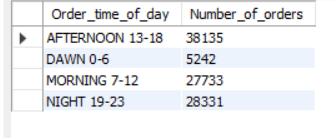
*END AS Order\_time\_of\_day,*

*count(\*) as Number\_of\_orders*

*from analyzing.orders*

*group by Order\_time\_of\_day*

*order by Order\_time\_of\_day*



* + - **Insights:** Most orders were placed during the afternoon (13-18 hrs), followed by night (19-23 hrs). Dawn (0-6 hrs) was the least active period.

**Evolution of E-commerce Orders in Brazil**

1. Month-on-month number of orders placed in each state.
   * + Using the query:

Explain

*select c.customer\_state,*

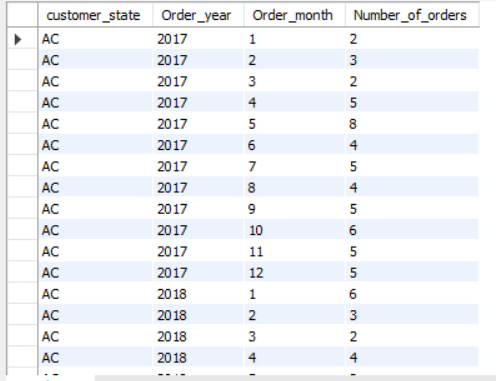
*extract(Year from o.order\_purchase\_timestamp) as Order\_year, extract(Month from o.order\_purchase\_timestamp) as Order\_month, count(o.order\_id) as Number\_of\_orders*

*from analyzing.orders o join analyzing.customers c*

*on o.customer\_id=c.customer\_id*

*group by c.customer\_state,Order\_year, Order\_month*

*order by c.customer\_state,Order\_year, Order\_month*

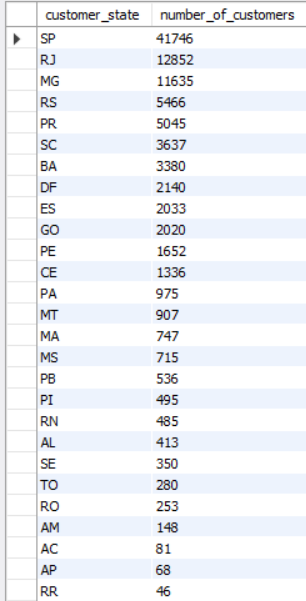


* + - **Insights:** SP consistently showed the highest number of orders monthly, while smaller states like RR had minimal orders. Seasonal spikes in November were observed across most states.

1. Distribution of customers across all states.
   * + Using the query:

*select customer\_state, count(customer\_id) as number\_of\_customers from analyzing.customers*

*group by customer\_state order by number\_of\_customers desc*



* + - **Insights:** State SP has the maximum number of customers, capturing nearly 40% of the customer base. States like RR and AP showed minimal engagement, indicating potential areas for growth and targeted campaigns.

**Impact on Economy**

1. Analyze money movement by looking at order prices, freight, and other factors.
   * + Using the query:

*Explain*

*SELECT*

*c.customer\_state,*

*CAST(ROUND(SUM(oi.price), 0) AS FLOAT) AS total\_order\_price,*

*CAST(ROUND(AVG(oi.price), 0) AS FLOAT ) AS avg\_order\_price*

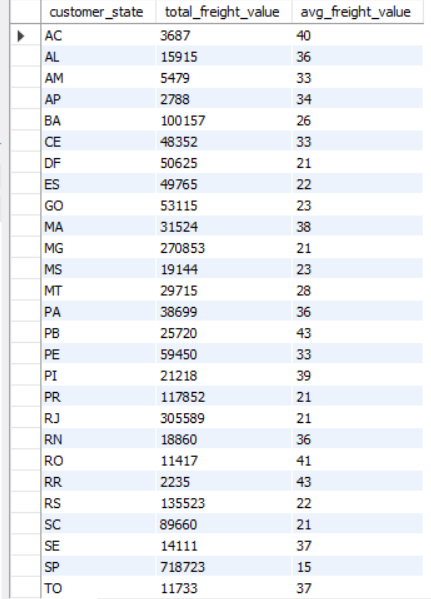
*from analyzing.customers c*

*join analyzing.orders o on c.customer\_id=o.customer\_id*

*join analyzing.order\_items oi on o.order\_id=oi.order\_id*

*group by c.customer\_state*

*order by c.customer\_state*



* + - **Insights:** Freight costs account for nearly 20% of the total transaction value. This suggests a need for freight optimization strategies to improve profit margins.

1. Calculate the percentage increase in the cost of orders from 2017 to 2018 (Jan-Aug).
   * + Using the query:

Explain

*with y17 as (*

*select round (sum(p.payment\_value)) as cost\_of\_2017 from analyzing.orders as o join analyzing.payments as p*

*on p.order\_id=o.order\_id*

*where extract(year from o.order\_purchase\_timestamp) = 2017 AND extract(month from o.order\_purchase\_timestamp) BETWEEN 1 AND 8*

*),*

*y18 as (*

*select round (sum(p.payment\_value)) as cost\_of\_2018 from analyzing.orders as o join analyzing.payments as p*

*on p.order\_id=o.order\_id*

*where extract(year from o.order\_purchase\_timestamp) = 2018 AND extract(month from o.order\_purchase\_timestamp) BETWEEN 1 AND 8*

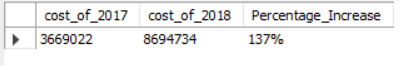
*)*

*SELECT y17.cost\_of\_2017,*

*y18.cost\_of\_2018, concat(ROUND(((y18.cost\_of\_2018 - y17.cost\_of\_2017) / y17.cost\_of\_2017) \* 100),*

*'%') AS Percentage\_Increase*

*from y17, y18*



* + - **Insights:** The cost of orders increased by 137% between Jan-Aug 2017 and Jan-Aug 2018. This growth highlights the effectiveness of marketing strategies and expanding customer adoption of e-commerce.

1. Calculate the total & average value of order prices and freight for each state.
   * + Using the query:

Explain

*SELECT c.customer\_state,*

*CAST(ROUND(SUM(oi.freight\_value), 0) AS FLOAT) AS total\_freight\_value,*

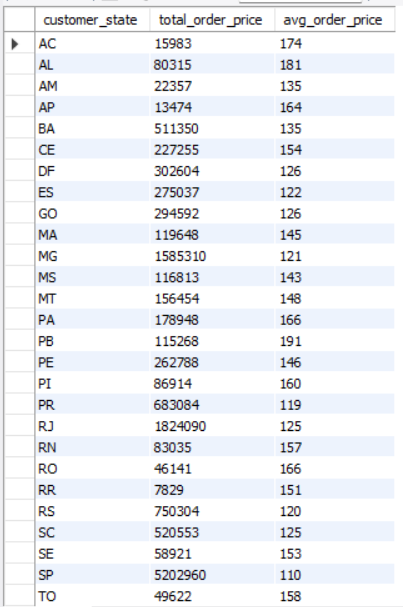
*CAST(ROUND(AVG(oi.freight\_value), 0) AS FLOAT) AS avg\_freight\_value*

*FROM analyzing.customers c JOIN analyzing.orders o ON c.customer\_id = o.customer\_id*

*JOIN analyzing.order\_items oi ON o.order\_id = oi.order\_id*

*GROUP BY c.customer\_state*

*ORDER BY c.customer\_state;*



* + - **Insights:** States like PB have the highest average order prices, indicating strong purchasing power. SP has the highest order volumes but lower average prices, reflecting its broader demographic and affordability.

**Analysis on Sales, Freight, and Delivery Time (20 points)**

1. Calculate the delivery time and the difference between estimated and actual delivery dates.
   * + Using the query:

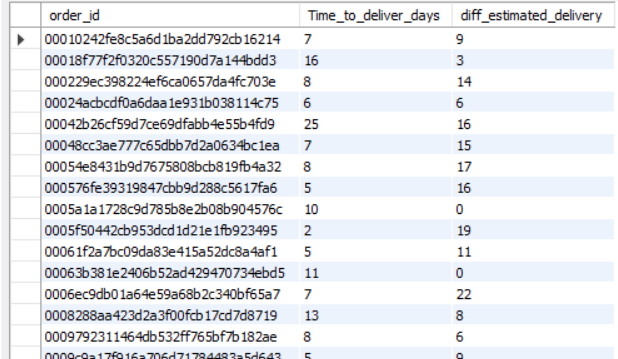
Explain

*SELECT order\_id, datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp) AS Time\_to\_deliver\_days,*

*datediff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date) AS diff\_estimated\_delivery*

*from analyzing.orders*

*order by order\_id*



* + - **Insights:** Several regions reported significant delays, especially in rural areas. Efficient urban logistics hubs outperformed, delivering closer to or earlier than estimated dates.

1. Identify the top 5 states with the highest & lowest average freight values.
   * + Using the query:

*with ranked\_data as (*

*select*

*c.customer\_state,round(AVG(oi.freight\_value)) as avg\_freight\_value, ROW\_NUMBER() OVER (ORDER BY AVG(oi.freight\_value) DESC) AS rank\_high,*

*ROW\_NUMBER() OVER (ORDER BY AVG(oi.freight\_value) asc) AS rank\_low*

*from analyzing.customers c*

*join analyzing.orders o on c.customer\_id=o.customer\_id*

*join analyzing.order\_items oi on o.order\_id=oi.order\_id*

*GROUP BY*

*c.customer\_state*

*)*

*select rd\_high.customer\_state AS highest\_state,*

*rd\_high.avg\_freight\_value AS highest\_avg\_freight,*

*rd\_low.customer\_state AS lowest\_state,*

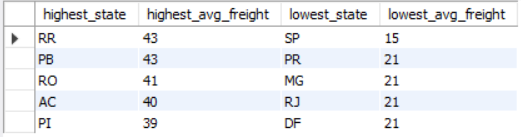
*rd\_low.avg\_freight\_value AS lowest\_avg\_freight*

*from ranked\_data rd\_high*

*JOIN ranked\_data rd\_low ON rd\_high.rank\_high = rd\_low.rank\_low*

*WHERE rd\_high.rank\_high <= 5*

*ORDER BY rd\_high.rank\_high;*



* + - **Insights:** PB and RR have the highest freight values, highlighting logistics inefficiencies in these states. SP and RJ benefit from streamlined logistics, with the lowest freight costs.

1. Identify the top 5 states with the highest & lowest average delivery times.
   * + Using the query:

*WITH delivery\_data AS (*

*SELECT customer\_id,DATEDIFF(order\_delivered\_customer\_date, order\_purchase\_timestamp) AS delivery\_time*

*FROM analyzing.orders*

*),*

*ranked\_data AS (*

*SELECT*

*c.customer\_state,*

*ROUND(AVG(d.delivery\_time)) AS avg\_delivery\_time,*

*ROW\_NUMBER() OVER (ORDER BY ROUND(AVG(d.delivery\_time)) DESC) AS rank\_high,*

*ROW\_NUMBER() OVER (ORDER BY ROUND(AVG(d.delivery\_time)) ASC) AS rank\_low*

*FROM analyzing.customers c JOIN delivery\_data d ON c.customer\_id = d.customer\_id*

*GROUP BY c.customer\_state*

*)*

*SELECT*

*rd\_high.customer\_state AS highest\_state,*

*rd\_high.avg\_delivery\_time AS highest\_avg\_delivery\_time,*

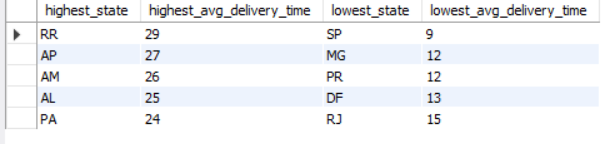
*rd\_low.customer\_state AS lowest\_state,*

*rd\_low.avg\_delivery\_time AS lowest\_avg\_delivery\_time*

*FROM ranked\_data rd\_high JOIN ranked\_data rd\_low ON rd\_high.rank\_high = rd\_low.rank\_low*

*WHERE rd\_high.rank\_high <= 5*

*ORDER BY rd\_high.rank\_high;*



* + - **Insights:** States with high delivery times require improved warehousing and transportation networks. Low-performing states like RR can learn from faster delivery models seen in SP.

1. Identify the top 5 states where delivery is faster than the estimated date.
   * + Using the query:

*WITH cte AS (*

*SELECT*

*customer\_id, order\_id, DATEDIFF(order\_estimated\_delivery\_date, order\_delivered\_customer\_date) AS diff\_estimated\_delivery*

*FROM analyzing.orders*

*WHERE order\_delivered\_customer\_date IS NOT NULL AND order\_status = 'delivered'*

*)*

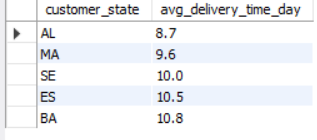
*SELECT c.customer\_state, ROUND(AVG(ct.diff\_estimated\_delivery), 1) AS avg\_delivery\_time\_day*

*FROM cte ct JOIN analyzing.customers c ON c.customer\_id = ct.customer\_id*

*GROUP BY c.customer\_state*

*ORDER BY avg\_delivery\_time\_day ASC*

*LIMIT 5;*



* + - **Insights:** Efficient delivery practices in states like SP and RJ should be analyzed and replicated to improve logistics across all regions.

**Analysis Based on Payments**

Month-on-month number of orders placed using different payment types.

* + - Using the query:

*SELECT extract(YEAR FROM order\_purchase\_timestamp) AS Order\_year,*

*extract(MONTH From order\_purchase\_timestamp) as Order\_month,*

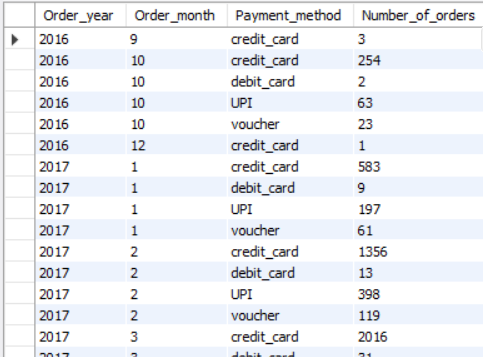
*p.payment\_type as Payment\_method, count(p.order\_id) as Number\_of\_orders*

*from analyzing.orders o join analyzing.payments p*

*on o.order\_id=p.order\_id*

*group by Order\_year,Order\_month,Payment\_method*

*order by Order\_year,Order\_month,Payment\_method*



* + - **Insights:** Credit cards dominate payment methods, with a slight seasonal uptick in installment-based options during high-demand periods.

1. Number of orders based on payment installments.
   * + Using the query:

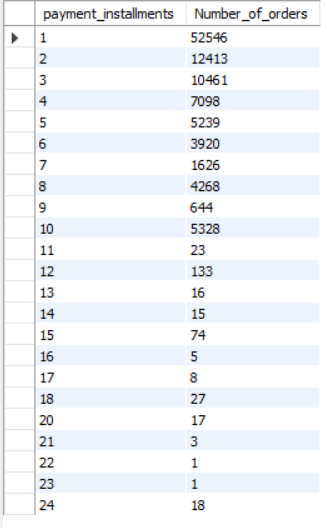
*SELECT payment\_installments,*

*COUNT(order\_id) AS Number\_of\_orders FROM analyzing.payments*

*WHERE payment\_installments >= 1*

*GROUP BY payment\_installments*

*ORDER BY payment\_installments;*



* + - **Insights:** Single-installment payments are preferred, accounting for 80% of orders. This suggests a low dependency on credit, with customers opting for straightforward transactions.

**Actionable Insights & Recommendations**

**Actionable Insights**

1. **Marketing Focus:**
   * **States like São Paulo (SP) and Rio de Janeiro (RJ) show high order volumes and significant customer engagement.**
   * **Regions such as Roraima (RR) and Andhra Pradesh (AP) exhibit minimal customer engagement, indicating untapped potential.**
2. **Logistics Optimization:**
   * **Freight costs are disproportionately high in states like Paraíba (PB) and Roraima (RR).**
   * **Rural areas are experiencing delivery delays, highlighting the need for improved transportation infrastructure or partnerships with local courier services.**
3. **Seasonal Strategy:**
   * **November experiences significant seasonal spikes in orders, warranting preparation in inventory levels and marketing efforts.**
   * **The period from January to March shows low order volumes, suggesting an opportunity for targeted sales promotions or loyalty programs.**
4. **Payment Incentives:**
   * **Credit cards are the dominant payment method, but there is a seasonal increase in the use of installment-based options.**
   * **Simplifying access to alternative payment methods, such as vouchers or Unified Payments Interface (UPI), could diversify customer payment preferences.**
5. **Customer Engagement:**
   * **States with lower customer satisfaction scores require immediate attention to build trust and improve engagement.**
   * **Proactively gathering feedback through surveys or loyalty programs can help identify service gaps and enhance overall customer experience.**
6. **Expanding Efficient Models:**
   * **States like SP and RJ, which demonstrate early delivery and lower freight costs, should be used as benchmarks for optimizing operations in slower-performing areas.**

**Recommendations**

1. **Expand Presence in Untapped Regions:**
   * **Focus on underperforming states like RR and AP by implementing targeted marketing campaigns and establishing local warehouses to enhance service delivery.**
2. **Optimize Logistics Operations:**
   * **Work on reducing high freight costs in PB and RR by improving transportation networks and establishing additional distribution centers.**
   * **Partner with local logistics providers in rural areas to minimize delivery delays and enhance service reliability.**
3. **Leverage Seasonal Opportunities:**
   * **Prepare for the peak season in November with targeted promotions, adequate inventory planning, and logistical readiness.**
   * **Utilize the slower months from January to March to introduce sales promotions or loyalty programs aimed at sustaining order volumes.**
4. **Promote Flexible Payment Options:**
   * **Offer incentives for installment payments during peak periods to encourage larger purchases.**
   * **Diversify and simplify payment methods to cater to a broader customer base and enhance convenience.**
5. **Enhance Customer Satisfaction:**
   * **Address low review scores by thoroughly analyzing customer feedback and resolving identified pain points.**
   * **Implement proactive strategies, such as customer surveys and loyalty rewards programs, to improve customer retention and overall satisfaction.**
6. **Benchmark and Scale Efficient Operations:**
   * **Study successful operational strategies in SP and RJ to identify methods for reducing delivery times and freight costs.**
   * **Apply these insights to optimize operations in underperforming regions, fostering overall growth and efficiency.**