

Business Case: Target SQL

About Target:

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

About Business Case:

This business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

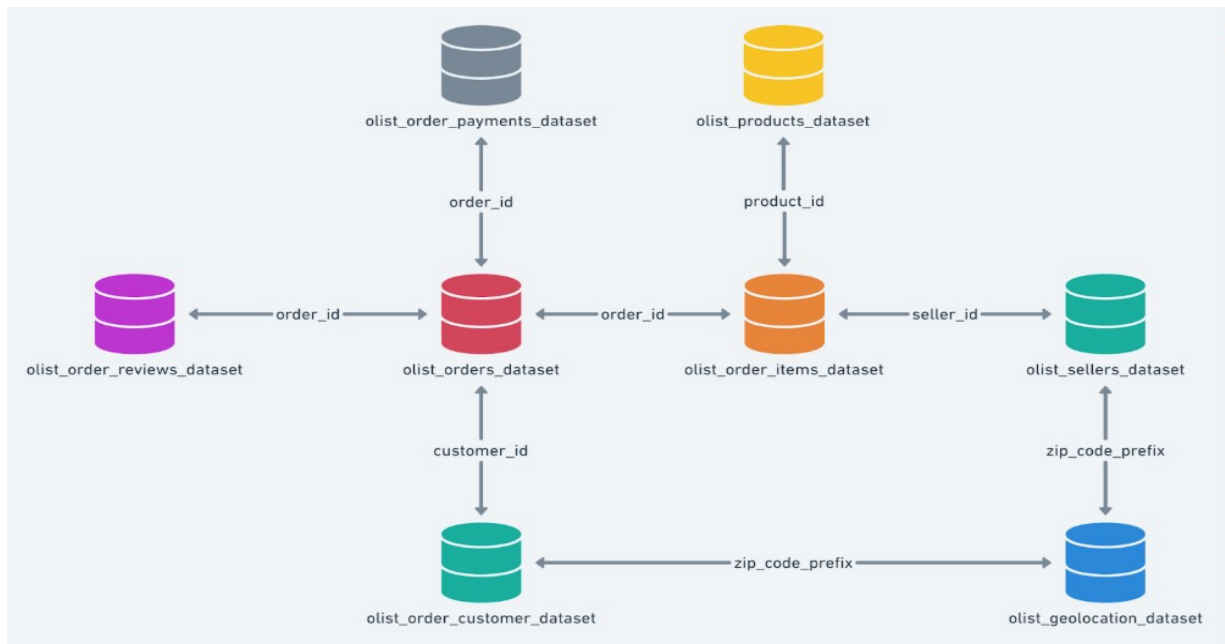
By analysing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

Tables in the Dataset:

The data is available in 8 csv files:

1. customers.csv
2. sellers.csv
3. order_items.csv
4. geolocation.csv
5. payments.csv
6. reviews.csv
7. orders.csv
8. products.csv

Dataset schema:



Analysis

1.Exploratory analysis

1.1 Data type of all columns in the “customers” table.

Query:

```
select column_name,data_type
from `Target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers'
ORDER BY ordinal_position;
```

Result:


Query results			Save results
Job information	Results	Chart	JSON
Execution details	Execution graph		
Row	column_name	data_type	
1	customer_id	STRING	
2	customer_unique_id	STRING	
3	customer_zip_code_prefix	INT64	
4	customer_city	STRING	
5	customer_state	STRING	

1.1.0 Data type of all columns in all tables from dataset.

Query:

```
select
table_schema,
table_name,
column_name,
data_type
from `Target.INFORMATION_SCHEMA.COLUMNS`
```

Result:

Query results						 Save results ▾
Job information						
Results						
Chart						
JSON						
Execution details						
Execution graph						
Row	table_schema ▾	table_name ▾	column_name ▾	data_type ▾		
1	Target	order_items	order_id	STRING		
2	Target	order_items	order_item_id	INT64		
3	Target	order_items	product_id	STRING		
4	Target	order_items	seller_id	STRING		
5	Target	order_items	shipping_limit_date	TIMESTAMP		
6	Target	order_items	price	FLOAT64		
7	Target	order_items	freight_value	FLOAT64		
8	Target	sellers	seller_id	STRING		
9	Target	sellers	seller_zip_code_prefix	INT64		
10	Target	sellers	seller_city	STRING		

Insights:


The dataset is clean and well-structured, allowing accurate analysis of trends, relationships, and distributions. Clear data types enable effective application of statistical methods and model building

1.2 Get the time range between which the orders were placed.

Query:

```
select min(order_purchase_timestamp) as first_order_date_time,
max(order_purchase_timestamp) as last_order_date_time
from `Target.orders`;
```

Result:

Query results				 Save results ▾
Job information				
Results				
Chart				
JSON				
Execution details				
Execution graph				
Row	first_order_date_time ▾	last_order_date_time ▾		
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC		

Insights:

From the results, Target in Brazil dataset has the details of orders that are placed between 2016 September and 2018 October

1.3 Count the Cities & States of customers who ordered during the given period

Query:

```
select
  c.customer_city,
  c.customer_state,
  count(*) AS orders_count
from
  Target.customers c
join Target.orders o ON c.customer_id = o.customer_id
group by
  c.customer_city, c.customer_state
order by
  orders_count desc
```

Result:

Row	customer_city	customer_state	orders_count
1	sao paulo	SP	15540
2	rio de janeiro	RJ	6882
3	belo horizonte	MG	2773
4	brasilia	DF	2131
5	curitiba	PR	1521
6	campinas	SP	1444
7	porto alegre	RS	1379
8	salvador	BA	1245
9	guarulhos	SP	1189
10	sao bernardo do campo	SP	938

Insights:

The City Sao Paulo plays the major role in Target market regards with purchase or orders placed from the given dataset

2. In-depth Exploration:

2.1 Growing trend in the no. of orders placed in each month over the past years

Query:

```
select format_datetime("%B",order_purchase_timestamp) as month_name,extract(month from
order_purchase_timestamp) as month,extract(year from order_purchase_timestamp) as
year,count(order_id) as total_order_placed
from `Target.orders`
group by month,year,month_name
order by year desc, month desc
```

Result:

Row	month_name	month	year	total_order_placed
1	October	10	2018	4
2	September	9	2018	16
3	August	8	2018	6512
4	July	7	2018	6292
5	June	6	2018	6167
6	May	5	2018	6873
7	April	4	2018	6939
8	March	3	2018	7211
9	February	2	2018	6728
10	January	1	2018	7269

Insights:

The dataset shows a strong upward trend in orders from 2016 to 2018, with order volumes growing significantly each year. Monthly order counts in 2017 and 2018 remain consistently high, reflecting sustained business growth over time. Monthly order volume trends reflect both seasonal buying behaviours and business cycle dynamics that can be leveraged for strategic planning.

2.2. Monthly seasonality in terms of the no. of orders being placed

Query:

```
select
month_name,
month,
avg(total_order_placed) as avg_monthly_orders
from
(select format_datetime("%B",order_purchase_timestamp) as month_name
,extract(month from order_purchase_timestamp) as month,
extract(year from order_purchase_timestamp) as year,count(order_id) as total_order_placed
from `Target.orders`
group by month,year,month_name
order by year desc, month desc) t

group by t.month_name,t.month
order by avg_monthly_orders desc
```

Result:

Row	month_name	month	avg_monthly_orders
1	November	11	7544.0
2	August	8	5421.5
3	May	5	5286.5
4	July	7	5159.0
5	March	3	4946.5
6	June	6	4706.0
7	April	4	4671.5
8	February	2	4254.0
9	January	1	4034.5
10	December	12	2837.0
11	October	10	1653.0
12	September	9	1435.0

Insights:

November is the peak month with the highest average orders (~7544), indicating strong seasonal demand. August, May, July, and March also show high average orders, suggesting these months consistently generate significant order volumes. The data indicates distinct monthly seasonality or festive in Brazil, in order volumes, with certain months consistently outperforming others.

2.2. Time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

Query:

```
select case when (extract(hour from order_purchase_timestamp)) between 0 and 6 then 'Dawn'
         when (extract(hour from order_purchase_timestamp)) between 7 and 12 then 'Mornings'
         when (extract(hour from order_purchase_timestamp)) between 13 and 18 then
'Afternoon'
         when (extract(hour from order_purchase_timestamp)) between 19 and 23 then 'Night'
         end as order_time,
        count(order_id) as total_order_placed
from `Target.orders`
group by order_time
order by total_order_placed desc
```

Result:

Row	order_time	total_order_placed
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insights:

Brazilian customers tend to order more during the afternoon and night hours than in the morning, but very few orders are placed during the dawn hours.

3. Evolution of E-commerce orders in the Brazil region

3.1 Month-on-month no. of orders placed in each state.

Query:

```
select c.customer_state,
       format_datetime("%B",o.order_purchase_timestamp) as month_name,
       extract(month from o.order_purchase_timestamp) as month,
       extract(year from o.order_purchase_timestamp) as year,
       count(o.order_id) as total_order_placed
from `Target.orders` o join
`Target.customers` c on
o.customer_id = c.customer_id
group by c.customer_state,month_name,month,year
order by total_order_placed desc
```

Result:

Row	customer_state	month_name	month	year	total_order_placed
1	SP	August	8	2018	3253
2	SP	May	5	2018	3207
3	SP	April	4	2018	3059
4	SP	January	1	2018	3052
5	SP	March	3	2018	3037
6	SP	November	11	2017	3012
7	SP	July	7	2018	2777
8	SP	June	6	2018	2773
9	SP	February	2	2018	2703
10	SP	December	12	2017	2357

3.1.1 Month-on-month no. of orders placed in each state.

Query:

```
select customer_state,time_period,total_orders,
       lag(total_orders) over(partition by customer_state order by year,month) as
previous_month_orders,
       round((((total_orders - lag(total_orders) over(partition by customer_state order by
year,month))
       /lag(total_orders) over(partition by customer_state order by year,month))*100,2) as
MoM_growth
from

(select t1.customer_state,
       t1.time_period,t1.year,t1.month,
       count(*) as total_orders from

(select o.order_id,o.order_purchase_timestamp,
       extract(year from o.order_purchase_timestamp) as year,
       extract(month from o.order_purchase_timestamp) as month,
       format_date('%b %Y',date(o.order_purchase_timestamp)) as time_period,
       c.customer_state
from `Target.orders` o join
`Target.customers` c on
o.customer_id = c.customer_id
order by year,month)t1
group by t1.customer_state,
t1.time_period,t1.year,t1.month) t2
order by MoM_growth desc
```

Result:

Row	customer_state	time_period	total_orders	previous_month...	MoM_growth
1	PR	Jan 2017	65	1	6400.0
2	SP	Oct 2016	113	2	5550.0
3	RS	Oct 2016	24	1	2300.0
4	TO	Aug 2017	15	1	1400.0
5	MS	Feb 2017	11	1	1000.0
6	PI	Jan 2017	7	1	600.0
7	BA	Jan 2017	25	4	525.0
8	AL	Feb 2017	12	2	500.0
9	PB	Feb 2017	12	2	500.0
10	AM	Jul 2017	5	1	400.0

Insights:

The Month-on-Month order placement count is high in the month Jan 2017 in state 'PR'. Customers from the state of 'SP' consistently rank among the top 10 in terms of monthly order placements. In contrast, several states such as 'RS', 'SC', and 'RJ' have customers who have made significantly fewer orders monthly

3.2 The customers distribution across all the states

Query:

```
select customer_state ,
       count(distinct customer_unique_id) as customer_count
from `Target.customers`
group by customer_state
order by customer_count desc
```

Result:

Row	customer_state	customer_count
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952

Insights:

The state of 'SP' holds the top position with a greater number of customers, followed by 'RJ' and 'MG'. Conversely, the states 'RR', 'AP', and 'AC' have a significantly lower number of customers making purchases at Target.

4. Impact on Economy

4.1 The % increase in the cost of orders from year 2017 to 2018 (months between Jan to Aug only).

Query:

```
with final_data as (  
  select order_id,  
         extract(year from order_purchase_timestamp) as year,  
         extract(month from order_purchase_timestamp) as month  
  from `Target.orders`  
  where extract(year from order_purchase_timestamp) in (2017,2018) and  
        extract(month from order_purchase_timestamp) between 1 and 8 and  
        order_status = 'delivered'  
  order by year  
)  
  
data1 as (  
  select year,sum(payment_value) as cost1  
  from `Target.payments` p join final_data f  
  on p.order_id = f.order_id  
  where year = 2017  
  group by year)  
  
data2 as (select year,sum(payment_value) as cost2  
  from `Target.payments` p join final_data f  
  on p.order_id = f.order_id  
  where year = 2018  
  group by year)  
  
select round(((d2.cost2 - d1.cost1)/d1.cost1) * 100,2) as percent_increase  
  from data1 d1 cross join data2 d2
```

Result:

Row	percent_increase
1	143.33

Insights:

There is 143% increase in cost of orders from the year 2017 to 2018 including the only months between Jan to Aug

4.2 Total & Average value of order price for each state

Query:

```
select c.customer_state, sum(o.price) as total_order_price, avg(o.price) as avg_order_price
from `Target.customers` c
join
`Target.orders` o on
c.customer_id = o.customer_id
join
`Target.order_items` oi on
o.order_id = oi.order_id
group by c.customer_state
order by total_order_price desc
```

Result:

Row	customer_state	total_order_price	avg_order_price
1	SP	5202955.050001...	109.6536291597...
2	RJ	1824092.669999...	125.1178180945...
3	MG	1585308.029999...	120.7485741488...
4	RS	750304.020000...	120.3374530874...
5	PR	683083.760000...	119.0041393728...
6	SC	520553.34000001	124.6535775862...
7	BA	511349.990000...	134.6012082126...
8	DF	302603.939999...	125.7705486284...
9	GO	294591.949999...	126.2717316759...
10	ES	275037.309999...	121.9137012411...

Insights:

The state of 'SP' has produced significant revenue for Target Brazil's operations, with an average order price of 109, followed by 'RJ' and 'MG'. In contrast, the state 'RR' has yielded low revenue, trailed by 'AP' and 'AC'.

4.3 The Total & Average value of order freight for each state.

Query:

```
select c.customer_state,  
       round(sum(oi.freight_value)) as total_freight_value,  
  
       round(avg(oi.freight_value)) as avg_frieght_value  
  
from `Target.customers` c  
join  
`Target.orders` o on  
c.customer_id = o.customer_id  
join  
`Target.order_items` oi on  
o.order_id = oi.order_id  
  
group by c.customer_state  
  
order by total_freight_value desc
```

Result:

Row	customer_state	total_freight_value	avg_frieght_value
1	SP	718723.0	15.0
2	RJ	305589.0	21.0
3	MG	270853.0	21.0
4	RS	135523.0	22.0
5	PR	117852.0	21.0
6	BA	100157.0	26.0
7	SC	89660.0	21.0
8	PE	59450.0	33.0
9	GO	53115.0	23.0
10	DF	50625.0	21.0

Insights:

The state of 'SP' has produced significant freight value for Target Brazil's operations, with an average order price of 15, followed by 'RJ' and 'MG'. In contrast, the state 'RR' has yielded low freight value, trailed by 'AP' and 'AC'.

5. Analysis based on sales, freight and delivery time.

5.1 The no. of days taken to deliver each order from the order's purchase date as delivery time. Also, the difference (in days) between the estimated & actual delivery date of an order.

Query:

```
select order_id,  
       timestamp_diff(order_delivered_customer_date, order_purchase_timestamp,day) as  
time_to_delivery,  
       timestamp_diff(order_estimated_delivery_date, order_delivered_customer_date,day) as  
diff_estimated_delivery,  
from `Target.orders`  
where order_status = 'delivered'
```

Result:

Row	order_id	time_to_delivery	diff_estimated_d...
1	bfb0f9bdef84302105ad712db...	54	-36
2	98974b076b01553d49ee64679...	43	6
3	c4b41c36dd589e901f6879f25a...	36	14
4	d2292ff2201e74c5db154d1b7a...	29	20
5	95e01270fcb9e986342340010...	30	19
6	ed8c7b1b3eb256c70ce0c7423...	44	5
7	5cc475c7c03290048eb2e742c...	68	-18
8	6b3ee7697a02619a0ace2b3f0a...	47	2
9	3b2ca3293a7ce539ea2379d70...	43	7
10	b2f92b2f7047cd8b35580d629d...	43	7

Insights:

It is found that the maximum days taken to deliver is 209 days and minimum is 0 days. The dataset shows a product is delivered 146 days ahead of estimated delivery date which is unusually early and might warrant further investigation for data accuracy or special cases.

5.2 The top 5 states with the highest & lowest average freight value

States with higher freight value

Query:

```
select c.customer_state,round(avg(oi.freight_value)) as top_5_freight_value,  
from `Target.customers` c join
```

```

`Target.orders` o on
c.customer_id = o.customer_id
join
`Target.order_items` oi on
o.order_id = oi.order_id
group by 1
order by 2 desc
limit 5

```

Result:

Row	customer_state	top_5_freight_value
1	RR	43.0
2	PB	43.0
3	RO	41.0
4	AC	40.0
5	PI	39.0

States with lowest freight value

Query:

```

select c.customer_state, round(avg(oi.freight_value)) as bottom_5_freight_value,
from `Target.customers` c join
`Target.orders` o on
c.customer_id = o.customer_id
join
`Target.order_items` oi on
o.order_id = oi.order_id
group by 1
order by 2
limit 5

```

Result:

Row	customer_state	bottom_5_freight...
1	SP	15.0
2	PR	21.0
3	RJ	21.0
4	SC	21.0
5	MG	21.0

Insights:

The state 'RR' is in the top 1 in accordance with freight value and 'SP' has the lowest freight value

5.3 The top 5 states with the highest & lowest average delivery time

States with highest average delivery time

Query:

```
select customer_state
from (
select
  customer_state,ROUND(AVG(time_to_delivery),2) AS avg_delivery_time
FROM (
select
  customer_state,
  TIMESTAMP_DIFF(order_delivered_customer_date, order_purchase_timestamp, day) AS
time_to_delivery,
FROM
`Target.orders` o
JOIN
`Target.customers` c
USING
(customer_id)
WHERE
order_status = "delivered") t1
GROUP BY
customer_state
ORDER BY
avg_delivery_time DESC) t2
LIMIT 5;
```

Result:

Row	customer_state	
1	RR	
2	AP	
3	AM	
4	AL	
5	PA	

States with lowest average delivery time

Query:

```
select customer_state
from (
select
  customer_state,ROUND(AVG(time_to_delivery),2) AS avg_delivery_time
FROM (
select
  customer_state,
  TIMESTAMP_DIFF(order_delivered_customer_date, order_purchase_timestamp, day) AS
time_to_delivery,
FROM
`Target.orders` o
JOIN
`Target.customers` c
USING
(customer_id)
WHERE
order_status = "delivered") t1
GROUP BY
customer_state
ORDER BY
avg_delivery_time ) t2
LIMIT 5;
```

Results:

Row	customer_state	
1	SP	
2	PR	
3	MG	
4	DF	
5	SC	

Insights:

The state 'RR' has the highest average delivery time and state 'SP' has the lowest average delivery time

5.4 The top 5 states where the order delivery is fast as compared to the estimated date of delivery

Query:

```
select customer_state,
       round(avg(t1.estimated_delivery_day - t1.actual_delivery_day),2) as delivery_time_diff
from

(select c.customer_state,
       timestamp_diff(o.order_delivered_customer_date,o.order_purchase_timestamp,day) as
actual_delivery_day,
       timestamp_diff(o.order_estimated_delivery_date,o.order_purchase_timestamp,day) as
estimated_delivery_day
from `Target.customers` c join
`Target.orders` o on
c.customer_id = o.customer_id
where o.order_status = 'delivered') t1
group by customer_state
order by delivery_time_diff desc
limit 5
```

Result:

Row	customer_state	delivery_time_diff
1	AC	20.09
2	RO	19.47
3	AP	19.13
4	AM	18.94
5	RR	16.66

Insights:

State 'AC' ranks first where orders are delivered fast compared to estimated delivery time

6. Analysis based on the payments

6.1 The month-on-month no. of orders placed using different payment types

Query:

```
select t.time_period, t.payment_type,
       count(*) as total_orders from

(select  extract(year from o.order_purchase_timestamp) as year,
        extract(month from o.order_purchase_timestamp) as month,
        format_date('%b %Y',date(o.order_purchase_timestamp)) as time_period,
        p.payment_type,o.order_id

from `Target.orders` o join
`Target.payments` p on
o.order_id = p.order_id) t
group by t.time_period,t.payment_type,t.year,t.month
order by t.year,t.month
```

Result:

Row	time_period	payment_type	total_orders
1	Sep 2016	credit_card	3
2	Oct 2016	voucher	23
3	Oct 2016	credit_card	254
4	Oct 2016	UPI	63
5	Oct 2016	debit_card	2
6	Dec 2016	credit_card	1
7	Jan 2017	credit_card	583
8	Jan 2017	UPI	197
9	Jan 2017	debit_card	9
10	Jan 2017	voucher	61

Insights:

It is found that credit card payment type is vastly used in placing orders

6.2 The no. of orders placed based on the payment installments that have been paid

Query:

```
select p.payment_installments, count(o.order_id) as total_orders
  from `Target.orders` o join
  `Target.payments` p on
  o.order_id = p.order_id
 where p.payment_installments > 0
 group by p.payment_installments
 order by total_orders desc
```

Result:

Row	payment_installm...	total_orders ▼	
1	1	52546	
2	2	12413	
3	3	10461	
4	4	7098	
5	10	5328	
6	5	5239	
7	8	4268	
8	6	3920	
9	7	1626	
10	9	644	

Insights:

Most customers prefer to pay in a single installment, while only a small portion opt for plans longer than 15 installments.

7.1. Actionable Insights:

Customer Geographical Distribution

Customers are predominantly urban and concentrated in southeastern Brazil, with São Paulo (SP) dominating due to its economic and population density. This suggests focusing marketing and logistics in high-density areas for efficiency.

SP accounts for over 40,000 units sold and the highest revenue (e.g., Pianco city noted for top revenue). Retention rates are highest in Acre/Rondônia (10%) and RJ (7%), indicating potential for loyalty programs in these areas.

Order Patterns and Behaviour

Key Statistics: ~99,441-100,000 orders total; monthly average ~4,143. Peaks in November (Black Friday, e.g., 2017 surge doubled orders) and August/May/July. Weekdays drive 77% of orders (Monday highest); repeat purchases by 2,962 customers (~6% of sales).

Insight: Orders grew from 2016-2017, stabilized in 2018, with forecasts predicting 60% growth by August 2019. Cancellation rate <1% (only 6 cancelled in sample), but undelivered orders (~15%) highlight logistics issues.

Payment Methods and Preferences

Key Statistics: Credit cards dominate (75-77% of transactions), followed by Boleto (20%), Voucher (~4%), Debit Card (~1%). Average order value (AOV) ~125 BRL; total revenue ~14-16M BRL (2016-2018).

Insight: Credit cards correlate with higher satisfaction (77% positive reviews) due to rewards/cashback. In samples, credit cards appear in ~80% of "payments.csv" entries, with installments up to 10.

Product Categories and Items

High-rated categories (>4/5 for 60% of products) drive 80% of orders/revenue. Low-margin items (e.g., small appliances ~10%) vs. high-margin (home comfort ~79%). Security/services lowest rated (3.74/5).

Delivery Performance

Key Statistics: Avg. delivery time 13.5 days; success rate ~85%. Delays common in categories like office furniture/fashion shoes (15-21 days). Processing time avg. hours; shipping to carrier avg. days.

Insight: Longer deliveries correlate with dissatisfaction (e.g., negative reviews rise with delays). In "orders.csv" sample, most delivered; some invoiced/unavailable.

6. Revenue and Satisfaction Trends

Key Statistics: Total revenue 14-16M BRL; peaks Q2 2018/November (Black Friday). Positive reviews 77-80%; negative 22-23%. High-rated products generate most revenue; imbalance in reviews (77% positive).

Insight: Freight costs higher for large items (e.g., office furniture 40.95 BRL avg.). Satisfaction tied to delivery/freight, low-cost items tolerant of freight if reasonable.

7.2 Recommendations for Targeting in Brazil

To leverage this dataset for targeted strategies, focus on personalization, regional optimization, and retention. Here are prioritized recommendations:

Geographical Targeting:

Allocate 50-60% of marketing budget to SP and RJ states, using geo-targeted ads on platforms like Google and Facebook. Offer free shipping in top cities (e.g., São Paulo, Rio) to boost conversion rates by 10-15%.

Expand to emerging areas like MG and PR with localized promotions (e.g., Portuguese-language campaigns highlighting regional products).

Timing and Promotion Optimization:

Schedule email/SMS campaigns and ads for peak hours (11 AM, 1-4 PM) to increase engagement by timing with customer availability.

Capitalize on seasonal peaks: Run Black Friday-style events in November, aiming for 20-30% sales uplift through flash sales on high-AOV categories like electronics.

Customer Segmentation and Retention:

Use RFM scores to personalize:

Loyalty programs (points, exclusive discounts) for Champions and Potential Loyalists to raise retention by 15-20%.

Win-back emails with incentives (e.g., 20% off) for At Risk and Lost segments to reduce churn.

implement surveys post-purchase to address negative reviews, targeting a 10% improvement in satisfaction scores.

Payment and Experience Enhancements:

Promote credit card usage with cashback or installment options to align with preferences and potentially increase AOV by 5-10%.

Simplify checkout and add upselling (e.g., product bundles) to reduce abandonment and encourage repeat buys.

Monitoring and Metrics:

Track KPIs like monthly active users, retention rate, AOV, and churn via dashboards. Supplement with additional data (e.g., registration details) for deeper retention analysis.

