

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).
3. 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.

	First_order_stamp	last_order_stamp
▶	2016-09-04 21:15:19	2018-10-17 17:30:18

- 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;
```

	Number of Cities	Number of States
▶	1114	4

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

In-depth Exploration

- 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
```

	Order_Year	Order_Month	Number_of_Orders
►	2016	9	4
	2016	10	324
	2016	12	1
	2017	1	800
	2017	2	1780
	2017	3	2682
	2017	4	2404
	2017	5	3700
	2017	6	3245
	2017	7	4026
	2017	8	4331
	2017	9	4285
	2017	10	4631
	2017	11	7544
	2017	12	5673
	2018	1	7269
	2018	2	6728
	2018	3	7211
	2018	4	6939
	2018	5	6873
	2018	6	6167
	2018	7	6292
	2018	8	6512
	2018	9	16
	2018	10	4

- **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;

```

	Order_year	Order_Month	Number_of_Orders	Months_Seasonality
▶	2017	11	7544	1
	2018	1	7269	1
	2018	3	7211	1
	2018	4	6939	1
	2018	5	6873	1
	2018	2	6728	2
	2018	8	6512	2
	2018	7	6292	2
	2018	6	6167	2
	2017	12	5673	2
	2017	10	4631	3
	2017	8	4331	3
	2017	9	4285	3
	2017	7	4026	3
	2017	5	3700	3
	2017	6	3245	4
	2017	3	2682	4
	2017	4	2404	4
	2017	2	1780	4
	2017	1	800	4
	2016	10	324	5
	2018	9	16	5
	2016	9	4	5
	2018	10	4	5
	2016	12	1	5

- Insights: November consistently showed high order volumes, indicating seasonal peaks. January through March were relatively low.
2. 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

	Order_time_of_day	Number_of_orders
▶	AFTERNOON 13-18	38135
	DAWN 0-6	5242
	MORNING 7-12	27733
	NIGHT 19-23	28331

- **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

	customer_state	Order_year	Order_month	Number_of_orders
▶	AC	2017	1	2
	AC	2017	2	3
	AC	2017	3	2
	AC	2017	4	5
	AC	2017	5	8
	AC	2017	6	4
	AC	2017	7	5
	AC	2017	8	4
	AC	2017	9	5
	AC	2017	10	6
	AC	2017	11	5
	AC	2017	12	5
	AC	2018	1	6
	AC	2018	2	3
	AC	2018	3	2
	AC	2018	4	4

- **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.

2. 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
```

	customer_state	number_of_customers
▶	SP	41746
	RJ	12852
	MG	11635
	RS	5466
	PR	5045
	SC	3637
	BA	3380
	DF	2140
	ES	2033
	GO	2020
	PE	1652
	CE	1336
	PA	975
	MT	907
	MA	747
	MS	715
	PB	536
	PI	495
	RN	485
	AL	413
	SE	350
	TO	280
	RO	253
	AM	148
	AC	81
	AP	68
	RR	46

- **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

	customer_state	total_freight_value	avg_freight_value
▶	AC	3687	40
	AL	15915	36
	AM	5479	33
	AP	2788	34
	BA	100157	26
	CE	48352	33
	DF	50625	21
	ES	49765	22
	GO	53115	23
	MA	31524	38
	MG	270853	21
	MS	19144	23
	MT	29715	28
	PA	38699	36
	PB	25720	43
	PE	59450	33
	PI	21218	39
	PR	117852	21
	RJ	305589	21
	RN	18860	36
	RO	11417	41
	RR	2235	43
	RS	135523	22
	SC	89660	21
	SE	14111	37
	SP	718723	15
	TO	11733	37

- **Insights:** Freight costs account for nearly 20% of the total transaction value. This suggests a need for freight optimization strategies to improve profit margins.
2. 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

```

	cost_of_2017	cost_of_2018	Percentage_Increase
▶	3669022	8694734	137%

- **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.

3. 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;

	customer_state	total_order_price	avg_order_price
▶	AC	15983	174
	AL	80315	181
	AM	22357	135
	AP	13474	164
	BA	511350	135
	CE	227255	154
	DF	302604	126
	ES	275037	122
	GO	294592	126
	MA	119648	145
	MG	1585310	121
	MS	116813	143
	MT	156454	148
	PA	178948	166
	PB	115268	191
	PE	262788	146
	PI	86914	160
	PR	683084	119
	RJ	1824090	125
	RN	83035	157
	RO	46141	166
	RR	7829	151
	RS	750304	120
	SC	520553	125
	SE	58921	153
	SP	5202960	110
	TO	49622	158

- **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

	order_id	Time_to_deliver_days	diff_estimated_delivery
▶	00010242fe8c5a6d1ba2dd792cb16214	7	9
	00018f77f2f0320c557190d7a144bdd3	16	3
	000229ec398224ef6ca0657da4fc703e	8	14
	00024acbcd0a6daa1e931b038114c75	6	6
	00042b26cf59d7ce69dfabb4e55b4fd9	25	16
	00048cc3ae777c65dbb7d2a0634bc1ea	7	15
	00054e8431b9d7675808bcb819fb4a32	8	17
	000576fe39319847cbb9d288c5617fa6	5	16
	0005a1a1728c9d785b8e2b08b904576c	10	0
	0005f50442cb953dcd1d21e1fb923495	2	19
	00061f2a7bc09da83e415a52dc8a4af1	5	11
	00063b381e2406b52ad429470734ebd5	11	0
	0006ec9db01a64e59a68b2c340bf65a7	7	22
	0008288aa423d2a3f00fcb17cd7d8719	13	8
	0009792311464db532ff765bf7b182ae	8	6
	0009c9a17fb16a706d71784483a5d643	5	0

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

2. 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;

```

	highest_state	highest_avg_freight	lowest_state	lowest_avg_freight
►	RR	43	SP	15
	PB	43	PR	21
	RO	41	MG	21
	AC	40	RJ	21
	PI	39	DF	21

- **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.

3. 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;

```

	highest_state	highest_avg_delivery_time	lowest_state	lowest_avg_delivery_time
▶	RR	29	SP	9
	AP	27	MG	12
	AM	26	PR	12
	AL	25	DF	13
	PA	24	RJ	15

- **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.

4. 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;

```

	customer_state	avg_delivery_time_day
▶	AL	8.7
	MA	9.6
	SE	10.0
	ES	10.5
	BA	10.8

- **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
```

	Order_year	Order_month	Payment_method	Number_of_orders
▶	2016	9	credit_card	3
	2016	10	credit_card	254
	2016	10	debit_card	2
	2016	10	UPI	63
	2016	10	voucher	23
	2016	12	credit_card	1
	2017	1	credit_card	583
	2017	1	debit_card	9
	2017	1	UPI	197
	2017	1	voucher	61
	2017	2	credit_card	1356
	2017	2	debit_card	13
	2017	2	UPI	398
	2017	2	voucher	119
	2017	3	credit_card	2016
	2017	3	debit_card	21

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

2. 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;
```

	payment_installments	Number_of_orders
▶	1	52546
	2	12413
	3	10461
	4	7098
	5	5239
	6	3920
	7	1626
	8	4268
	9	644
	10	5328
	11	23
	12	133
	13	16
	14	15
	15	74
	16	5
	17	8
	18	27
	20	17
	21	3
	22	1
	23	1
	24	18

- 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.