

Analyzing Target's Operations in Brazil - Extracting Valuable Insights and Actionable Recommendations

Dear Stakeholders,

I hope this message finds you well. I am writing to address a significant undertaking that has the potential to unlock valuable insights and provide actionable recommendations for Target's operations in Brazil. I have been assigned the task of analyzing an extensive dataset encompassing 100,000 orders placed between 2016 and 2018. This dataset offers a comprehensive view of various dimensions, including order status, pricing, payment and freight performance, customer location, product attributes, and customer reviews.

Our goal in undertaking this analysis is to gain a deeper understanding of Target's operations in Brazil, with a focus on several crucial aspects of the business. By meticulously examining this dataset, we aim to extract meaningful insights and offer actionable recommendations that can enhance our strategies, drive operational efficiency, and elevate the overall guest experience.

Ultimately, the objective of this endeavor is to empower our stakeholders with evidence-backed knowledge, enabling them to make informed decisions and drive Target's success in Brazil. We aim to provide you with:

- A comprehensive report encompassing our findings
- Actionable recommendations
- An implementation roadmap to guide the execution of suggested strategies.

I look forward to embarking on this analysis and delivering valuable insights that will contribute to the growth and prosperity of Target in Brazil.

Should you have any questions or suggestions throughout this process, please do not hesitate to reach out.

Thank you for your support.

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset

- a. Data type of all columns in the "customers" table.

```
SELECT column_name,data_type
FROM sqlcase1-
target.Brazil_Market.INFORMATION_SCHEMA.COLUMNS
WHERE table_name = 'customers';
```

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

- b. Get the time range between which the orders were placed.

```
select min(order_purchase_timestamp) as
First_order_timestamp,max(order_purchase_timestamp) as
Last_order_timestamp
from `Brazil_Market.orders`;
```

Row	First_order_timestamp	Last_order_timestamp
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

- c. Count the number of Cities and States in our dataset.

```
select count(distinct(geolocation_city)) as
Number_of_Cities,count(distinct(geolocation_state)) as
Number_of_States
from `Brazil_Market.geolocation`
```

Row	Number_of_Cities	Number_of_States
1	8011	27

2. In-depth Exploration

- a. Is there a growing trend in the no. of orders placed over the past years?

```
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 `Brazil_Market.orders`
GROUP BY order_year,order_month
ORDER BY order_year,order_month
```

Row	Order_Year	Order_Month	Number_of_Orders
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026

Insights:

- Number of orders has relatively increased month on month since start of operations in sep,2016 till Nov,2017 touching highest orders with 7544. Thereafter, relative downward slope is observed till the end of timeline in Oct,2018.
- 2018 is pretty consistent in terms of number of orders uptill the month of August.
- Need to plan with sales and marketing team on how to increase sales and align all teams to achieve the set target. New target setting can be done considering the numbers achieved so far and potential of the market further discussed in the report.

- b. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

```
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 `Brazil_Market.orders`
GROUP BY order_year,order_month
ORDER BY order_year,order_month) tbl
order by Number_of_Orders desc;
```

Row	Order_year	Order_Month	Number_of_Orders	Months_Seasonality
1	2017	11	7544	1
2	2018	1	7269	1
3	2018	3	7211	1
4	2018	4	6939	1
5	2018	5	6873	1
6	2018	2	6728	2
7	2018	8	6512	2
8	2018	7	6292	2
9	2018	6	6167	2
10	2017	12	5673	2
11	2017	10	4631	3

Insights:

- 2017 ended with good numbers in the last quarter with November 2017 touching highest numbers.
- 2018 started with good numbers as shown from jan to june 2018 showing relatively consistent performance.

- c. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

```
With cte as
(select order_id, order_purchase_timestamp,
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
from `Brazil_Market.orders`)

select Order_time_of_day,
count(*) as Number_of_orders
from cte
group by Order_time_of_day
order by Number_of_orders desc;
```

Row	Order_time_of_day	Number_of_orders
1	Afternoon (13-18)	38135
2	Night (19-23)	28331
3	Morning (7-12)	27733
4	Dawn (0-6)	5242

Insights:

- Afternoon time (13-18 hrs) is the time where most orders are placed followed by night time (19-23 hrs) closely followed by morning time (7-12 hrs).
- Dawn (0-6 hrs) is the least preferred time for the customers to place orders

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

- a. Get the month on month no. of orders placed in each state.

```
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 `Brazil_Market.orders` o join `Brazil_Market.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;
```

Row	customer_state	order_year	order_month	Number_of_orders
1	AC	2017	1	2
2	AC	2017	2	3
3	AC	2017	3	2
4	AC	2017	4	5
5	AC	2017	5	8
6	AC	2017	6	4
7	AC	2017	7	5
8	AC	2017	8	4
9	AC	2017	9	5
10	AC	2017	10	6
11	AC	2017	11	5

Insights:

- Can clearly observe month on month performance of each which can help management to set targets and plan operations accordingly.

b. How are the customers distributed across all the states?

```
select customer_state, count(customer_unique_id) as  
Number_of_customers  
from `Brazil_Market.customers`  
group by customer_state  
order by Number_of_customers desc;
```

Row	customer_state	Number_of_custome
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020
11	PE	1652

Insights:

- State code SP has got the maximum number of customers followed by RJ with big difference. Basis this data resources can be allocated accordingly to check cost to profit ratio in these states.
- RR has got the least number of customers. Need to device a strategy to increase footprint in those states.

4. **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**

- a. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only)

```
with y17 as
(select round(sum(p.payment_value)) as cost_of_orders_2017
 from `Brazil_Market.orders` o join
 `Brazil_Market.payments` 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_orders_2018
 from `Brazil_Market.orders` o join
 `Brazil_Market.payments` 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 cost_of_orders_2017,y18.cost_of_orders_2018,
concat(round(((y18.cost_of_orders_2018-
y17.cost_of_orders_2017)/y17.cost_of_orders_2017)*100), '%') as
Percent_increase
from y17,y18;
```

Row	cost_of_orders_2017	cost_of_orders_2018	Percent_increase
1	3669022.0	8694734.0	137%

Insights:

- The value of orders has increased significantly from 2017 to 2018 by 137%
- Clearly shows the potential business has got in this category.

- b. Calculate the Total & Average value of order price for each state.

```
select c.customer_state,cast(sum(oi.price) as INT) as
Total_order_price,cast(avg(oi.price) as INT) as
Avg_order_price
from `Brazil_Market.customers` c
join `Brazil_Market.orders` o on c.customer_id=o.customer_id
join `Brazil_Market.order_items` oi on o.order_id=oi.order_id
group by c.customer_state
order by c.customer_state;
```

Row	customer_state	Total_order_price	Avg_order_price
1	AC	15983	174
2	AL	80315	181
3	AM	22357	135
4	AP	13474	164
5	BA	511350	135
6	CE	227255	154
7	DF	302604	126
8	ES	275037	122
9	GO	294592	126
10	MA	119648	145
11	MG	1585308	121

Insights:

- State code PB has got the highest average order price with value of 191 which shows the potential purchasing power of that state. This potential can be used to pour in new products and services for that market.
- State code SP has got the lowest average order price of 110 which could infer the low purchasing power in that state or we don't have products or services relevant to that market.

- c. Calculate the Total & Average value of order freight for each state.

```
select c.customer_state,cast(sum(oi.freight_value) as INT) as
Total_freight_value,cast(avg(oi.freight_value) as INT) as
Avg_freight_value
from `Brazil_Market.customers` c
join `Brazil_Market.orders` o on c.customer_id=o.customer_id
join `Brazil_Market.order_items` oi on o.order_id=oi.order_id
group by c.customer_state
order by c.customer_state;
```

Row	customer_state	Total_freight_value	Avg_freight_value
1	AC	3687	40
2	AL	15915	36
3	AM	5479	33
4	AP	2789	34
5	BA	100157	26
6	CE	48352	33
7	DF	50625	21
8	ES	49765	22
9	GO	53115	23
10	MA	31524	38
11	MG	270853	21

Insights:

- The data shows the total and average freight value for each state.
- Which furthers down to understand how efficient is one state to other w.r.t cost involved in delivering the goods to customers.
- Required actions to be taken to improve freight cost of the states wherever its higher than expected.

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

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

```
select
order_id,date_diff(order_delivered_customer_date,order_purchase_time
stamp,day) as Time_to_deliver_days,
date_diff(order_estimated_delivery_date,order_delivered_customer_dat
e,day) as diff_estimated_delivery
from `Brazil_Market.orders`
order by order_id;
```

Row	order_id	Time_to_deliver_days	diff_estimated_delivery
1	00010242fe8c5a6d1ba2dd792...	7	8
2	00018f77f2f0320c557190d7a1...	16	2
3	000229ec398224ef6ca0657da...	7	13
4	00024acbcd0a6daa1e931b03...	6	5
5	00042b26cf59d7ce69dfabb4e...	25	15
6	00048cc3ae777c65dbb7d2a06...	6	14
7	00054e8431b9d7675808bcb8...	8	16
8	000576fe39319847cbb9d288c...	5	15
9	0005a1a1728c9d785b8e2b08...	9	0
10	0005f50442cb953dcd1d21e1f...	2	18
11	00061f2a7bc09da83e415a52d...	4	10

Insights:

- The data shows the difference between actual delivery time versus estimated delivery time of an order.
- The difference signifies the initiatives to be taken in the direction of improving this gap whether to increase seller base in those locations or maintaining enough stock at the local warehouse.

b. Find out the top 5 states with the highest & lowest average freight value.

```
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 `Brazil_Market.customers` c  
  JOIN `Brazil_Market.orders` o ON c.customer_id = o.customer_id  
  JOIN `Brazil_Market.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;
```

Row	highest_state	highest_avg_freight	lowest_state	lowest_avg_freight
1	RR	43.0	SP	15.0
2	PB	43.0	PR	21.0
3	RO	41.0	MG	21.0
4	AC	40.0	RJ	21.0
5	PI	39.0	DF	21.0

Insights:

- Average freight value is found to be highest for states like PB and RR. Which signifies the cost involved is high for delivering the goods to customers when compared to other states. There is a scope of improvement here while looking at the details where this factor and cost can be improved.
- State code SP has got the lowest average freight value which signifies how efficient it is w.r.t other states.

c. Find out the top 5 states with the highest & lowest average delivery time.

```
WITH a AS
(select
customer_id,date_diff(order_delivered_customer_date,order_purchase_time
stamp,day) as Delivery_time
from `Brazil_Market.orders`
),
ranked_data AS (
SELECT c.customer_state, ROUND(AVG(a.Delivery_time)) AS
avg_delivery_time,
ROW_NUMBER() OVER (ORDER BY AVG(a.Delivery_time) DESC) AS rank_high,
ROW_NUMBER() OVER (ORDER BY AVG(a.Delivery_time) ASC) AS rank_low
FROM `Brazil_Market.customers` c
JOIN a ON a.customer_id = c.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;
```

Row	highest_state	highest_avg_delivery_time	lowest_state	lowest_avg_delivery_time
1	RR	29.0	SP	8.0
2	AP	27.0	PR	12.0
3	AM	26.0	MG	12.0
4	AL	24.0	DF	13.0
5	PA	23.0	SC	14.0

Insights:

- The data shows the difference between actual delivery time versus estimated delivery time of an order.
- The states with high average delivery time are the ones which need attention w.r.t resources and planning in those states. If not improved can negatively impact customer experience.
- The states with low average delivery time can be the model to those with high delivery time. Strategy can be devised to further bring it down w.r.t best in the industry.

- d. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

```
with cte as
(select customer_id,order_id,
date_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as diff_estimated_delivery
from `Brazil_Market.orders`
where order_delivered_customer_date is not null or
order_status= 'delivered')

select
c.customer_state,round(avg(ct.diff_estimated_delivery),1) as
Avg_delivery_time_day
from cte ct join
`Brazil_Market.customers` c on c.customer_id=ct.customer_id
group by c.customer_state
order by avg_delivery_time_day asc
limit 5
```

Row	customer_state	Avg_delivery_time_day
1	AL	7.9
2	MA	8.8
3	SE	9.2
4	ES	9.6
5	BA	9.9

Insights:

- The data display the states where the order delivery is faster with reference to estimated time of delivery
- The data displays the states with most efficient delivery time.

6. Analysis based on the payments

- a. Find the month on month no. of orders placed using different payment types.

```
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 `Brazil_Market.orders` o join `Brazil_Market.payments` p
on o.order_id=p.order_id
GROUP BY Order_year,order_month,p.payment_type
ORDER BY Order_year,order_month,p.payment_type
```

Row	Order_Year	Order_Month	Payment_method	Number_of_Orders
1	2016	9	credit_card	3
2	2016	10	UPI	63
3	2016	10	credit_card	254
4	2016	10	debit_card	2
5	2016	10	voucher	23
6	2016	12	credit_card	1
7	2017	1	UPI	197
8	2017	1	credit_card	583
9	2017	1	debit_card	9
10	2017	1	voucher	61
11	2017	2	UPI	398

Insights:

- The data displays the various payment methods used by the customers for payment of the orders in each month.
- As observed, the most used methods include credit card, UPI, debit card and vouchers.

- b. Find the no. of orders placed on the basis of the payment installments that have been paid

```
select payment_installments, count(order_id) as  
Number_of_orders  
from `Brazil_Market.payments`  
where payment_installments >= 1  
group by payment_installments  
order by payment_installments
```

Row	payment_installment	Number_of_orders
1	1	52546
2	2	12413
3	3	10461
4	4	7098
5	5	5239
6	6	3920
7	7	1626
8	8	4268
9	9	644
10	10	5328

Insights:

- The data shows the payments made basis the number of installments paid till that time period.
- As observed, the maximum number of orders belong to the first installment paid by the customers.

ADDITIONAL QUESTIONS BY ME

Q1. Display top 3 categories of products ordered in each state.

```
with cte as
(
select c.customer_state,p.product_category,count(*) as
No_of_orders
from `Brazil_Market.customers` c join
`Brazil_Market.orders` o on c.customer_id=o.customer_id join
`Brazil_Market.order_items` oi on o.order_id=oi.order_id join
`Brazil_Market.products` p on oi.product_id=p.product_id
group by c.customer_state,p.product_category
order by c.customer_state,p.product_category,No_of_orders desc
)
select customer_state,product_category,No_of_orders,top_category
from
(select customer_state,product_category,No_of_orders,
dense_rank() over(partition by customer_state order by
No_of_orders desc) as top_category
from cte
)tbl
where top_category<=3
order by customer_state,No_of_orders desc,top_category asc
```

Row	customer_state ▼	product_category ▼	No_of_orders ▼	top_category ▼
1	AC	Furniture Decoration	12	1
2	AC	computer accessories	9	2
3	AC	sport leisure	9	2
4	AC	HEALTH BEAUTY	7	3
5	AL	HEALTH BEAUTY	63	1
6	AL	computer accessories	41	2
7	AL	Watches present	36	3
8	AM	HEALTH BEAUTY	20	1
9	AM	computer accessories	17	2
10	AM	telephony	15	3
11	AP	HEALTH BEAUTY	10	1

Insights:

- Purpose of this data is to identify most popular category of products (Top 3) in each state and to understand requirements of the customers in that state.
- If we know that computer accessories is among top selling product categories in state code AC, then there is higher possibility of selling softwares also like antivirus, windows installation..etc. We can push such products along with accessories.
- There is a category named 'Technical books' or 'Imported books'. If this category is among top in any state then it reflects there is a big student base to which we can push stationary stuff , spectacles, cool t-shirts..etc along with books.
- This data basically helps understanding demography in a state and then push new products to that market and enhancing product base and revenue in the existing channel.

Q2. Identify top 10 states with Lowest score of reviews on the placed orders.

```
With a as
(select c.customer_state, count(r.review_id) as Lowest_score
from `Brazil_Market.customers` c join
`Brazil_Market.orders` o on c.customer_id=o.customer_id join
`Brazil_Market.order_reviews` r on o.order_id=r.order_id
where r.review_score<3
group by c.customer_state),
b as
(select c.customer_state, count(r.review_id) as Total_score
from `Brazil_Market.customers` c join
`Brazil_Market.orders` o on c.customer_id=o.customer_id join
`Brazil_Market.order_reviews` r on o.order_id=r.order_id
group by c.customer_state)

select a.customer_state,
round((a.Lowest_score/b.Total_score)*100,2) as lowscore_percent
from a join b on a.customer_state=b.customer_state
order by lowscore_percent desc
limit 10
```

Row	customer_state	lowscore_percent
1	RR	23.91
2	AL	23.91
3	SE	21.78
4	MA	21.72
5	RJ	20.74
6	CE	19.86
7	PA	19.63
8	BA	18.89
9	PI	18.33
10	PE	16.65

Insights:

- The data displays top 10 states with lowest review scores.
- Depending on the low scores acceptable by management, correct course of action can be taken in that region
- Identifying exact reason for such low scores can be difficult since most customers don't comment. But this can be an indicator to identify that there is definitely some issue in a particular state which need attention.
- Direct customer feedback would be a great option where review comments are not available.
- Such data can be an eye opener as such share of low scores will need investigation and correction in terms of product or services.

Recommendations:

- ❖ With the monthly seasonality report it can be observed the potential of sales during ending of 2017 and starting of 2018. We can capitalize on our strengths and put some more marketing effort to boost sales during these potential months.
- ❖ Need to prepare all teams and channel partners, upkeep SKUs and stock to set ourselves ready for these potential months.
- ❖ For the months not doing well, need to understand the reasons whether its internal or external. If internal, we need to identify the gap and gear up in terms of marketing efforts or pricing of products or service related issues..
- ❖ If the reasons are external, need to identify them and understand what customer wants during those times whether its related to products or services. Since we already have a channel in place, we can always come up with some innovative service or product to reach out to our customers. Which would help us stay relevant in the market at any point of time and will help build confidence and trust in market.
- ❖ For ex. State code RJ shows good order numbers in 2017-2018. We can plan to allocate resources accordingly and ensure our Cost to Profit ratio is under check.
- ❖ A strategy needs to be planned for the states showing least numbers in order to increase our footprint there.
- ❖ There are specific months during which the sale in a particular state is very high. Possibilities can be festival or other reasons. We will need to ensure our preparation in terms of stock keeping and resources during those times and in that state.
- ❖ Resource allocation including man power, service support..etc can be done according to the state wise customer base.
- ❖ For the states having less customers, a strategy can be figured in order to increase customer base in those states.
- ❖ For the states having high customer base, cost to profit ratio can be checked while improving profits and reducing identified overheads

