

Business Case: Target SQL

Analysis of Target's operations in Brazil between
2016 and 2018.

By

AJAY KUMAR

ajay.saroj04@gmail.com

DSML-June-23-beginner-tue-batch

Index:

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset. ----- Page 3
2. In-depth Exploration. ----- Page 5
3. Evolution of E-commerce orders in the Brazil region. ----- Page 8
4. Impact on Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others. ----- Page 11
5. Analysis based on sales, freight and delivery time. ----- Page 16
6. Analysis based on the payments. ----- Page 22
7. Actionable Insights & Recommendations. ----- Page 24

Problem Statement:

Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

What does 'good' look like?

Identifying and measuring different parameters and KPIs from the given Dataset related to Target's operations in Brazil.

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

1. Data type of all columns in the "customers" table.

```
select column_name, data_type
from Target.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

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

```
select
  min(order_purchase_timestamp) as First_Order_Date_and_Time,
  max(order_purchase_timestamp) as Last_Order_Date_and_Time
from `Target.orders`
```

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

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

Since here we need to find the details only of customers and not of sellers, we will refer customers table and not geolocation table.

```
select
  count(distinct(customer_city)) as Total_Customers_Cities,
  count(distinct(customer_state)) as Total_Customers_States
from `Target.customers`
```

Row	Total_Customers_Cities	Total_Customers_States
1	4119	27

Insight:

The available dataset provides information regarding operations of Target in Brazil between September 2016 and October 2018. During this time period Target received orders of customers from 27 States and 4119 cities.

2. In-depth Exploration:

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

```
select
  FORMAT_DATE("%Y-%m", order_purchase_timestamp) as Year_Month,
  count(order_id) as Number_of_Orders
from `Target.orders`
group by Year_Month
order by Year_Month
```

Row	Year_Month	Number_of_Orders
1	2016-09	4
2	2016-10	324
3	2016-12	1
4	2017-01	800
5	2017-02	1780
6	2017-03	2682
7	2017-04	2404
8	2017-05	3700
9	2017-06	3245
10	2017-07	4026

Insight:

- It is being observed that it took initial five months from Sep 2016 to Jan 2017 for Target to establish their business in Brazil and gain customer trust, as we can observe during this period “Number of Order” received were very low.
- After this we can observe consistent rise in Targets business throughout year 2017. “Number of Order” data from Feb 2017 to Oct 2017 can be observed to confirm this rising trend.
- After this for next 10 months from Nov 2017 to Aug 2018 Target did a consistent business in Brazil with reasonable monthly variation.
- For month Sep 2018 and Oct 2018 very unusual sudden dip in “Number of Order” was observed which might be because of some extreme event resulting in business disruption or loss of data for that time period.

Recommendations:

- Customer trust building and marketing should be strengthened so that the initial time period of business take off can be reduced. In our case it was 5 months which is lot of time in the age of digital marketing.
- After Oct 2017 no month-to-month growth in the business was observed. Further investigation is required to find if the cause of this stagnation was market saturation or lack of initiative and marketing from Target.

- Cause of sudden dip of “Number of Order” in dataset for month Sep 2018 and Oct 2018 needs to be investigated.

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

```
select
  FORMAT_DATE("%Y-%m", order_purchase_timestamp) as Year_Month,
  count(order_id) as Number_of_Orders
from `Target.orders`
group by Year_Month
order by Year_Month
```

Row	Year_Month	Number_of_Orders
13	2017-10	4631
14	2017-11	7544
15	2017-12	5673
16	2018-01	7269
17	2018-02	6728
18	2018-03	7211
19	2018-04	6939
20	2018-05	6873
21	2018-06	6167
22	2018-07	6292

Insight:

Although we can observe some months with comparatively high volume of orders between Nov 2017 and May 2018. We don't have enough data available of multiple years to establish any type of seasonality for number of orders.

Recommendation:

Data spanning multiple years needs to be collected to establish any existing seasonality in terms of the no. of orders being placed.

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

- 0-6 hrs : Dawn
- 7-12 hrs : Mornings
- 13-18 hrs : Afternoon
- 19-23 hrs : Night

```
select Time_of_Day, count(order_id) as No_of_Order_Placed
from
(
  select order_id,
  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 'Morning'
    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 Time_of_Day
  from `Target.orders`
)
group by Time_of_day
order by No_of_Order_Placed desc
```

Row	Time_of_Day	No_of_Order_Placed
1	Afternoon	38135
2	Night	28331
3	Morning	27733
4	Dawn	5242

Insight:

It is being observed that most orders are being placed by the Brazilian customers during Afternoon followed by Night and Morning. During Dawn time least orders are placed.

Recommendation:

Any system upgrade or maintenance activity shall be planned during Dawn time i.e., between 0 and 6 Hrs.

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

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

```
select customer_state,  
FORMAT_DATE("%b-%Y", order_purchase_timestamp) as Month_Year,  
count(order_id) as Number_of_Orders  
from  
(  
  select order_id, order_purchase_timestamp, customer_state  
  from `Target.orders` as o  
  left join  
  `Target.customers` as c  
  on o.customer_id = c.customer_id  
)  
group by customer_state, Month_Year  
order by customer_state, PARSE_DATE("%b-%Y", Month_Year)
```

Row	customer_state	Month_Year	Number_of_Orders
1	AC	Jan-2017	2
2	AC	Feb-2017	3
3	AC	Mar-2017	2
4	AC	Apr-2017	5
5	AC	May-2017	8
6	AC	Jun-2017	4
7	AC	Jul-2017	5
8	AC	Aug-2017	4
9	AC	Sep-2017	5
10	AC	Oct-2017	6

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

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

Row	customer_state	Number_of_Customers
1	AC	77
2	AL	401
3	AM	143
4	AP	67
5	BA	3277
6	CE	1313
7	DF	2075
8	ES	1964
9	GO	1952
10	MA	726

Insight:

- The customers distributed across all the state is highly uneven. With very few states having good customer base,

Row	customer_state	Number_of_Customers
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882

- And a large number of states are having very few customer

Row	customer_state	Number_of_Customers
1	RR	45
2	AP	67
3	AC	77
4	AM	143
5	RO	240
6	TO	273
7	SE	342
8	AL	401
9	RN	474
10	PI	482

- Low customer base in most of the state shows untapped business potential for Target in Brazil therefore indicates an opportunity of further business growth for Target in Brazil.

Recommendation:

A team should be formed to investigate the cause of low customer number in such a large number of Brazilian states. This untapped customer base and business potential needs to be explored using best practices and learning from the states having large customer base.

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

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

```
select *,
round((((Cost_of_Order_2018 - Cost_of_Order_2017)/
Cost_of_Order_2017)*100),2) as Percent_inc_in_Cost_of_Orders
from
(
  select FORMAT_DATE("%B", order_purchase_timestamp) as Month,
  round(sum(case when EXTRACT(year from order_purchase_timestamp) = 2017
then payment_value else 0 end), 2) as Cost_of_Order_2017,
  round(sum(case when EXTRACT(year from order_purchase_timestamp) = 2018
then payment_value else 0 end), 2) as Cost_of_Order_2018
  from `Target.orders` as o
  join `Target.payments` as p
  on o.order_id = p.order_id
  where EXTRACT(month from order_purchase_timestamp) between 1 and 8
  group by Month
  order by PARSE_DATE("%B", Month)
)
```

Row	Month	Cost_of_Order_2017	Cost_of_Order_2018	Percent_inc_in_Cost_of_Orders
1	January	138488.04	1115004.18	705.13
2	February	291908.01	992463.34	239.99
3	March	449863.6	1159652.12	157.78
4	April	417788.03	1160785.48	177.84
5	May	592918.82	1153982.15	94.63
6	June	511276.38	1023880.5	100.26
7	July	592382.92	1066540.75	80.04
8	August	674396.32	1022425.32	51.61

Insight:

Even after ignoring the outlier value of January there was on an average more than 100% growth in cost of order from year 2017 to 2018. This indicates high positive effect of Target on Brazil's GDP and money circulation in Economy.

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

```
select customer_state,  
round(sum(payment_value), 2) as Total_Order_Value,  
round(avg(payment_value), 2) as Average_Order_Value  
  
from  
`Target.customers` as c  
join  
`Target.orders` as o  
on c.customer_id = o.customer_id  
join  
(  
select order_id, sum(payment_value) as payment_value  
from `Target.payments`  
group by order_id  
) as p  
on p.order_id = o.order_id  
  
group by customer_state  
order by customer_state  
--order by Total_Order_Value desc
```

Row	customer_state	Total_Order_Value	Average_Order_Value
1	AC	19680.62	242.97
2	AL	96962.06	234.77
3	AM	27966.93	188.97
4	AP	16262.8	239.16
5	BA	616645.82	182.44
6	CE	279464.03	209.18
7	DF	355141.08	165.95
8	ES	325967.55	160.34
9	GO	350092.31	173.31
10	MA	152523.02	204.18

Insight:

- There is direct relation between total Order value and number of customers. The higher total Order Value are from states where Target have larger number of customer base.

Row	customer_state	Number_of_Customers
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882

Row	customer_state	Total_Order_Value	Average_Order_Value
1	SP	5998226.96	143.69
2	RJ	2144379.69	166.85
3	MG	1872257.26	160.92
4	RS	890898.54	162.99
5	PR	811156.38	160.78

- Also, data shows that in general States having higher “Total Order Value” have lower “Average Order Value” indicating inclusion of customers from all economic strata.

Recommendation:

- The positive impact of Target’s business on Brazil’s economy and GDP shall be marketed to create better brand image value of Target in Brazil.
- To increase customer base, we should target population from all economic section of society. This will help Target to increase overall order value from that state.

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

```
select customer_state,  
round(sum(freight_value), 2) as Total_Order_Freight_Value,  
round(avg(freight_value), 2) as Average_Order_Freight_Value  
  
from  
  `Target.customers` as c  
join  
  `Target.orders` as o  
on c.customer_id = o.customer_id  
join  
  (  
    select order_id, sum(freight_value) as freight_value  
    from `Target.order_items`  
    group by order_id  
  ) oi  
on oi.order_id = o.order_id  
  
group by customer_state  
order by customer_state  
--order by Total_Order_Freight_Value desc
```

Row	customer_state	Total_Order_Freight_Value	Average_Order_Freight_Value
1	AC	3686.75	45.52
2	AL	15914.59	38.72
3	AM	5478.89	37.27
4	AP	2788.5	41.01
5	BA	100156.68	29.83
6	CE	48351.59	36.44
7	DF	50625.5	23.82
8	ES	49764.6	24.58
9	GO	53114.98	26.46
10	MA	31523.77	42.6

Insight:

It is being observed that in general larger volume of order results in less “Average Order Freight Value” indicating more cost efficiency in transportation of goods.

Row	customer_state	Total_Order_Freight_Value	Average_Order_Freight_Value
1	SP	718723.07	17.37
2	RJ	305589.31	23.95
3	MG	270853.46	23.46
4	RS	135522.74	24.95
5	PR	117851.68	23.58
22	TO	11732.68	42.05
23	RO	11417.38	46.22
24	AM	5478.89	37.27
25	AC	3686.75	45.52
26	AP	2788.5	41.01
27	RR	2235.19	48.59

Recommendation:

Possibility to implement learnings from states with less average order freight value in states with more average freight value should be explored to improve overall cost efficiency of order delivery. This will help Target to improve its profit margin.

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

1. 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.

Do this in a single query.

```
select
  order_id,
  DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) as
  Delivery_Time,
  DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date,
  DAY) as Diff_bw_Estimated_and_Delivery
from `Target.orders`
where order_status = "delivered"
```

Row	order_id	Delivery_Time	Diff_bw_Estimated_and_Delivery
1	635c894d068ac37e6e03dc54eccb6189	30	1
2	3b97562c3aee8bdedcb5c2e45a50d5e1	32	0
3	68f47f50f04c4cb6774570cfde3a9aa7	29	1
4	276e9ec344d3bf029ff83a161c6b3ce9	43	-4
5	54e1a3c2b97fb0809da548a59f64c813	40	-4
6	fd04fa4105ee8045f6a0139ca5b49f27	37	-1
7	302bb8109d097a9fc6e9cefc5917d1f3	33	-5
8	66057d37308e787052a32828cd007e58	38	-6
9	19135c945c554eebfd7576c733d5ebdd	36	-2
10	4493e45e7ca1084efcd38ddebff174dda	34	0

Insight:

There are many orders which missed the estimated delivery date.

Recommendation:

Target needs to ensure that very less to no order misses their delivery date as it creates very bad impression in mind of customer and erodes their trust and seriously affects company's reputation.

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

```
(
select customer_state,
round(avg(freight_value), 2) as Average_Order_Freight_Value

from
`Target.customers` as c
join
`Target.orders` as o
on c.customer_id = o.customer_id
join
(
select order_id, sum(freight_value) as freight_value
from `Target.order_items`
group by order_id
) oi
on oi.order_id = o.order_id

group by customer_state
order by Average_Order_Freight_Value asc
limit 5
)
union all
(
select customer_state,
round(avg(freight_value), 2) as Average_Order_Freight_Value

from
`Target.customers` as c
join
`Target.orders` as o
on c.customer_id = o.customer_id
join
(
select order_id, sum(freight_value) as freight_value
from `Target.order_items`
group by order_id
) oi
on oi.order_id = o.order_id

group by customer_state
order by Average_Order_Freight_Value asc
limit 5
offset 22
)
order by Average_Order_Freight_Value
```

Row	customer_state	Average_Order_Freight_Value
1	SP	17.37
2	MG	23.46
3	PR	23.58
4	DF	23.82
5	RJ	23.95
6	PI	43.04
7	AC	45.52
8	RO	46.22
9	PB	48.35
10	RR	48.59

Insight:

Average order freight value of bottom 5 states is almost more than double than that of top 5 states.

Recommendation:

Learning from top five states with low average order freight value should be used to improve the average freight value of other states.

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

```
(
select customer_state,
round(avg(Delivery_Time), 2) as Average_Delivery_Time

from
`Target.customers` as c
join
(
select order_id, customer_id,
DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
as Delivery_Time,
from `Target.orders`
where order_status = "delivered"
) as o
on c.customer_id = o.customer_id

group by customer_state
order by Average_Delivery_Time asc
limit 5
)

union all

(
select customer_state,
round(avg(Delivery_Time), 2) as Average_Delivery_Time

from
`Target.customers` as c
join
(
select order_id, customer_id,
DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY)
as Delivery_Time,
from `Target.orders`
where order_status = "delivered"
) as o
on c.customer_id = o.customer_id

group by customer_state
order by Average_Delivery_Time asc
limit 5
offset 22
)
order by Average_Delivery_Time
```

Row	customer_state	Average_Delivery_Time
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48
6	PA	23.32
7	AL	24.04
8	AM	25.99
9	AP	26.73
10	RR	28.98

Insight:

Average delivery time of bottom 5 states is almost double than that of top 5 states.

Recommendation:

Learning from top five states with low Average delivery time should be used to improve the Average delivery time of other states.

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

```
select customer_state,
round(avg(Diff_bw_Estimated_and_Delivery), 2) as
Avg_Diff_bw_Estimated_and_Delivery

from
`Target.customers` as c
join
(
select order_id, customer_id,
DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date,
DAY) as Diff_bw_Estimated_and_Delivery,
from `Target.orders`
where order_status = "delivered"
) as o
on c.customer_id = o.customer_id

group by customer_state
order by Avg_Diff_bw_Estimated_and_Delivery desc
limit 5
```

Row	customer_state	Avg_Diff_bw_Estimated_and_Delivery
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41

Insight:

Top five states with fast order delivery have really impressive record of on an average delivering order before the estimated delivery date.

Recommendation:

Wherever possible we may reduce the estimated delivery date of identified Products, as customers are more prone to buy Product from retailers where expected delivery time is less.

6. Analysis based on the payments:

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

```
select
FORMAT_DATE("%b-%Y", order_purchase_timestamp) as Month_Year,
payment_type,
count(order_id) as Number_of_Orders
from
(
    select p.order_id, order_purchase_timestamp, payment_type
    from `Target.orders` as o
    join
    `Target.payments` as p
    on o.order_id = p.order_id
)
group by payment_type, Month_Year
order by PARSE_DATE("%b-%Y", Month_Year), payment_type
```

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

Insight:

Four different payment type (UPI, credit card, debit card, and voucher) has been used to make payments.

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

```
select
  payment_installments,
  count(distinct order_id) as number_of_orders

from `Target.payments`

where payment_installments > 0
group by payment_installments
```

Row	payment_installments	number_of_orders
1	1	49060
2	2	12389
3	3	10443
4	4	7088
5	5	5234
6	6	3916
7	7	1623
8	8	4253
9	9	644
10	10	5315

7. Actionable Insights & Recommendations

1. Customer trust building and marketing should be strengthened so that the initial time period of business take off can be reduced. In our case it was 5 months which is lot of time in the age of digital marketing.
2. After Oct 2017 no month-to-month growth in the business was observed. Further investigation is required to find if the cause of this stagnation was market saturation or lack of initiative and marketing form Target.
3. Cause of sudden dip of “Number of Order” in dataset for month Sep 2018 and Oct 2018 needs to be investigated.
4. Data spanning multiple years needs to be collected to establish any existing seasonality in terms of the no. of orders being placed. This will help to predict expected volume of orders for those months thus giving time for additional capacity building if required.
5. Any system upgrade or maintenance activity shall be planned during Dawn time i.e., between 0 and 6 Hrs. to minimize impact on business as there is least order request during this time period.
6. Low customer base in most of the state shows untapped business potential for Target in Brazil therefore indicates an opportunity of further business growth for Target in Brazil. A team should be formed to investigate the cause of low customer number in such a large number of Brazilian states. This untapped customer base and business potential needs to be explored using best practices and learning from the states having large customer base.
7. There was on an average more than 100% growth in cost of order from year 2017 to 2018. This indicates high positive effect of Target on Brazil’s GDP and money circulation in Economy. This positive impact of Target’s business on Brazil’s economy and GDP shall be marketed to create better brand image value of Target in Brazil.
8. Data shows that in general States having higher “Total Order Value” have lower “Average Order Value” indicating inclusion of customers from all economic strata. Thus, to increase customer base, we should target population from all economic section of society. This will help Target to increase overall order value from that state.
9. Possibility to implement learnings from states with less average order freight value in states with more average freight value should be explored to improve overall cost efficiency of order delivery. This will help Target to improve its profit margin.
10. There are many orders which missed the estimated delivery date. Target needs to ensure that very less to no order misses their delivery date as it creates very bad impression in mind of customer and erodes their trust and seriously affects company’s reputation.
11. Average order freight value of bottom 5 states is almost more than double than that of top 5 states. Learning from top five states with low average order freight value should be used to improve the average freight value of other states.

12. Average delivery time of bottom 5 states is almost double than that of top 5 states. Learning from top five states with low Average delivery time should be used to improve the Average delivery time of other states.
13. Top five states with fast order delivery have really impressive record of on an average delivering order before the estimated delivery date. Wherever possible we may reduce the estimated delivery date of identified Products, as customers are more prone to buy Product from retailers where expected delivery time is less.