

Target Business Analysis

Target is one of the world's most recognized brands and one of America's leading retailers. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This business case has information of 100k orders from 2016 to 2018 made at Target in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

Q1.Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset

a. Data type of columns in a table

Answer-

Query –

```
select table_name, column_name, data_type from target.INFORMATION_SCHEMA.COLUMNS
order by table_name
```

Output-

Row	table_name ▼	column_name ▼	data_type ▼
1	customer	customer_id	STRING
2	customer	customer_unique_id	STRING
3	customer	customer_zip_code_prefix	INT64
4	customer	customer_city	STRING
5	customer	customer_state	STRING
6	geolocation	geolocation_zip_code_prefix	INT64
7	geolocation	geolocation_lat	FLOAT64
8	geolocation	geolocation_lng	FLOAT64
9	geolocation	geolocation_city	STRING
10	geolocation	geolocation_state	STRING
11	order_items	order_id	STRING
12	order_items	order_item_id	INT64
13	order_items	product_id	STRING
14	order items	seller id	STRING

b. Time period for which the data is given

Answer-

Query-

```
select DATE_DIFF(MAX(DATE(order_delivered_customer_date)),MIN(DATE(order_purchase_timestamp)),
,day) AS time_period_in_days,

DATE_DIFF(MAX(DATE(order_delivered_customer_date)),MIN(DATE(order_purchase_timestamp)) ,month)
AS time_period_in_months
from `target.orders`
```

Output-

Row	time_period_in_days	time_period_in_mont
1	773	25

THIS DATA SET HAS 773 DAYS OR 25 MONTHS OF DATA

c. Cities and States of customers ordered during the given period

Answer-

Query

```
select c.customer_city ,c.customer_state, count(o.order_id) Number_of_order
from target.customer as c
inner join target.orders as o
on c.customer_id=o.customer_id
group by customer_city,customer_state
order by count(o.order_id) desc
```

Output-

Row	customer_city	customer_state	Number_of_orde
1	sao paulo	SP	15540
2	rio de janeiro	RJ	6882
3	belo horizonte	MG	2773
4	brasilgia	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
11	niteroi	RJ	849
12	santo andre	SP	796
13	osasco	SP	746
14	santos	SP	713
15	goiania	GO	692
16	sao jose dos campos	SP	691
17	fortaleza	CE	654
18	sorocaba	SP	633

City-Sao Paulo ,State-SP ,where highest numbers of orders came

Q2. In-depth Exploration:

a. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario?
Can we see some seasonality with peaks at specific months?

Answer-

Trend-

Query-

```
with order_items_sales as (SELECT order_id,product_id,seller_id,
shipping_limit_date,price,freight_value,
ROUND(((x.price + x.freight_value) * x.quantity),2) as Total_amt_to_Pay
FROM
(
SELECT DISTINCT order_id,product_id,seller_id,shipping_limit_date, price,freight_value,
COUNT(product_id) OVER (PARTITION BY order_id, product_id) AS quantity
FROM `target.order_items`) as x
)

SELECT EXTRACT(YEAR FROM order_purchase_timestamp) as year, ROUND(SUM(Total_amt_to_Pay), 2) AS
revenue, COUNT(o.order_id) AS number_of_Orders
FROM `target.orders` as o JOIN order_items_sales AS ois
ON o.order_id = ois.order_id GROUP BY year
ORDER BY year
```

Output-

Row	year	revenue	number_of_Orders
1	2016	57183.21	332
2	2017	7142672.43	46314
3	2018	8643697.6	55779

There is a year on year growth of number of orders and total revenue

Total revenue growth 21%

Total number of orders growth 20.4%

.Monthly trend:

Query –

```
SELECT x.month,AVG(x.num_of_orders) AS average_orders_per_month
from ( SELECT extract(year from order_purchase_timestamp) as year, extract(month from
order_purchase_timestamp) as month,
count(order_id) as num_of_orders
from `target.orders`
group by year,month order by year,month
) as x group by x.month
order by x.month
```

Output-

Row	month	average_orders_per...
1	1	4034.5
2	2	4254.0
3	3	4946.5
4	4	4671.5
5	5	5286.5
6	6	4706.0
7	7	5159.0
8	8	5421.5
9	9	1435.0
10	10	1653.0
11	11	7544.0
12	12	2837.0

During the month of November month the sales increases.

September, October, December are the months where the sales are lower than average.

B.What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

Answer-

Query-

```
WITH order_time AS (SELECT X.HOUR,COUNT(X.order_id) as number_of_orders,
CASE
WHEN X.HOUR BETWEEN 0 AND 6 THEN 'Dawn'
WHEN X.HOUR BETWEEN 7 AND 12 THEN 'Morning'
WHEN X.hour BETWEEN 13 AND 18 THEN 'Afternoon'
WHEN X.hour BETWEEN 19 and 23 THEN 'night'
END AS time
FROM (SELECT order_id, EXTRACT (HOUR FROM order_purchase_timestamp) AS HOUR FROM
`target.orders`) AS X GROUP BY X.HOUR)
SELECT time,SUM(number_of_orders) AS total_order
FROM order_time
GROUP BY time
ORDER BY total_order DESC
```

Output-

Row	time	total_order
1	Afternoon	38135
2	night	28331
3	Morning	27733
4	Dawn	5242

In afternoon the number of orders increases

during Dawn number of orders are significantly low

Q3. Evolution of E-commerce orders in the Brazil region:

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

Answer-

```
Query- select customer_state,extract(month from order_purchase_timestamp) as month,
count(o.order_id) Number_of_order
from `target.customer` as c
inner join `target.orders` as o
on c.customer_id = o.customer_id
group by customer_state,month
order by count(o.order_id) desc
```

Output-

Row	customer_state	month	Number_of_order
1	SP	8	4982
2	SP	5	4632
3	SP	7	4381
4	SP	6	4104
5	SP	3	4047
6	SP	4	3967
7	SP	2	3357
8	SP	1	3351
9	SP	11	3012
10	SP	12	2357
11	SP	10	1908

State SP makes the highest number of orders

B. Distribution of customers across the states in Brazil

Answer-

Query-

```
with state_wise_customer as
(select distinct customer_state, COUNT(customer_id) over (partition by customer_state) as
Number_of_customers from `target.customer`), customer_summ as (select count(customer_id)
total_customer from `target.customer`)

select customer_state Customer_state,
Number_of_customers,round((Number_of_customers/(select total_customer from customer_summ)
)*100,2) as percentage_of_customer
from state_wise_customer
```

Output-

Row	Customer_state ▼	Number_of_custome	percentage_of_custo
1	SP	41746	41.98
2	RJ	12852	12.92
3	MG	11635	11.7
4	RS	5466	5.5
5	PR	5045	5.07
6	SC	3637	3.66
7	BA	3380	3.4
8	DF	2140	2.15
9	ES	2033	2.04
10	GO	2020	2.03
11	PE	1652	1.66
12	CE	1336	1.34

Q4-Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

- Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use "payment_value" column in payments table

Answer-

Query-

```
with order_month_year as
(select order_id, extract(year from order_purchase_timestamp) year
from `target.orders`
where extract(year from order_purchase_timestamp) in (2017,2018) and extract(month from
order_purchase_timestamp) >= 1 and extract(month from order_purchase_timestamp) <= 8)

select year, round(sum(payment_value),2) cost_of_orders
from `target.payments` p
join order_month_year o
on o.order_id=p.order_id group by o.year
order by o.year
```

Output-

Row	year ▼	cost_of_orders ▼
1	2017	3669022.12
2	2018	8694733.84

There is a 137% growth of cost of orders from 2017 to 2018

- Mean & Sum of price and freight value by customer state

Answer-

Query-

```
select c.customer_state, round(avg(oi.price),2) mean_price, round(sum(oi.price)) sum_of_price,
round(avg(oi.freight_value),2) mean_freight_value, round(sum(oi.freight_value))
sum_of_freight_value
from `target.customer` c
join `target.orders` o on c.customer_id = o.customer_id join `target.order_items` oi on
oi.order_id=o.order_id
group by c.customer_state
order by c.customer_state
```

Output-

Row	customer_state	mean_price	sum_of_price	mean_freight_value	sum_of_freight_value
1	AC	173.73	15983.0	40.07	3687.0
2	AL	180.89	80315.0	35.84	15915.0
3	AM	135.5	22357.0	33.21	5479.0
4	AP	164.32	13474.0	34.01	2789.0
5	BA	134.6	511350.0	26.36	100157.0
6	CE	153.76	227255.0	32.71	48352.0
7	DF	125.77	302604.0	21.04	50625.0
8	ES	121.91	275037.0	22.06	49765.0
9	GO	126.27	294592.0	22.77	53115.0
10	MA	145.2	119648.0	38.26	31524.0
11	MG	120.75	1585308.0	20.63	270853.0
12	MS	142.63	116813.0	23.37	19144.0

Q5. Analysis on sales, freight and delivery time

a. Calculate days between purchasing, delivering and estimated delivery

Answer-

Query-

```
select order_id, date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as
days_to_deliver,
date_diff(order_estimated_delivery_date,order_purchase_timestamp,day) as
estimated_day_to_deliver
from `target.orders`
WHERE order_status = 'delivered'
```

Answer-

Row	order_id	days_to_deliver	estimated_day_to_de
1	635c894d068ac37e6e03dc54e...	30	32
2	3b97562c3aee8bdedcb5c2e45...	32	33
3	68f47f50f04c4cb6774570cfde...	29	31
4	276e9ec344d3bf029ff83a161c...	43	39
5	54e1a3c2b97fb0809da548a59...	40	36
6	fd04fa4105ee8045f6a0139ca5...	37	35
7	302bb8109d097a9fc6e9cefc5...	33	28
8	66057d37308e787052a32828...	38	32
9	19135c945c554eebfd7576c73...	36	33
10	4493e45e7ca1084efcd38ddeb...	34	33

b.Find time_to_delivery & diff_estimated_delivery.

Answer-

```
Query- select order_id, date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as
time_to_delivery, date_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as
diff_estimated_delivery
from `target.orders`
WHERE order_status = 'delivered'
```

Output-

Row	order_id	time_to_delivery	diff_estimated_delive
1	635c894d068ac37e6e03dc54e...	30	1
2	3b97562c3aee8bdedcb5c2e45...	32	0
3	68f47f50f04c4cb6774570cfde...	29	1
4	276e9ec344d3bf029ff83a161c...	43	-4
5	54e1a3c2b97fb0809da548a59...	40	-4
6	fd04fa4105ee8045f6a0139ca5...	37	-1
7	302bb8109d097a9fc6e9cefc5...	33	-5
8	66057d37308e787052a32828...	38	-6
9	19135c945c554eebfd7576c73...	36	-2
10	4493e45e7ca1084efcd38ddeb...	34	0

C. Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

Answer-

Query-

```
select c.customer_state,round(avg(oi.freight_value),2) as mean_freight_value,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
mean_time_to_delivery,
round(avg(date_diff(order_estimated_delivery_date,order_delivered_customer_date,day))) as
mean_diff_estimated_delivery
from `target.customer` c join `target.orders` o on c.customer_id = o.customer_id
join `target.order_items` oi on oi.order_id=o.order_id
WHERE order_status = 'delivered' group by c.customer_state
order by customer_state;
```


Output-

Row	customer_state	mean_freight_value	mean_time_to_delive	mean_diff_estimated
1	AC	40.05	20.0	20.0
2	AL	35.87	24.0	8.0
3	AM	33.31	26.0	19.0
4	AP	34.16	28.0	17.0
5	BA	26.49	19.0	10.0
6	CE	32.73	21.0	10.0
7	DF	21.07	13.0	11.0
8	ES	22.03	15.0	10.0
9	GO	22.56	15.0	11.0
10	MA	38.49	21.0	9.0
11	MG	20.63	12.0	12.0

d.Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

Answer-

Top 5 states with highest average freight value

Query-

```
select c.customer_state, round(avg(oi.freight_value),2) average_freight_value
from `target.customer` as c
join `target.orders` as o on c.customer_id = o.customer_id
join `target.order_items` oi on oi.order_id=o.order_id
WHERE order_status = 'delivered'
group by c.customer_state order by average_freight_value desc
limit 5;
```

Output-

Row	customer_state	average_freight_valu
1	PB	43.09
2	RR	43.09
3	RO	41.33
4	AC	40.05
5	PI	39.12

Top 5 states with lowest average freight value

Query- `select c.customer_state, round(avg(oi.freight_value),2) average_freight_value`

```
from `target.customer` as c
join `target.orders` as o on c.customer_id = o.customer_id
join `target.order_items` oi on oi.order_id=o.order_id
WHERE order_status = 'delivered'
group by c.customer_state order by average_freight_value asc
limit 5;
```

Output-

Row	customer_state	average_freight_valu
1	SP	15.12
2	PR	20.47
3	MG	20.63
4	RJ	20.91
5	DF	21.07

e.Top 5 states with highest/lowest average time to delivery

Answer-

Query-

Highest average time to delivery

```
select c.customer_state,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
avg_time_to_delivery
from `target.customer` c join `target.orders` o
on c.customer_id=o.customer_id
WHERE order_status = 'delivered'
group by c.customer_state order by avg_time_to_delivery asc
limit 5
```

Output-

Row	customer_state	avg_time_to_delivery
1	SP	8.0
2	PR	12.0
3	MG	12.0
4	DF	13.0
5	SC	14.0

Lowest average time to delivery

```
select c.customer_state,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
avg_time_to_delivery
from `target.customer` c join `target.orders` o
on c.customer_id=o.customer_id
WHERE order_status = 'delivered'
group by c.customer_state order by avg_time_to_delivery desc
limit 5
```

Row	customer_state	avg_time_to_delivery
1	RR	29.0
2	AP	27.0
3	AM	26.0
4	AL	24.0
5	PA	23.0

Top 5 states where delivery is really fast compared to estimated date

Answer-

Query-

```
select c.customer_state,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
avg_time_to_delivery,round(avg(date_diff(order_estimated_delivery_date,order_purchase_timestamp,
day))) as
avg_est_to_delivery
from `target.customer` c join `target.orders` o
on c.customer_id=o.customer_id
WHERE order_status = 'delivered'
group by c.customer_state order by avg_time_to_delivery asc
limit 5
```

Output-

Row	customer_state	avg_time_to_delivery	avg_est_to_delivery
1	SP	8.0	19.0
2	PR	12.0	24.0
3	MG	12.0	24.0
4	DF	13.0	24.0
5	SC	14.0	25.0

Top 5 states where delivery is really not so fast compared to estimated date

Answer-

Query-

```
select c.customer_state,
round(avg(date_diff(order_delivered_customer_date,order_purchase_timestamp,day))) as
avg_time_to_delivery,round(avg(date_diff(order_estimated_delivery_date,order_purchase_timestamp,
day))) as
avg_est_to_delivery
from `target.customer` c join `target.orders` o
on c.customer_id=o.customer_id
WHERE order_status = 'delivered'
group by c.customer_state order by avg_time_to_delivery desc
limit 5
```

Output-

Row	customer_state	avg_time_to_delivery	avg_est_to_delivery
1	RR	29.0	46.0
2	AP	27.0	46.0
3	AM	26.0	45.0
4	AL	24.0	32.0
5	PA	23.0	37.0

Q6. Payment type analysis:

1. a. Month over Month count of orders for different payment types

```
with month_payment_type
as
(select extract(month from order_purchase_timestamp) as month, p.payment_type,o.order_id from
`target.orders` o join `target.payments` p on o.order_id=p.order_id)
```

```
select * from month_payment_type pivot (count(order_id) for payment_type in
('credit_card','UPI','voucher','debit_card','not_defined'))
order by month
```

Output-

Row	month	credit_card	UPI	voucher	debit_card	not_defined
1	1	6103	1715	477	118	0
2	2	6609	1723	424	82	0
3	3	7707	1942	591	109	0
4	4	7301	1783	572	124	0
5	5	8350	2035	613	81	0
6	6	7276	1807	563	209	0
7	7	7841	2074	645	264	0
8	8	8269	2077	589	311	2
9	9	3286	903	302	43	1
10	10	3778	1056	318	54	0
11	11	5897	1509	387	70	0
12	12	4378	1160	294	64	0

People makes most transactions with credit cards and then via upi

c. Count of orders based on the no. of payment instalments.

Answer-

Query-

```
with pay_by_installments as
(select distinct payment_installments, COUNT(order_id) over (partition by payment_installments)
as no_of_order
from `target.payments` ), order_summ as (select count(order_id) total_order
from `target.payments`)

select payment_installments, no_of_order
from pay_by_installments
order by no_of_order desc
```

Output-

Row	payment_installment	no_of_order
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:

This data set contains 25 months of data or 773 days of data

There are 99441 number of customer data available in the order data set.

There are 27 states and 4119 cities from these orders are made.

There is a growing trend of e-commerce business in Brazil where there was 21% growth in revenue and 20.4% growth in number of orders was observed from 2017 to 2018

Sao Paulo made highest number of purchases

In Brazil during afternoon the number of orders were most and least in dawn

During the month of November the sells increased and September,October, December the sales were lower than average

For states like RR,AP,AM,AL,PA the average delivery time is high

for states like SP,PR,MG,DF,SC the average delivery time is low.

For states like PB ,RR,RO,AC the average freight cost is high

For states like SP,PR,MG,RJ,DF the average freight cost is low

People prefer Credit cards most for making transactions

People mostly pay in 1 installments

Recommendations:

People mostly active on phones during afternoon so pushing ads during afternoon might increase sells

As the sales are low in September, October , December some promotional offer can be given to increase sells

for some states the freight values are too high,creating more partnerships with shipping companies can reduce it

Delivery time is high as compared to other e-commerce ,which can be reduced by the above mentioned method.

Delivering products faster than estimated time can enhance customer experience.

People prefer credit cards ,so creating partnerships with credit card companies can create more discount for customers.