Business Case: Target SQL

Target Corporation is an American retail corporation that operates a of and , headquartered in , Minnesota. It is the seventh-largest retailer in the United States.Target is notable for its focus on upscale, trend-forward merchandise at l ower costs.Its stores typically sell —including clothing, household goods, electronics, toys, and more—as well as .Its m ain competitors include and .

PROBLEM STATEMENT

This business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 order s 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.

GOAL

Analyzing the given dataset to extract valuable insights:

Evaluate Operational Efficiency

Understand Customer Behavior

Optimize Product Performance

Enhance Customer Satisfaction

Drive Data-Driven Decision Making

—----------

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

—----------

ER-Diagram

The schema represents Target Brazil’s e-commerce operations, detailing the order lifecycle, customer information, produc t details, payment data, and logistics.

Entity Relationships:

Customers ↔ Orders: Linked via customer_id (Customer purchases)

Orders ↔ Order Items: Linked via order_id (Each order can have multiple items)

Orders ↔ Order Reviews: Linked via order_id (Customer feedback for each order)

Orders ↔ Payments: Linked via order_id (Multiple payment installments for a single order)

Order Items ↔ Products: Linked via product_id (Product characteristics in each order)

Order Items ↔ Sellers: Linked via seller_id (Identify the seller for each product)

Customers ↔ Geolocation: Linked via zip_code_prefix (Customer location mapping)

Sellers ↔ Geolocation: Linked via zip_code_prefix (Seller location mapping)

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

```
select column_name, data_type
from `sql-scaler-450218.Target.INFORMATION_SCHEMA.COLUMNS`
where table_name = 'customers'
```

ANALYSIS:

The customers table contains customer-related information

Get the time range between which the orders were placed.

```
select min(timestamp(order_purchase_timestamp)) as First_date_time,
max(timestamp(order_purchase_timestamp)) as Last_date_time,
```

extract(date from min(timestamp(order_purchase_timestamp))) as First_date,

extract(date from max(timestamp(order_purchase_timestamp))) as Last_date

from `Target.orders`

ANALYSIS:

The first order was placed on 4th September 2016.

The last order was placed on 17th October 2018.

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

select count(distinct customer_city) as Cities,

count(distinct customer_state) as States

from `Target.customers`

No of Customers per city per state

select customer_state as State,

count(distinct customer_city) as Cities,

count(distinct customer_id) as No_of_Customers_per_State,

from `Target.customers`

group by 1

select c.customer_city as City,

c.customer_state as State,

count(distinct o.order_id) as No_of_Customers_per_City

from `Target.customers` c

join `Target.orders` o on o.customer_id = c.customer_id

group by 1,2

order by 3 desc

limit 5

ANALYSIS:

The query indicates customers are broadly from 4119 cities belonging to 27 states.

Maximum Customers are from ‘SP’ state, followed by ‘RJ’ and ‘MG’.

Same ranking is followed by the states as well -

Sao-paulo from ‘SP’ state tops the chart.

Rio De Janeiro from ‘RJ’ comes in second.

Belo Horizonte from ‘MG’ ranks 3rd.

—----------

IN DEPTH EXPLORATION

—----------

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

select extract(year from order_purchase_timestamp) as Year,

count(*) as Orders_placed

from `Target.orders`

group by 1

order by 1

ANALYSIS:

The data depicts an increasing upward trend in the number of orders being placed every year.

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

```
select extract(year from order_purchase_timestamp) as Year,
extract(month from order_purchase_timestamp) as Month,
count(*) as Orders_placed
from `Target.orders`
group by 1,2
order by 1,2
```

Top 3 months for orders_placed Year Wise

```
with data_distribution as
(
select extract(year from order_purchase_timestamp) as Year,
FORMAT_DATETIME("%B", DATETIME(order_purchase_timestamp)) as Month,
count(*) as Orders_placed
from `Target.orders`
group by 1,2
order by 1,2 desc
)
select *
from
(
select *,
dense_rank()over(partition by d.Year order by Orders_placed desc) as TOP_3
from data_distribution d
) e
where TOP_5 <=3
order by 1
```

ANALYSIS:

Orders increased significantly from 2016 to 2018.

In 2016, there were very few orders, indicating the initial phase of operations.

2017 showed substantial growth, with a peak in November (7,544 orders), likely due to promotional events like Black Friday.

2018 continued strong growth with the highest orders in January (7,269 orders).

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

```
with zone_cte as
(
select count(*) as Number_of_orders,
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 Zone
from `Target.orders`
group by 2
order by 1
)
select *,
(z.Number_of_orders/(select count(*) from `Target.orders`))*100 as Percentage_share
from zone_cte z
```

ANALYSIS:

Afternoon Dominates: 38% of orders occur in the afternoon, making it the most active period.This suggests a preference f or midday shopping, possibly due to breaks or post-lunch browsing.

Morning Surge: 28% of orders are placed in the morning, indicating strong early-day demand.This indicates customers plac ing orders before starting their day.

Night Activity: 28% of orders happen at night, showing that customers actively shop after typical work hours.This could reflect a second wave of online shopping post-dinner.

Low Dawn Orders: Only 5% of orders occur during dawn, suggesting minimal customer activity in the early morning hours.

—----------

Evolution of E-commerce orders in the Brazil region.

—----------

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

```
With state_vs_month_cte as
(
select c.customer_state as State, format_datetime("%B",o.order_purchase_timestamp) as Month, count(o.customer_id) as No
of_orders_placed
from `Target.orders` o
join `Target.customers` c
on o.customer_id = c.customer_id
group by 1,2,extract(month from o.order_purchase_timestamp)
order by 1,extract(month from o.order_purchase_timestamp)
)
select * from
(select *,row_number()over(partition by State order by No_of_orders_placed desc) as TOP_3
```

from state_vs_month_cte s) e

where TOP_3 <=3

order by 1

ANALYSIS:

The highest performing states are:

SP :

August: 4,982 orders (Rank 1)

May: 4,632 orders (Rank 2)

July: 4,381 orders (Rank 3)

BA :

July: 405 orders (Rank 1)

May: 368 orders (Rank 2)

March: 340 orders (Rank 3)

RS :

August: 599 orders (Rank 1)

March: 569 orders (Rank 2)

July: 565 orders (Rank 3)

The lowest performing states are:

RR :

March: 8 orders (Rank 1)

June: 8 orders (Rank 2)

February: 7 orders (Rank 3)

RO :

March: 29 orders (Rank 1)

July: 27 orders (Rank 2)

May: 26 orders (Rank 3)

TO :

May: 34 orders (Rank 1)

April: 33 orders (Rank 2)

February: 28 orders (Rank 3)

How are the customers distributed across all the states?

select customer_state as State, count(customer_id) as No_of_customers

from `Target.customers`

group by customer_state

order by 2 desc

ANALYSIS:

Top 5 States by Number of Customers

SP – 41,746 customers

RJ – 12,852 customers

MG – 11,635 customers

RS – 5,466 customers

PR – 5,045 customers

SP has a significant lead with over 41k customers, accounting for a substantial portion of the total customer base.

RJ and MG are secondary hubs, but their customer volume is significantly lower compared to SP.

RS and PR also exhibit strong customer bases, suggesting these regions are key for business growth.

Bottom 5 States by Number of Customers

RR – 46 customers

AP – 68 customers

AC – 81 customers

AM – 148 customers

RO – 253 customers

Northern states like RR, AP, and AC have very low customer engagement, highlighting a potential service gap.

These states combined account for less than 1% of the total customer base.

—----------

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

—----------

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

```
with orders_total as
(
select extract(year from o.order_purchase_timestamp) as Year,
round(sum(p.payment_value),2) as Cost_of_Orders,
from `Target.payments` p
join `Target.orders` o on o.order_id = p.order_id
where extract(year from o.order_purchase_timestamp) in (2017,2018)
and extract(month from o.order_purchase_timestamp) in (1,2,3,4,5,6,7,8)
group by 1
order by 1
),
first_year as
(
select ord.Year,
ord.Cost_of_Orders
from orders_total ord
where ord.Year = 2017
),
second_year as
(
select ord.Year,
```

ord.Cost_of_Orders

from orders_total ord

where ord.Year = 2018)

select * from first_year

UNION ALL

Select s.year,

round(((s.Cost_of_Orders-f.Cost_of_Orders)*100)/f.Cost_of_Orders),2)

from

first_year f

join second_year s

on f.year<s.Year

UNION ALL

select * from second_year

ANALYSIS:

The data indicates an increase of 137% in the yearly cost of orders in 2018 as compared to 2017.

Cost_of_orders in 2017 = 3669022

Cost_of_orders in 2018 = 8694734

Increase in cost_of_orders = 8694734 - 3669022 = 5025712

% change = (5025712 / 3669022) * 100 = 136.9%

This positive change owes to the increase in number of orders during 2018 as compared to 2017 as calculated earlier.

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

select c.customer_state,

count(c.customer_state) as Count_of_order,

round(sum(oi.price),2) as Total_value_Order_price,

round(avg(oi.price),2) as Average_value_Order_price

from `Target.order_items` oi

join `Target.orders` o on o.order_id=oi.order_id

join `Target.customers` c on c.customer_id=o.customer_id

group by 1

order by 4 desc

ANALYSIS:

Average Order Value (AOV) – High vs. Low:

Highest AOV:

PB: $191.48.

AL: $180.89.

Lowest AOV:

SP: $109.65 – Despite being the largest by volume, SP has the lowest average order value.

PB, AL have higher AOV but fewer orders overall.

SP, RJ dominate in volume but have lower AOV.

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

```
select c.customer_state,
count(c.customer_state) as Count_of_freight,
round(sum(oi.freight_value),2) as Total_value_Freight_value,
round(avg(oi.freight_value),2) as Average_value_Freight_value
from `Target.order_items` oi
join `Target.orders` o on o.order_id=oi.order_id
join `Target.customers` c on c.customer_id=o.customer_id
group by 1
order by 4 desc
```

ANALYSIS:

Average Freight Value (AFV) – High vs. Low:

Highest AFV:

RR : $42.98 – Smallest volume (52 shipments) but the highest freight cost, likely due to geographic remoteness.

Lowest AOV:

SP : $15.15 – Economies of scale from high shipment volumes.

SP and RJ maintain low average freight values despite their large shipment sizes.

Northern regions like RR and PB show a significant increase in average freight cost relative to their shipment count.

—----------

Analysis based on sales, freight and delivery time.

—----------

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

Do this in a single query.

```
with deliverytime_cte as
(select order_id,
extract(date from order_delivered_customer_date) as Delivered_date,
extract(date from order_purchase_timestamp) as Ordered_date,
extract(date from order_estimated_delivery_date) as Expected_delivery_date,
date_diff(extract(date from order_delivered_customer_date),extract(date from order_purchase_timestamp),day) as Delivery_ time,
date_diff(extract(date from order_estimated_delivery_date),extract(date from order_delivered_customer_date),day) as Diff _estimated_and_actual_delivery
from `Target.orders`
where order_delivered_customer_date is not null
order by 6 desc)
(select *,
'Top_5' as Ranking
from deliverytime_cte
```

```
limit 5)
UNION ALL
(select *,
'Bottom_5' as Ranking
from deliverytime_cte
order by 6
limit 5)
```

ANALYSIS:

Top 5 Fastest Deliveries:

The fastest delivery was completed in 3 days (Order ID: 0607f0...), which was 147 days earlier than the estimated date.

All top deliveries were completed at least 109 days ahead of schedule, highlighting efficiency in handling these orders.

The average difference for the Top 5 deliveries was approximately +131 days (early).

Bottom 5 Slowest Deliveries:

The slowest delivery took 210 days, which was 188 days late beyond the estimated delivery date.

All bottom deliveries were at least 166 days late, suggesting major fulfillment delays.

The average difference for the Bottom 5 deliveries was approximately -180 days (late).

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

```
with Top_5_cte as
(
select customer_state,round(average_freight_value,2) as Avg_Freight_val
,'Top 5' as Ranking
from
(select c.customer_state,avg(oi.freight_value) as average_freight_value,
dense_rank()over(order by avg(oi.freight_value)desc) as ranking
from `Target.order_items` oi
join `Target.orders` o on o.order_id = oi.order_id
join `Target.customers` c on c.customer_id = o.customer_id
group by c.customer_state
)tb1
where ranking <=5),
Bottom_5_cte as
(
select customer_state,round(average_freight_value,2) as Avg_Freight_val
,'Bottom 5' as Ranking
from
(select c.customer_state,avg(oi.freight_value) as average_freight_value,
dense_rank()over(order by avg(oi.freight_value) asc) as ranking
from `Target.order_items` oi
join `Target.orders` o on o.order_id = oi.order_id
```

```
join `Target.customers` c on c.customer_id = o.customer_id
group by c.customer_state
)tb2
where ranking <=5
order by 2 asc
)
select * from Top_5_cte
union ALL
select * from Bottom_5_cte
order by 2 desc
```

ANALYSIS:

Top 5 States (Highest Freight Costs):

RR, PB, RO, AC, PI have the highest average freight values, ranging from 39.15 to 42.98.

These states may be remote areas or difficult to access, leading to higher transportation costs.

Bottom 5 States (Lowest Freight Costs):

DF, RJ, MG, PR, SP have the lowest freight values, ranging from 15.15 to 21.04.

These states might be major urban hubs or well-connected regions, resulting in lower logistics expenses.

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

```
with delivery_time_calc as
(select o.order_id,
o.customer_id,
c.customer_state,
extract(date from o.order_purchase_timestamp) as Ordered_date,
extract(date from o.order_delivered_customer_date) as Delivered_date,
date_diff(extract(date from o.order_delivered_customer_date),
extract(date from o.order_purchase_timestamp),day) as Delivery_time
from `Target.orders` o
join `Target.customers` c
on o.customer_id = c.customer_id
where extract(date from o.order_delivered_customer_date) is not null
order by 4,5,6
)
(select customer_state,round(avg(Delivery_time),2) as Avg_Delivery_time,'Top 5' as Ranking
from delivery_time_calc
group by customer_state
order by 2 desc
limit 5)
UNION ALL
(select customer_state,round(avg(Delivery_time),2),'Bottom 5' as Ranking
```

```
from delivery_time_calc
group by customer_state
order by 2 asc
limit 5)
```

ANALYSIS:

Top 5 states with Very Low Average Delivery Time:

SP, PR, MG, DF, SC have the lowest average delivery times ranging from 9 to 15 days.

These states have better logistics and faster delivery networks, likely due to their urban locations or centralized infr astructure.

SP has the quickest average delivery time, reflecting its position as a major hub.

Bottom 5 States with Very High Average Delivery Time:

RR, AP, AM, AL, PA have the highest average delivery times, ranging from 24 to 29 days.

These states are likely remote or less connected regions, contributing to longer delivery times.

RR experiences the slowest deliveries, aligning with its geographical position in northern Brazil and challenging transp ort routes.

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

```
(select c.customer_state,
count(*) as No_of_orders,
'Top 5'
from `Target.orders` o join `Target.customers` c on c.customer_id = o.customer_id
where extract(date from o.order_delivered_customer_date) < extract(date from o.order_estimated_delivery_date)
group by 1
order by 2 desc
limit 5)
UNION ALL
(select c.customer_state,
count(*) as No_of_orders,
'Bottom 5'
from `Target.orders` o join `Target.customers` c on c.customer_id = o.customer_id
where extract(date from o.order_delivered_customer_date) < extract(date from o.order_estimated_delivery_date)
group by 1
order by 2 asc
limit 5)
order by 2 desc
```

ANALYSIS:

Top 5 states with Highest orders :

SP, MG, RJ, RS, PR have the highest orders placed.

These states represent economic hubs with high population density and stronger logistical capabilities, driving higher o rder volumes.

SP significantly outperforms other states, handling over 3.5x more orders than MG or RJ.

MG and RJ are relatively close in order volume, reflecting a balanced demand across these regions.

Bottom 5 States with Lowest Orders:

RO, AM, AC, AP, RR have the lowest Orders.

These regions exhibit very low order volumes, especially RR with only 36 orders.

Geographical isolation, lower population, and logistical challenges likely contribute to reduced customer demand.

AM and RO, while larger in area, still receive fewer orders, indicating limited market penetration.

—----------

Analysis based on the payments:

—----------

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

```
with payments_cte as
(select p.payment_type,FORMAT_DATETIME("%B", DATETIME(order_purchase_timestamp)) as Month,count(p.order_id) as
ers,round(sum(payment_value)) as Total_payment_value,
dense_rank()over(partition by p.payment_type order by count(p.order_id))as rank_of_orders
from `Target.payments` p
join `Target.orders` o on o.order_id = p.order_id
group by 1,2)
select payment_type,Month,No_of_Orders,Total_payment_value
from payments_cte
where rank_of_orders <= 5
order by 1,3 desc
```

ANALYSIS:

By Total Value

Credit Card is the most popular payment method by a large margin, contributing to 80% of the total payment value.

UPI is the second-highest but far behind credit cards, accounting for 19% of the total payment value.

Debit Cards show low adoption, contributing the least in both order volume and payment value.

Vouchers play a minor role in transactions but have moderate usage.

Month vs payment Type

UPI:

Peak Month: January - 1715 Orders

Trend: Consistent decline from January to September.

UPI usage is seasonal, with a spike at the beginning of the year and a gradual decrease over time.

Credit Card:

Peak Month: January - 6103 orders

Trend: Consistently high with November (5897 orders)

Credit card payments are the dominant choice across all months, with a year-end boost likely due to holiday sales.

Debit Card:

Peak Month: May - 81 orders

Trend: Minimal with no clear pattern.

Debit card usage is infrequent and shows low customer preference.

Voucher:

Peak Month: February - 424 orders

Trend: Steady but with a slight decline from February to November.

Vouchers experience higher activity during the early year promotions but slow down later.

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

```
select
Payment_installments,
count(order_id) as No_of_Orders,
sum(payment_value) as Total_value_of_payments
from `Target.payments`
where payment_value>0
group by 1
order by 3 desc
```

ANALYSIS:

Single Payment Dominance:

51.4% of the orders (52,537 out of 102,164) are made with 1 installment, contributing to 38.3% of the total payment valu e.

Customers prefer paying in full rather than in multiple installments.

High-Value Installments:

10 Installments account for 5,328 orders but contribute 14.4% of the total payment value, making it the second-highest p ayment method.

2 Installments also show a significant contribution with 12,413 orders.

Longer Installments Are Rare:

Installments beyond 12 are not common, with orders dropping sharply.

Orders with 22 and 23 installments have only 1 record each, indicating very low adoption.

Mid-Range Preferences (3-10 Installments):

3-10 installments cover a substantial share of the orders and payments, accounting for approximately 37% of total orders .

Customers may prefer mid-range payment flexibility over extended installments.

OBSERVATIONS

State vs Total Orders.

Top 5 States (SP, MG, RJ, RS, PR) contribute to the bulk of the orders, with SP (38,108 orders) accounting for over 50% of the total volume.

Bottom 5 States (RO, AM, AC, AP, RR) reflect low consumer activity, with RR contributing the least (36 orders).

State vs Freight Charges

States with lower order volumes (such as RR, AP, and AC) are likely to incur higher freight costs per order due to dista

nce and less frequent deliveries.

High-order states (SP, MG) benefit from economies of scale and reduced average shipping costs.

State vs Customer Base

SP is the largest customer base, implying higher urban density and purchasing power.

States like RR and AP have limited customer penetration, suggesting room for growth through customer acquisition campaig ns.

Payment Insights

Payment Method Preferences

UPI leads both order volume and payment value, particularly in high-order states.

Credit Cards are the second most popular method, especially for high-ticket items.

Debit Cards and Vouchers account for a small fraction of total transactions.

Payment Installments

Single payments dominate (over 50% of orders).

Installments between 3-10 months are popular, while 12+ month plans have minimal adoption.

Seasonal Trends

January, November, and December are the busiest months for both UPI and Credit Card payments.

September reflects moderate activity, suggesting potential for a pre-holiday push.

RECOMMENDATIONS

High-Performing States (SP, MG, RJ, RS, PR)

Exclusive Loyalty Programs: Implement tier-based reward programs to encourage frequent purchases.

Personalized Promotions: Offer state-specific deals aligned with their preferred payment methods.

Priority Delivery: Introduce same-day or next-day delivery for these regions to maintain a competitive edge.

Low-Performing States (RO, AM, AC, AP, RR)

Localized Campaigns: Design hyper-local advertising in regional languages to engage underrepresented areas.

Strategic Partnerships: Partner with local delivery services to reduce freight costs and improve delivery speed.

Community Outreach: Use community influencers to build trust and promote the brand in smaller markets.

Reduce Delivery Time in Key Regions

Set up localized warehouses in regions with high order volumes (SP, MG).

Implement AI-based delivery routing to shorten transit times.

Smart Inventory Allocation

UPI & Credit Card Optimization

Exclusive Discounts: Offer UPI-exclusive cashback or credit card EMI discounts to drive usage.

Promote Contactless Payments: Push UPI-first marketing during festive seasons and flash sales.

Increase Adoption of Debit Cards and Vouchers

Limited-Time Offers: Provide exclusive product access or extra discounts for voucher users.

Introduce zero-cost EMIs for debit card transactions to boost large-order adoption.

Customer Acquisition & Retention

Increase Customer Base in Low-Performing States.

Drive Customer Loyalty in High-Performing Regions

CONCLUSION

In conclusion, Target's data-driven insights reveal clear opportunities to enhance operational efficiency and customer s atisfaction. By focusing on high-performing states, optimizing logistics in low-demand regions, and leveraging popular p ayment methods, Target can drive sustainable growth. Strengthening seasonal marketing strategies and tailoring payment o ptions will further boost customer engagement and sales. With a strategic, data-backed approach, Target is well-position ed to maximize market share and deliver exceptional value to its customers.