TARGET BUSINESS CASE

I. 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.

QUERY:

```
SELECT
column_name,data_type
FROM `practicesql-459007.Target.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers'
```

OUTPUT:

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

INSIGHTS:

- Identifying data types helps determine if any columns need conversion (e.g., dates stored as strings).

- Checking for missing values ensures data completeness
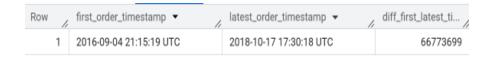
BUSINESS RECOMMENDATIONS:

- **Indexing**: Add indexes to email, registration_date, is_active or other frequently queried columns.
- **Constraints**: Enforce NOT NULL, UNIQUE, and DEFAULT values as needed to maintain data integrity.

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

QUERY:

```sql
Select min(order_purchase_timestamp) as
first_order_timestamp,max(order_purchase_timestamp) as
latest_order_timestamp,
timestamp_diff(max(order_purchase_timestamp),min(order_purchase_t
imestamp),second) as diff_first_latest_time
from `Target.orders`
```

OUTPUT:

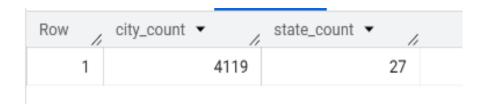| Row	first_order_timestamp ▼	latest_order_timestamp ▼	diff_first_latest_ti... ▼
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC	66773699

INSIGHTS:

- Finding the earliest and latest order dates gives the business timeline.
- Seasonal trends can be inferred if data spans multiple years.

BUSINESS RECOMMENDATION:

- Plan inventory proactively for seasonal spikes. Develop targeted marketing campaigns for upcoming peak seasons. Offer seasonal promotions or product bundles.
- Analyze average lead times. If lead times are long, investigate bottlenecks in your fulfillment process (e.g., inventory management, shipping partners, production if applicable). Communicate realistic delivery expectations to customers. Faster and more transparent lead times can significantly improve customer satisfaction.

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

QUERY:

```sql
select count(distinct customer_city) as city_count,count(distinct
customer_state) as state_count
from `Target.customers`
```

OUTPUT:

Row	city_count	state_count
1	4119	27

INSIGHTS:

- The most common cities and states help identify key markets.
- Comparing order frequency by state/city can highlight high-demand region.

BUSINESS RECOMMENDATION:

- Indicates broad market penetration. Consider targeted marketing campaigns to deepen engagement in existing strongholds and develop strategies for emerging areas.
- Evaluate your delivery partners' effectiveness in specific states/cities.

II.   In-depth Exploration:

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

QUERY:

```
select extract(year from order_purchase_timestamp) as past_year,
count(distinct order_id) as no_of_orders
from `Target.orders`
group by 1
order by 1;
```

OUTPUT:

Row	past_year	no_of_orders
1	2016	329
2	2017	45101
3	2018	54011

INSIGHTS:

- Orders jumped from **329 to 45,101**, which is a **~13,600% increase**.
- 2018 saw another **~20% increase** in orders (from 45,101 → 54,011)
- The trend is consistently upward, which is ideal for long-term forecasting and investor confidence.

BUSINESS RECOMMENDATION:

- Ensure sufficient inventory levels to prevent stockouts during peak periods. Consider automation for inventory tracking and reordering.
- Invest in robust e-commerce platforms, CRM systems, and analytics tools that can handle higher volumes and provide deeper insights. Process automation can become crucial.

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

QUERY:

```
select extract(year from order_purchase_timestamp) as year,extract(month from
order_purchase_timestamp) as month,
count(distinct order_id) as no_of_orders
from `Target.orders`
group by 1,2
order by 1,2
```

OUTPUT:

Row	year	month	no_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:

- If there's strong seasonality, marketing campaigns should align with peak shopping periods.

BUSINESS RECOMMENDATION:

- Count the total number of orders for each month (e.g., total orders in January, total orders in February, etc., *summed across all years*).
- Calculate the average number of orders for each month across all years in your dataset. This helps smooth out individual year variations and highlights the typical monthly pattern.

3. **During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)**
    1. **0-6 hrs : Dawn**
    2. **7-12 hrs : Mornings**
    3. **13-18 hrs : Afternoon**
    4. **19-23 hrs : Night**

QUERY:

```
select 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'
else 'Night' end as braz_time,
count(distinct order_id) as no_orders
from `Target.orders`
group by 1
order by 2 desc
```

OUTPUT:

Row	braz_time	no_orders
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

<u>**INSIGHTS**</u>:

- o Understanding time-of-day preferences helps in timing promotions and customer engagement effectively
- o Afternoon is Peak Order Time
- o Dawn Has Very Low Activity
- o Night and Morning Are Almost Equal

<u>**BUSINESS RECOMMENDATIONS**</u>:

- Run Promotions During Afternoon Peak
- Ensure servers, recommendation engines, and payment gateways are **scalable** and responsive in the afternoon.
- If customers browse at night but don't buy, **retarget them in the afternoon** (e.g., abandoned cart emails at noon).

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

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

<u>**QUERY**</u>:

```
select extract(month from o.order_purchase_timestamp) as
order_month,c.customer_state,count(distinct o.order_id) as no_of_orders

from `Target.orders` o

join `Target.customers` c

using(customer_id)

group by 1,2

order by no_of_orders;
```

OUTPUT:

Row	order_month	customer_state	no_of_orders
1	1	RR	2
2	9	AP	2
3	9	RR	2
4	11	RR	2
5	5	RR	3
6	10	AP	3
7	10	AM	3
8	10	RR	4
9	2	AP	4

<u>**INSIGHTS**</u>:

- Some states may show consistent growth, indicating increasing e-commerce penetration.
- If orders drop in certain months, it could indicate seasonal effects or economic factors.

BUSINESS RECOMMENDATIONS:

- Offer First-Time Buyer Incentives
- High delivery times or freight costs may be deterring repeat buyers.
- While volumes are low, small increases month-over-month could indicate successful outreach.
- Set quarterly targets for acquisition in these regions.
- Look for repeat customer_ids in these states and create **VIP or ambassador programs** to drive organic growth.

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

QUERY:

```
select customer_state, count(distinct customer_id) as no_of_customers
from `Target.customers`
group by 1
order by no_of_customers desc;
```

OUTPUT:

Row	customer_state	no_of_customers
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
12	CE	1336

INSIGHTS:

- Major cities (São Paulo, Rio de Janeiro) likely have the highest number of customers.

- Smaller states may show emerging trends, indicating new market potential.
- Some states may have high order volume but fewer customers, meaning customers place multiple order.

BUSINESS RECOMMENDATIONS:

- RJ, MG, RS, PR, these 4 states combined contribute **~30,000+ customers**
- Launch:
  - ➢ **Localized promotions** (e.g., regional festivals or holidays)
  - ➢ **Geo-targeted ads**
  - ➢ Partner with **local influencers**
- For states like **PE, CE, GO**, explore:
  - ➢ Referral programs
  - ➢ Regional onboarding offers (e.g., free delivery for 1st 3 orders)
  - ➢ **Language/local culture alignment** in ads

IV. 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).**

QUERY:

```
with year_cost as
(
  select extract(year from o.order_purchase_timestamp) as order_year,
  FORMAT_DATE('%B', o.order_purchase_timestamp) AS order_month,
  sum(p.payment_value) as cost_of_orders
  from `Target.payments` p
  join `Target.orders` o
  on p.order_id=o.order_id
  where extract(year from o.order_purchase_timestamp) in (2017,2018) and
   extract(month from o.order_purchase_timestamp) between 1 and 8
  group by 1,2
)
select order_year,order_month,cost_of_orders,
 round(100*((cost_of_orders - lag(cost_of_orders) over(order by
order_year,order_month))/lag(cost_of_orders) over(order by
order_year,order_month)),2) AS percentage_increase
from year_cost
ORDER BY order_year, order_month, percentage_increase
```

OUTPUT:

Row	order_year ▼	order_month ▼	cost_of_orders ▼	percentage_incre...
1	2017	April	417788.0300000...	null
2	2017	August	674396.3200000...	61.42
3	2017	February	291908.0099999...	-56.72
4	2017	January	138488.0399999...	-52.56
5	2017	July	592382.9200000...	327.75
6	2017	June	511276.3800000...	-13.69
7	2017	March	449863.6000000...	-12.01
8	2017	May	592918.8200000...	31.8
9	2018	April	1160785.48	95.77
10	2018	August	1022425.319999...	-11.92

INSIGHTS:

- If the percentage increase is positive, it indicates higher spending on orders in 2018 compared to 2017.
- A high percentage increase suggests growth in e-commerce activity, possibly due to more customers, higher order values, or inflation
- Although August 2018 saw strong order value, growth declined YoY — possibly due to saturation or poor campaign performance.

BUSINESS RECOMMENDATIONS:

- April-May, these months consistently show high performance or strong growth
- Recommend investing in:
  - ➢ Product launches
  - ➢ Targeted ads
  - ➢ Loyalty/discount programs
- Consider **email re-engagement** or influencer campaigns to build early momentum

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

QUERY:

```
Select distinct c.customer_state,round(sum(oi.price),2) as Total_order_price,
round(avg(oi.price),2) as Avg_order_price
from `Target.customers` c
join `Target.orders` o
on c.customer_id=o.customer_id
join `Target.order_items` oi
on o.order_id=oi.order_id
group by 1
```

```
order by 1;
```

OUTPUT:

Row	customer_state	Total_order_price	Avg_order_price
1	AC	15982.95	173.73
2	AL	80314.81	180.89
3	AM	22356.84	135.5
4	AP	13474.3	164.32
5	BA	511349.99	134.6
6	CE	227254.71	153.76
7	DF	302603.94	125.77
8	ES	275037.31	121.91
9	GO	294591.95	126.27
10	MA	119648.22	145.2

INSIGHTS:

- If some less-populated states have high total order values, it suggests a high adoption rate of online shopping.
- These states represent **potential growth markets** for businesses to target with better logistics, promotions, and localized strategies.
- If certain states have a **high average but low total order value**, it means that **fewer customers are making large purchases**

BUSINESS RECOMMENDATIONS:

- For states with high total and average order value:

    o Launch **personalized offers** and **VIP customer programs**
    o Introduce **same-day delivery** and loyalty per

- For states with low total and average value:

    o Run **geo-targeted ad campaigns** to boost visibility
    o Offer incentives like **free shipping** or **first-time discounts**

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

QUERY:

```
select distinct c.customer_state,round(sum(oi.freight_value),2) as
Total_freight_order_price,
round(avg(oi.freight_value),2) as Avg_freight_order_price
from `Target.customers` c
join `Target.orders` o
on c.customer_id=o.customer_id
join `Target.order_items` oi
on o.order_id=oi.order_id
group by 1
order by 1;
```

OUTPUT:

Row	customer_state ▼	Total_freight_ord...	Avg_freight_order...
1	AC	3686.75	40.07
2	AL	15914.59	35.84
3	AM	5478.89	33.21
4	AP	2788.5	34.01
5	BA	100156.68	26.36
6	CE	48351.59	32.71
7	DF	50625.5	21.04
8	ES	49764.6	22.06
9	GO	53114.98	22.77
10	MA	31523.77	38.26

INSIGHTS:

- States far from key logistics hubs (Amazonas, Acre, Roraima, Amapá) often have high total and average freight costs due to longer distances and higher delivery challenges.
- Customers in these regions might abandon carts due to high delivery fees.
- If some states have low total freight costs despite many orders, it indicates efficient shipping logistics or subsidized delivery cost.

BUSINESS RECOMMENDATIONS:

- For states with high average freight:
  - Partner with local third party providers
  - Explore pickup point models to reduce last-mile costs

- For states with low average freight:
  - Promote **same-day or 24hr delivery**
  - Run **express delivery upsell campaigns**
- Adjust Free Shipping Thresholds by Region
- For high-cost states, encourage **bulk orders**

V.   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.**

   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
   - **time_to_deliver = order_delivered_customer_date - order_purchase_timestamp**
   - **diff_estimated_delivery = order_delivered_customer_date - order_estimated_delivery_date**

QUERY:

```
Select order_id,
date_diff(order_delivered_customer_date,order_purchase_timestamp,day) as
time_to_deliver,
date_diff(order_delivered_customer_date,order_estimated_delivery_date,day) as
diff_estimated_delivery
from `Target.orders`
group by 1,2,3
order by 1
```

OUTPUT:

Row	order_id ▼	time_to_deliver ▼	diff_estimated_d...
1	00010242fe8c5a6d1ba2dd792...	7	-8
2	00018f77f2f0320c557190d7a1...	16	-2
3	000229ec398224ef6ca0657da...	7	-13
4	00024acbcdf0a6daa1e931b038...	6	-5
5	00042b26cf59d7ce69dfabb4e5...	25	-15
6	00048cc3ae777c65dbb7d2a06...	6	-14
7	00054e8431b9d7675808bcb81...	8	-16
8	000576fe39319847cbb9d288c...	5	-15
9	0005a1a1728c9d785b8e2b08b...	9	0
10	0005f50442cb953dcd1d21e1fb...	2	-18

INSIGHTS:

- Zero Difference (On-Time Deliveries): If the difference is zero, the logistics system is accurate and efficient.
- Positive Difference (Late Deliveries): If there's a high positive difference, it suggests delays due to issues like weather, supply chain disruptions, or delivery inefficiencies.
- Frequent late deliveries can reduce customer trust and lead to cart abandonment or negative reviews.
- Early/on-time deliveries increase customer satisfaction, leading to higher retention & repeat purchases.
- Companies might need better tracking, warehouse distribution, or delivery partnerships to improve delivery speed.

BUSINESS RECOMMENDATIONS:

- Reducing the buffer from 15–18 days to 5–7 days in many regions
- Use messages like "Customers in your area receive items 8 days early on average."
- For customers who need urgency, offer:
  - ➢ **Standard (5–9 days)**: free or discounted
  - ➢ **Express (1–3 days)**: at a premium

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

QUERY:

```sql
with highest_freight as
    (
    Select c.customer_state,round(avg(oi.freight_value),2) as
    Avg_freight_value,'Highest' as Top_5_States_Category
      from `Target.customers` c
      join `Target.orders` o
      using(customer_id)
      join `Target.order_items` oi
      using(order_id)
      group by 1
      order by 2 desc
      limit 5
    ),

lowest_freight as
    (
      select c.customer_state,round(avg(oi.freight_value),2) as
    Avg_freight_value,'Lowest' as Top_5_States_Category
      from `Target.customers` c
      join `Target.orders` o
      using(customer_id)
      join `Target.order_items` oi
      using(order_id)
      group by 1
```

```
      order by 2
      limit 5
   )
   select customer_state,Avg_freight_value,Top_5_States_Category
   from highest_freight
   union all
   select customer_state,Avg_freight_value,Top_5_States_Category
   from lowest_freight
```

OUTPUT:

Row	customer_state	Avg_freight_value	Top_5_States_Category
1	SP	15.15	Lowest
2	PR	20.53	Lowest
3	MG	20.63	Lowest
4	RJ	20.96	Lowest
5	DF	21.04	Lowest
6	RR	42.98	Highest
7	PB	42.72	Highest
8	RO	41.07	Highest
9	AC	40.07	Highest
10	PI	39.15	Highest

INSIGHTS:

- Customers in these states may abandon carts due to high shipping costs.
- Businesses may need to offer free shipping promotions to attract customers
- Businesses can offer fast & low-cost shipping, improving customer satisfaction & retention.
- These states are ideal for e-commerce warehouses & distribution centers.

BUSINESS RECOMMENDATIONS:

- Establish **micro-fulfillment hubs** closer to demand zones
- Consider **shipping subsidies** to stay competitive
- Introduce **minimum cart value thresholds** for free shipping based on customer state.
- Expand operations in low-cost regions like **SP, PR, MG** to maximize ROI.

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

QUERY:

```sql
with highest_state_avg as
(
  select
c.customer_state,round(AVG(DATE_DIFF(o.order_delivered_customer_date,o
.order_purchase_timestamp, DAY)),2) AS Avg_Delivery_Time,
  'Highest' as Top_5_States_Category
  from `Target.customers` c
  join `Target.orders` o
  using(customer_id)
  join `Target.order_items` oi
  using(order_id)
  group by 1
  order by 2 desc
  limit 5

),
lowest_state_avg as
(
  select
c.customer_state,round(AVG(DATE_DIFF(o.order_delivered_customer_date,o
.order_purchase_timestamp, DAY)),2) as Avg_Delivery_Time,
  'Lowest' as Top_5_States_Category
  from `Target.customers` c
  join `Target.orders` o
  using(customer_id)
  join `Target.order_items` oi
  using(order_id)
  group by 1
  order by 2
  limit 5
)
select customer_state,Avg_Delivery_Time,Top_5_States_Category
from highest_state_avg
union all
select customer_state,Avg_Delivery_Time,Top_5_States_Category
from lowest_state_avg;
```

OUTPUT:

Row	customer_state ▼	Avg_Delivery_Time ▼	Top_5_States_Category ▼
1	RR	27.83	Highest
2	AP	27.75	Highest
3	AM	25.96	Highest
4	AL	23.99	Highest
5	PA	23.3	Highest
6	SP	8.26	Lowest
7	PR	11.48	Lowest
8	MG	11.52	Lowest
9	DF	12.5	Lowest
10	SC	14.52	Lowest

INSIGHTS:

- Long delivery times can cause customer dissatisfaction and order cancellations.
- Businesses might need regional warehouses or faster shipping options
- Faster deliveries increase customer satisfaction and retention.

BUSINESS RECOMMENDATIONS:

- Launch flash sales or guaranteed next-day delivery **only in fast regions** like SP, PR, MG.
- Promote premium services like **same-day or weekend delivery** in those areas.
- Gather feedback from high-delay regions to assess **satisfaction impact**.
- Offer incentives or **free shipping vouchers** in RR, AP, and AM to improve experience.

4. **Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

QUERY:

```
select
c.customer_state,round(avg(date_diff(o.order_delivered_customer_date,o
.order_estimated_delivery_date,day)),2) as Avg_Diff_Estimated_Delivery
from `Target.customers` c
join `Target.orders` o
using(customer_id)
group by 1
```

```
order by 2
limit 5;
```

OUTPUT:

Row	customer_state ▼	Avg_Diff_Estimat...
1	AC	-19.76
2	RO	-19.13
3	AP	-18.73
4	AM	-18.61
5	RR	-16.41

INSIGHTS:

- Orders in states like **AC (-19.76 days)**, **RO (-19.13)**, and **AP (-18.73)** are delivered **on average 16–20 days earlier** than estimated.
- An average negative difference of nearly 20 days suggests that **delivery estimates are significantly off** — likely **padded too heavily** to avoid late delivery penalties.

- Customers receiving packages much earlier than expected might be **pleasantly surprised**, but also **less prepared.**

BUSINESS RECOMMENDATIONS:

- Use actual historical delivery performance to **update the estimated delivery window**
- Highlight early or on-time delivery rates in marketing for these regions
- **Shift logistics focus** or resources to regions with frequent delays.

VI. Analysis based on the payments:

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

QUERY:

```sql
select format_timestamp('%Y',o.order_purchase_timestamp) as
order_year,
format_timestamp('%B',o.order_purchase_timestamp) as
order_month,p.payment_type,
count(o.order_id) as order_count
from `Target.orders` o
join `Target.payments` p
```

```
        on o.order_id=p.order_id
        group by 1,2,3
        order by 1,2,3
```

OUTPUT:

Row	order_year ▼	order_month ▼	payment_type ▼	order_count ▼
1	2016	December	credit_card	1
2	2016	October	UPI	63
3	2016	October	credit_card	254
4	2016	October	debit_card	2
5	2016	October	voucher	23
6	2016	September	credit_card	3
7	2017	April	UPI	496
8	2017	April	credit_card	1846
9	2017	April	debit_card	27
10	2017	April	voucher	202

INSIGHTS:

- UPI adoption is rapidly increasing among users, showing a clear **growth trend**.
- Credit cards remain the **most preferred payment method** across both years.
- Vouchers and Debit Cards payment types may cater to a **small or specific customer segment**.
- April 2017 might have seen a **promotional campaign, new product launch**, or a **seasonal trend** (e.g., summer sale).

BUSINESS RECOMMENDATIONS:

- Consider faster checkout experience
- Exclusive discounts for UPI users
- Offer **cashback or EMI options** for large purchases
- Reverse engineer marketing, product, or external factors that boosted April 2017 orders.
- Survey customers for friction in debit/voucher use

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

QUERY:

```
select payment_installments,count(distinct o.order_id) as no_of_orders
from `Target.orders` o
join `Target.payments` p
on o.order_id=p.order_id
group by 1
order by 1;
```

OUTPUT:

Row	payment_installm...	no_of_orders
1	0	2
2	1	49060
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644

INSIGHTS:

- A very high proportion of orders are paid in a single installment.
- The number of orders drastically drops as the installment count increases.
- While small in number, some customers consistently use 2–5 installment options.

BUSINESS RECOMMENDATIONS:

- Display installment benefits upfront, especially for high-ticket items.
- Introduce 0% EMI or Low-Interest Plans for High-Value Orders
- **One-time payers** with loyalty discounts or up-sell offers
- **Installment users** with flexible payment plans or bundled offers