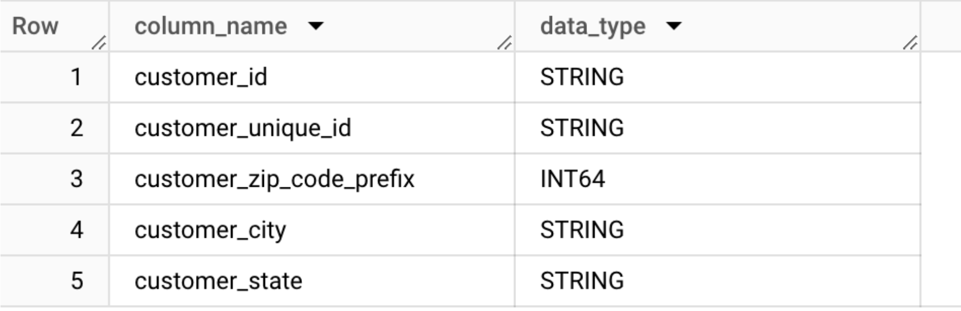
Business Case: Target

Ques 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.

Ans.

```
SELECT
column_name,
data_type
FROM target-401004.target.INFORMATION_SCHEMA.COLUMNS
WHERE table_name = 'customers'
```



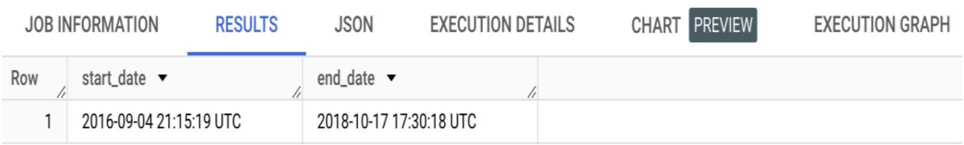
Row	column_name	data_type
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

- By looking at the output of the query, it is clear that customer_id , customer_unique_id, customer_city, customer_state is of string data type ,where in customer_zip_code_prefix is integer data type.

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

Ans.

```
SELECT
MIN (order_purchase_timestamp) AS start_date,
MAX (order_purchase_timestamp) AS end_date
FROM `target.orders`
```



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

- By looking at the output of the query, it is clear that orders placing was started from “September 4 ,2016” and ended on “August 17,2018”.

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

Ans.

```
select
count(distinct customer_city) as city,
count(distinct customer_state) as state
from `target.orders` o
inner join `target.customers` c
on c.customer_id = o.customer_id
```

- The output of the query provides us the different cities and different states of different customers who have ordered the products.

Ques.2 In-depth Exploration:

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

Ans.

```
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 1,2
order by 1,2
```

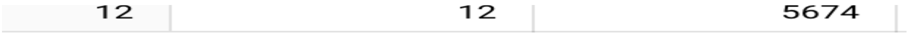
10	2017	7	4026

- The output shows that orders per month and per year are growing gradually.So yes, we can say that there a growing trend in the no. of orders placed over the past years .

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

Ans.

```
select
extract(month from order_purchase_timestamp) as month,
count(order_id) as num_of_orders
from `target.orders`
group by 1
order by 1
```



12	12	5674

- There is indeed some type of monthly seasonality in the number of orders being placed

- There seems to be increase in the number of orders during certain months, followed by a decrease in other. Understanding these patterns can help businesses
- Months 5 (May) and 8 (August) have the highest number of orders (10573 and 10843, respectively)

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

Ans.

```
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'
when extract(hour from order_purchase_timestamp ) between 19 and 24 then 'Night'
else 'unknown'
end as order_time_of_day,
count(order_id) as num_of_orders
from `target.orders`
group by 1
order by 2 desc
```

- During 'Afternoon', the Brazilian customers mostly place their orders. Businesses can use this information to time marketing campaigns or promotions during peak order times.

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

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

Ans.

```
select
extract(month from order_purchase_timestamp ) as month,
c.customer_state,
count(order_id) as num_of_orders
from `target.orders` o
inner join `target.customers` c
on o.customer_id = c.customer_id
group by 1,2
order by 2,1
```

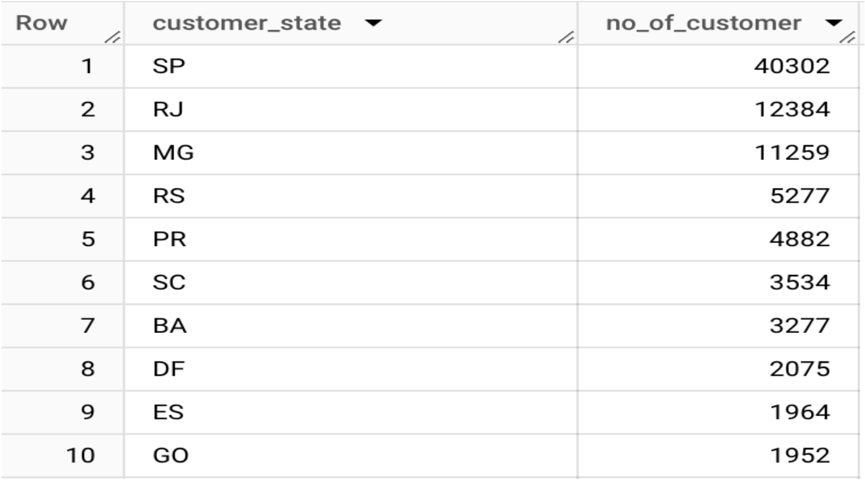
10	10	AC	6

- The output shows that SP, RJ, MG have the highest numbers of orders month on month, and AP, RR, MM have the lowest numbers of orders month on month.
- The high numbers of orders month on month indicate strong and stable demand.

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

ANS.

```
select
customer_state,
count(distinct customer_unique_id) as no_of_customer
from `target.customers`
group by 1
order by 2 desc
```



Row	customer_state	no_of_customer
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952

- The output provides that states SP, RJ, MG have the most number of customers. This can be a positive for a business.
- States RR, AP, AC have the lowest number of customers, a negative from the company point of view. These states require a more targeted or region-specific marketing and customer service approach.
-

QUES 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 2017 to 2018 (include months between Jan to Aug only). You can use the "payment value" column in the payments table to get the cost of orders.

ANS;

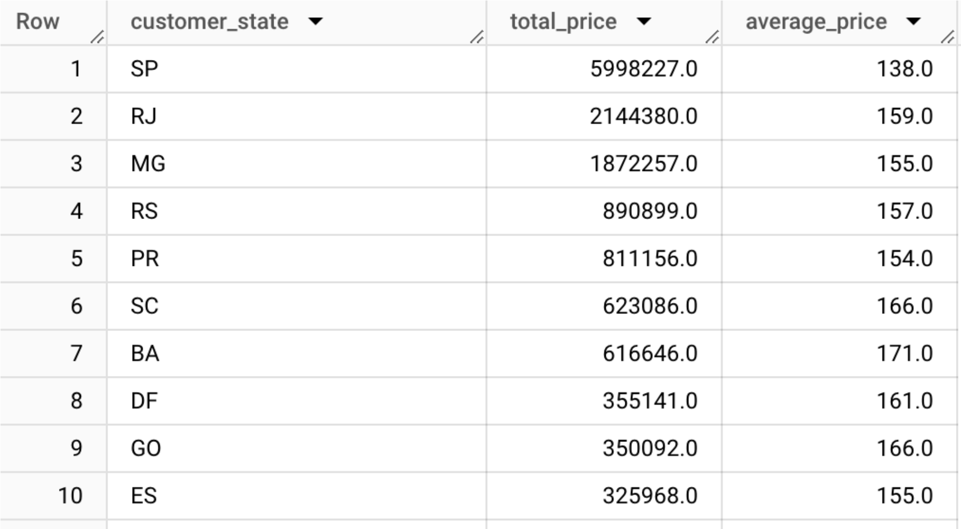
```
with final as(
select format_date('%Y',
      order_purchase_timestamp)as date,
      sum(p.payment_value) as cost_of_orders
from `target.payments` p
inner join `target.orders` o
on p.order_id = o.order_id
where extract(year from o.order_purchase_timestamp) between 2017 and 2018 and
     extract(month from o.order_purchase_timestamp)between 1 and 8
group by 1
order by 1)
select * ,lag(cost_of_orders) over(order by date) as cost_of_orders_previous_month,
round(100*(cost_of_orders - lag(cost_of_orders)over(order by date))/lag(cost_of_orders)over(order by date),1) as
percent_increase
from final
order by date
```

- The cost of orders in 2018 increased by approximately 137.0% compared to 2017. This indicates a substantial growth in orders between the two years, indicate positive business growth or increased customer demand

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

ANS:

```
select
c.customer_state ,
round(sum(p.payment_value))as total_price,
round(avg(p.payment_value))as average_price
from `target.payments` p
inner join `target.orders` o
on p.order_id = o.order_id
inner join `target.customers` c
on o.customer_id = c.customer_id
group by 1
order by 2 desc
```



Row	customer_state	total_price	average_price
1	SP	5998227.0	138.0
2	RJ	2144380.0	159.0
3	MG	1872257.0	155.0
4	RS	890899.0	157.0
5	PR	811156.0	154.0
6	SC	623086.0	166.0
7	BA	616646.0	171.0
8	DF	355141.0	161.0
9	GO	350092.0	166.0
10	ES	325968.0	155.0

- SP, RJ, MG remain the states with the highest total price.

- While SP has the lowest average price, this suggests that SP customers are purchasing lower value.

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

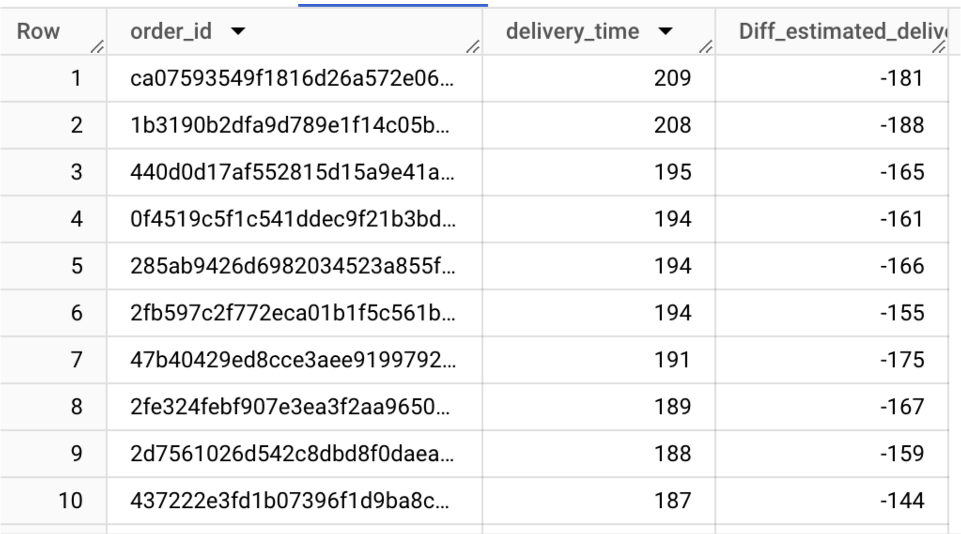
ANS:
```
select c.customer_state as State,
    round(sum(i.freight_value)) as Total_Price,
    round(avg(i.freight_value)) as Avg_Price
    from `target.customers`c
    inner join `target.orders`o
    on c.customer_id = o.customer_id
    inner join `target.order_items`i
    on o.order_id = i.order_id
    group by c.customer_state
    order by 2 desc
```

- The variations in freight value can be influenced by factors such as geographical location, distance from the distribution center, transportation infrastructure.
- AC, AP, RR have the lowest total freight values which implies that shipping to and from these states is less expensive.
- SP, RJ, MG have the highest total freight values, which means that shipping to and from these states is more expensive.

QUES 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.

ANS:
```
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_estimated_delivery
from `target.orders`
order by 2 desc
```

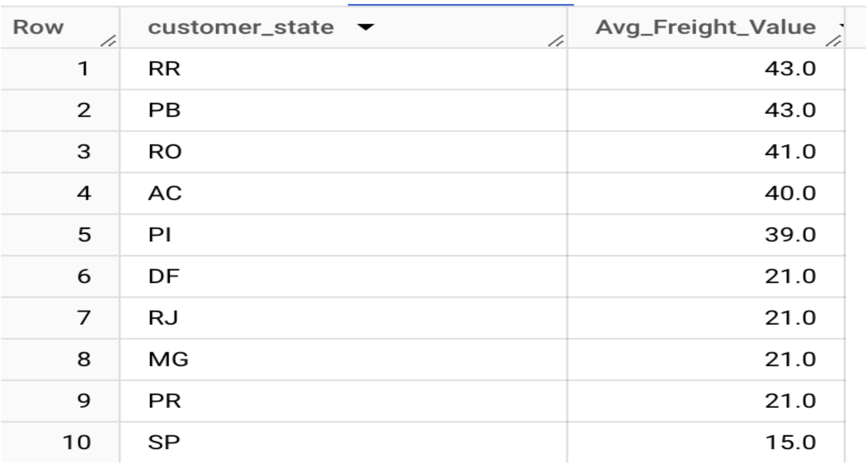


Row	order_id	delivery_time	Diff_estimated_deliv
1	ca07593549f1816d26a572e06...	209	-181
2	1b3190b2dfa9d789e1f14c05b...	208	-188
3	440d0d17af552815d15a9e41a...	195	-165
4	0f4519c5f1c541ddec9f21b3bd...	194	-161
5	285ab9426d6982034523a855f...	194	-166
6	2fb597c2f772eca01b1f5c561b...	194	-155
7	47b40429ed8cce3aee9199792...	191	-175
8	2fe324febf907e3ea3f2aa9650...	189	-167
9	2d7561026d542c8dbd8f0daea...	188	-159
10	437222e3fd1b07396f1d9ba8c...	187	-144

- The results show that it is taking a long time to deliver the orders to the customer, this could be the concern for the customers' satisfaction and it's essential to investigate the reason behind such extended delivery time.
- Many orders were delivered earlier than the estimated delivery date, as indicated by negative values in the ‘Diff_estimated_delivery’ column. Early deliveries can be a positive customer experience.

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

```
ANS: WITH AvgFreightValues AS (
SELECT
c.customer_state,
ROUND(AVG(i.freight_value)) AS Avg_Freight_Value,
ROW_NUMBER() OVER (ORDER BY AVG(i.freight_value) DESC) AS HighRank,
ROW_NUMBER() OVER (ORDER BY AVG(i.freight_value) ASC) AS LowRank
FROM `target.customers` c
INNER JOIN`target.orders` o
ON c.customer_id = o.customer_id
INNER JOIN `target.order_items` i
ON o.order_id = i.order_id
GROUP BY c.customer_state
)
SELECT
customer_state,
Avg_Freight_Value
FROM AvgFreightValues
WHERE HighRank <= 5 OR LowRank <= 5
ORDER BY HighRank, LowRank;
```



Row	customer_state	Avg_Freight_Value
1	RR	43.0
2	PB	43.0
3	RO	41.0
4	AC	40.0
5	PI	39.0
6	DF	21.0
7	RJ	21.0
8	MG	21.0
9	PR	21.0
10	SP	15.0

- States like RR, PB, RO have the heighest average freight value indicating that customers in these states tend to have higher shipping costs per order.
- States like SP, PR, MG have the lowest average freight value indicating that customers in these states tend to have lower shipping costs per order.

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

ANS:

```
WITH AvgDeliveryTime AS (
SELECT
c.customer_state,
ROUND(AVG(date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp, day))) AS
Avg_Delivery_Time,
ROW_NUMBER() OVER (ORDER BY AVG(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day)) DESC) AS HighRank,
ROW_NUMBER() OVER (ORDER BY AVG(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, day)) ASC) AS LowRank

FROM `target.customers` c
INNER JOIN`target.orders` o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state)
SELECT
customer_state,
Avg_Delivery_Time
FROM AvgDeliveryTime
WHERE HighRank <= 5 OR LowRank <= 5
ORDER BY HighRank , LowRank ;
```

- States like SP, PR, MG have the lowest average delivery time, indicating that customers in these states receive their orders quickly.
- States like RR, AP, AM have the highest average delivery time, indicating that customers in these states receive their orders late compared to others.

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

ANS.

```
SELECT c.customer_state, ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_estimated_delivery_date, day))) AS Delivery_Time
FROM `target.customers`c
JOIN `target.orders`o
ON c.customer_id = o.customer_id
WHERE o.order_status = 'delivered'
FROUP BY c.customer_state
ORDER BY 2 ASC
LIMIT 5
```

- In these states orders were consistently delivered ahead of their estimated delivery time as indicated by negative signs in the “Delivery_Time” column.
- Data implies that logistics and operations in these states are well organized. This can contribute to a positive customer experience.

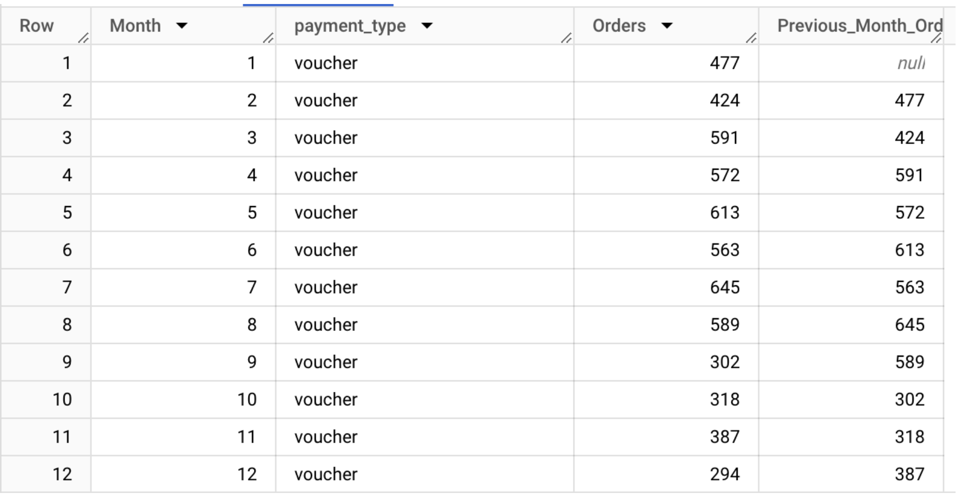
QUES 6. Analysis based on the payments:

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

ANS.

```
with final as(
SELECT
EXTRACT(month FROM o.order_purchase_timestamp) AS Month,
p.payment_type,
COUNT(p.order_id) AS Orders,
FROM `target.orders`o
JOIN `target.payments`p
ON o.order_id = p.order_id
GROUP BY 1,2
ORDER BY 1 ASC)
```

```
select * ,LAG(Orders) OVER (PARTITION BY payment_type ORDER BY Month) AS
Previous_Month_Orders
from final
```



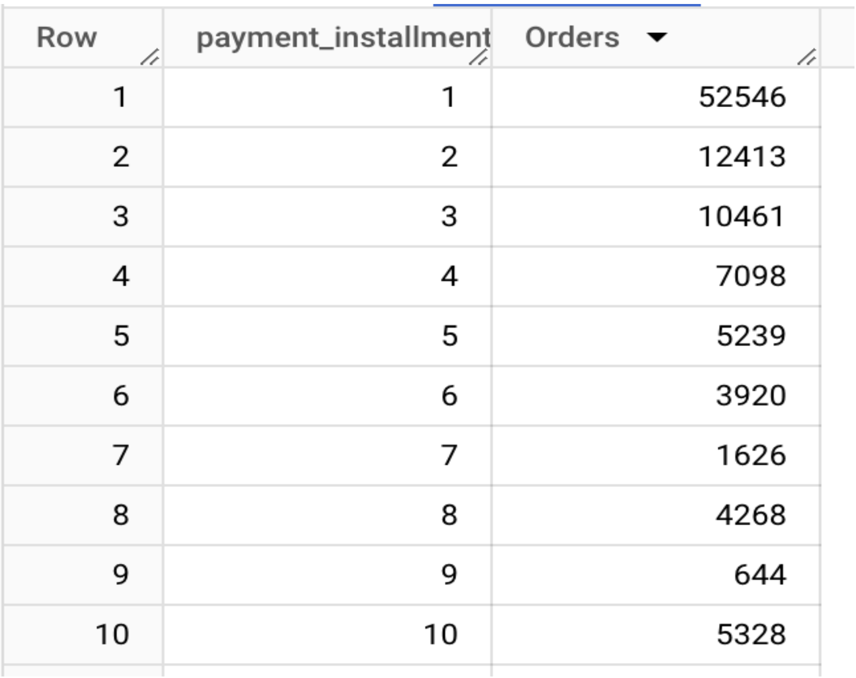
Row	Month	payment_type	Orders	Previous_Month_Ord
1	1	voucher	477	null
2	2	voucher	424	477
3	3	voucher	591	424
4	4	voucher	572	591
5	5	voucher	613	572
6	6	voucher	563	613
7	7	voucher	645	563
8	8	voucher	589	645
9	9	voucher	302	589
10	10	voucher	318	302
11	11	voucher	387	318
12	12	voucher	294	387

- The data shows trends in the number of orders for different payment types, including "credit_card," "UPI," "voucher," and "debit_card.
- "credit_card" appears to be a consistently popular payment method, with high order counts in most months.
- "UPI" also exhibits a significant number of orders, particularly in the later months.

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

Ans:

```
SELECT
payment_installments,
COUNT(order_id) AS Orders
FROM `target.payments`
WHERE payment_installments>=1
GROUP BY 1
ORDER BY 1 ASC;
```



Row	payment_installment	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

- There are significantly more orders with fewer payment installments, especially with 1, 2, and 3 installments.
- As the number of installments increases, the number of orders decreases. This suggests that many customers prefer to make payments in fewer installments.