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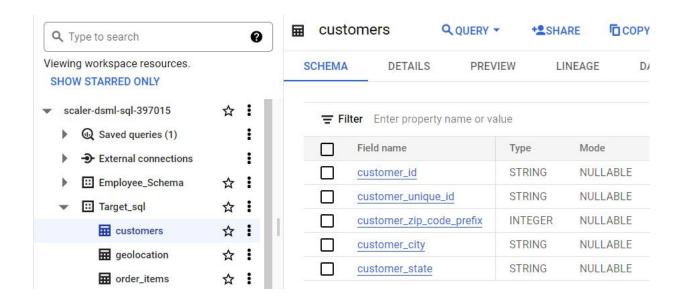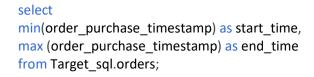Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**

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

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

Ans:

```
select
min(order_purchase_timestamp) as start_time,
max (order_purchase_timestamp) as end_time
from Target_sql.orders;
```
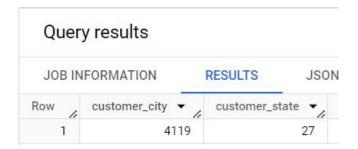

Insights:

In Target, customer orders were initiated starting on September 4, 2016, at 21:15:19 UTC, and concluded on October 17, 2018, at 17:30:18 UTC.

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

Ans:

```
select
count(distinct c.customer_city) as customer_city , count(distinct c.customer_state)
as customer_state
from Target_sql.customers c join Target_sql.orders o using (customer_id);
```


Insights:

During the given time period, a total of 4119 cities and 27 states of customers placed orders within Brazil. Determining the cities and states with the highest and lowest customers is valuable for pinpointing regions to focus marketing efforts on or consider for expansion.

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 order_year,
format_timestamp('%B', order_purchase_timestamp) as order_month, count(*) as
order_count
from Target_sql.orders
group by order_year, order_month
order by order_count desc;
```

Row	order_year	order_month	order_count
1	2017	November	7544
2	2018	January	7269
3	2018	March	7211
4	2018	April	6939
5	2018	May	6873
6	2018	February	6728
7	2018	August	6512
8	2018	July	6292
9	2018	June	6167
10	2017	December	5673

Insights:

The peak order counts were observed in November 2017, January 2018, and March 2018. However, during the early months of 2017, specifically in January and February, there was a notable drop in the number of orders compared to the rest of the year. The increase in orders from 2016 to 2017 was substantial, and it is crucial to maintain or even enhance this growth rate to remain competitive in the current market. To achieve this, the company should consider implementing strategies to retain customers and encourage them to make repeat purchases during months with lower order volumes.

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

Ans:

```
select format_timestamp('%B', order_purchase_timestamp) as month,
count(*) as orders
from Target_sql.orders
group by month
order by 2 desc;
```

Query results

	JOB INFORMATION	RESULTS	JSON	EX

Row	month ▾	orders ▾
1	August	10843
2	May	10573
3	July	10318
4	March	9893
5	June	9412
6	April	9343
7	February	8508
8	January	8069
9	November	7544
10	December	5674

Insights:

Many orders are made during the months of August, May, and July because these are winter months in Brazil. During this period, online shopping becomes particularly popular as people prefer to stay indoors due to the cold weather and a general sense of laziness that discourages them from venturing to physical stores. Additionally, these months often coincide with holidays like Easter, which tend to bring about promotional deals, thus making them attractive for shopping at Target. There isn't a significant disparity when compared to other months, as there can be appealing offers tailored to various seasons that may motivate higher levels of shopping.

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**

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 23 then
'Night'
end as time_of_day, count(*) as order_count
FROM
 Target_sql.orders
group by time_of_day
 order by order_count desc;
```

Query results

	JOB INFORMATION	RESULTS	JSON	EXECUTIO

Row	time_of_day ▼	order_count ▼
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insights:

At Target, Brazilian customers tend to place orders during the afternoon, maybe this time coincides with their lunch breaks or when they do their grocery shopping for meals, whether it's for lunch or dinner. Another peak time for orders is during the night, as individuals frequently browse shopping apps and make purchases before going to sleep. Orders also remain relatively high in the morning, with little difference compared to the nighttime. However, during dawn, there is a notable drop in orders, likely because people are busy commuting from work to home during that time. Target can increase marketing efforts during peak hours in the afternoon and at night by using targeted promotions to attract more customers.

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
 FORMAT_TIMESTAMP('%B', o.order_purchase_timestamp) AS months, count(*) as
orders, c.customer_state
from Target_sql.orders o left join Target_sql.customers c using(customer_id)
group by months, c.customer_state
order by orders desc;
```

Row	months	orders	customer_state
1	August	4982	SP
2	May	4632	SP
3	July	4381	SP
4	June	4104	SP
5	March	4047	SP
6	April	3967	SP
7	February	3357	SP
8	January	3351	SP
9	November	3012	SP
10	December	2357	SP

Insights:

In Target, there is a clear seasonal variation in the number of orders placed in different states. In SP (Sao Paulo) the number of orders peaks in August and gradually decreases towards the end of the year. SP have higher order volumes   compared to others like MG and RJ throughout the year. May is a month with relatively high orders for RJ and MG. December seems to have a lower number of orders across all states. Focusing on customer engagement and retention strategies to encourage repeat purchases.

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

 Ans:

```
select
 customer_state, count(*) as no_of_customers
from Target_sql.customers
group by  customer_state
order by 2 desc;
```

Query results

| | JOB INFORMATION | RESULTS | JSON | EXECUTI |

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

Insights:

Sao Paulo (SP) has the highest number of customers with 41,746, followed by Rio de Janeiro (RJ) and Minas Gerais (MG). This indicates that a significant portion of the customer base is concentrated in a few states. States like SP, RJ, and MG have a much larger customer base compared to smaller states like Sergipe (SE) and Alagoas (AL). The company can consider expanding its operations or marketing efforts in states with lower customer numbers but significant market potential, such as the Northeastern states like Pernambuco (PE), Ceara (CE), and Paraiba (PB). Collaborating with local businesses or organizations in specific states to strengthen the company's presence.

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).
You can use the "payment_value" column in the payments table to get the cost of orders.

Ans:

```
select
    ((sum(case when extract(YEAR from order_purchase_timestamp) = 2018 and
extract(MONTH from order_purchase_timestamp) between 1 and 8 then
payment_value else 0 end) -
    sum(case when extract(YEAR from order_purchase_timestamp) = 2017 and
extract(MONTH from order_purchase_timestamp) between 1 and 8 then
payment_value else 0 end)) /
```

```
    sum(case when extract(YEAR from order_purchase_timestamp) = 2017 AND
extract(MONTH from order_purchase_timestamp) between 1 and 8 then
payment_value else 0 end)) * 100 as percentage_increase
from
    Target_sql.payments p join Target_sql.orders o using (order_id)
where
 (extract(YEAR from order_purchase_timestamp) = 2017 AND extract(MONTH from
order_purchase_timestamp) between 1 and 8)
    or
    (extract(YEAR from order_purchase_timestamp) = 2018 AND extract(MONTH from
order_purchase_timestamp) between 1 and 8);
```

Query results

	JOB INFORMATION	RESU
Row	percentage_increase	
1	136.9768716466…	

Insights:

Orders cost increased by 137% from Jan to Aug 2017 to the same period in 2018, indicating rapid e-commerce growth. Enhancing promotional efforts and actively engaging with customers can further enhance the growth potential in this context.

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

Ans:

```
select round(sum(oi.price),2) as total_price,
round(avg(oi.price),2) as avg_price,
cu.customer_state
from Target_sql.order_items oi left join Target_sql.orders o using (order_id)
left join Target_sql.customers cu using (customer_id)
group by cu.customer_state
order by 1 desc;
```

Row	total_price	avg_price	customer_state
1	5202955.05	109.65	SP
2	1824092.67	125.12	RJ
3	1585308.03	120.75	MG
4	750304.02	120.34	RS
5	683083.76	119.0	PR
6	520553.34	124.65	SC
7	511349.99	134.6	BA
8	302603.94	125.77	DF
9	294591.95	126.27	GO
10	275037.31	121.91	ES

Insights:

Sao Paulo (SP) stands out as the top revenue generator for Target retail stores in Brazil due to its highest total order price. This is likely because it is one of the country's largest industrial areas, attracting a significant workforce that places orders with Target. However, the average order price in SP is comparatively lower, suggesting a greater number of orders with moderate values.

Rio de Janeiro (RJ) also contributes significantly to the total order price, as it is the next most famous city after SP. In RJ, the average order price is relatively higher, indicating a mixture of both high-value and moderate-value orders.

Minas Gerais (MG) shows a decent total order price, but the average order price is slightly lower, indicating a larger volume of orders with moderate values.

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

Ans:

```
select round(sum(oi.freight_value),2) as total_freight_value,
round(avg(oi.freight_value),2) as avg_freight_value,
cu.customer_state
from Target_sql.order_items oi left join Target_sql.orders o using (order_id)
left join Target_sql.customers cu using (customer_id)
group by cu.customer_state
order by 1 desc;
```

Row	total_freight_value	avg_freight_value	customer_state
1	718723.07	15.15	SP
2	305589.31	20.96	RJ
3	270853.46	20.63	MG
4	135522.74	21.74	RS
5	117851.68	20.53	PR
6	100156.68	26.36	BA
7	89660.26	21.47	SC
8	59449.66	32.92	PE
9	53114.98	22.77	GO
10	50625.5	21.04	DF

Insights:

The company's shipping costs vary significantly by state, with some places like Sao Paulo (SP) having lower costs (around 15.15), while others like Paraiba (PA) have much higher costs (around 42.72), possibly due to their location and infrastructure challenges. Then states like Santa Catarina (SC) and Rio Grande do Sul (RS) have lower costs, this could be because of better transportation infrastructure and proximity to manufacturing centers. Target should consider expanding its distribution centers or warehouses in regions with lower freight costs to serve nearby states more efficiently.

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.
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_estimated_delivery_date - order_delivered_customer_date

Ans:

```
select
order_id,
timestamp_diff(order_delivered_customer_date, order_purchase_timestamp,
day)as time_to_deliver,
timestamp_diff(order_estimated_delivery_date, order_delivered_customer_date,
day)as diff_estimated_delivery
from Target_sql.orders
```

order by 1;

Query results

Row	order_id ▼	time_to_deliver ▼	diff_estimated_delive
1	00010242fe8c5a6d1ba2dd792...	7	8
2	00018f77f2f0320c557190d7a1...	16	2
3	000229ec398224ef6ca0657da...	7	13
4	00024acbcdf0a6daa1e931b03...	6	5
5	00042b26cf59d7ce69dfabb4e...	25	15
6	00048cc3ae777c65dbb7d2a06...	6	14
7	00054e8431b9d7675808bcb8...	8	16
8	000576fe39319847cbb9d288c...	5	15
9	0005a1a1728c9d785b8e2b08...	9	0
10	0005f50442cb953dcd1d21e1f...	2	18

Insights:

The majority of orders were delivered ranging from 5 to 9 days, and in some cases, they were delivered after more than 100 days. This is highly detrimental to our customer satisfaction and needs immediate improvement. Ensuring that orders are consistently delivered within a 14-day time frame would greatly benefit of Target.

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

Ans:

```
with AvgFreight as (
 select
   c.customer_state,
   round(avg(oi.freight_value),2) as avg_freight
 from
 Target_sql.order_items oi left join
   Target_sql.orders o using (order_id)
   left join Target_sql.customers c using (customer_id)
 group by
   c.customer_state
),
Ranked as (
 select
   customer_state,
   avg_freight,
```

```
    row_number() over (order by avg_freight desc) as top_rank,
    row_number() over (order by avg_freight asc) as bottom_rank
  from
    AvgFreight
)
select
  customer_state,
  avg_freight
from
  Ranked
where
  top_rank <= 5 or bottom_rank <= 5
order by
  avg_freight desc;
```

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

Insights:

States in the North (e.g., RR, PB, RO, AC) have notably higher average freight costs compared to states in the South and Southeast (e.g., SP, PR, MG, RJ). This suggests that there might be logistical challenges or differences in transportation infrastructure in these regions. Conducting a market analysis in high-cost states can help identify factors contributing to higher freight costs, such as remoteness, lack of competition among carriers, or regulatory issues.

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

Ans:
```
with AvgFreight as (
select
  c.customer_state,
```

```
round(avg(TIMESTAMP_DIFF(o.order_estimated_delivery_date,
o.order_delivered_customer_date, DAY)),2) as delivery_time
from
Target_sql.orders o
left join
 Target_sql.customers c
using
(customer_id)
group by
 c.customer_state
),
Ranked as (
  select
 customer_state,
delivery_time,
row_number() over (order by delivery_time desc) as top_rank,
row_number() over (order by delivery_time asc)  as bottom_rank
 from
 AvgFreight
)
select
customer_state,
delivery_time
from
 Ranked
where
 top_rank <= 5 or bottom_rank <= 5
order by
delivery_time desc;
```

Row	customer_state	delivery_time
1	AC	19.76
2	RO	19.13
3	AP	18.73
4	AM	18.61
5	RR	16.41
6	BA	9.93
7	ES	9.62
8	SE	9.17
9	MA	8.77
10	AL	7.95

Insights:

The delivery times vary from 7 to 10 days in the states with the quickest deliveries
and from 16 to 20 days in the states with the slowest deliveries. The northern states

(e.g., AC, RO, AP) have longer delivery times than the northeast (e.g., AL, MA, SE) and southeast (e.g., ES, BA). To make deliveries faster, the company should check how well its supply chain works.

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.

Ans:

```
select
c.customer_state,
round(avg(timestamp_diff(o.order_estimated_delivery_date,
order_delivered_customer_date, day)),2)        as fast_delivery
from
Target_sql.orders o
left join Target_sql.customers c using (customer_id)
group by
c.customer_state
order by
fast_delivery ASC
limit 5;
```

Row	customer_state	fast_delivery
1	AL	7.95
2	MA	8.77
3	SE	9.17
4	ES	9.62
5	BA	9.93

Insights:

The top 5 states with delivery times ranging from 7 to 10 days faster than the estimated dates are AL, MA, SE, ES, BA demonstrate effective logistics and delivery operations. Examining the procedures and strategies employed in these states for achieving swift deliveries can serve as a model for implementing similar best practices in areas experiencing slower delivery times.

6.Analysis based on the payments:

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

Ans:

```
select
count(*) as orders, FORMAT_TIMESTAMP('%B', o.order_purchase_timestamp) AS
months, p.payment_type
```

```
from Target_sql.payments p left join Target_sql.orders o using(order_id)
group by  months, p.payment_type
order by orders desc;
```

Query results

	JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS

Row	orders ▼	months ▼	payment_type ▼
1	8350	May	credit_card
2	8269	August	credit_card
3	7841	July	credit_card
4	7707	March	credit_card
5	7301	April	credit_card
6	7276	June	credit_card
7	6609	February	credit_card
8	6103	January	credit_card
9	5897	November	credit_card
10	4378	December	credit_card

Insights:

> May stands out as the month with the highest number of credit card orders. This
> could be due to special promotions, events, or external factors specific to May.
> The number of orders placed using UPI remains relatively consistent throughout
> the year, with no significant seasonal variation. This suggests that customers
> using UPI as their payment method have a more stable pattern of ordering.
> Investigate why May consistently has the highest number of credit card orders.
> If there are specific reasons such as holidays or product launches, consider
> replicating those strategies in other months. Encourage the use of other
> payment options like UPI by offering incentives or discounts to customers who
> choose these methods.

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

Ans:

```
select
count(*) as orders, payment_installments
from Target_sql.payments
where payment_installments > 1
group by payment_installments;
```

Row	orders	payment_installment
1	12413	2
2	10461	3
3	7098	4
4	5239	5
5	3920	6
6	1626	7
7	4268	8
8	644	9
9	5328	10
10	23	11

Insights:

It's evident that as the number of payment installments increases, the number of orders decreases significantly. A large number of orders especially in the 2nd and 3rd installments indicate that customers are more comfortable with fewer initial payment commitments.