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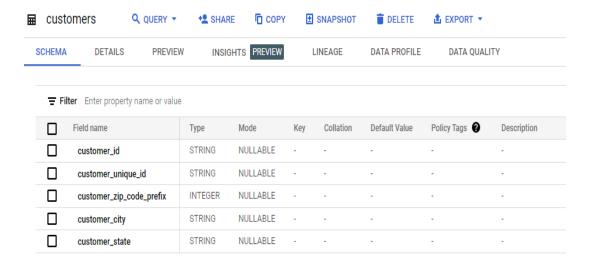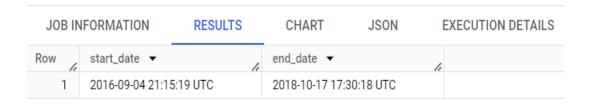
1.Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

Field name	Type	Mode	Key	Collation	Default Value	Policy Tags	Description
customer_id	STRING	NULLABLE	-	-	-	-	-
customer_unique_id	STRING	NULLABLE	-	-	-	-	-
customer_zip_code_prefix	INTEGER	NULLABLE	-	-	-	-	-
customer_city	STRING	NULLABLE	-	-	-	-	-
customer_state	STRING	NULLABLE	-	-	-	-	-

INFERENCE:
- This table provides information about customers.

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

select min(order_purchase_timestamp) as start_date,
    max(order_purchase_timestamp) as end_date
from `dsml-target-case.business_case1.orders`

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

INFERENCE:

- The time range during which the orders were placed extends from the earliest order on "2016-09-04" to the latest order on "2018-10-17"
- This range provides insight into the period of data collection or the timeframe covered by the dataset, which is crucial for understanding the temporal scope of the data analysis.

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

```
select count(distinct(geolocation_city)) as city,
       count(distinct(geolocation_state)) as state
from `dsml-target-case.business_case1.geolocation`
```

Query results

JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH

Row	city ▼	state ▼
1	8011	27

INFERENCE:

- The data reveals that orders were placed from 8011 unique cities and 27 unique states.
- This indicates a wide geographical distribution of the customer base, highlighting the reach and coverage of the business across different regions.
- The data collected will reflect the company's customer distribution within the country, which helps Target company to tailor marketing strategies, plan logistics, and allocate resources according to the geographical concentration of customers.

2.In-depth Exploration:

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

```sql
select
 extract(year from order_purchase_timestamp) as past_years,
 count(*) as num_of_orders,
 from `dsml-target-case.business_case1.orders`
 group by  extract(year from order_purchase_timestamp)
 order by past_years
```

Query results

	JOB INFORMATION	RESULTS	CHART	J

Row	past_years ▼	num_of_orders ▼
1	2016	329
2	2017	45101
3	2018	54011

INFERENCE:

- There is a substantial increase in the number of orders from 2016 to 2017, with orders jumping from 329 to 45,101. This indicates a significant growth in customer activity or market reach during this period.
- The trend continues with another notable increase from 2017 to 2018, where the number of orders rises from 45,101 to 54,011. This suggests sustained growth and an expanding customer base or increased customer purchasing behavior.

- The data demonstrates a strong and positive growth trend in the number of orders from 2016 to 2018. This suggests a healthy upward trajectory in the company's performance, likely driven by successful strategic initiatives and increasing customer engagement.
- Therefore, the company can use this information to further capitalize on growth opportunities and optimize its operations to sustain and enhance this upward trend.

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

```
select extract(month from order_purchase_timestamp)
month_number,
 format_timestamp ('%B',
timestamp_trunc(order_purchase_timestamp, month))
month_name,
count (*)  no_of_orders
from `dsml-target-case.business_case1.orders`
group by month_name, month_number
order by  month_number;
```

JOB INFORMATION		RESULTS	CHART	JSON	EXECUTI
Row	month_number	month_name			no_of_orders
1	1	January			8069
2	2	February			8508
3	3	March			9893
4	4	April			9343
5	5	May			10573
6	6	June			9412
7	7	July			10318
8	8	August			10843
9	9	September			4305
10	10	October			4959

INFERENCE:

- **Peak Season:** The data suggests a peak in orders during the late spring and summer months (May to August). This could be attributed to various factors such as seasonal promotions, holidays, or weather conditions that influence consumer purchasing behavior.

- **Off-Peak Season:** A significant drop in orders is observed starting from September, reaching its lowest in October, before a minor recovery in November and December. This pattern might indicate a post-peak season slowdown, possibly due to factors like end-of-summer, back-to-school expenses, or less consumer spending during these months.

- This seasonal variation should be considered in strategic planning, operational management, and marketing efforts to align with consumer purchasing behaviors and optimize business performance throughout the year.

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

```
select order_placed_time,
     count(*)as no_of_orders
from
(select *,
     extract(hour from order_purchase_timestamp) as hours,
     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 order_placed_time
        from `dsml-target-case.business_case1.orders`
) x
group by order_placed_time
order by no_of_orders
```

Row	order_placed_time ▼	no_of_orders ▼
1	dawn	5242
2	mornings	27733
3	night	28331
4	afternoon	38135

INFERENCE:

- **Afternoon:** The highest number of orders, 38,135, occurs in the afternoon. This indicates that customers are most active during this period, possibly due to increased leisure time or a midday break where they are more likely to engage in online shopping.
- **Night:** The second-highest volume of orders, 28,331, occurs at night. This suggests that many customers prefer shopping during the evening, potentially after work hours when they have more free time.
- **Morning:** With 27,733 orders, mornings also see substantial shopping activity, indicating a trend of customers placing orders early in the day, possibly before starting their daily routines.

- **Dawn:** The fewest orders (5,242) are placed during the dawn hours, which is consistent with typical human sleep patterns and suggests that online shopping activity is minimal during this time.
- Understanding these peak times can help in optimizing customer service and support operations. For example, increasing customer support availability during the afternoon and night could improve service quality and customer satisfaction.
- Inventory management can also be adjusted to ensure that popular products are well-stocked during peak shopping hours.

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

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

```
 select  o.month_number,o.year,c.customer_state,o.month_name,
      count(*) as no_of_orders
from
(select *,
      extract(year from order_purchase_timestamp) as year,
      extract(month from order_purchase_timestamp) month_number,
      format_timestamp('%B', timestamp_trunc(order_purchase_timestamp, month)) month_name,
from `dsml-target-case.business_case1.orders` )o join
`dsml-target-case.business_case1.customers` c
on o.customer_id = c.customer_id
group by o.month_number,o.month_name, c.customer_state,o.year
order by o.year,c.customer_state,o.month_number
```

Row	month_number ▼	year ▼	customer_state ▼	month_name ▼	no_of_orders ▼
1	10	2016	AL	October	2
2	10	2016	BA	October	4
3	10	2016	CE	October	8
4	10	2016	DF	October	6
5	10	2016	ES	October	4
6	10	2016	GO	October	9
7	10	2016	MA	October	4
8	10	2016	MG	October	40
9	10	2016	MT	October	3
10	10	2016	PA	October	4

INFERENCE:

- .The month-on-month order data provides a comprehensive overview of consumer behavior and operational demands across different states in Brazil. By leveraging these insights, the Target company can optimize various aspects of its business operations, from marketing and inventory management to logistics and customer engagement, thereby enhancing overall efficiency and customer satisfaction.

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

select count(*) as number_of_customers,
    customer_state
from `dsml-target-case.business_case1.customers`
group by customer_state
order by customer_state

Row	number_of_customer	customer_state
1	81	AC
2	413	AL
3	148	AM
4	68	AP
5	3380	BA
6	1336	CE
7	2140	DF
8	2033	ES
9	2020	GO
10	747	MA

INFERENCE:

- The customer distribution spans all states, indicating a nationwide reach of the Target company's products and services. This broad presence suggests that the company has successfully penetrated various regional markets, accommodating diverse consumer needs.
- The distribution of customers across Brazilian states provides critical insights into the market dynamics faced by the Target company. By leveraging this data, the company can optimize its marketing, sales, and operational strategies to better serve its diverse customer base, capitalize on strong markets, and explore new growth opportunities in less penetrated areas.

- This data-driven approach ensures a more efficient allocation of resources and enhances the company's ability to meet the needs of its customers across the country.

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.

```sql
with cte_1 as (select
        round(sum(p.payment_value),2) as order_cost,
        extract(year from o.order_purchase_timestamp) as year
         from `dsml-target-case.business_case1.payments`p join
`dsml-target-case.business_case1.orders`o
         on p.order_id = o.order_id
         where  extract(year from o.order_purchase_timestamp)=2017
and extract(month from o.order_purchase_timestamp) between 1 and 8
         group by extract(year from o.order_purchase_timestamp)),
      cte_2 as(select
        round(sum(p.payment_value),2) as order_cost,
        extract(year from o.order_purchase_timestamp) as year
         from `dsml-target-case.business_case1.payments`p join
`dsml-target-case.business_case1.orders`o
         on p.order_id = o.order_id
         where  extract(year from o.order_purchase_timestamp)=2018
and extract(month from o.order_purchase_timestamp) between 1 and 8
         group by extract(year from o.order_purchase_timestamp)),
      cte_3 as (select round(((cte_2.order_cost - cte_1.order_cost) /
    cte_1.order_cost * 100),2)  percentage_increase
    from cte_1 cross join cte_2 )
```

select * from cte_3

Row	percentage_increase
1	136.98

INFERENCE:

- The 136% increase in the cost of orders from 2017 to 2018 reflects a period of significant growth for the Target company in Brazil.
- This impressive rise points to successful business strategies and possibly favorable market conditions.
- Moving forward, the company should leverage this growth momentum while strategically planning for continued expansion and addressing any operational challenges that accompany such rapid growth.
- Understanding the factors contributing to this increase will be crucial for sustaining and capitalizing on this upward trend.

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

```sql
select geolocation_state,
    round(sum(price),2) as total_value,
    round(avg(price),2) as avg_value
from `dsml-target-case.business_case1.order_items` oi
join `dsml-target-case.business_case1.sellers` s on oi.seller_id = s.seller_id
join `dsml-target-case.business_case1.geolocation` g
on g.geolocation_zip_code_prefix= s.seller_zip_code_prefix
group by geolocation_state
order by geolocation_state
```

Query results

JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS

Row	geolocation_state ▼	total_value ▼	avg_value ▼
1	AC	43788.0	267.0
2	AM	31779.0	392.33
3	BA	23385841.45	351.61
4	CE	740073.63	246.12
5	DF	4674257.66	72.56
6	ES	3211486.26	127.31
7	GO	4444926.59	164.13
8	MA	3604486.05	89.9
9	MG	253958733.43	122.97
10	MS	724755.24	165.43
11	MT	961121.8	116.85

INFERENCE:

- The analysis of total and average order values across different states in Brazil provides a detailed view of the economic activity and consumer spending behavior associated with the Target company's operations.
- The data highlights both the cumulative revenue generated and the typical order size in each state, offering insights into regional market dynamics.
- By leveraging this data, the company can refine its marketing strategies, optimize resource allocation, and explore growth opportunities in underperforming regions. Understanding these regional variations is essential for maximizing revenue and enhancing customer satisfaction across diverse markets.

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

```
select customer_state,
    round(sum(price),2) as total_value,
    round(avg(price),2) as avg_value
```

```
from `dsml-target-case.business_case1.order_items` oi
join `dsml-target-case.business_case1.orders` o on oi.order_id =
o.order_id
join `dsml-target-case.business_case1.customers` c on c.customer_id =
o.customer_id
group by c.customer_state
order by c.customer_state
```

Row	customer_state	total_freight_value	avg_freight_value
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

INFERENCE:

.

- This data provides insights into the variations in freight charges based on geographic location and potentially reflects differences in logistical efficiency, distance from distribution centers, and infrastructure quality.
- By understanding these regional differences, the company can develop targeted strategies to manage and optimize delivery costs, enhance customer experience, and efficiently allocate resources. Addressing the challenges in high-cost areas while capitalizing on efficiencies in others will be crucial for sustaining profitability and competitive advantage.

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.

```sql
select order_id,
date_diff(date(order_delivered_customer_date),date(order_purchase_time
stamp),day) as delivery_day,
date_diff(date(order_estimated_delivery_date),date(order_purchase_times
tamp),day) as estimated_day,
date_diff(date(order_delivered_customer_date),date(order_purchase_time
stamp),day) -
date_diff(date(order_estimated_delivery_date),date(order_purchase_times
tamp),day) as delivery_difference
from `dsml-target-case.business_case1.orders`
where order_delivered_customer_date is not null
order by delivery_difference desc
```

Row	order_id ▼	delivery_day ▼	estimated_day ▼	delivery_difference
1	1b3190b2dfa9d789e1f14c05b...	208	20	188
2	ca07593549f1816d26a572e06...	210	29	181
3	47b40429ed8cce3aee9199792...	191	16	175
4	2fe324febf907e3ea3f2aa9650...	190	23	167
5	285ab9426d6982034523a855f...	195	29	166
6	440d0d17af552815d15a9e41a...	196	31	165
7	c27815f7e3dd0b926b5855262...	188	26	162
8	0f4519c5f1c541ddec9f21b3bd...	194	33	161
9	d24e8541128cea179a11a6517...	175	14	161
10	2d7561026d542c8dbd8f0daea...	188	29	159
11	6e82dcfb5eada6283dba34f16...	183	28	155

INFERENCE:

- The analysis reveals significant delays in the delivery process, with a consistently large gap between estimated and actual delivery times.
- Addressing these delays is crucial for improving customer satisfaction and maintaining a positive brand image.
- By optimizing logistics operations, enhancing estimation accuracy, and ensuring effective communication with customers, the Target company can work towards reducing these delivery discrepancies and better meeting customer expectations.

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

```
with cte as (select c.customer_state,
            round(avg(freight_value),2) as avg_value
        from `dsml-target-case.business_case1.order_items` oi
            join `dsml-target-case.business_case1.orders`o on oi.order_id =
o.order_id
            join `dsml-target-case.business_case1.customers` c on
o.customer_id= c.customer_id
            group by customer_state),
    highest as (select *,
             row_number()over(order by avg_value desc) as rn,
             rank()over(order by avg_value) as highest_freight
         from cte
         order by avg_value desc
         limit 5),
    lowest as(select *,
             row_number()over(order by avg_value ) as rn,
             rank()over(order by avg_value) as lowest_freight
         from cte
         order by avg_value
```

```
        limit 5)
SELECT *
FROM highest h join lowest l
on l.rn = h.rn
```

Row	customer_state	avg_value	rn	highest_freight	customer_state_1	avg_value_1	rn_1	lowest_freight
1	RR	42.98	1	27	SP	15.15	1	1
2	PB	42.72	2	26	PR	20.53	2	2
3	RO	41.07	3	25	MG	20.63	3	3
4	AC	40.07	4	24	RJ	20.96	4	4
5	PI	39.15	5	23	DF	21.04	5	5

INFERENCE:

- The states exhibit the highest average freight values, which could be attributed to several factors such as distance from major distribution centers, challenging geographical terrains, less developed transportation infrastructure, or higher operational costs. The elevated freight costs in these regions suggest that logistics expenses are a significant factor for the company's pricing strategy in these areas.

- In contrast, some states demonstrate the lowest average freight values. These regions likely benefit from well-developed infrastructure, proximity to major distribution hubs, and efficient transportation networks. The lower freight costs suggest more cost-effective logistics and distribution processes in these areas, which may contribute to competitive pricing and enhanced market reach.

- The disparities in average freight values across Brazilian states highlight the logistical challenges and opportunities for the Target company. By addressing the high costs in certain regions and leveraging efficiencies in others, the company can optimize its logistics strategy, enhance customer satisfaction, and potentially improve profitability.

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

```
with cte_1 as (select c.customer_state,
        extract(hour from o.order_delivered_carrier_date) as order_deliver_time
from `dsml-target-case.business_case1.orders` o join
`dsml-target-case.business_case1.customers`c
            on o.customer_id = c.customer_id),
    cte_2 as (select customer_state,
                round(avg(order_deliver_time),2) as avg_time
            from cte_1
            group by customer_state ),
    highest as(select *,
                row_number() over(order by avg_time desc) as rn,
                rank()over (order by avg_time desc)as highest_delivery_time
                from cte_2
                order by highest_delivery_time
                limit 5),
    lowest as(select *,
                row_number() over(order by avg_time) as rn,
                rank()over (order by avg_time)as lowest_delivery_time
                from cte_2
                order by lowest_delivery_time
                limit 5)
    select *
        from highest h join lowest l
            on h.rn = l.rn
```

JOB INFORMATION	RESULTS	CHART	JSON	EXECUTION DETAILS	EXECUTION GRAPH			

Row	customer_state	avg_time	rn	highest_delivery_time	customer_state_1	avg_time_1	rn_1	lowest_delivery_time
1	AP	15.82	1	1	AC	14.0	1	1
2	RN	15.78	2	2	TO	14.38	2	2
3	RR	15.6	3	3	SE	14.63	3	3
4	AL	15.35	4	4	AM	14.67	4	4
5	PA	15.26	5	5	RO	14.91	5	5

INFERENCE:

- Top 5 States with the Highest Average Delivery Time: These states exhibit the longest average delivery times, indicating potential challenges in the logistics and distribution network. Factors contributing to these delays may include geographical distance from main distribution centers, infrastructure limitations, or regional logistical constraints.
- Top 5 States with the Lowest Average Delivery Time: These states demonstrate the shortest average delivery times, suggesting more efficient delivery processes or favorable logistical conditions. These regions likely benefit from better infrastructure, more efficient supply chain routes, or closer proximity to distribution hubs.
- The variations in average delivery times across different Brazilian states provide valuable insights into the operational efficiency of the Target company's logistics network. By addressing the challenges in states with longer delivery times and leveraging the strengths of regions with shorter delivery times, the company can enhance its overall delivery performance, leading to improved customer satisfaction and competitive advantage.

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.

```
with cte_1 as(select c.customer_state,
  date (order_purchase_timestamp) as purchase_date,
  date (o.order_delivered_customer_date) as delivery_date,
  date (o.order_estimated_delivery_date) as estimated_date
from `dsml-target-case.business_case1.orders` o join
`dsml-target-case.business_case1.customers`c
```

```sql
        on o.customer_id = c.customer_id),
cte_2 as(select customer_state,
        round(avg(date_diff(delivery_date, purchase_date, day)),2) as
actual_delivery_time,
         round(avg(date_diff(estimated_date, purchase_date, day)),2)as
estimated_delivery_time
    from cte_1
    group by customer_state)
select *,
round(estimated_delivery_time -actual_delivery_time,2)as delivery_difference
from cte_2
order by delivery_difference desc
limit 5
```

JOB INFORMATION	**RESULTS**	CHART	JSON	EXECUTION DETAILS	EXECUT

Row	customer_state ▼	actual_delivery_time	estimated_delivery_ti	delivery_difference
1	AC	21.0	41.77	20.77
2	RO	19.28	39.41	20.13
3	AP	27.18	46.71	19.53
4	AM	26.36	45.76	19.4
5	RR	29.34	47.17	17.83

INFERENCE:

- The results show that these states consistently deliver orders quicker than anticipated, suggesting an efficient logistics and supply chain system in these regions.
- This efficiency could be due to a variety of factors, including better infrastructure, more efficient local distribution centers, or favorable regulatory environments.
- Such findings are crucial for understanding regional performance and can guide the company's logistics strategy. For instance, best practices from

these states can be analyzed and potentially replicated in other regions to improve overall delivery performance.

- Moreover, highlighting these states can also be a marketing advantage, as promoting faster-than-expected delivery times can enhance customer satisfaction and brand loyalty.
- Understanding and leveraging these insights can lead to operational optimizations and strategic initiatives to maintain and improve delivery performance across the country.

6.Analysis based on the payments:

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

```
select extract(month from order_purchase_timestamp) as month,
    sum(case when payment_type ="voucher" then 1 else 0 end)as
voucher,
    sum(case when payment_type ="UPI" then 1 else 0 end) as UPI,
    sum(case when payment_type="credit_card" then 1 else 0 end) as
credit_card,
    sum(case when payment_type ="debit_card" then 1 else 0 end )as
debit_card
 from `business_case1.payments` p join `business_case1.orders`o
        on p.order_id = o.order_id
group by month
order by month
```

row	month	voucher	UPI	credit_card	debit_card
1	1	477	1715	6103	118
2	2	424	1723	6609	82
3	3	591	1942	7707	109
4	4	572	1783	7301	124
5	5	613	2035	8350	81
6	6	563	1807	7276	209
7	7	645	2074	7841	264
8	8	589	2077	8269	311
9	9	302	903	3286	43
10	10	318	1056	3778	54

INFERENCE:
- The analysis of payment types used for placing orders reveals significant insights into customer behavior and preferences.
- The dominance of credit card payments, alongside the steady use of UPI and vouchers, highlights the importance of maintaining a diverse and efficient payment ecosystem.
- By understanding these trends, the Target company can enhance customer experience, optimize its payment processing systems, and tailor its promotional strategies to align with consumer payment preferences.

2. 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
from `business_case1.payments`
group by payment_installments
order by payment_installments
```

Row	payment_installment	no_of_orders
1	0	2
2	1	52546
3	2	12413
4	3	10461
5	4	7098
6	5	5239
7	6	3920

INFERENCE:

- The analysis of payment installments reveals clear preferences for single and short-term installment plans among customers.
- By understanding these preferences, the company can optimize its payment options, improve customer experience, and manage financial risk effectively.
- The focus should be on enhancing the attractiveness and simplicity of popular payment plans while evaluating the potential for promoting longer-term installment options.