TITLE: Business Case Study – TARGET SQL

Description: Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

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

By analysing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

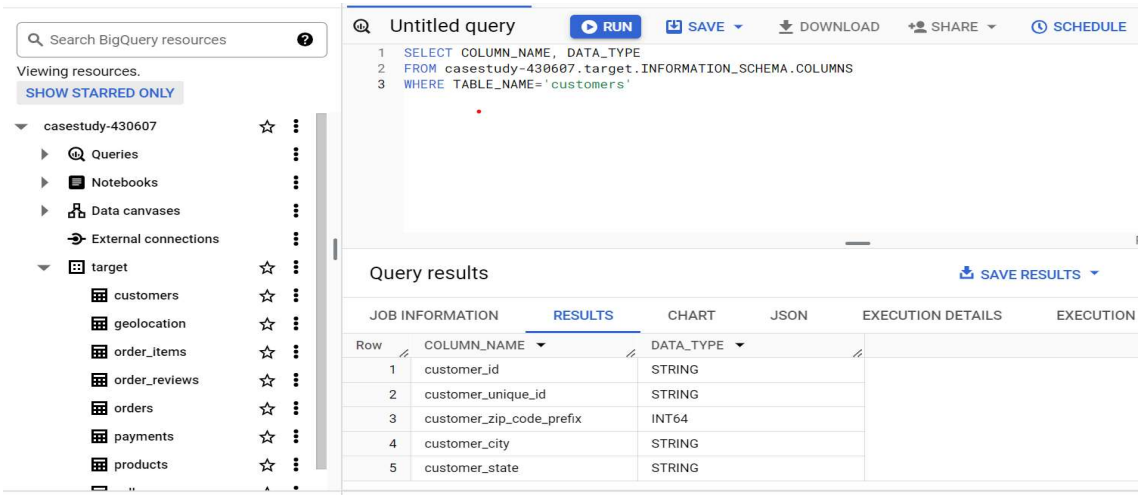
Q1) Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

Ans:

```
SELECT COLUMN_NAME, DATA_TYPE
FROM casestudy-430607.target.INFORMATION_SCHEMA.COLUMNS
WHERE TABLE_NAME='customers'
```

Output:



Insights:

In the above output screenshot the first column is column names and the second column is corresponding data types of customers table.All columns in this example are saved as strings, except customer_zip_code_prefix which is integer.

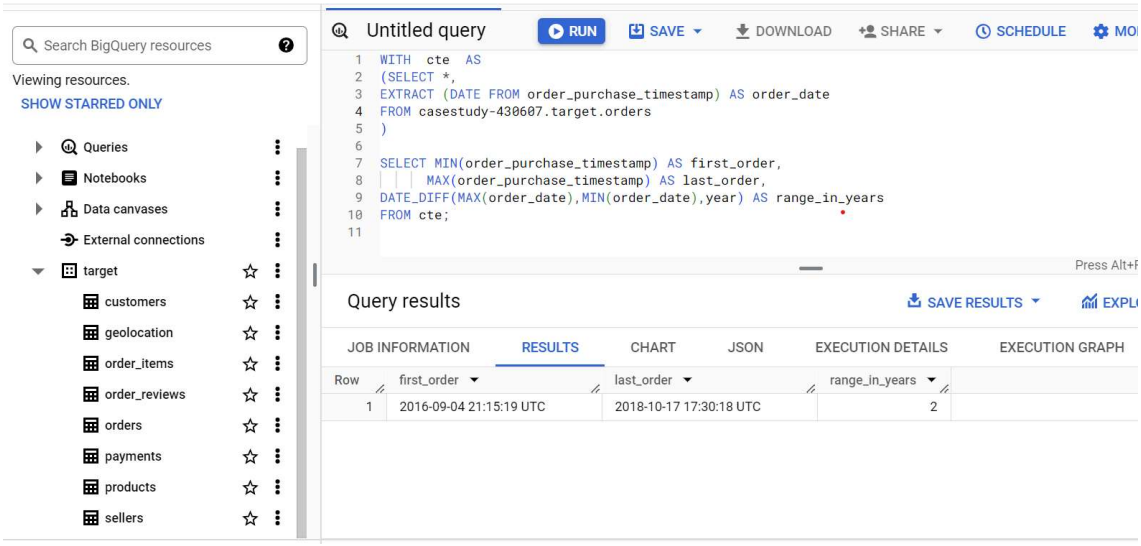
2.Get the time range between which the orders were placed.

Ans:

```
WITH  cte  AS
(SELECT *,
EXTRACT (DATE FROM order_purchase_timestamp) AS order_date
FROM casestudy-430607.target.orders
)

SELECT MIN(order_purchase_timestamp) AS first_order,
MAX(order_purchase_timestamp) AS last_order,
DATE_DIFF(MAX(order_date),MIN(order_date),year) AS range_in_years
FROM cte;
```

Output:



Insights:

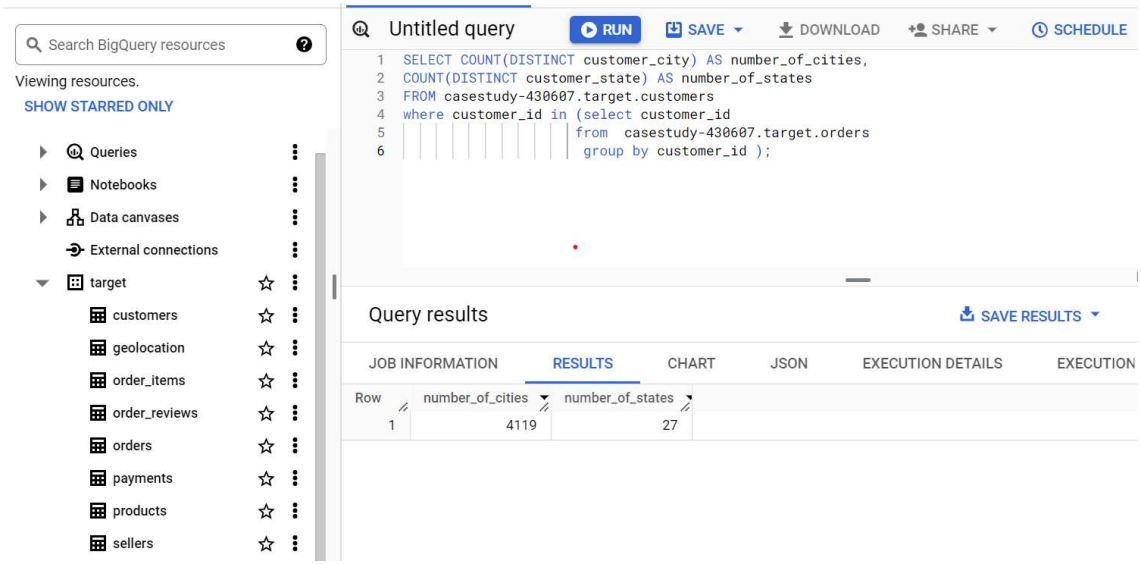
The time range between which the orders were placed is 2 years.First column is timestamp of first order placed and second column is timestamp of last order placed in the given range.

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

Ans:

```
SELECT COUNT(DISTINCT customer_city) AS number_of_cities,
COUNT(DISTINCT customer_state) AS number_of_states
FROM casestudy-430607.target.customers
where customer_id in (select customer_id
                      from  casestudy-430607.target.orders
                       group by customer_id );
```

Output:



Insights:

Total 4119 number of cities and 27 states of customers ordered in the given period of time. In the above query we have used the sub query for finding the customer_id's of customers who ordered during the given period.

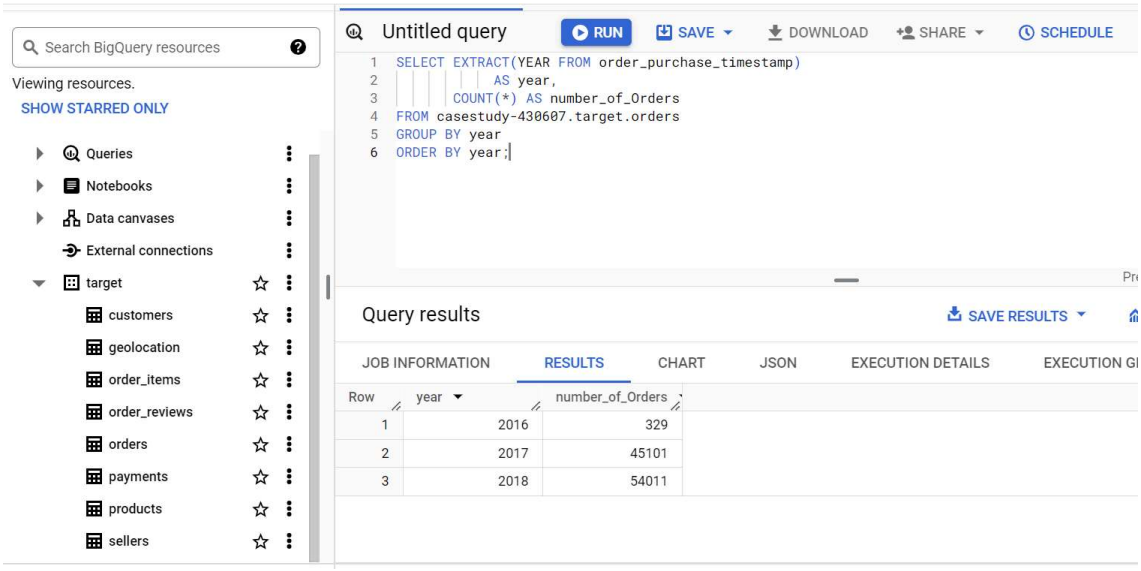
Q2) In-depth Exploration:

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,
  COUNT(*) AS number_of_Orders
FROM casestudy-430607.target.orders
GROUP BY year
ORDER BY year;
```

Output:



Insights:

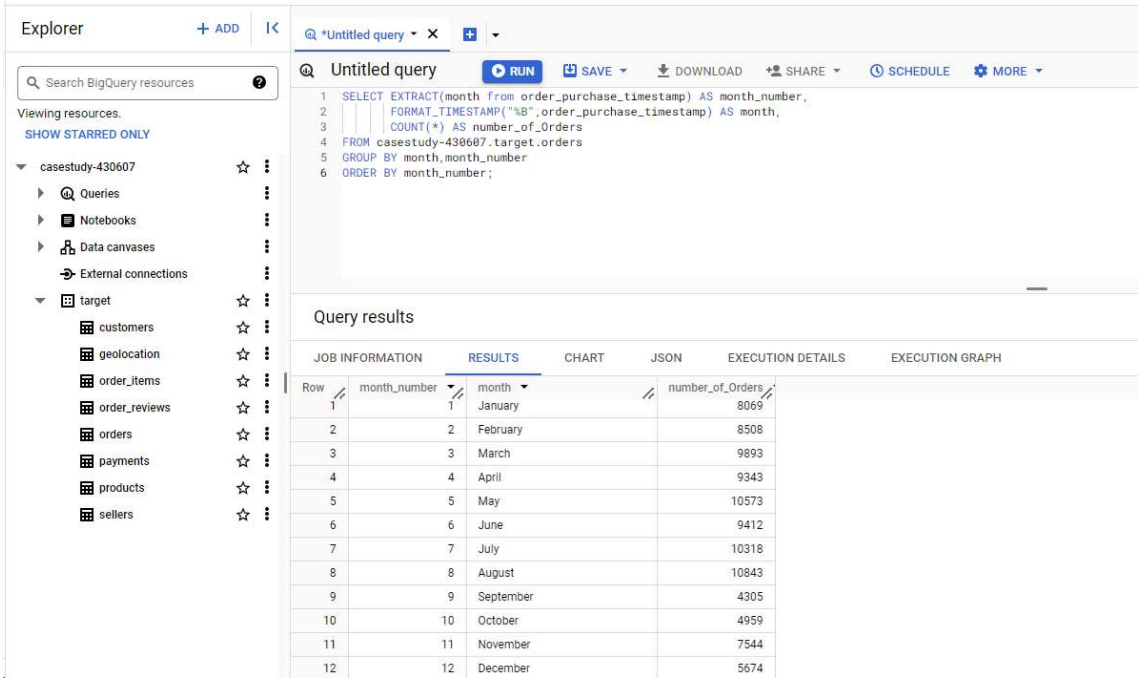
In the above output we can see the number of orders placed every year.After analysing the output there is a growing trend in the number of orders placed over the past year.

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_number,
       FORMAT_TIMESTAMP("%B",order_purchase_timestamp) AS month,
       COUNT(*) AS number_of_Orders
FROM casestudy-430607.target.orders
GROUP BY month,month_number
ORDER BY month_number;
```

Output:



Row	month_number	month	number_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
11	11	November	7544
12	12	December	5674

Insights:

In the above output screenshot we can see some kind of monthly seasonality in terms of the number of orders being placed.In the month of august highest number of orders were placed.And in the month of september lowest number of orders were placed.

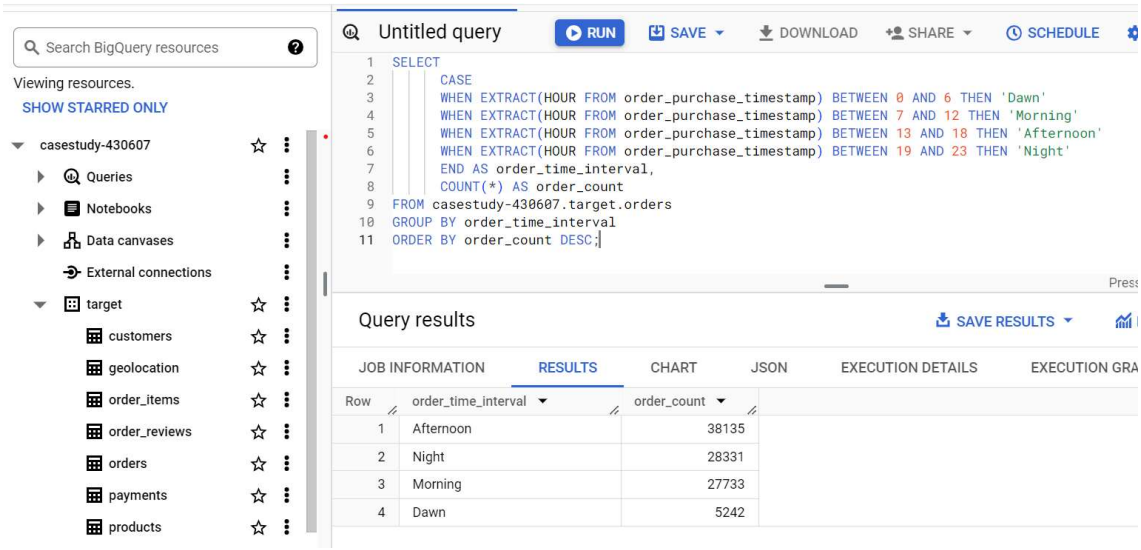
3.During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

- 0-6 hrs : Dawn
- 7-12 hrs : Mornings
- 13-18 hrs : Afternoon
- 19-23 hrs : Night

Ans:

```sql
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 'Morning'
      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 order_time_interval,
      COUNT(*) AS order_count
FROM casestudy-430607.target.orders
GROUP BY order_time_interval
ORDER BY order_count DESC;
```

Output:



Insights:

Based on the hour component of the timestamp, the query divides the order timestamps into various time groups (Dawn, Morning, Afternoon, Night). The results are then sorted based on how many orders fell inside each time frame.Brazilian customers frequently order more in the afternoons and customers placed least orders in Dawn time.

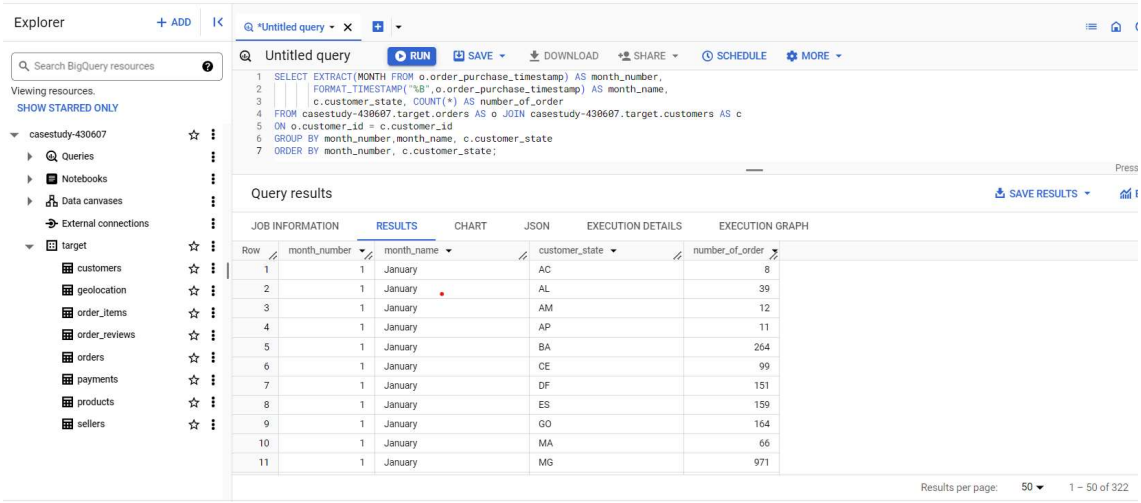
Q3)Evolution of E-commerce orders in the Brazil region:

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

Ans:

```
SELECT EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month_number,
       FORMAT_TIMESTAMP("%B",o.order_purchase_timestamp) AS month_name,
       c.customer_state, COUNT(*) AS number_of_order
FROM casestudy-430607.target.orders AS o JOIN
casestudy-430607.target.customers AS c
ON o.customer_id = c.customer_id
GROUP BY month_number,month_name, c.customer_state
ORDER BY month_number, c.customer_state;
```

Output:



Insights:

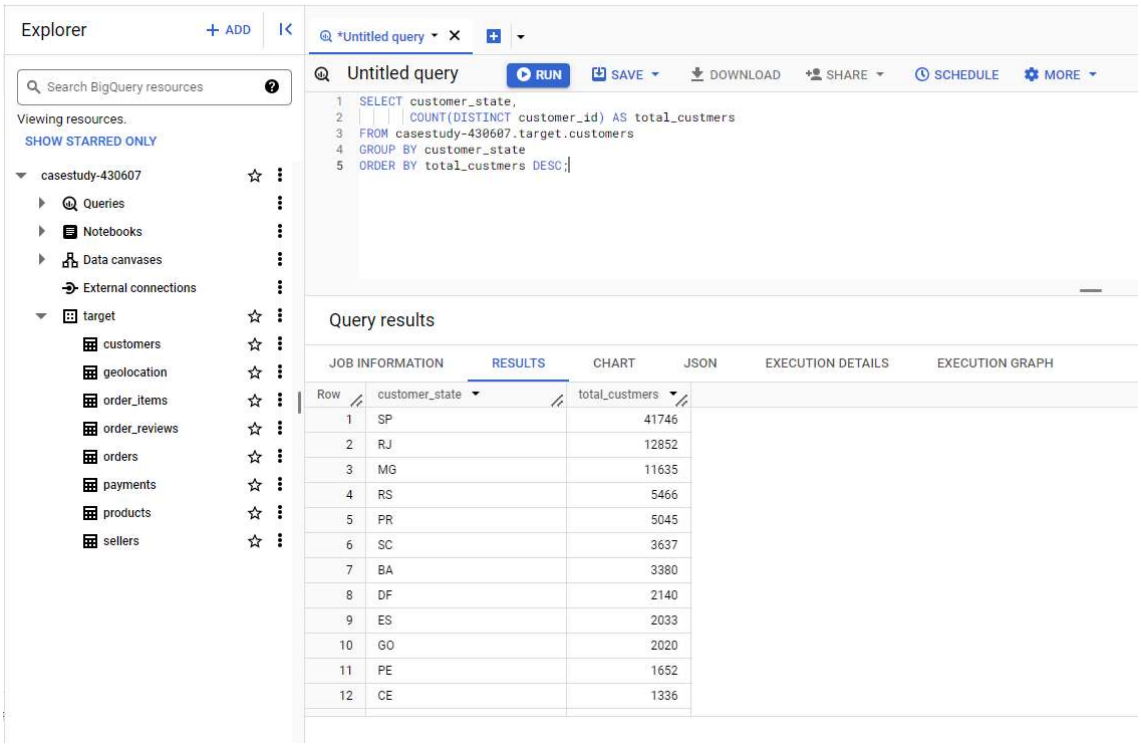
In the above output we can find the month on month number of orders placed in each state. After analysing the query result we can find that for every month the state called "SP" has the highest number of orders.

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

Ans:

```
SELECT customer_state,
       COUNT(DISTINCT customer_id) AS total_custmers
FROM casestudy-430607.target.customers
GROUP BY customer_state
ORDER BY total_custmers DESC;
```

Output:



Insights:

In the above query result we can see the total customers of each state along with their states. After analysing the query result we can find that the state called "SP" has the highest number of customers and the state called "RR" has the fewest number of customers.

Q4) Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

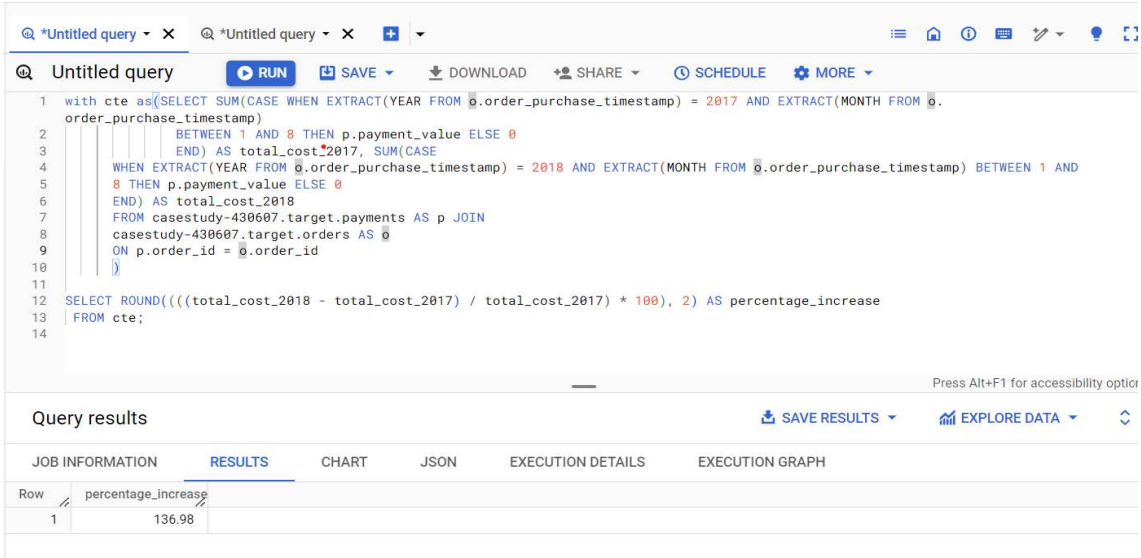
1.Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).
You can use the "payment_value" column in the payments table to get the cost of orders.

Ans:

```
With cte as(
      SELECT
      SUM(CASE WHEN EXTRACT(YEAR FROM   o.order_purchase_timestamp) =
      2017 AND EXTRACT(MONTH FROM o.order_purchase_timestamp)
      BETWEEN 1 AND 8 THEN p.payment_value ELSE 0
      END) AS total_cost_2017,
      SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
      2018 AND EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN
      1 AND 8 THEN p.payment_value ELSE 0 END) AS total_cost_2018
      FROM casestudy-430607.target.payments AS p JOIN
      casestudy-430607.target.orders AS o
      ON p.order_id = o.order_id
      )

SELECT
      ROUND((((total_cost_2018 - total_cost_2017) / total_cost_2017)
      * 100), 2) AS percentage_increase
FROM cte;
```

Output:



Insights:

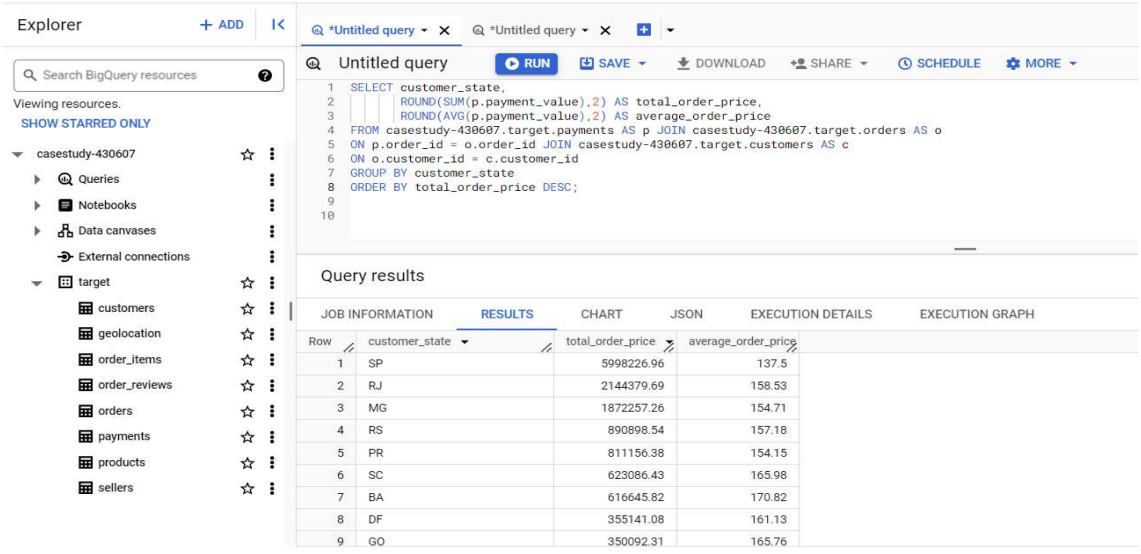
In the above query first we have written cte for finding the cost of orders placed from January to August in the year 2017 and 2018 individually.After analysing the result we can find the growth rate approximately 137% from 2017 to 2018.

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

Ans:

```
SELECT customer_state,
      ROUND(SUM(p.payment_value),2) AS total_order_price,
      ROUND(AVG(p.payment_value),2) AS average_order_price
FROM casestudy-430607.target.paymentsAS p JOIN
casestudy-430607.target.orders AS o
ON p.order_id = o.order_id JOIN
casestudy-430607.target.customers AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state
ORDER BY total_order_price DESC;
```

Output:



Insights:

In the above result the sum of all order prices for each state is displayed in the "total_order_price" column, which represents the total revenue in each state.The "average_order_price" column shows the average order price of customers in each state.

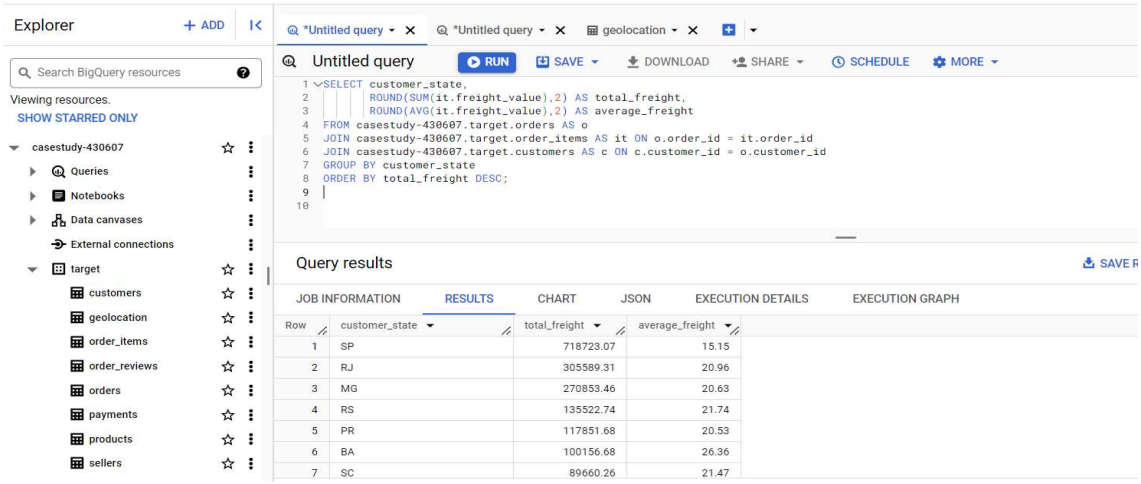
After analysing the query result we can find that the state called "SP" has the highest revenue and the state called "RR" has the lowest revenue.In the state of "RR" the number of customers are less compared to other states.while increasing the customers we can get more revenue because "RR" state has good average_order_price.

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

Ans:

```
SELECT customer_state,
       ROUND(SUM(it.freight_value),2) AS total_freight,
       ROUND(AVG(it.freight_value),2) AS average_freight
FROM casestudy-430607.target.orders AS o
JOIN casestudy-430607.target.order_items AS it ON o.order_id = it.order_id
JOIN casestudy-430607.target.customers AS c ON c.customer_id = o.customer_id
GROUP BY customer_state
ORDER BY total_freight DESC;
```

Output:



Insights:

By analysing the Query result We can find the state called "SP" has the highest total freight costs which could point to regions with higher shipping prices.Understanding the differences in order freight rates between states can offer information about local shipping habits, supplier locations, or client preferences that can be used to optimize processes and cut costs.

Q5)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.

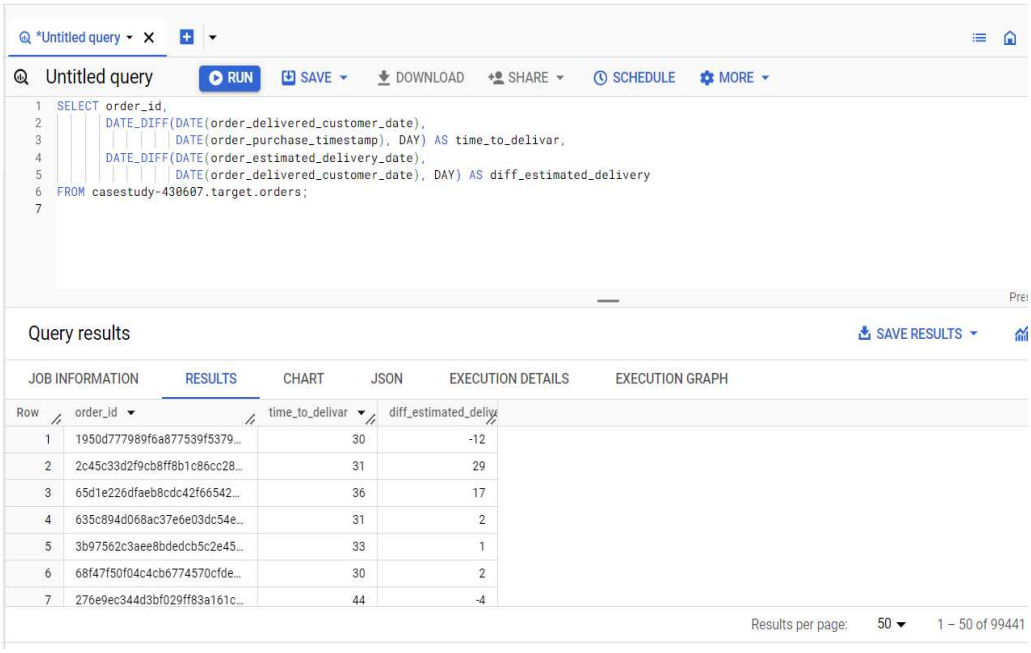
You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

- time_to_deliver = order_delivered_customer_date - order_purchase_timestamp
- diff_estimated_delivery = order_delivered_customer_date - order_estimated_delivery_date

Ans:

```
SELECT order_id,
DATE_DIFF(DATE(order_delivered_customer_date),
          DATE(order_purchase_timestamp), DAY) AS time_to_delivar,
DATE_DIFF(DATE(order_estimated_delivery_date),
        DATE(order_delivered_customer_date), DAY) AS diff_estimated_delivery
FROM casestudy-430607.target.orders;
```

Output:



Insights:

By analysing the above query result we can find out the effectiveness of the delivery process, including any delays or early deliveries.It can be applied to manage customer expectations, enhance customer satisfaction, optimize the delivery process.

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

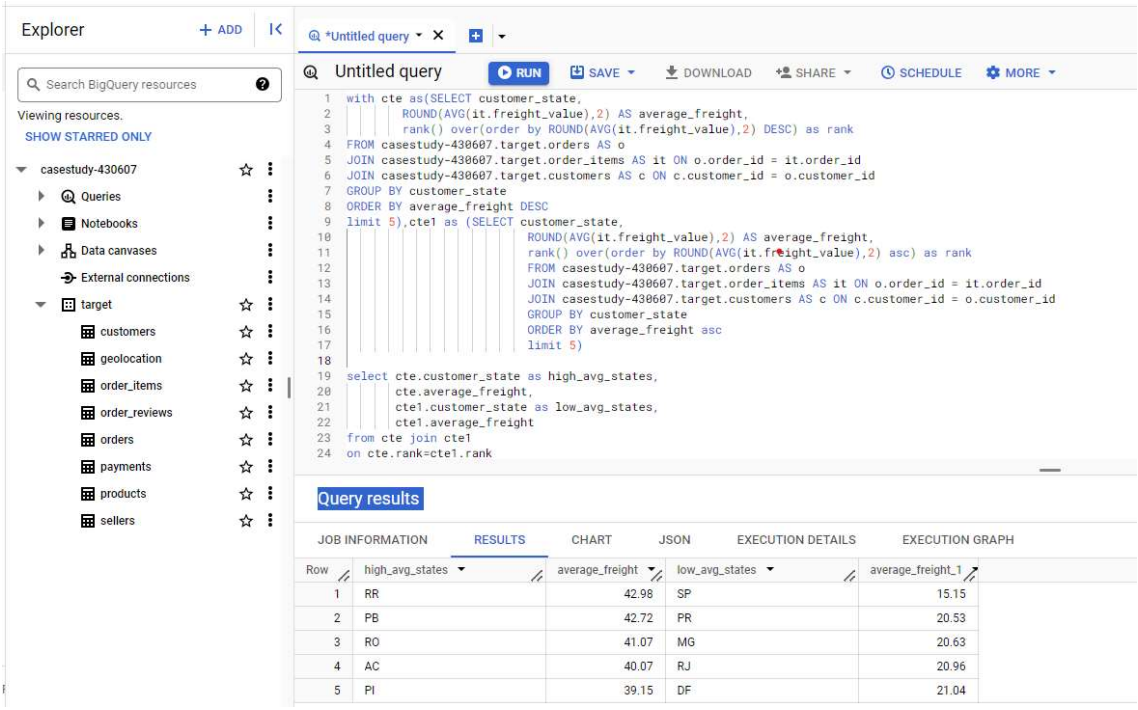
Ans:

```
with cte as(
SELECT customer_state,
    ROUND(AVG(it.freight_value),2) AS average_freight,
    rank() over(order by ROUND(AVG(it.freight_value),2) DESC) as rank
FROM casestudy-430607.target.orders AS o
JOIN casestudy-430607.target.order_items AS it ON o.order_id = it.order_id
```

```
JOIN casestudy-430607.target.customers AS c ON c.customer_id = o.customer_id
GROUP BY customer_state
ORDER BY average_freight DESC
limit 5),
cte1 as (
SELECT customer_state,
       ROUND(AVG(it.freight_value),2) AS average_freight,
       rank() over(order by ROUND(AVG(it.freight_value),2) asc) as rank
FROM casestudy-430607.target.orders AS o
JOIN casestudy-430607.target.order_items AS it
 ON o.order_id = it.order_id
JOIN casestudy-430607.target.customers AS c ON c.customer_id = o.customer_id
GROUP BY customer_state
ORDER BY average_freight asc
limit 5)

select cte.customer_state as high_avg_states,
       cte.average_freight,
       cte1.customer_state as low_avg_states,
       cte1.average_freight
from cte join cte1
on cte.rank=cte1.rank
```

Output:



Insights:

In the above query result we can find that The states with the highest average freight values like states called RR and PB may experience greater shipping prices due to various reasons.And the states with the lowest average freight values like states such as SP and PR relatively reduced shipping prices by looking at the average freight values.

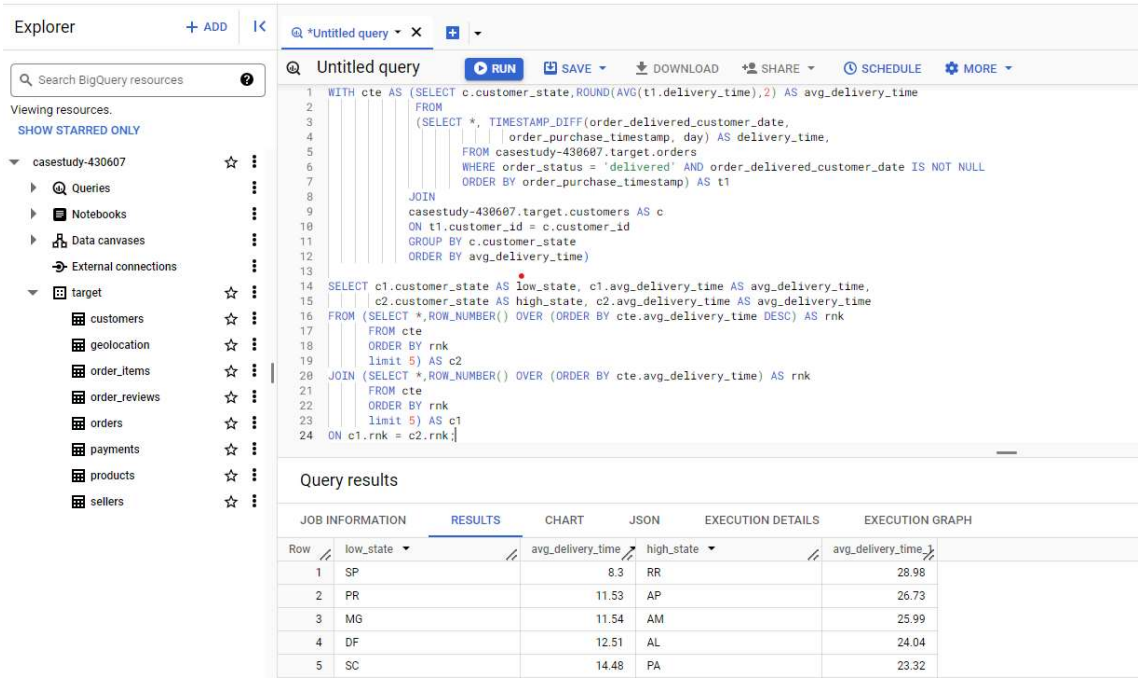
3.Find out the top 5 states with the highest & lowest average delivery time.

Ans:

```
WITH cte AS
(SELECT c.customer_state,ROUND(AVG(t1.delivery_time),2) AS avg_delivery_time
 FROM (SELECT *, TIMESTAMP_DIFF(order_delivered_customer_date,
                order_purchase_timestamp, day) AS delivery_time,
       FROM casestudy-430607.target.orders
       WHERE order_status = 'delivered'
             AND order_delivered_customer_date IS NOT NULL
       ORDER BY order_purchase_timestamp) AS t1
 JOIN casestudy-430607.target.customers AS c
 ON t1.customer_id = c.customer_id
 GROUP BY c.customer_state
 ORDER BY avg_delivery_time)

SELECT c1.customer_state AS low_state,
       c1.avg_delivery_time AS avg_delivery_time,
       c2.customer_state AS high_state,
       c2.avg_delivery_time AS avg_delivery_time
FROM
(SELECT *,ROW_NUMBER() OVER (ORDER BY cte.avg_delivery_time DESC) AS rnk
 FROM cte
ORDER BY rnk LIMIT 5) AS c2
JOIN
(SELECT *,ROW_NUMBER() OVER (ORDER BY cte.avg_delivery_time) AS rnk
FROM cte
ORDER BY rnk LIMIT 5) AS c1
ON c1.rnk = c2.rnk;
```

Output:



Insights:

In the above query result we can find that the states with the highest average delivery time like states called RR and AP and the states with the lowest average delivery time like states called SP and PR.These insights can be helpful for our company looking to improve customer satisfaction, operational efficiency, delivery process optimization, and setting reasonable expectations for customers based on regional delivery time.

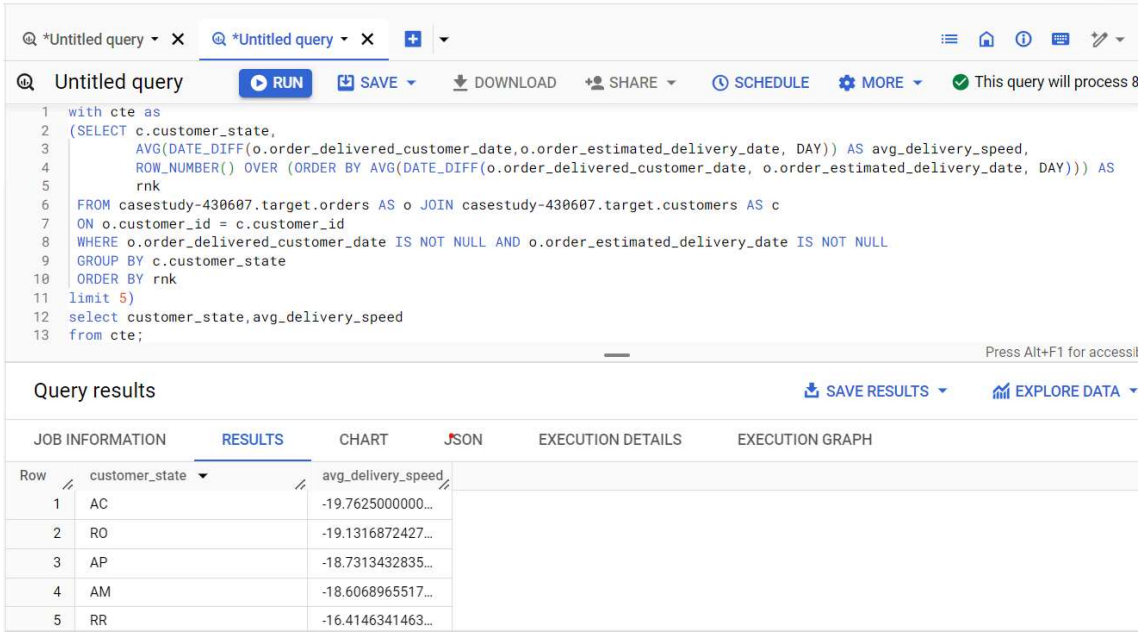
4.Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

Ans:

```sql
with cte as
(SELECT c.customer_state,
AVG(DATE_DIFF(o.order_delivered_customer_dat,
o.order_estimated_delivery_date, DAY)) AS avg_delivery_speed,
ROW_NUMBER() OVER (ORDER BY AVG(
DATE_DIFF(o.order_delivered_customer_date,
o.order_estimated_delivery_date, DAY))) AS rnk
FROM casestudy-430607.target.orders AS o
JOIN casestudy-430607.target.customers AS c
ON o.customer_id = c.customer_id
WHERE o.order_delivered_customer_date IS NOT NULL AND
o.order_estimated_delivery_date IS NOT NULL
GROUP BY c.customer_state
ORDER BY rnk
limit 5)

select customer_state,avg_delivery_speed
from cte;
```

Output:



Insights:

In the above query result we can find that AC,RO,AP,AM and RR are the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.we can use this data to get more revenue from the customers of these states through increasing their orders.

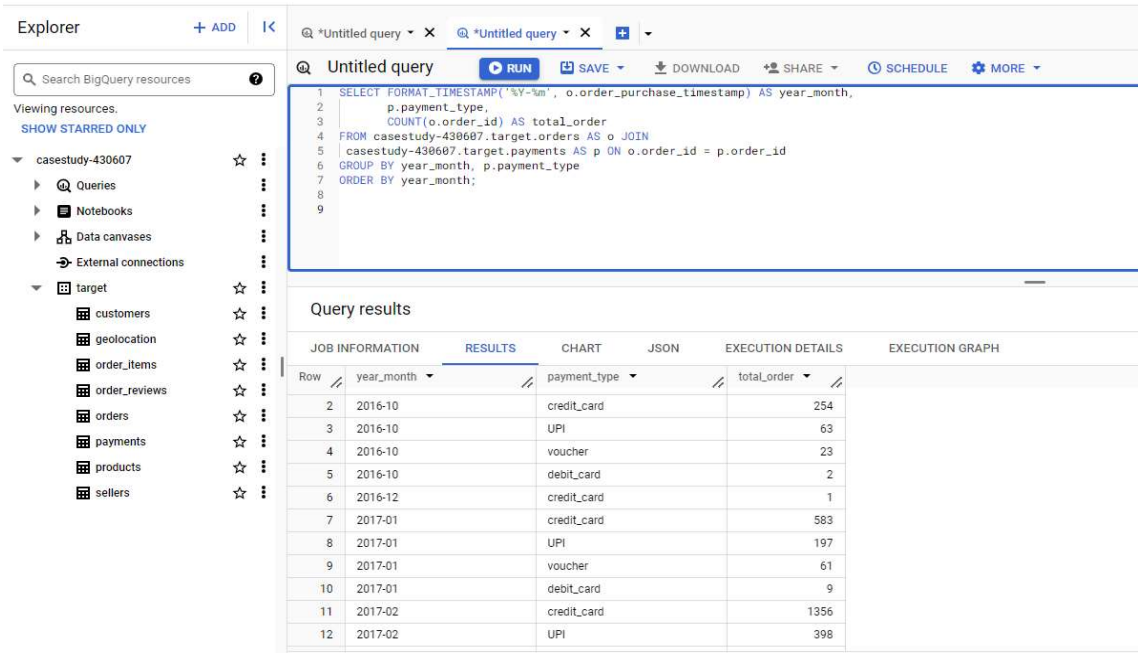
Q6)Analysis based on the payments:

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

Ans:

```
SELECT FORMAT_TIMESTAMP('%Y-%m', o.order_purchase_timestamp) AS year_month,
      p.payment_type,
      COUNT(o.order_id) AS total_order
FROM casestudy-430607.target.orders AS o JOIN
casestudy-430607.target.payments AS p ON o.order_id = p.order_id
GROUP BY year_month, p.payment_type
ORDER BY year_month;
```

Output:



Insights:

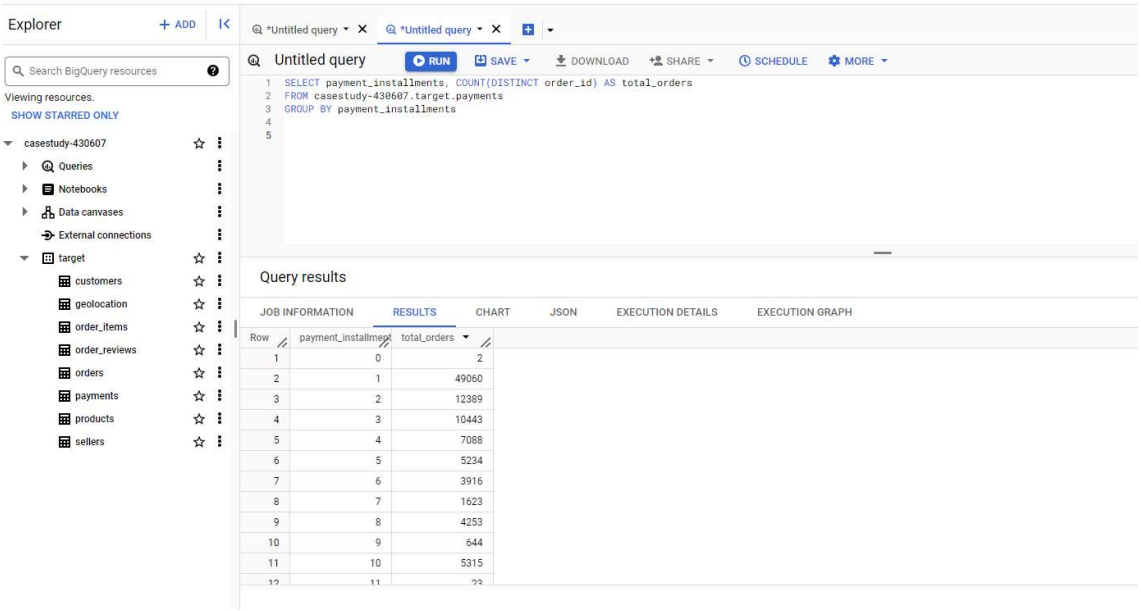
In the above query result we can find that credit card as a payment method was most used in November 2017.Because highest number of orders placed in the november 2017 through credit card.

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

Ans:

```
SELECT payment_installments,
      COUNT(DISTINCT order_id) AS total_orders
FROM casestudy-430607.target.payments
GROUP BY payment_installments
```

Output:



Insights:

In the above query result we can find that 49060 orders were placed where payment instalment was 1.According to this analysis we can say that most of the customers are prefer Or interested in the one payment instalment process.