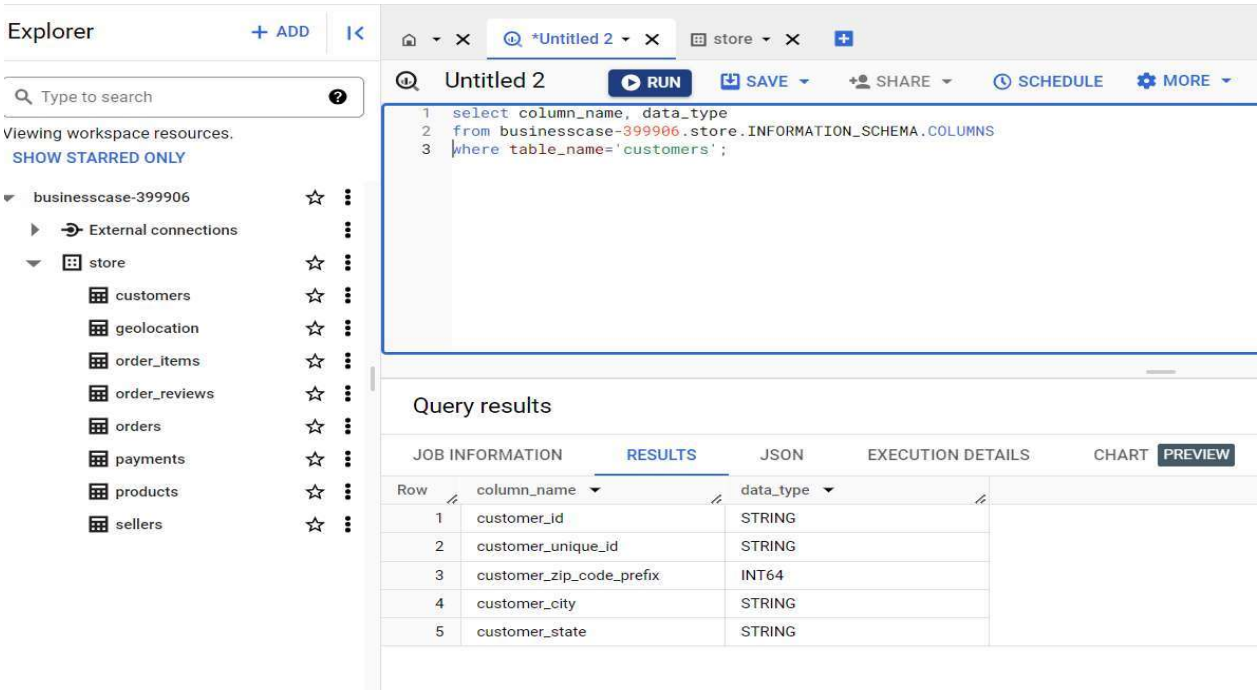
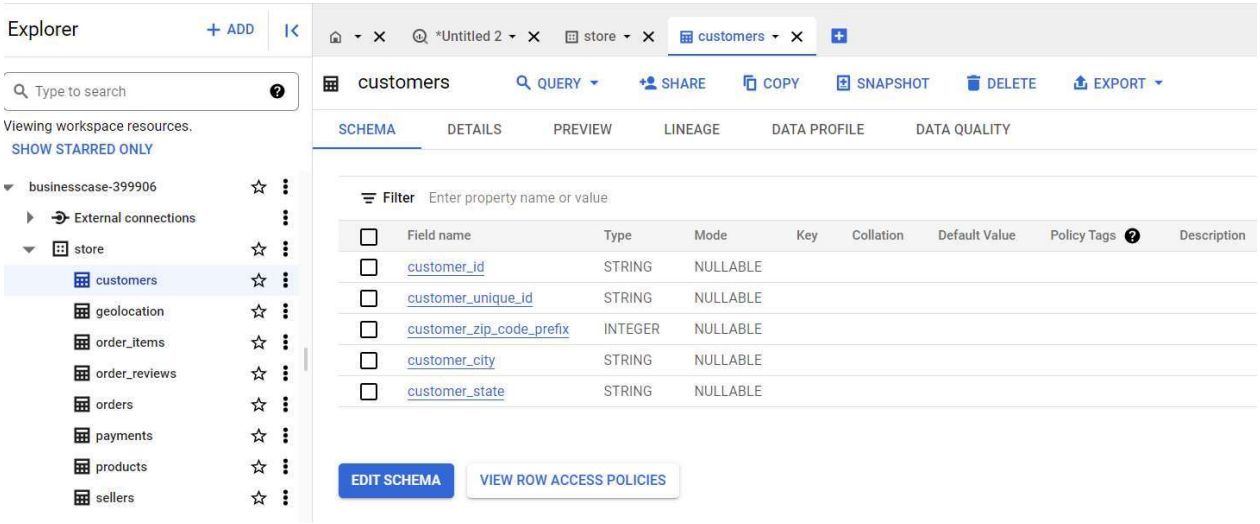
1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:
    1. Data type of all columns in the "customers" table.

```
select column_name, data_type from
businesscase-399906.store.INFORMATION_SCHEMA.COLUMNS where
table_name='customers';
```

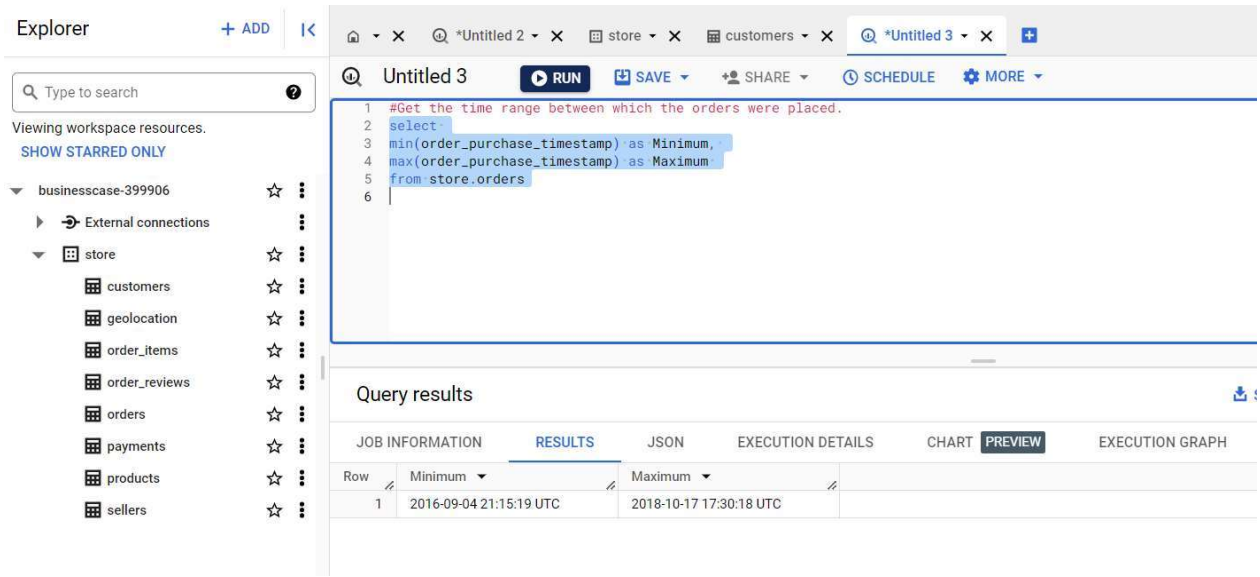




Insights:- *From above we can see that customer table has 5 columns namely, customer_id, customer_unique_id,customer_zip_code_prefix,Customer_city,customer_state. Apart from customer_zip_code_prefix which is of type Integer, all other columns are of String data type. Also, all columns can contain Null.*

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

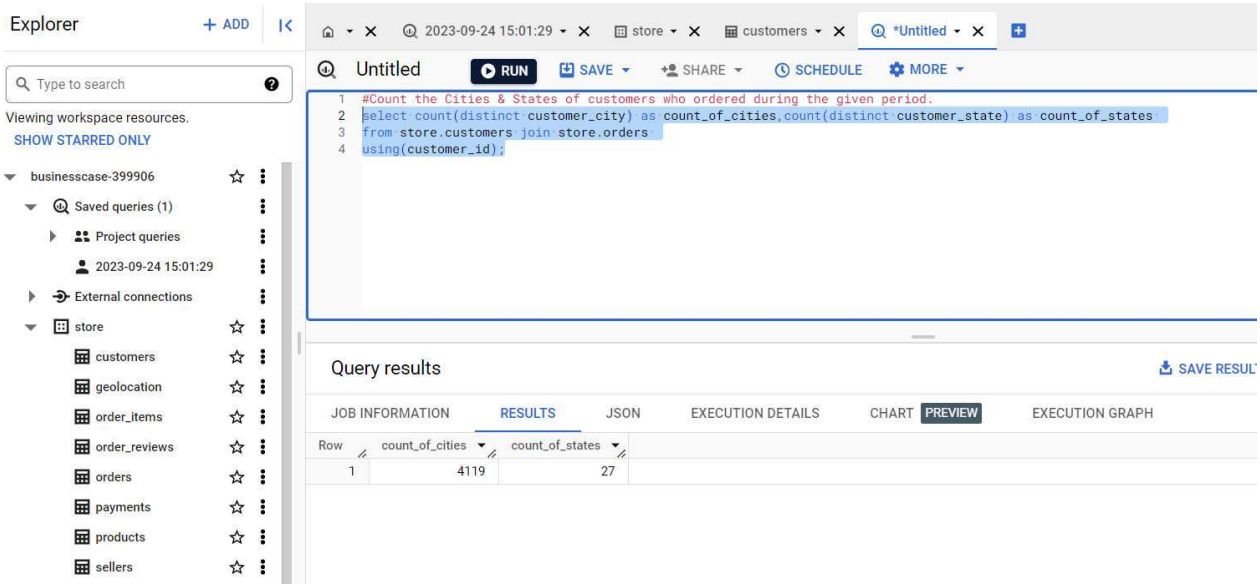
```
select
min(order_purchase_timestamp) as Minimum,
max(order_purchase_timestamp) as Maximum
from store.orders
```



Insights:- *From above we can conclude that the orders were placed in the range of 772 days approx.*

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

```
select count(distinct customer_city) as count_of_cities,count(distinct
customer_state) as count_of_states
from store.customers join store.orders
using(customer_id);
```

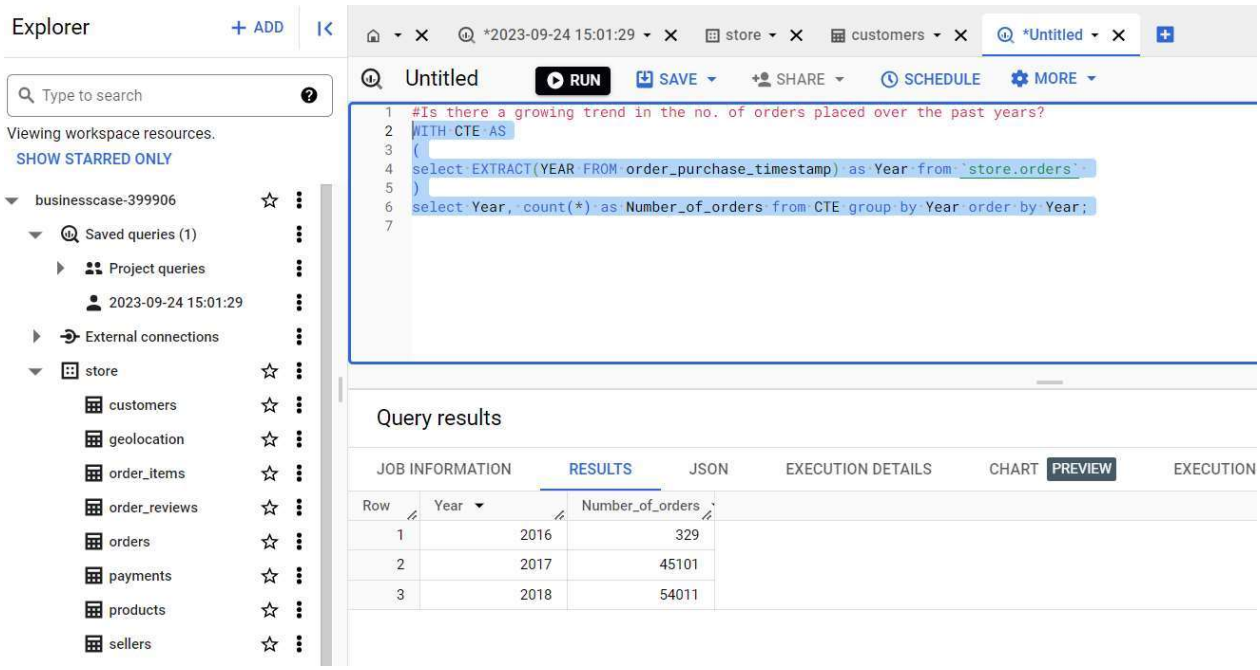


Insights:- *From above we can conclude that the orders were placed by customers belonging to 4119 different cities covering all the 26 states and 1 federal district of Brazil.*

2. In-depth Exploration:

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

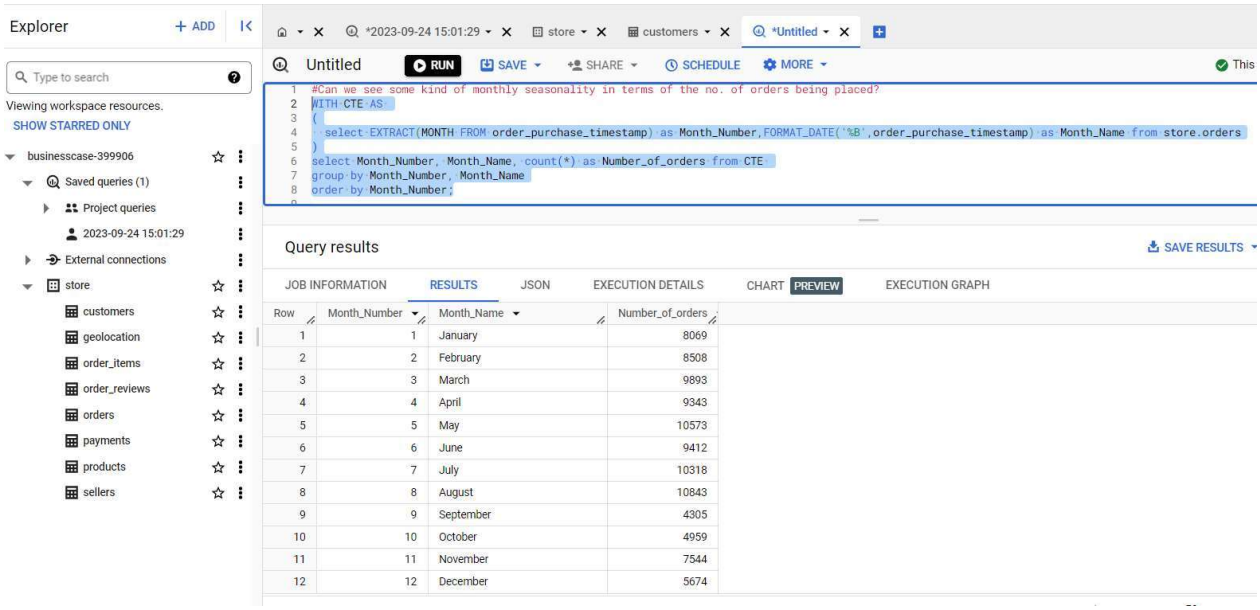
```
WITH CTE AS
(
select EXTRACT(YEAR FROM order_purchase_timestamp) as Year from
store.orders
)
select Year, count(*) as Number_of_orders from CTE group by Year order by
Year;
```



Insights:- *From the analysis of data we can see that the company started taking orders at the end of 2nd last quarter of 2016 so the number of orders in 2016 is significantly low. Number of orders are consecutively increasing from the year 2017.*

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

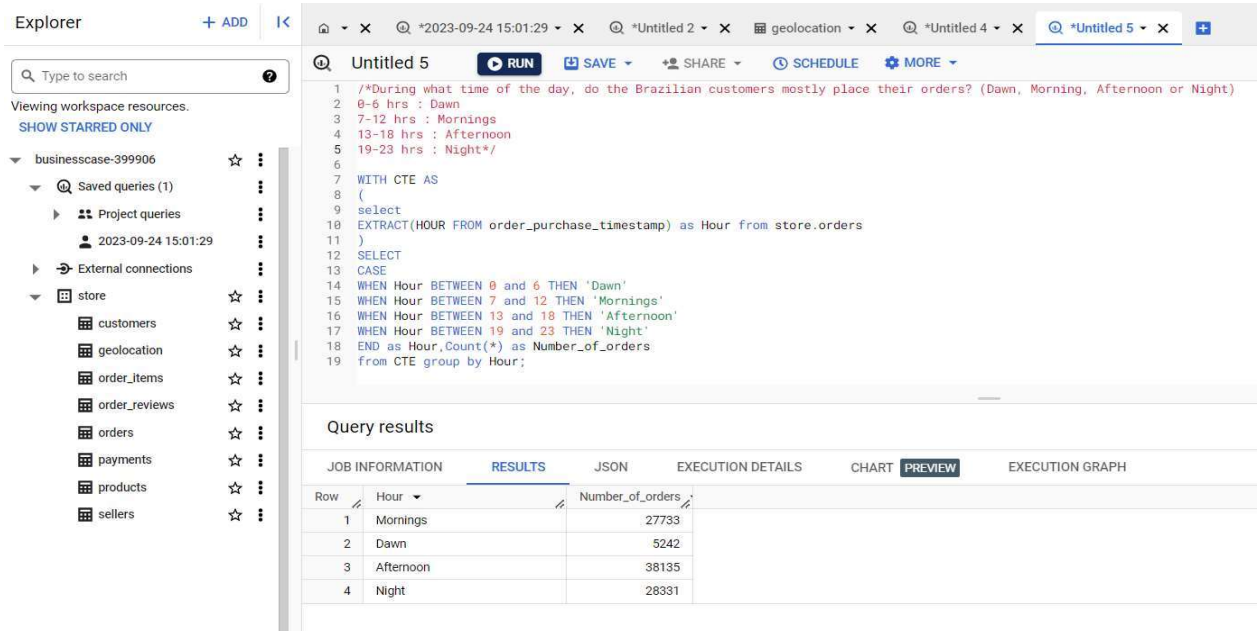
```
WITH CTE AS
(
select EXTRACT(MONTH FROM order_purchase_timestamp) as
Month_Number,FORMAT_DATE('%B',order_purchase_timestamp) as Month_Name
from store.orders
)
select Month_Number, Month_Name, count(*) as Number_of_orders from CTE
group by Month_Number, Month_Name
order by Month_Number;
```



Insights:- *We can see that the lowest number of orders were placed during the Spring season and highest number of orders were placed in Winter.*
In Brazil the year starts with the summer season followed by Autumn, Winter and Spring. The number of orders placed shows an increasing trend in the first three seasons but there is a dip in the number of orders being placed in the last season.

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

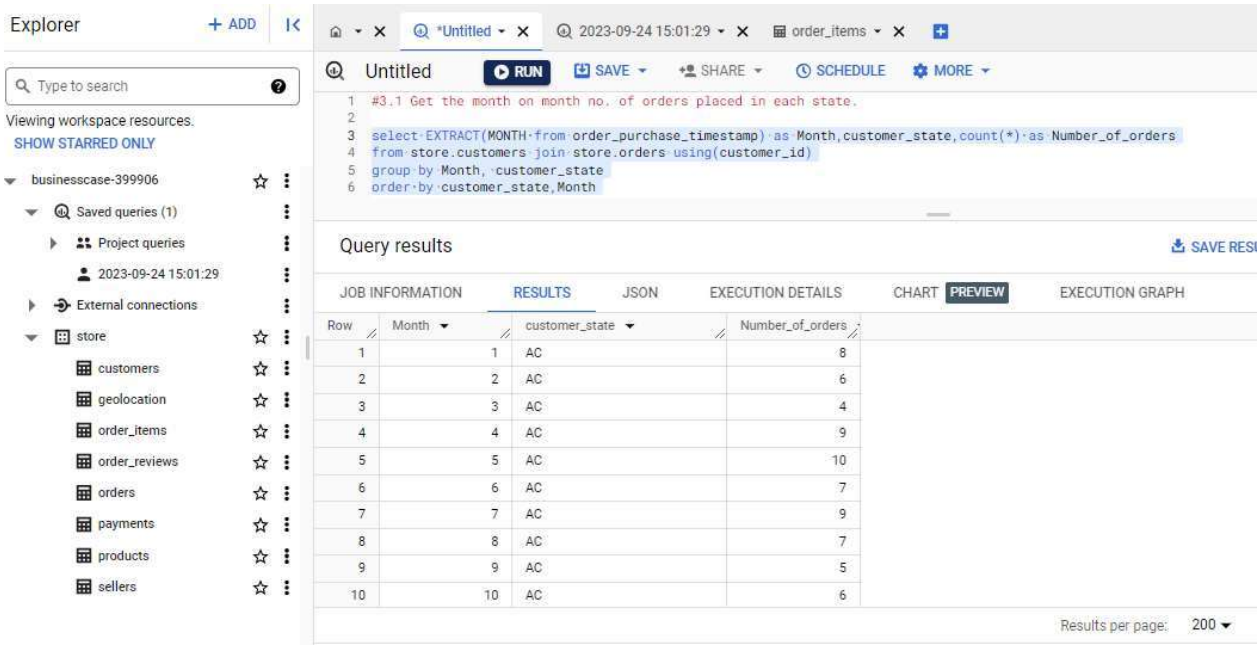
```sql
WITH CTE AS
(
select
EXTRACT(HOUR FROM order_purchase_timestamp) as Hour from
store.orders
)
SELECT
CASE
WHEN Hour BETWEEN 0 and 6 THEN 'Dawn'
WHEN Hour BETWEEN 7 and 12 THEN 'Mornings'
WHEN Hour BETWEEN 13 and 18 THEN 'Afternoon'
WHEN Hour BETWEEN 19 and 23 THEN 'Night'
END as Hour,Count(*) as Number_of_orders
from CTE group by Hour;
```



Insights:- *Above data shows that the Brazilian customers mostly prefer placing their orders during afternoon hours and less during the dawn hours. Number of orders placed during morning and night hours are also significant compared to the number of orders placed during afternoon hours.*

3. Evolution of E-commerce orders in the Brazil region:
    1. Get the month on month no. of orders placed in each state.

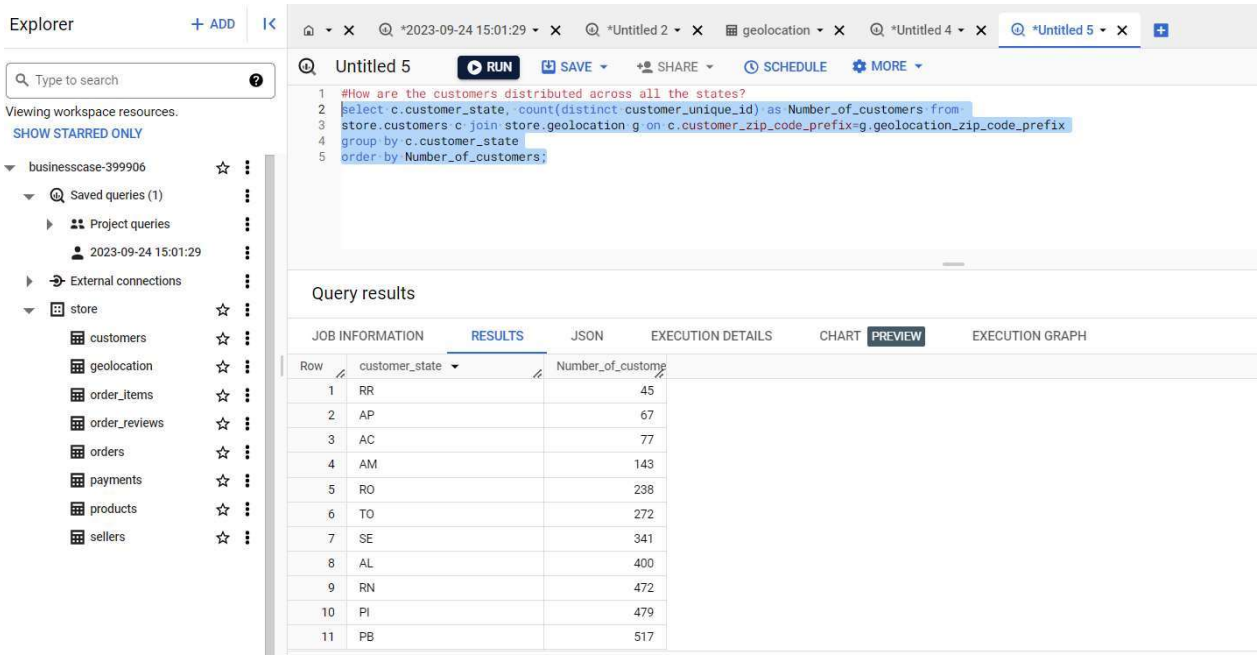
```
select EXTRACT(MONTH from order_purchase_timestamp) as
Month,customer_state,count(*) as Number_of_orders
from store.customers join store.orders using(customer_id)
group by Month, customer_state
order by customer_state,Month
```



Insights:- *Data above shows the number of orders placed each month in each state.*

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

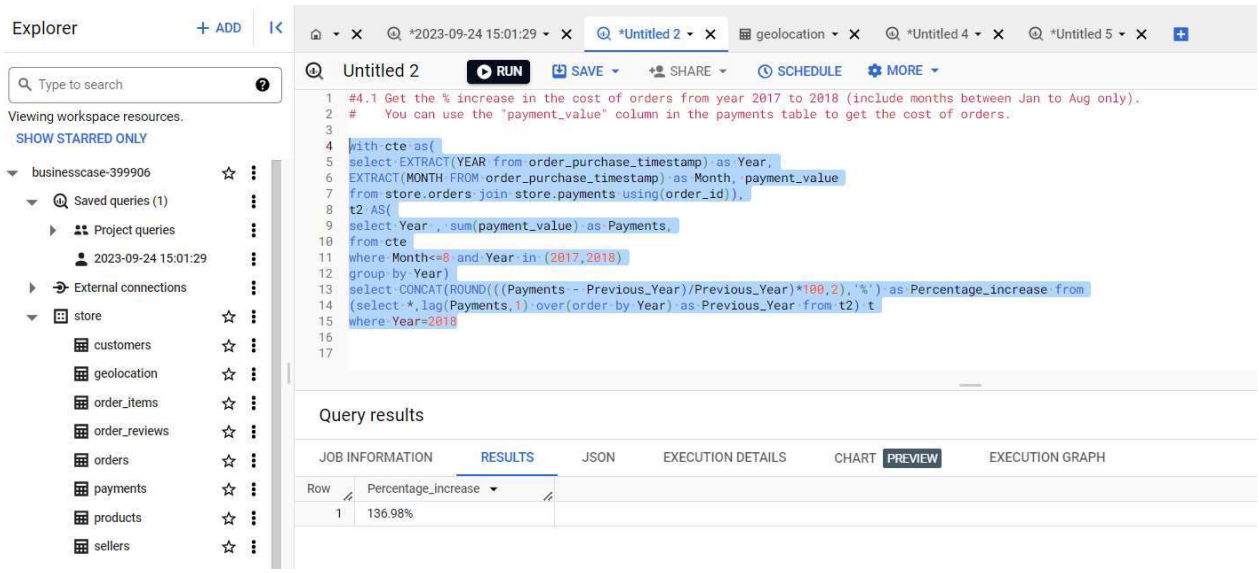
```
select c.customer_state, count(distinct customer_unique_id) as
Number_of_customers from
store.customers c join store.geolocation g on
c.customer_zip_code_prefix=g.geolocation_zip_code_prefix
group by c.customer_state
order by Number_of_customers;
```



Insights:- *Data above shows that the highest numbers of orders were placed from state SP and lowest were placed from RR. Number of orders placed from each state shows that there is quite a variation in numbers.*

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.

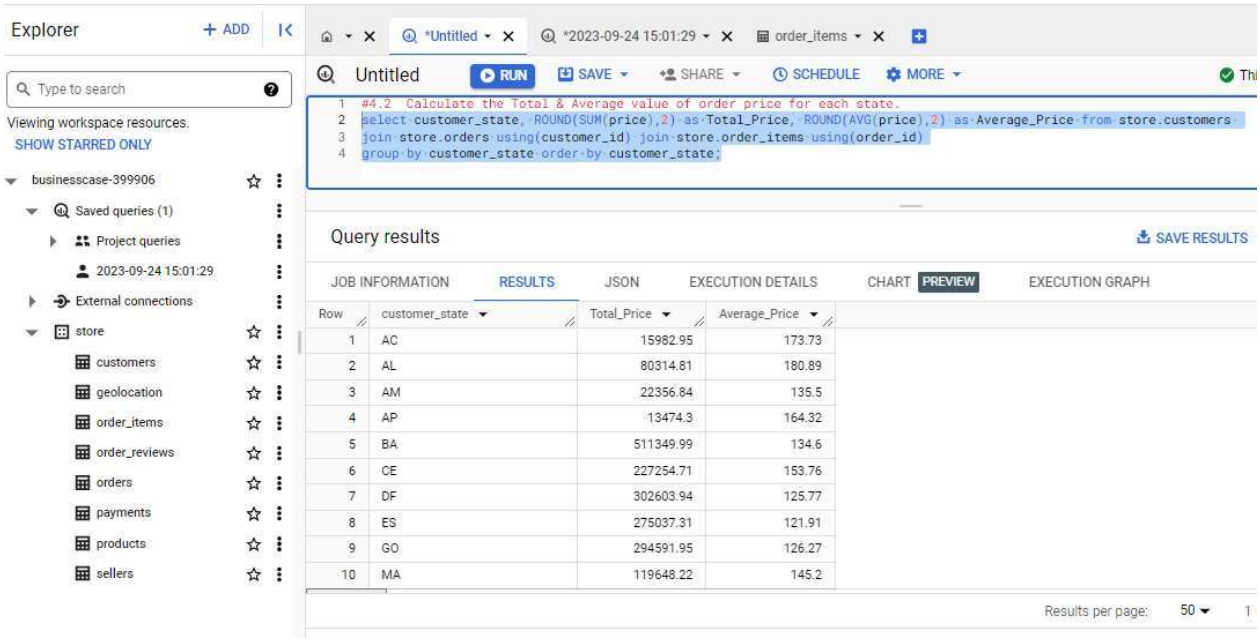
```
with cte as(
select EXTRACT(YEAR from order_purchase_timestamp) as Year,
EXTRACT(MONTH FROM order_purchase_timestamp) as Month, payment_value
from store.orders join store.payments using(order_id)),
t2 AS(
select Year , sum(payment_value) as Payments,
from cte
where Month<=8 and Year in (2017,2018)
group by Year)
select CONCAT(ROUND(((Payments -
Previous_Year)/Previous_Year)*100,2),'%') as Percentage_increase from
(select *,lag(Payments,1) over(order by Year) as Previous_Year from t2) t
where Year=2018
```



Insights:- *Above analysis shows that there is a great increase in percentage of cost of orders between 2017 and 2018 considering the orders between month of January and August.*

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

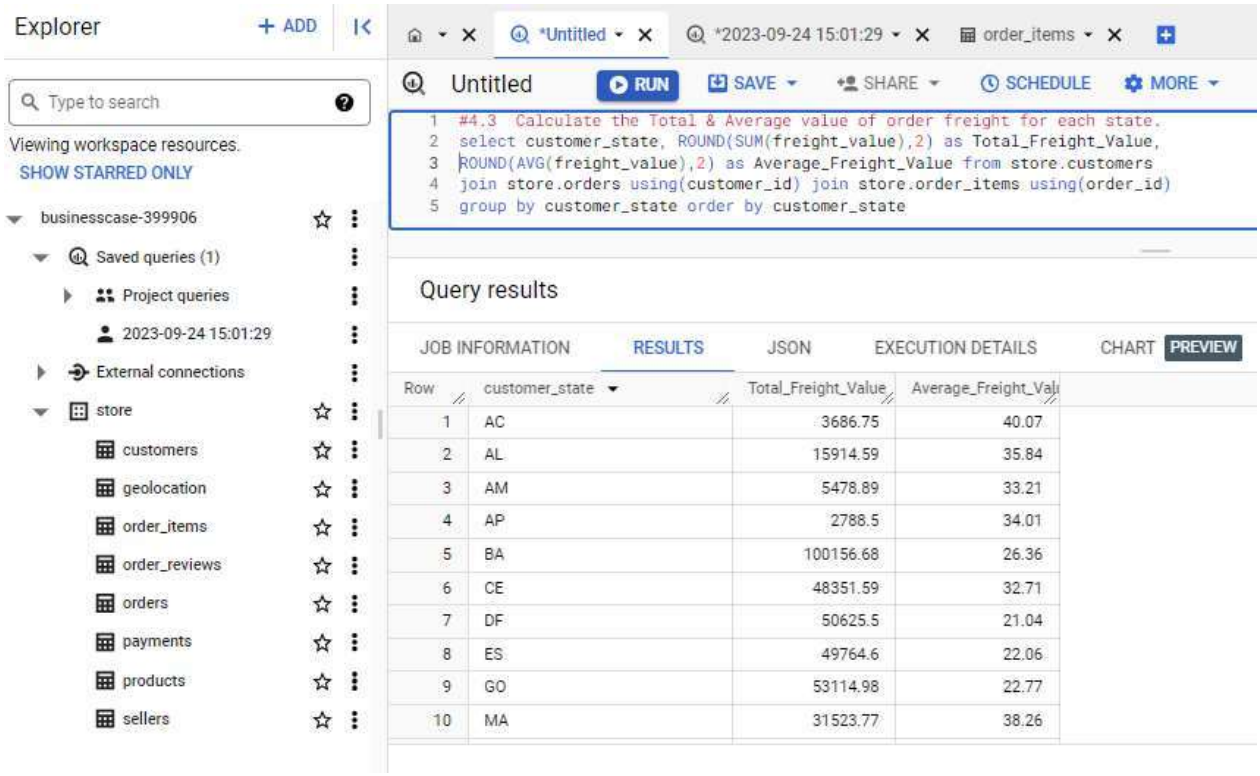
```
select customer_state, ROUND(SUM(price),2) as Total_Price,
ROUND(AVG(price),2) as Average_Price from store.customers
join store.orders using(customer_id) join store.order_items
using(order_id)
group by customer_state order by customer_state;
```



Insights:- *From the above analysis we can say that state SP is leading in terms of total price of orders and RR is at the bottom of the table.*
Also, in terms of average price of orders SP has the lowest average price and PB has the highest.

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

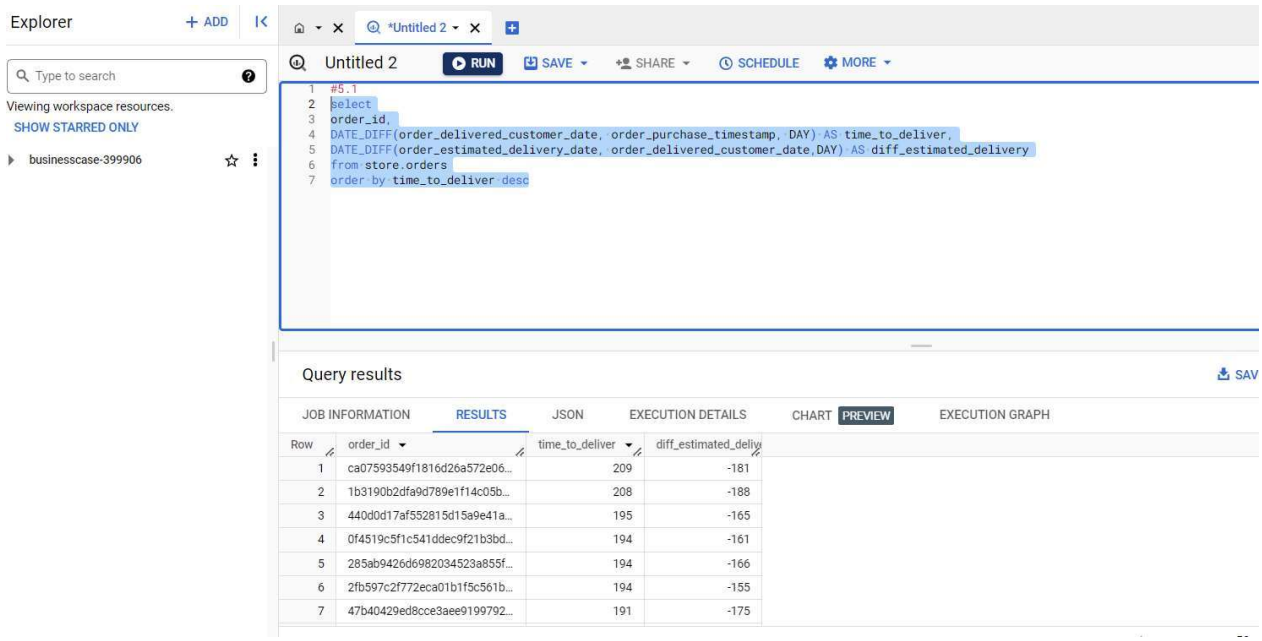
```
select customer_state, ROUND(SUM(freight_value),2) as
Total_Freight_Value, ROUND(AVG(freight_value),2) as Average_Freight_Value
from store.customers
join store.orders using(customer_id)
join store.order_items using(order_id)
group by customer_state order by customer_state
```



Insights:- *From the above analysis we can say that state SP is also leading in terms of total freight of orders and RR is at the bottom of the table.*
Also, in terms of average freight value of orders SP has the lowest and RR has the highest.

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

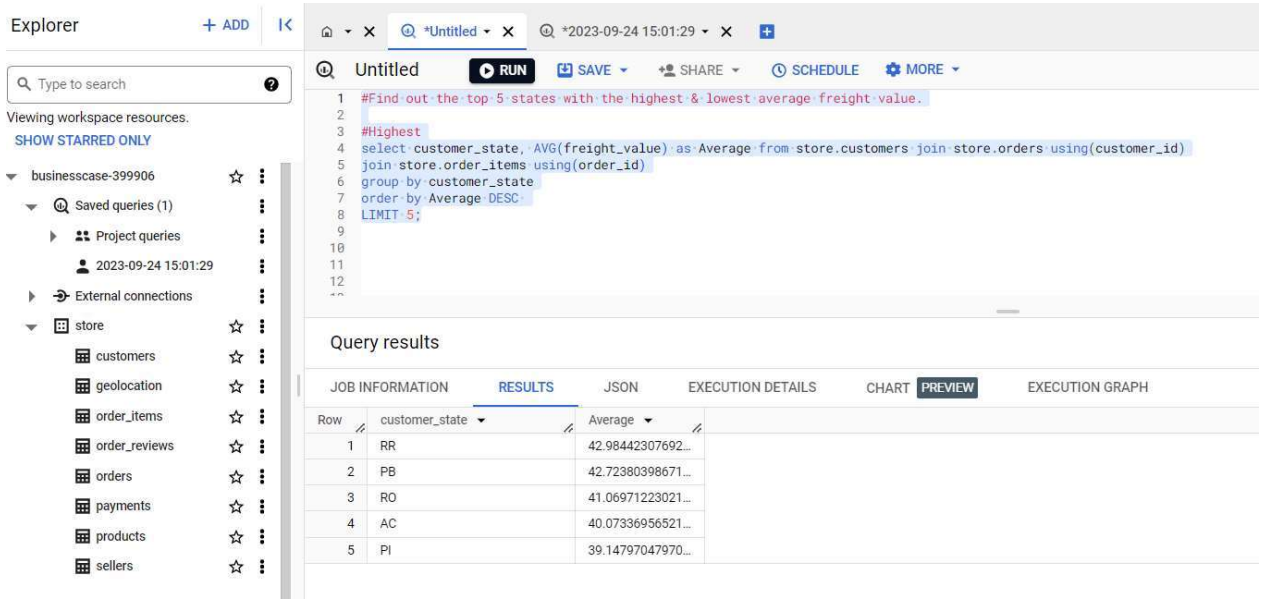
```
select
order_id,
DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,
DAY) AS time_to_deliver,
DATE_DIFF(order_estimated_delivery_date,
order_delivered_customer_date,DAY) AS diff_estimated_delivery
from store.orders
order by time_to_deliver desc
```



Insights:- *From the above analysis we can see that there are a very high number of orders which took more days than usual to deliver them. But most of the orders were delivered before the expected date of delivery.*

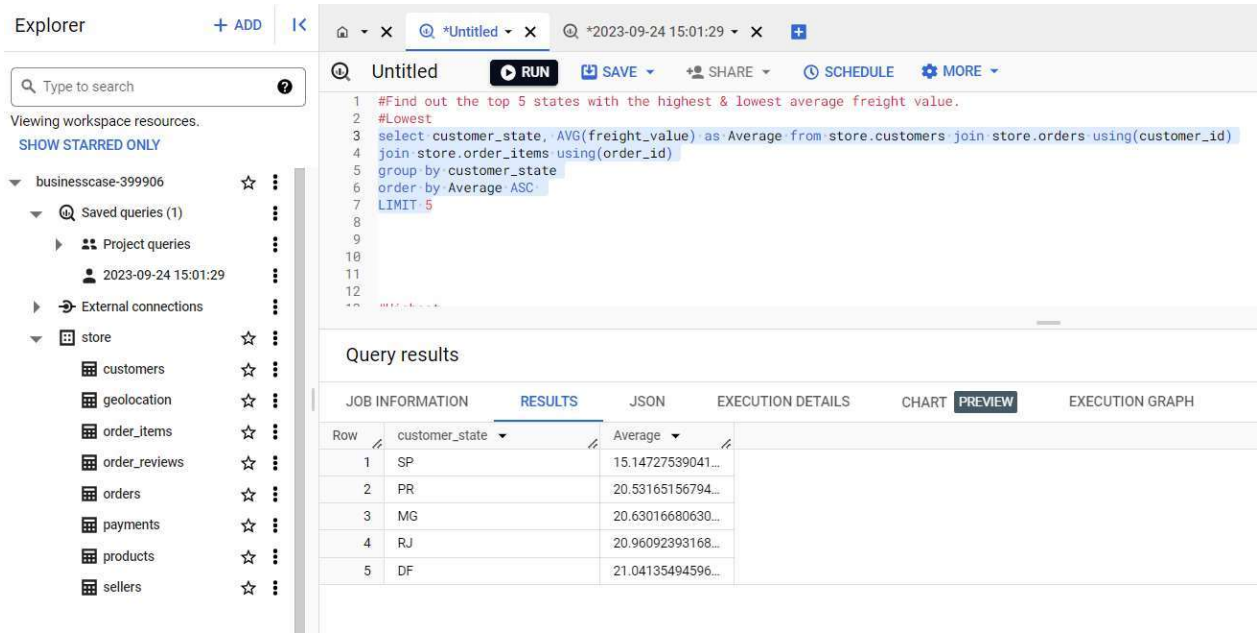
2. Find out the top 5 states with the highest & lowest average freight value.

```
#Highest
select customer_state, AVG(freight_value) as Average from store.customers
join store.orders using(customer_id)
join store.order_items using(order_id)
group by customer_state
order by Average DESC
LIMIT 5;
```



Insights:- *From the above analysis we can say that state RR is having the highest average in terms of freight value, followed by PB,RO,AC and PI.*

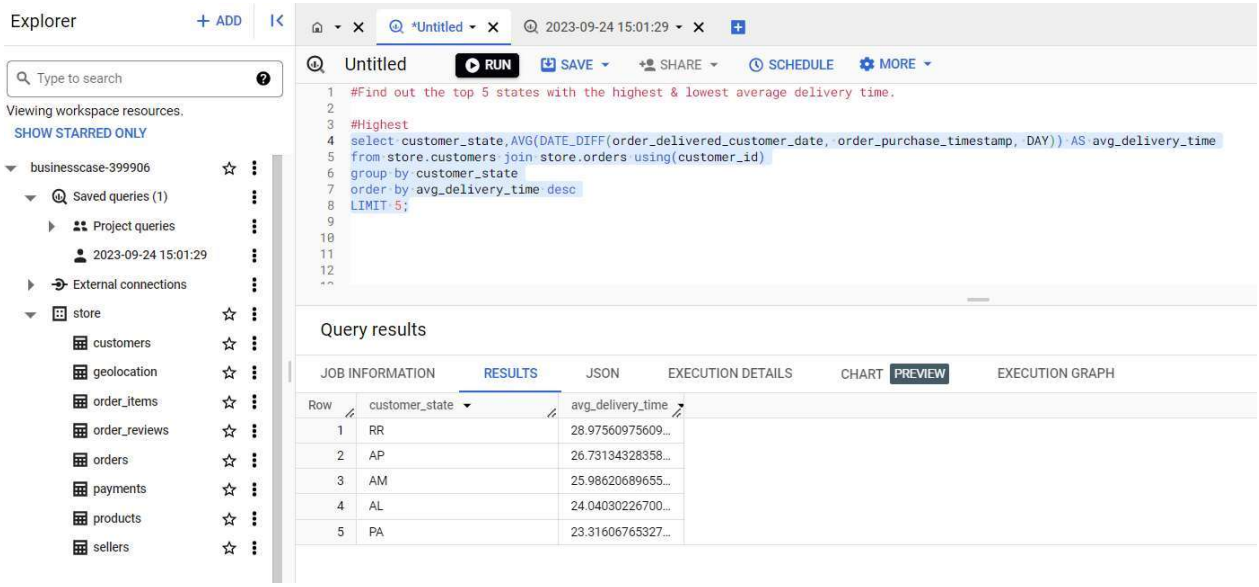
```sql
#Lowest
select customer_state, AVG(freight_value) as Average from store.customers
join store.orders using(customer_id)
join store.order_items using(order_id)
group by customer_state
order by Average ASC
LIMIT 5
```



Insights:- *From the above analysis we can say that state SP is having the lowest average in terms of freight value, followed by PR,MG,RJ and DF.*

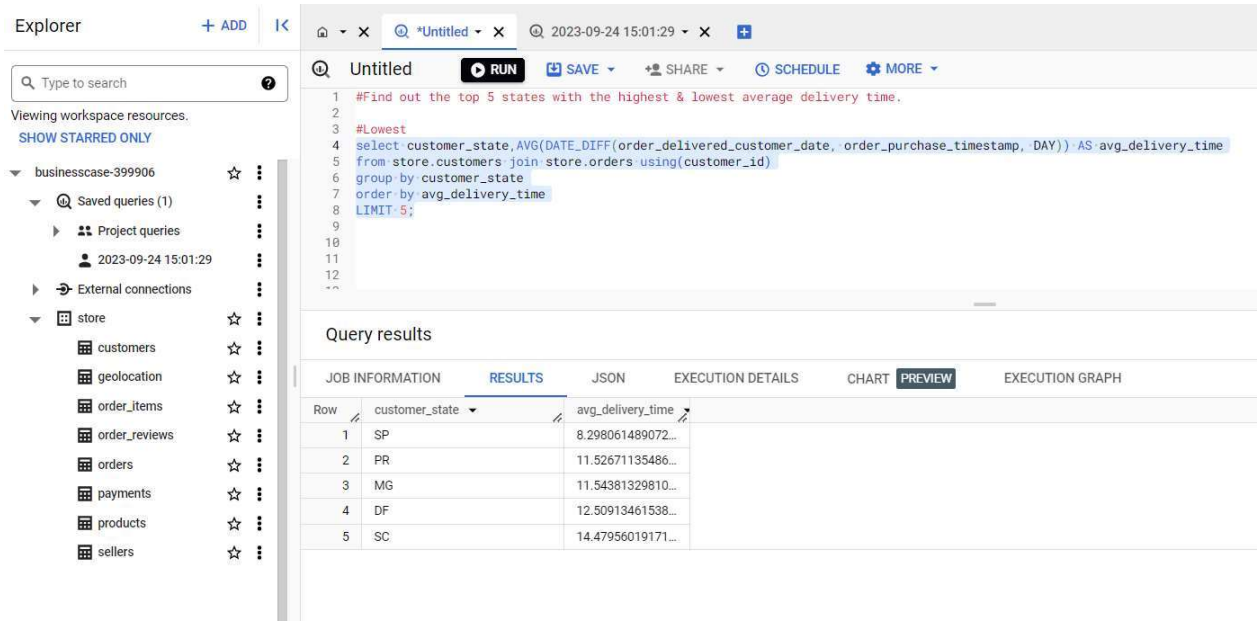
3. Find out the top 5 states with the highest & lowest average delivery time.

```
#Highest
select customer_state,AVG(DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)) AS avg_delivery_time
from store.customers join store.orders using(customer_id)
group by customer_state
order by avg_delivery_time desc
LIMIT 5;
```



Insights:- *RR is leading in terms of highest delivery time, followed by AP, AM.AL and PA.*

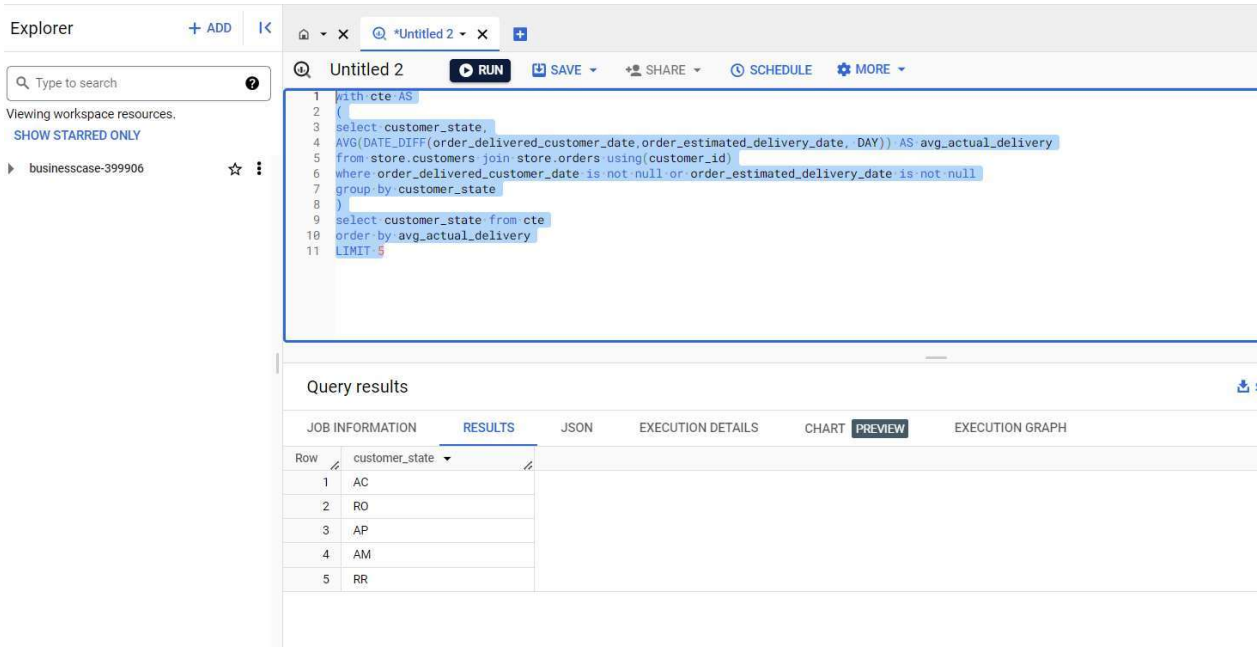
```sql
#Lowest
select customer_state,AVG(DATE_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)) AS avg_delivery_time
from store.customers join store.orders using(customer_id)
group by customer_state
order by avg_delivery_time
LIMIT 5;
```



Insights:- *SP is leading in terms of lowest delivery time, followed by PR,MG,DF and SC.*

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.

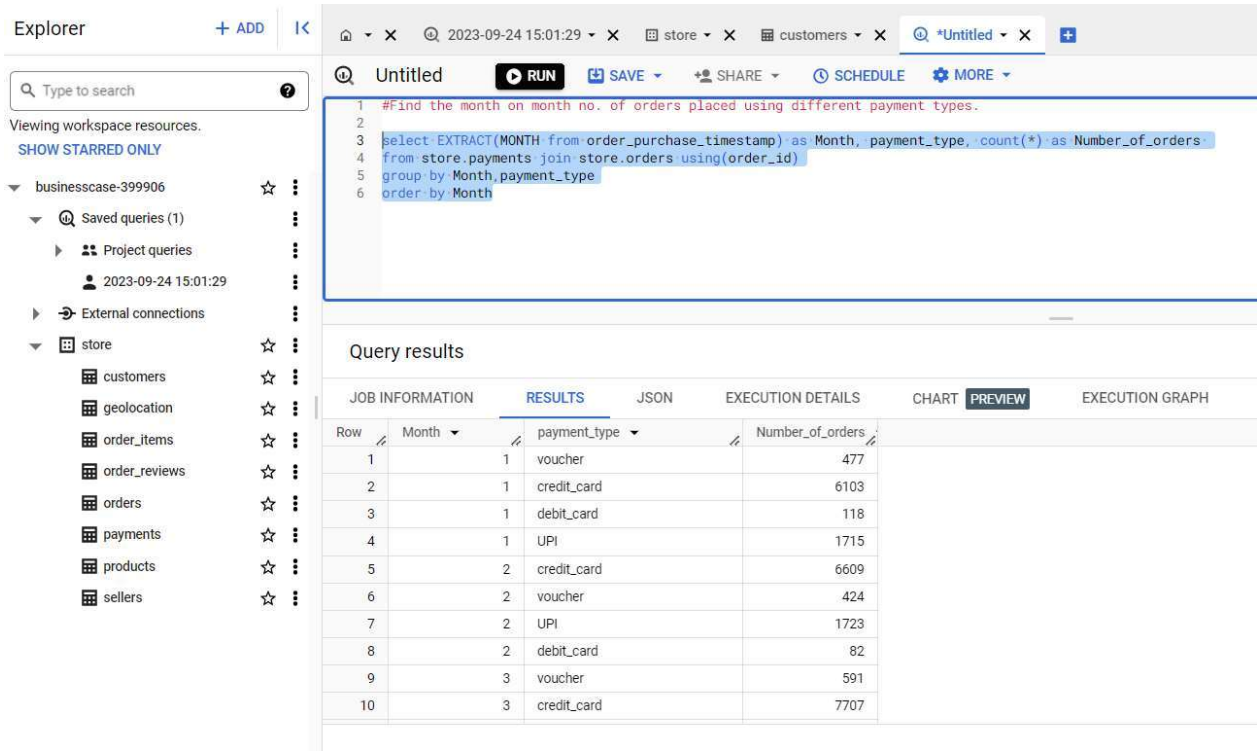
```
with cte AS
(
select customer_state,
AVG(DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_date
, DAY)) AS avg_actual_delivery
from store.customers join store.orders using(customer_id)
where order_delivered_customer_date is not null or
order_estimated_delivery_date is not null
group by customer_state
)
select customer_state from cte
order by avg_actual_delivery
LIMIT 5
```



Insights:- *AL is leading in terms of order being delivered before the expected delivery followed by MA, SE,ES and CE.*

6. Analysis based on the payments:
    1. Find the month on month no. of orders placed using different payment types.

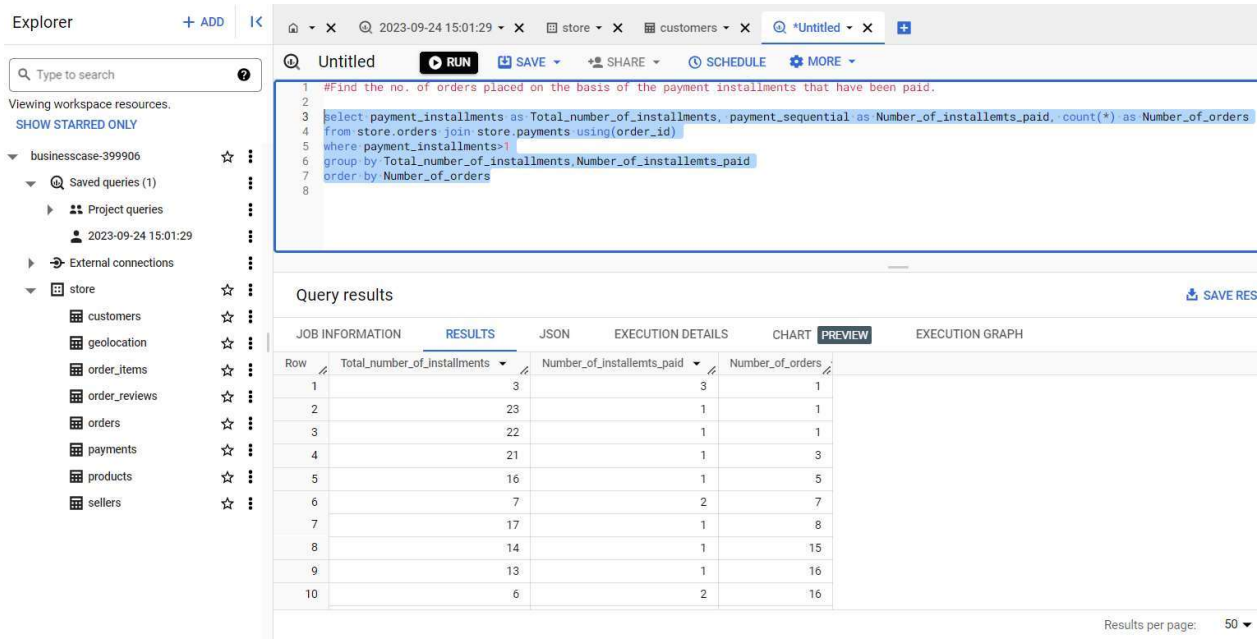
```
select EXTRACT(MONTH from order_purchase_timestamp) as Month,
payment_type, count(*) as Number_of_orders
from store.payments join store.orders using(order_id)
group by Month,payment_type
order by Month;
```



Insights:- *From above analysis we can see that the most preferred payment type for orders is credit card and least preferred is debit card.*
Also we can see there are 2 orders in August and 1 in September where payment type is not defined.

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

```
select payment_installments as Total_number_of_installments,
payment_sequential as Number_of_installemts_paid, count(*) as
Number_of_orders
from store.orders join store.payments using(order_id)
where payment_installments>1
group by Total_number_of_installments,Number_of_installemts_paid
order by Number_of_orders
```



Insights:- *From above analysis we can see that there are significantly higher numbers of orders placed using EMI where the number of EMI is 10 or less. Also, there are very few orders for which EMI is fully paid.*

7. Actionable Insights & Recommendations

a. We can see that there are fewer orders in the last quarter of the year so company can provide good offers during that time to increase the sales during this quarter.

b. There are few states where the customer base is good but many states are having a number of customers below median count so the company should come up with strategies to attract them.

c.There are states where the average order price and freight is high compared to others. The company should try to build networks with local vendors and delivery partners which can reduce the cost to customers.

d. Overall the delivery time is high in most of the states. Good and reliable delivery partner collaboration is recommended.

e. We can see that the order placed using credit cards and on EMI is very high. Company can try to bring offers on credit cards like No cost emi and cashbacks on partnered bank cards.