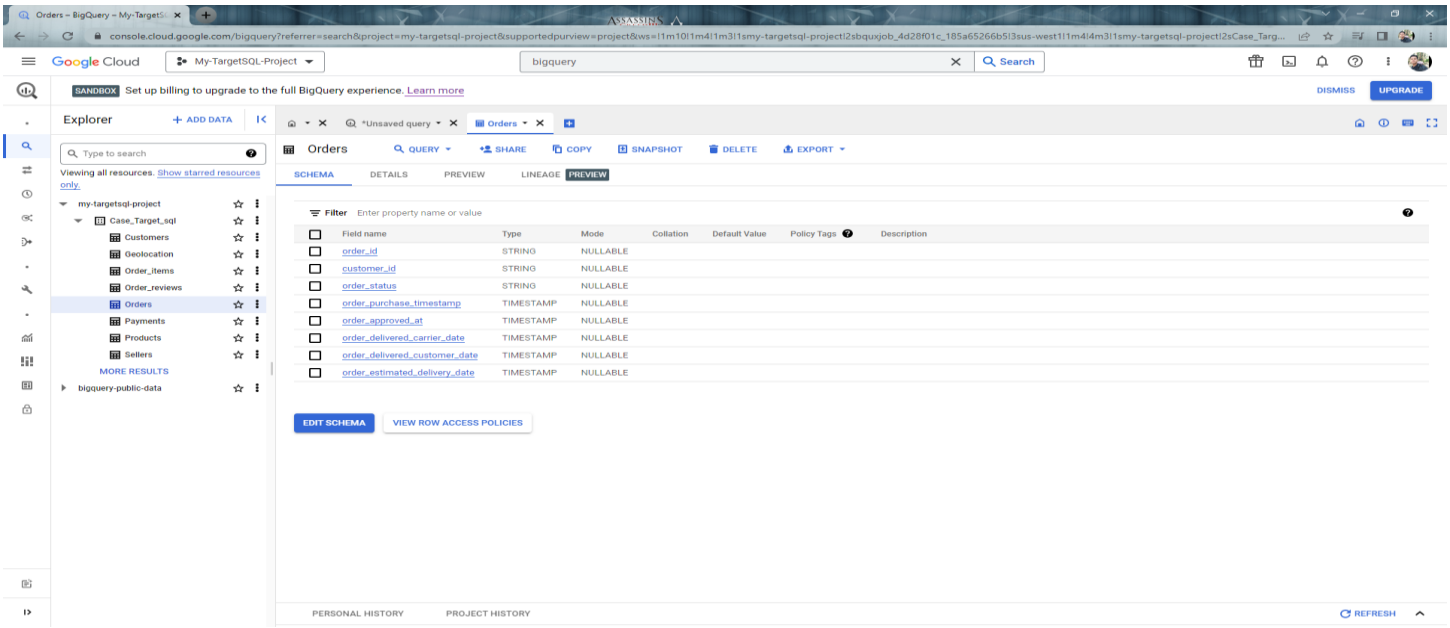
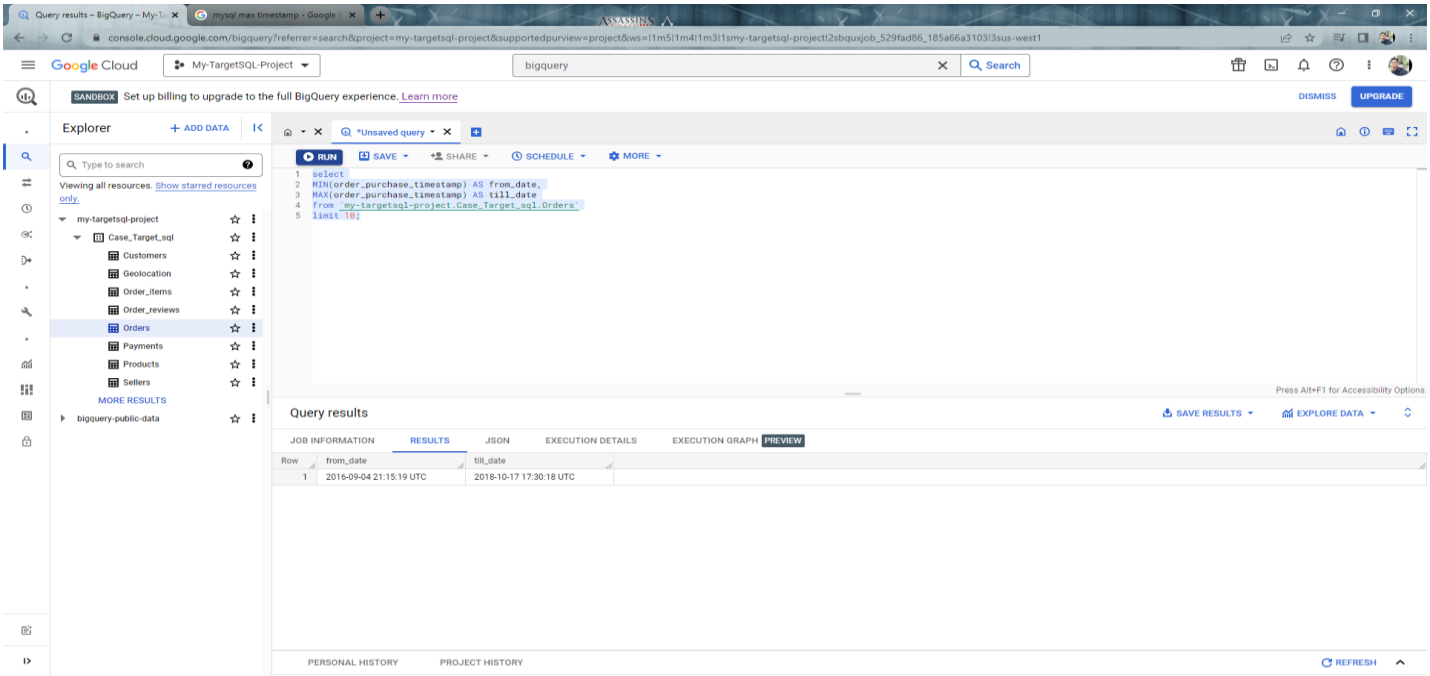
Project_name: my-targetsql-project

1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset
    1. Data type of columns in a table – Orders



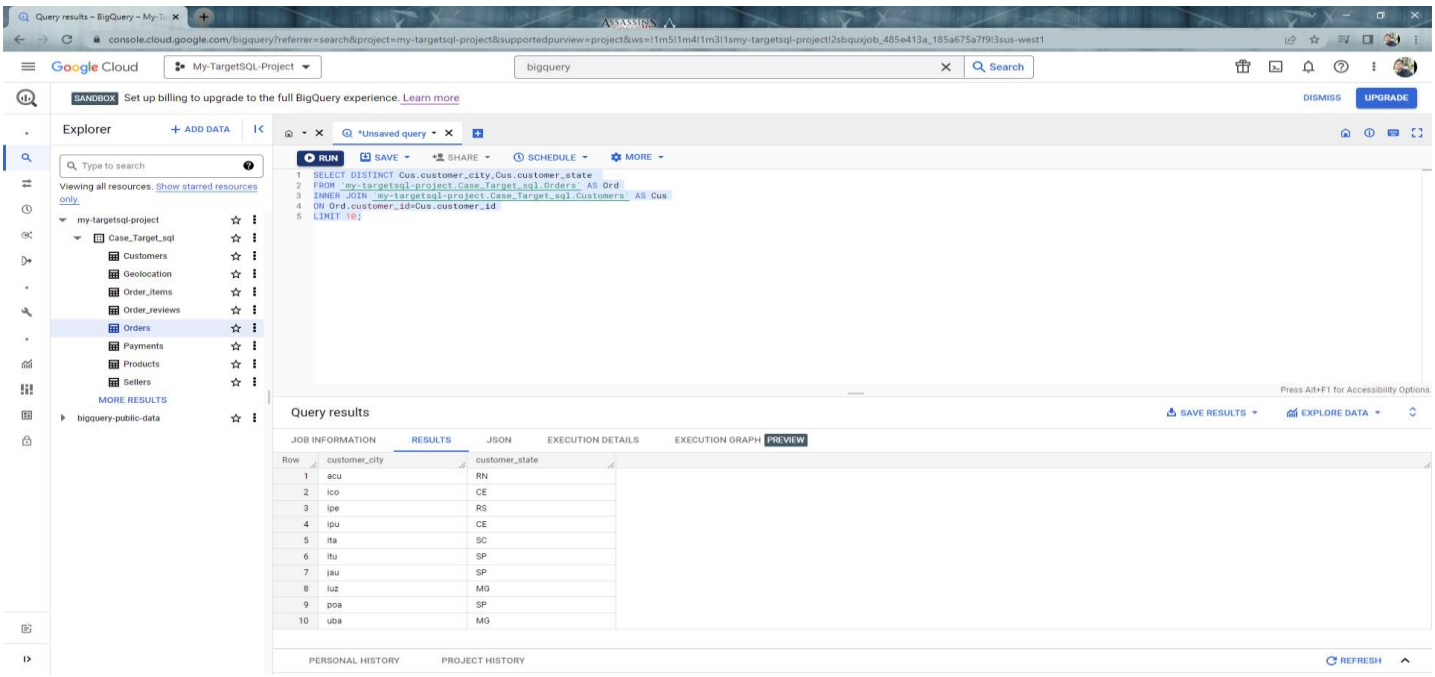
    2. Time period for which the data is given

```
select
MIN(order_purchase_timestamp) AS from_date,
MAX(order_purchase_timestamp) AS till_date
from `my-targetsql-project.Case_Target_sql.Orders`
limit 10;
```



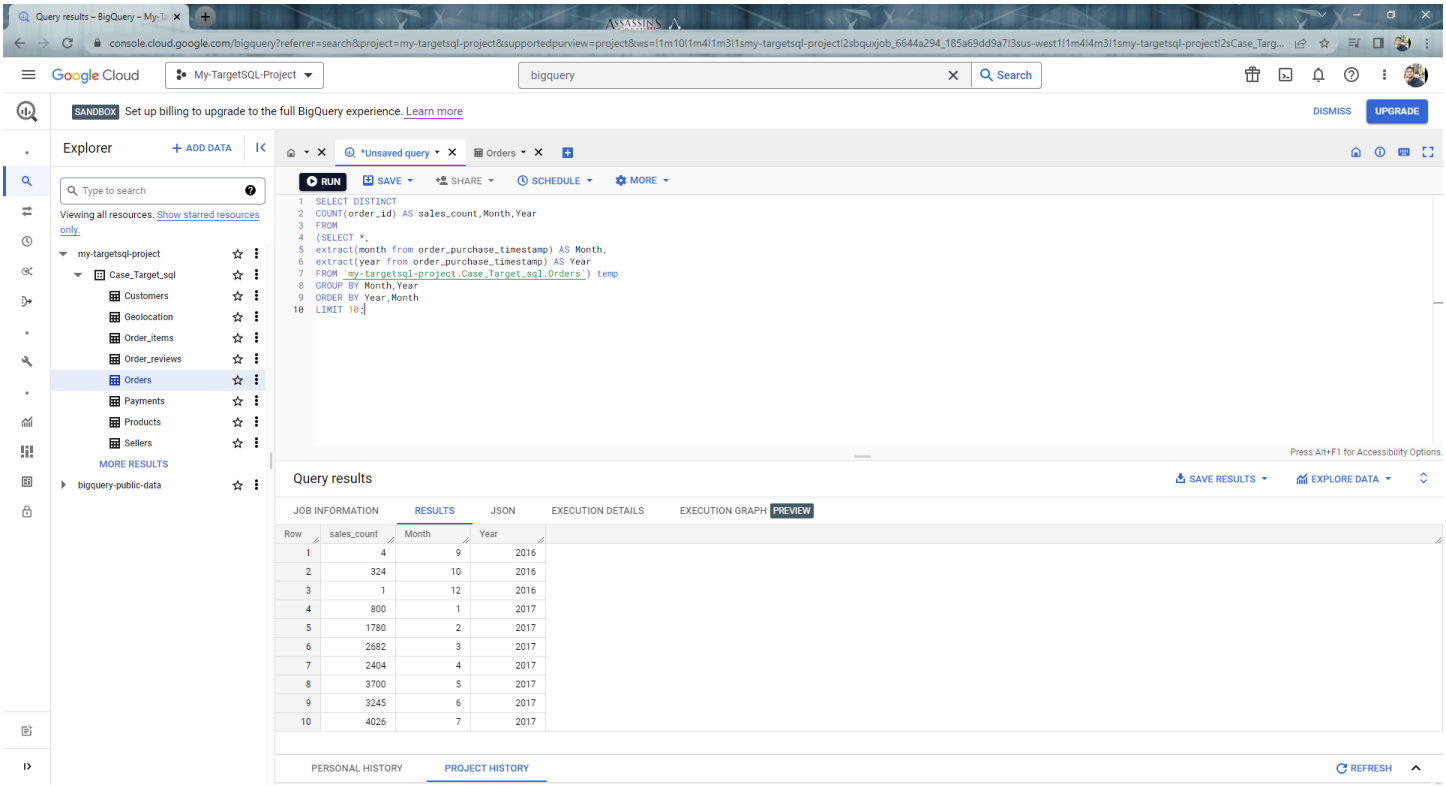
3. Cities and States of customers ordered during the given period

```
SELECT DISTINCT Cus.customer_city,Cus.customer_state
FROM `my-targetsql-project.Case_Target_sql.Orders` AS Ord
INNER JOIN `my-targetsql-project.Case_Target_sql.Customers` AS Cus
ON Ord.customer_id=Cus.customer_id
LIMIT 10;
```



2. In-depth Exploration:
    1. Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?

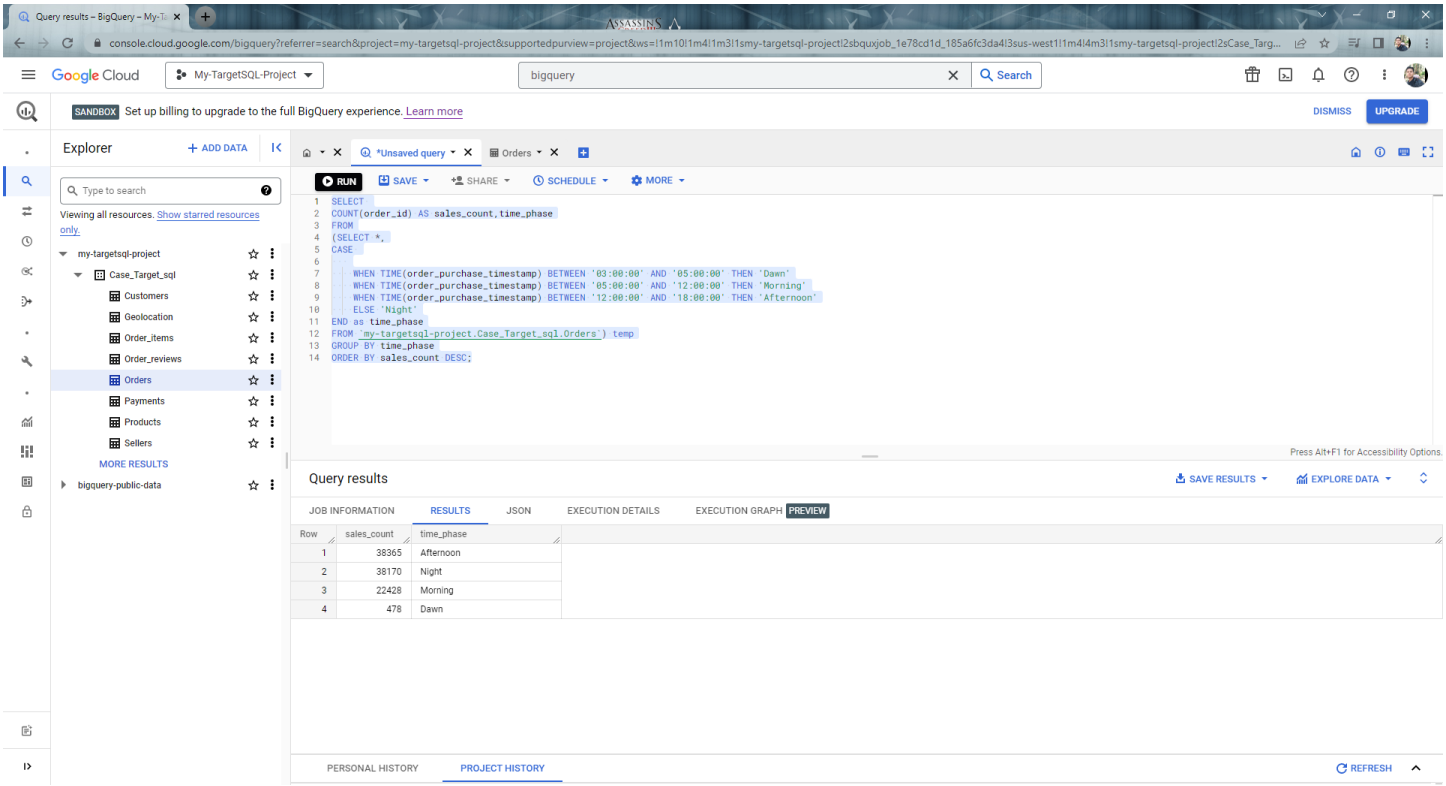
```sql
SELECT DISTINCT
COUNT(order_id) AS sales_count,Month,Year
FROM
(SELECT *,
extract(month from order_purchase_timestamp) AS Month,
extract(year from order_purchase_timestamp) AS Year
FROM `my-targetsql-project.Case_Target_sql.Orders`) temp
GROUP BY Month,Year
ORDER BY Year,Month
LIMIT 10;
```



>>> Initially in **October 2016** there is increase in order purchases/sales which again declines for next 2 months. However, during **year 2017** there was upward trend in order purchases/sales for whole year. Also observed some sudden spike during Month of **November 2017** as its month with **Black Friday sale** and it can be considered seasonal. Also, **1st quarter** of the **year 2018** shows high sales which slowly declines in next quarter and sharply declines further from the end of **3rd quarter**. Overall, we can say there is growing trend on e-commerce in Brazil since 2016 till 2018 as it has grown quite a lot than initial stage.

2. What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?

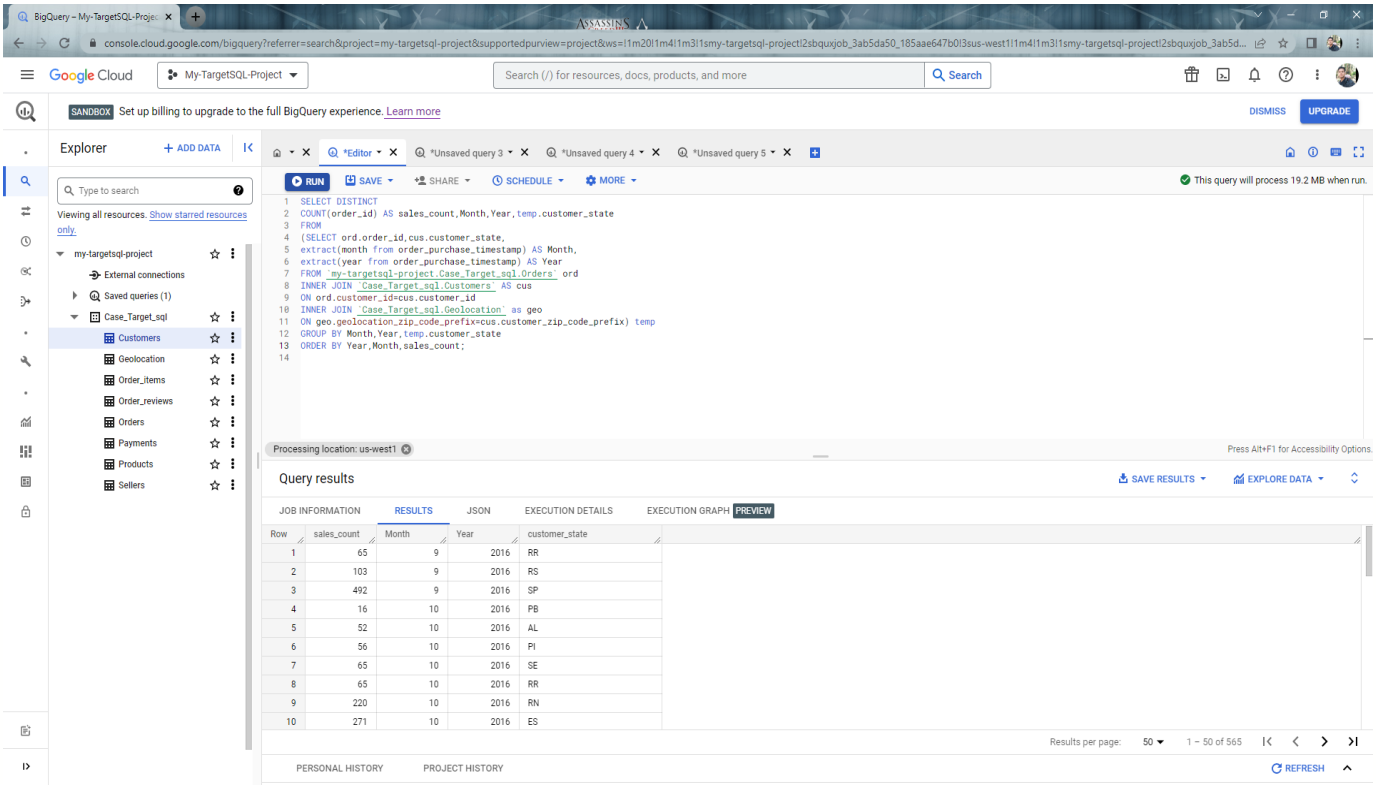
```
SELECT
COUNT(order_id) AS sales_count,time_phase
FROM
(SELECT *,
CASE

    WHEN TIME(order_purchase_timestamp) BETWEEN '03:00:00' AND '05:00:00' THEN 'Dawn'
    WHEN TIME(order_purchase_timestamp) BETWEEN '05:00:00' AND '12:00:00' THEN 'Morning'
    WHEN TIME(order_purchase_timestamp) BETWEEN '12:00:00' AND '18:00:00' THEN 'Afternoon'
    ELSE 'Night'
END as time_phase
FROM `my-targetsql-project.Case_Target_sql.Orders`) temp
GROUP BY time_phase
ORDER BY sales_count DESC;
```



>>>Overall customers are active during daytime and maximum purchases are made during afternoon time following nighttime at second place. Morning time purchases are still made at commendable rate however dawn phase is the least to have activity from customers.

3. Evolution of E-commerce orders in the Brazil region:
    1. Get month on month orders by states

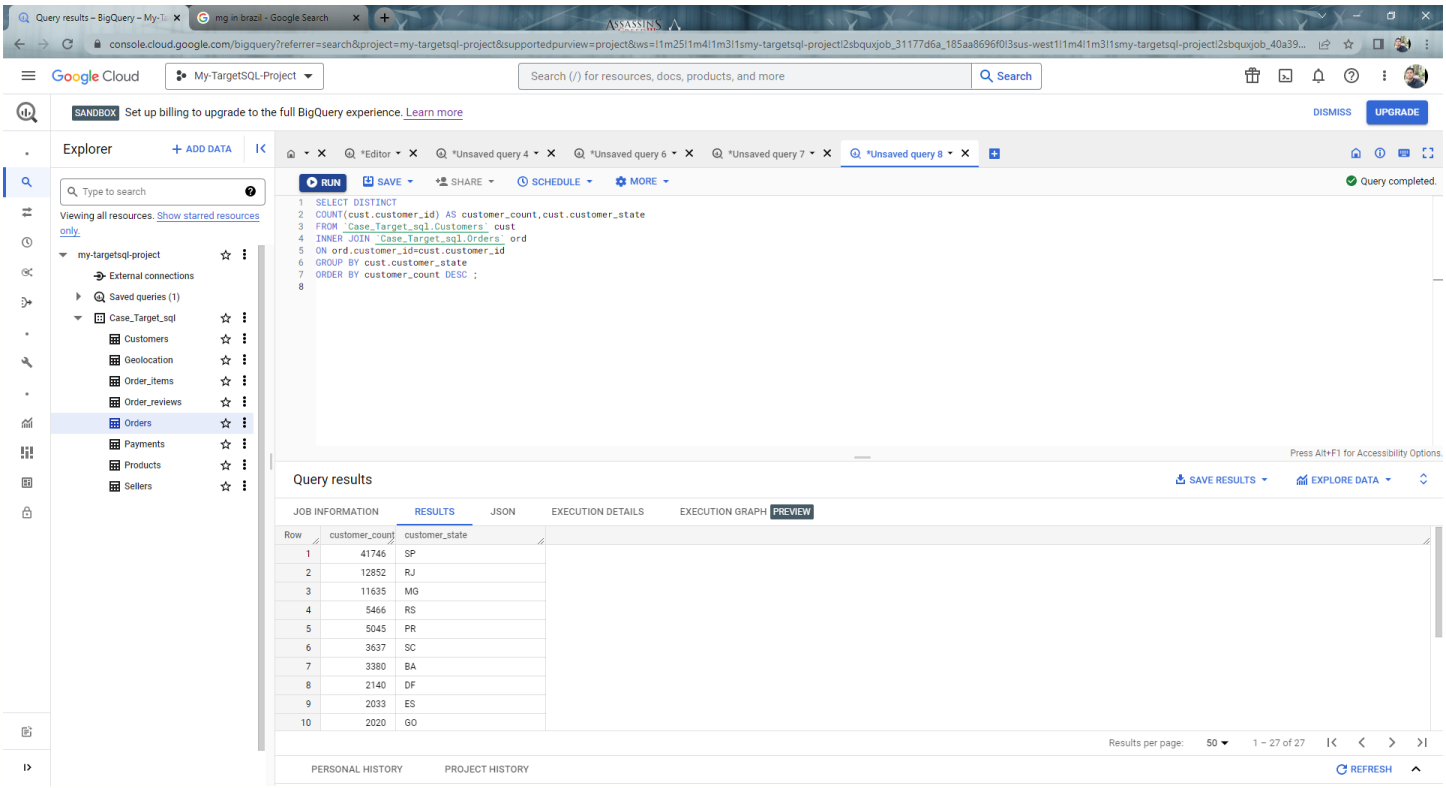
```
SELECT DISTINCT
COUNT(order_id) AS sales_count,Month,Year,temp.customer_state
FROM
(SELECT ord.order_id,cus.customer_state,
extract(month from order_purchase_timestamp) AS Month,
extract(year from order_purchase_timestamp) AS Year
FROM `my-targetsql-project.Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Customers` AS cus
ON ord.customer_id=cus.customer_id
INNER JOIN `Case_Target_sql.Geolocation` as geo
ON geo.geolocation_zip_code_prefix=cus.customer_zip_code_prefix) temp
GROUP BY Month,Year,temp.customer_state
ORDER BY Year,Month,sales_count;
```



>>> Here data shows state wise and month on month sales count is highlighted.

2. Distribution of customers across the states in Brazil

```
SELECT DISTINCT
COUNT(cust.customer_id) AS customer_count,cust.customer_state
FROM `Case_Target_sql.Customers` cust
INNER JOIN `Case_Target_sql.Orders` ord
ON ord.customer_id=cust.customer_id
GROUP BY cust.customer_state
ORDER BY customer_count DESC ;
```

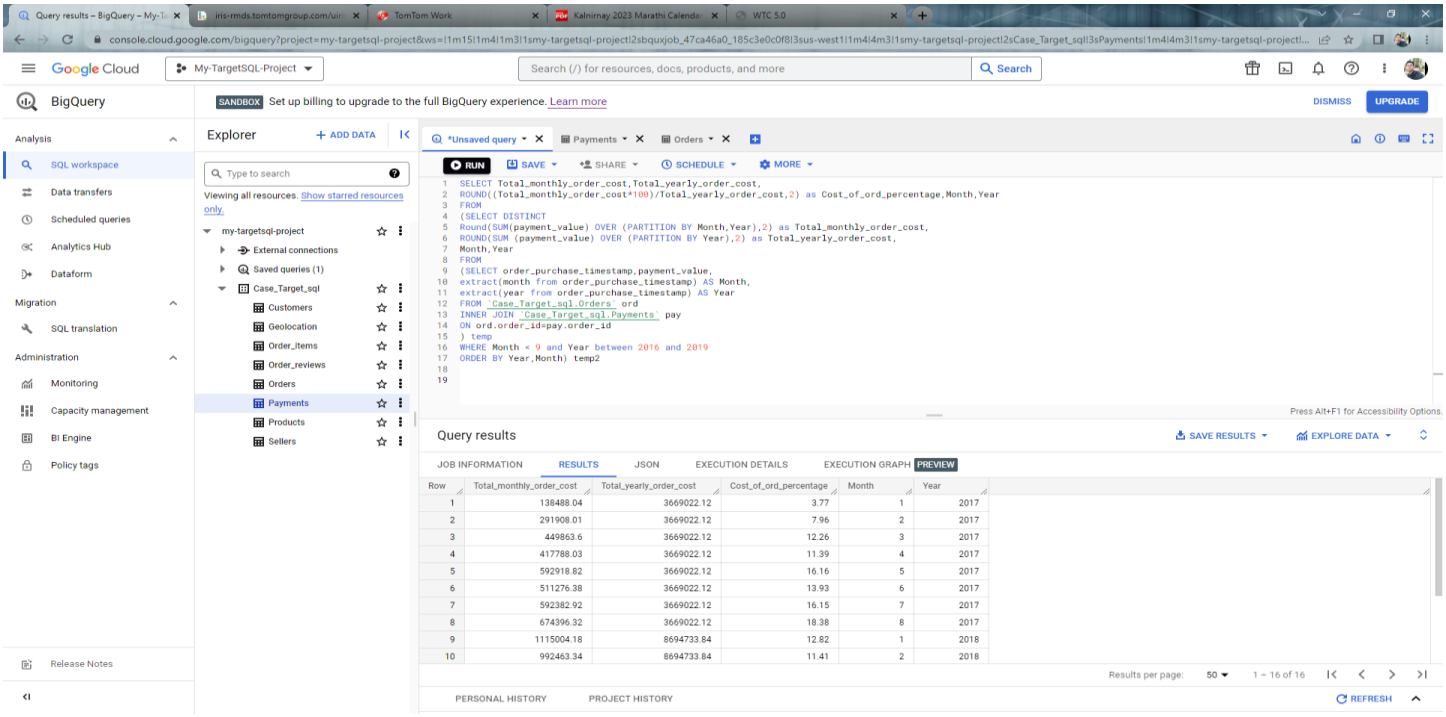


>>> Out of 27 Brazilian states Sao Paulo leads with the number of maximum customers purchases of around 40% of the total orders. Rio de Jannero and Minas Gerais following Sao Paulo where together 3 of these states have customers of around 66% out of total. SP and MG are the two major states with highest number of populations.

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

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

```
SELECT Total_monthly_order_cost,Total_yearly_order_cost,
ROUND((Total_monthly_order_cost*100)/Total_yearly_order_cost,2) as Cost_of_ord_percentage,Month,Year
FROM
(SELECT DISTINCT
Round(SUM(payment_value) OVER (PARTITION BY Month,Year),2) as Total_monthly_order_cost,
ROUND(SUM (payment_value) OVER (PARTITION BY Year),2) as Total_yearly_order_cost,
Month,Year
FROM
(SELECT order_purchase_timestamp,payment_value,
extract(month from order_purchase_timestamp) AS Month,
extract(year from order_purchase_timestamp) AS Year
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Payments` pay
ON ord.order_id=pay.order_id
) temp
WHERE Month < 9 and Year between 2016 and 2019
ORDER BY Year,Month) temp2
```

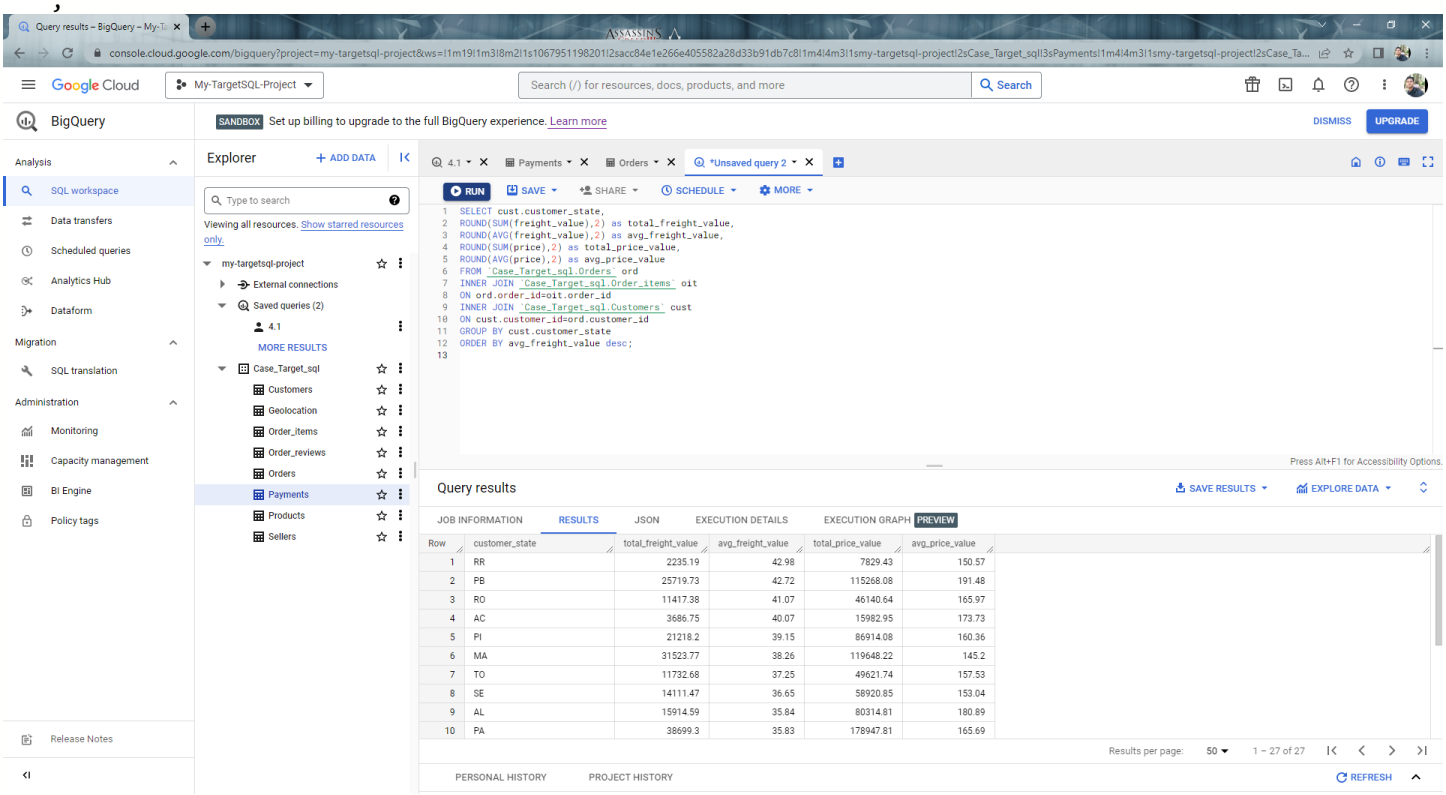


>>> Total cost of orders between month Jan to Aug for years 2017 & 2018. Shows significant increase after first two months where it reached maximum of 18% of total cost during August 2017. For an entire year 2018 it is quite steady trend compared to year 2017 with an average 12.5% of total cost of orders.

2. Mean & Sum of price and freight value by customer state

```
SELECT cust.customer_state,
ROUND(SUM(freight_value),2) as total_freight_value,
```

```
ROUND(AVG(freight_value),2) as avg_freight_value,
ROUND(SUM(price),2) as total_price_value,
ROUND(AVG(price),2) as avg_price_value
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Order_items` oit
ON ord.order_id=oit.order_id
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY avg_freight_value desc;
5.
;
```



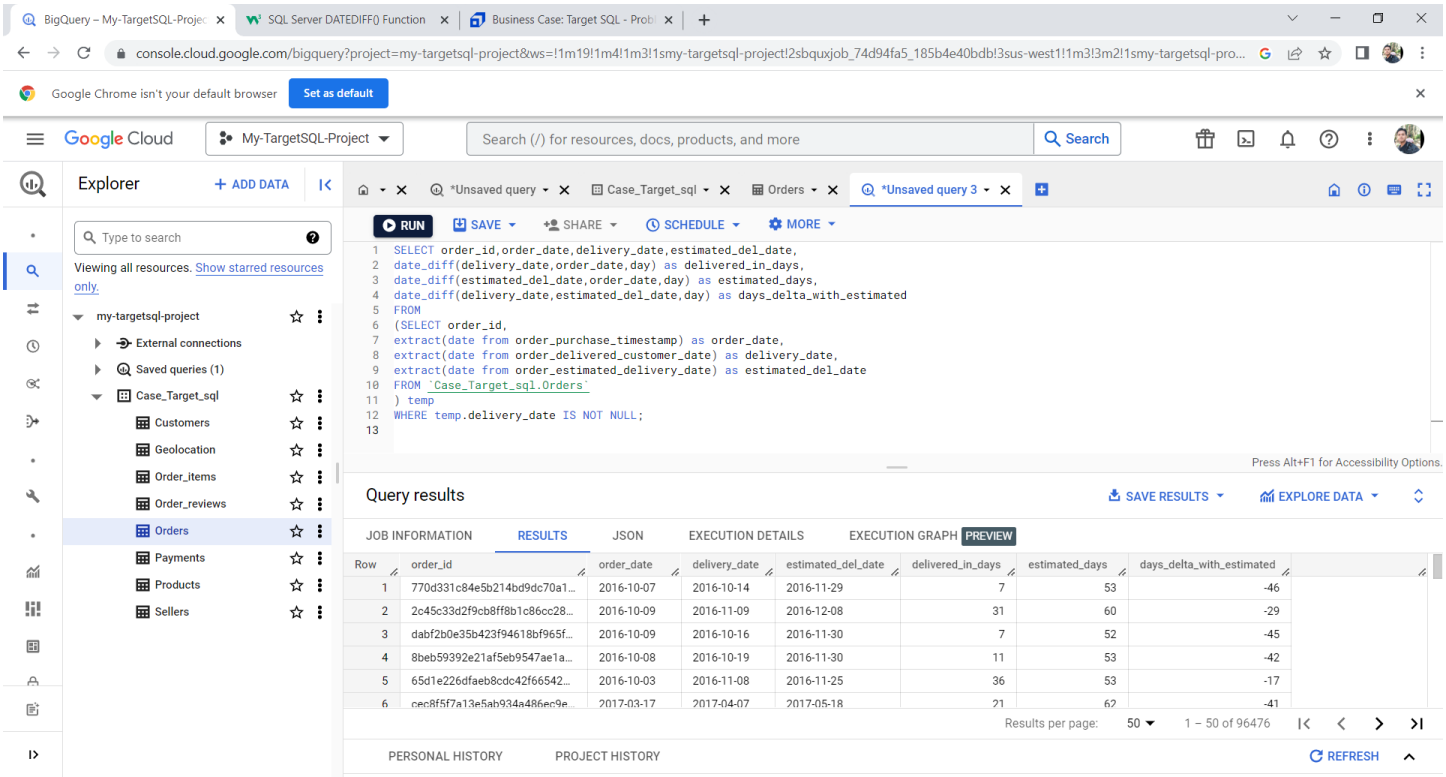
>>> Here just a view for the states where average freight cost is higher than rest of the states. Where the more populated states have lower freight cost. Average Freight value in the states where it is highest like RR, PB, RO & AC is twice more than compared to states with lower freight cost.

5. Analysis on sales, freight and delivery time

1.Calculate days between purchasing, delivering and estimated delivery

```
SELECT order_id,order_date,delivery_date,estimated_del_date,
```

```
date_diff(delivery_date,order_date,day) as delivered_in_days,
date_diff(estimated_del_date,order_date,day) as estimated_days,
date_diff(delivery_date,estimated_del_date,day) as days_delta_with_estimated
FROM
(SELECT order_id,
extract(date from order_purchase_timestamp) as order_date,
extract(date from order_delivered_customer_date) as delivery_date,
extract(date from order_estimated_delivery_date) as estimated_del_date
FROM `Case_Target_sql.Orders`
) temp
WHERE temp.delivery_date IS NOT NULL;
```

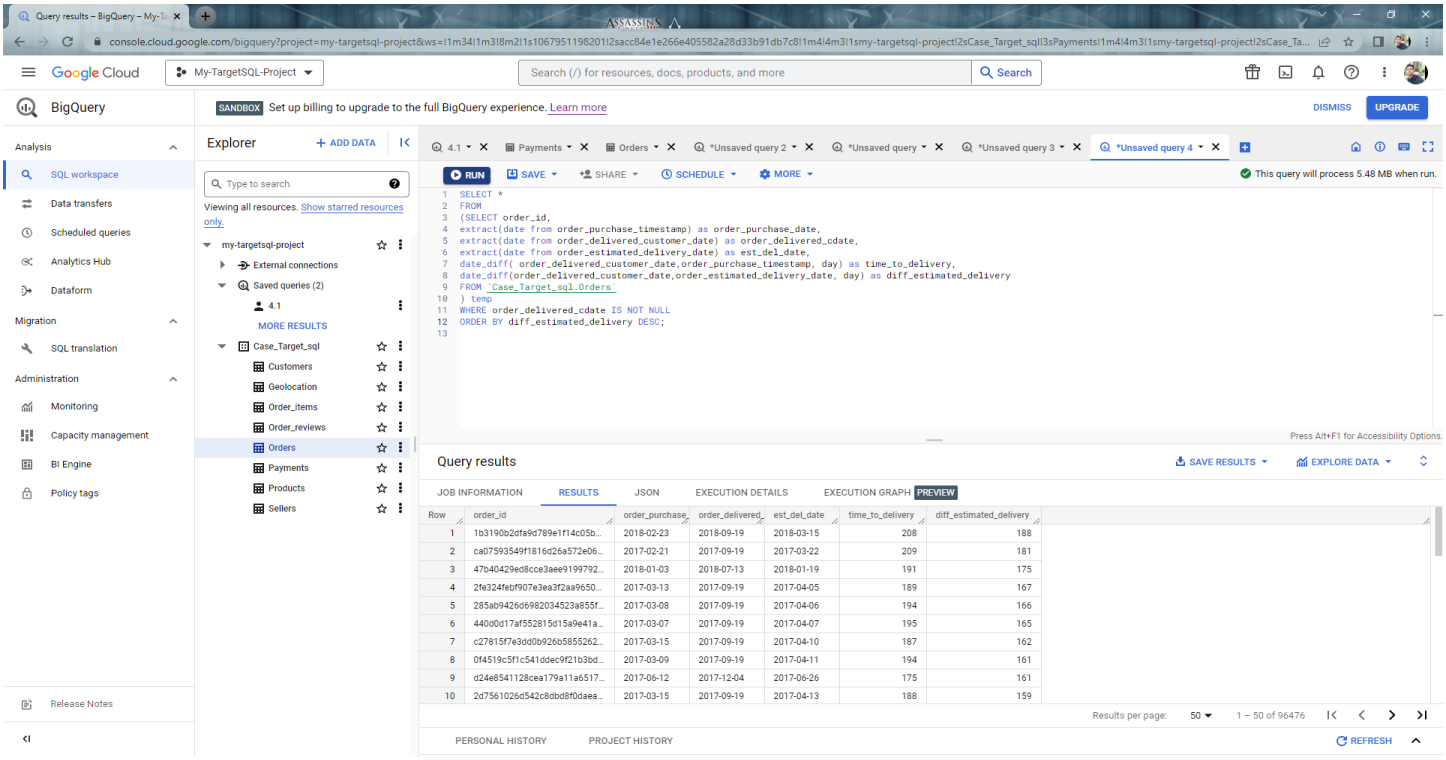


2. Find time_to_delivery & diff_estimated_delivery. Formula for the same given below:

    - time_to_delivery = order_purchase_timestamp-order_delivered_customer_date
    - diff_estimated_delivery = order_estimated_delivery_date-order_delivered_customer_date

```
SELECT *
```

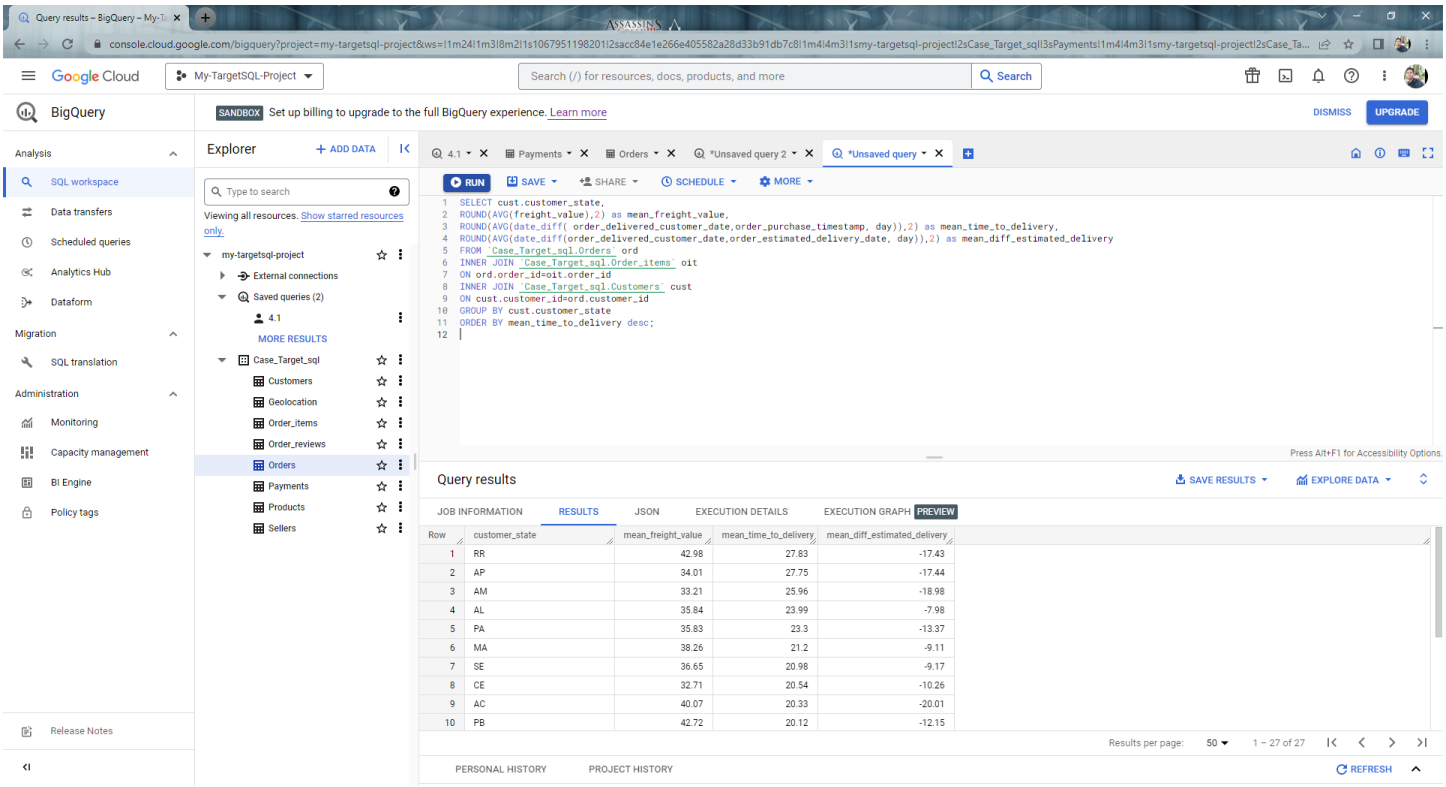
```
FROM
(SELECT order_id,
extract(date from order_purchase_timestamp) as order_purchase_date,
extract(date from order_delivered_customer_date) as order_delivered_cdate,
extract(date from order_estimated_delivery_date) as est_del_date,
date_diff( order_delivered_customer_date,order_purchase_timestamp, day) as time_to_delivery,
date_diff(order_delivered_customer_date,order_estimated_delivery_date, day) as diff_estimated_delivery
FROM `Case_Target_sql.Orders`
) temp
WHERE order_delivered_cdate IS NOT NULL
ORDER BY diff_estimated_delivery DESC;
```



>>>Here required date with just added with some order to showcase what is the maximum `diff_estimated_delivery`.

3.Group data by state, take mean of freight_value, time_to_delivery, diff_estimated_delivery

```sql
SELECT cust.customer_state,
ROUND(AVG(freight_value),2) as mean_freight_value,
ROUND(AVG(date_diff( order_delivered_customer_date,order_purchase_timestamp, day)),2) as mean_time_to_deliver
y,
ROUND(AVG(date_diff(order_delivered_customer_date,order_estimated_delivery_date, day)),2) as mean_diff_estima
ted_delivery
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Order_items` oit
ON ord.order_id=oit.order_id
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY mean_time_to_delivery desc;
```



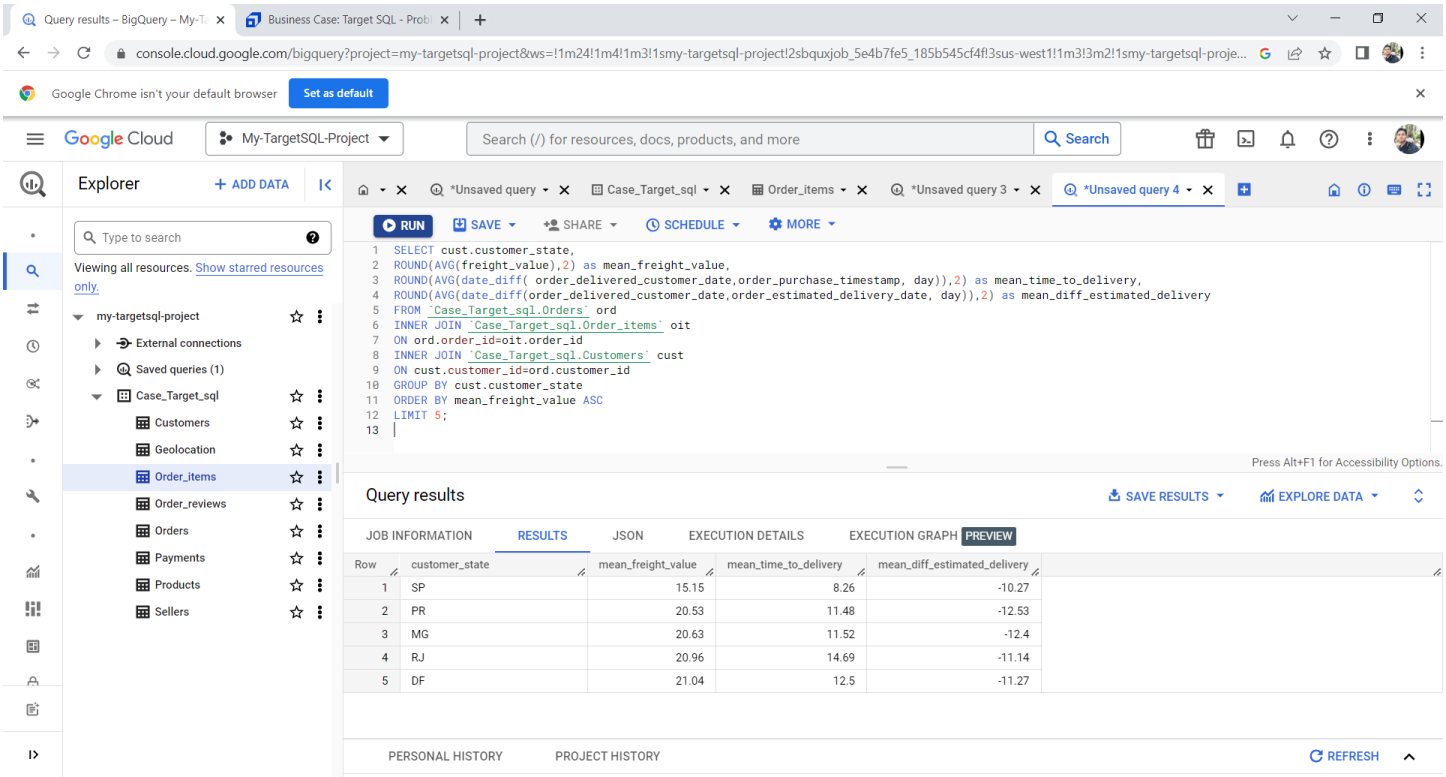
>>> Average time for the delivery around state like RR, AP, AM looks higher, also average fright value looks higher in these states.

4.Sort the data to get the following:

5.Top 5 states with lowest/highest average freight value - sort in desc/asc limit 5

```sql
SELECT cust.customer_state,
ROUND(AVG(freight_value),2) as mean_freight_value,
```

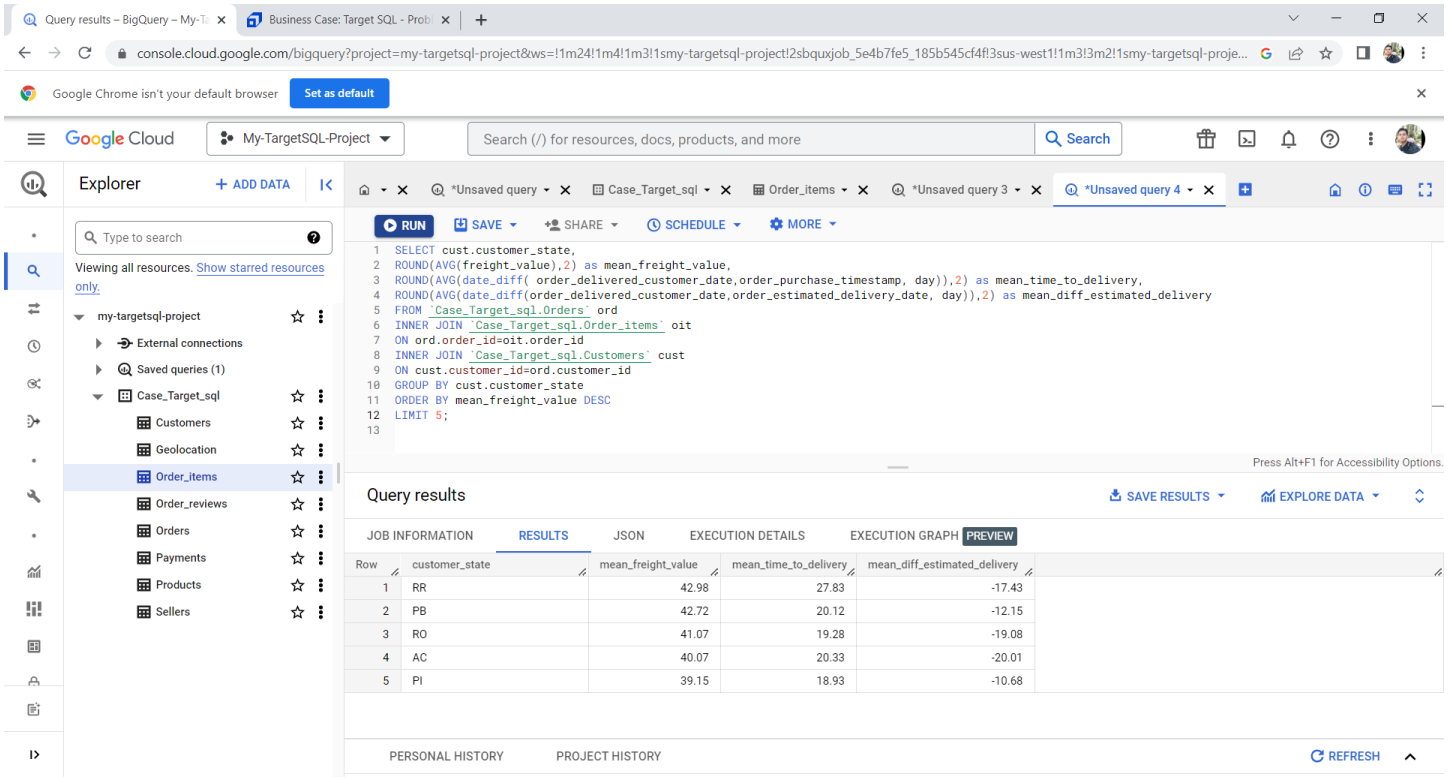
```
ROUND(AVG(date_diff( order_delivered_customer_date,order_purchase_timestamp, day)),2) as mean_time_to_deliver
y,
ROUND(AVG(date_diff(order_delivered_customer_date,order_estimated_delivery_date, day)),2) as mean_diff_estima
ted_delivery
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Order_items` oit
ON ord.order_id=oit.order_id
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY mean_freight_value ASC
LIMIT 5;
```



>>>The states SP, PR, MG & RJ with less average freight value are one of the most populated states where most customers are located in.

```
SELECT cust.customer_state,
ROUND(AVG(freight_value),2) as mean_freight_value,
ROUND(AVG(date_diff( order_delivered_customer_date,order_purchase_timestamp, day)),2) as mean_time_to_deliver
y,
ROUND(AVG(date_diff(order_delivered_customer_date,order_estimated_delivery_date, day)),2) as mean_diff_estima
ted_delivery
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Order_items` oit
```

```
ON ord.order_id=oit.order_id
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY mean_freight_value DESC
LIMIT 5;
```

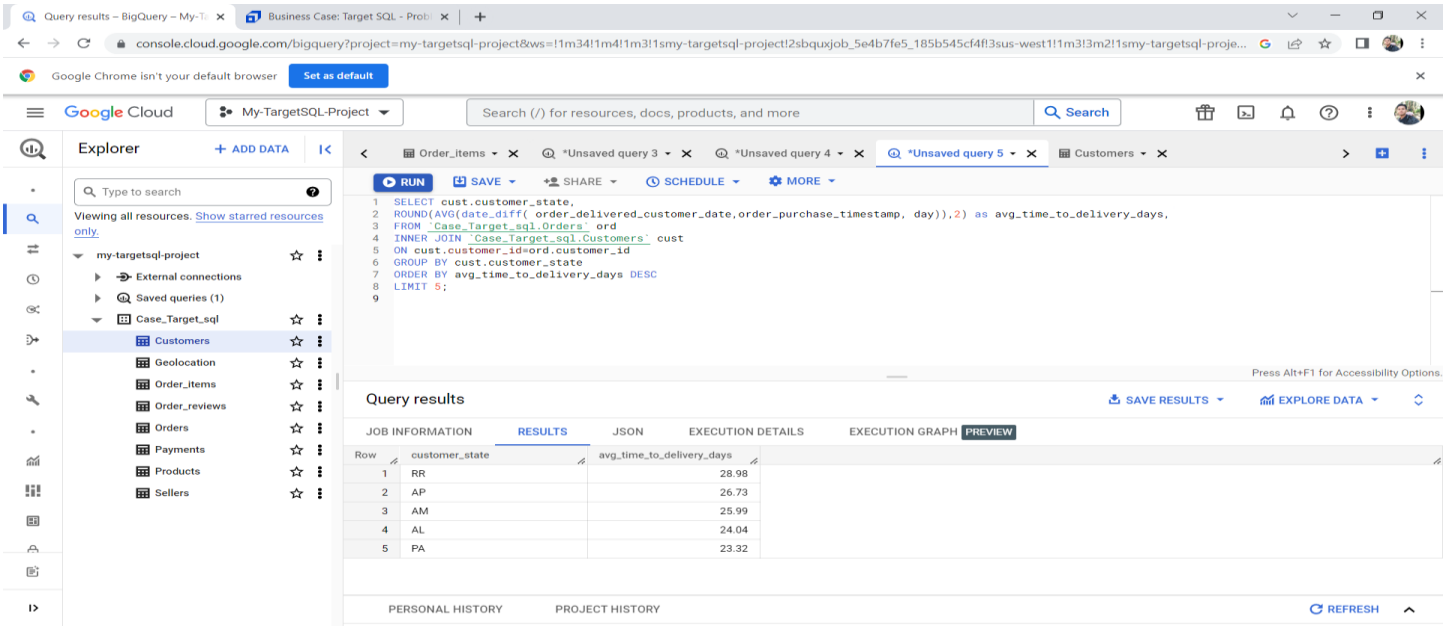


>>>The state like RR,PB,RO & AC with most average freight cost is having twice more than the states with lowest freight cost. Similar strategies like states with low cost, shall be planned to reduce the freight cost.

6.Top 5 states with highest/lowest average time to delivery

```
SELECT cust.customer_state,
ROUND(AVG(date_diff( order_delivered_customer_date,order_purchase_timestamp, day)),2) as avg_time_to_delivery
_days,
FROM `Case_Target_sql.Orders` ord
```

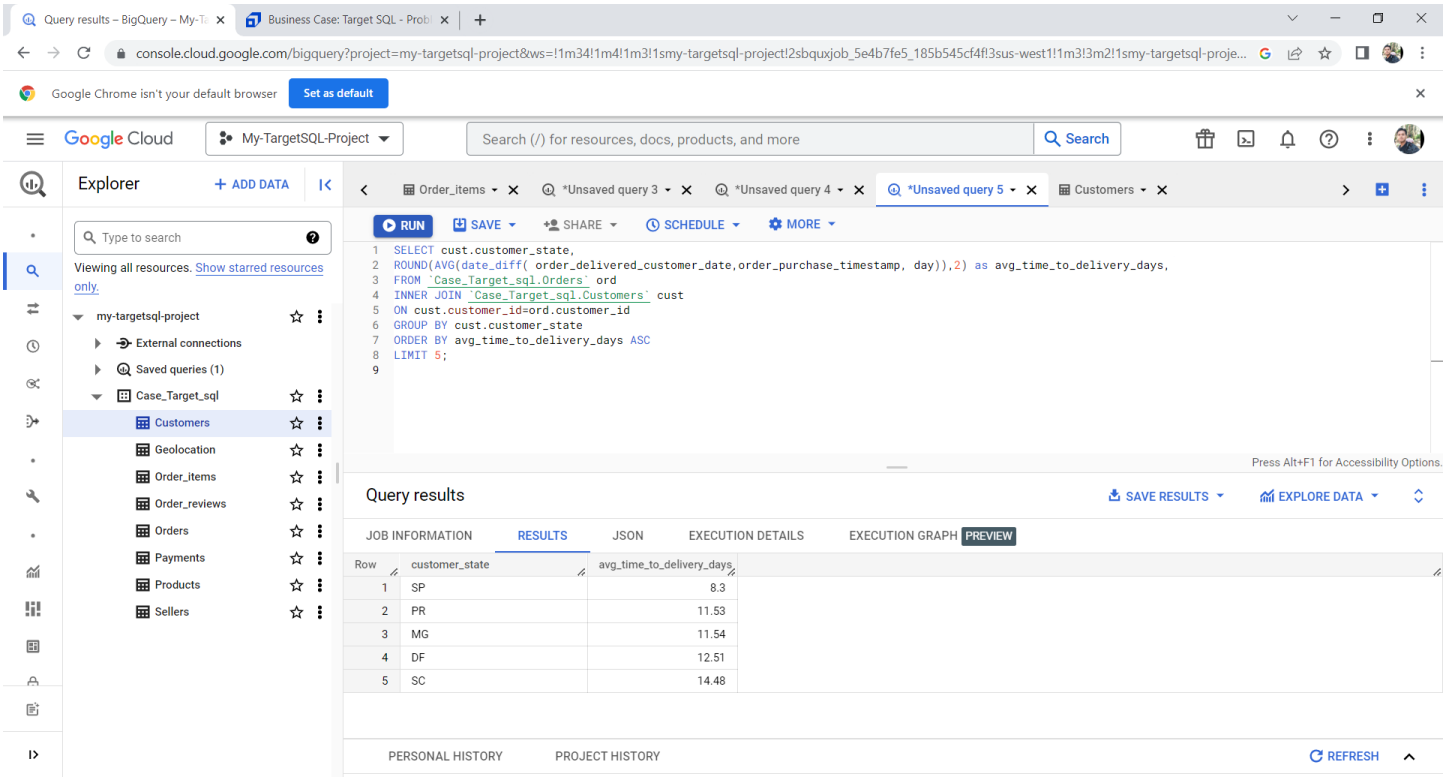
```
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY avg_time_to_delivery_days DESC
LIMIT 5;
```



>>>States RR, AP, AM, AL & PA are the topmost states where average delivery time is highest up to 29 days.

```
SELECT cust.customer_state,
ROUND(AVG(date_diff( order_delivered_customer_date,order_purchase_timestamp, day)),2) as avg_time_to_delivery
_days,
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
```

```
GROUP BY cust.customer_state
ORDER BY avg_time_to_delivery_days ASC
LIMIT 5;
```

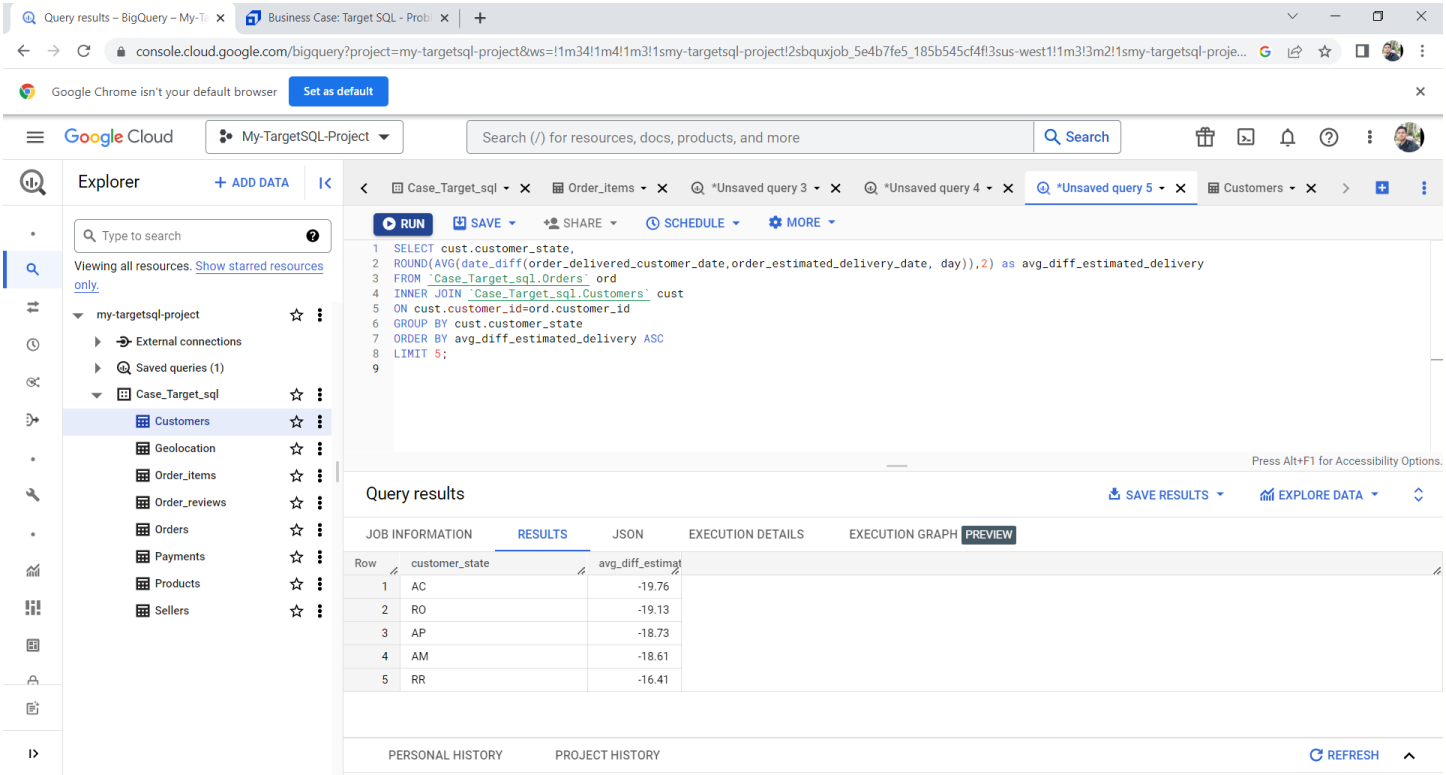


>>>These states are those where the avg delivery time is less than compared to rest of the states. State Sao Paulo with lowest average delivery time of 1 week only.

7.Top 5 states where delivery is really fast/ not so fast compared to estimated date

```
SELECT cust.customer_state,
ROUND(AVG(date_diff(order_delivered_customer_date,order_estimated_delivery_date, day)),2) as avg_diff_estimated_delivery
FROM `Case_Target_sql.Orders` ord
```

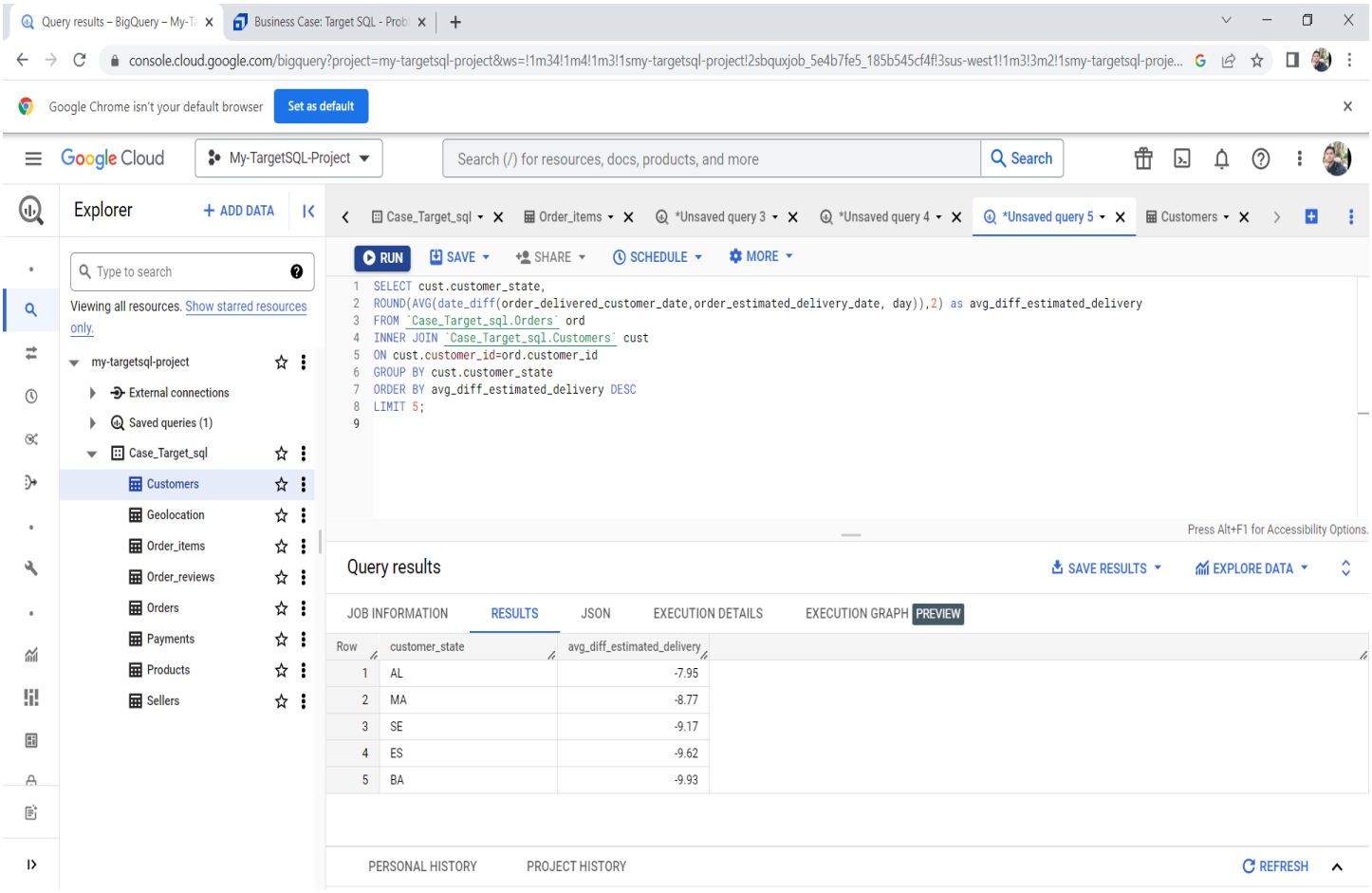
```
INNER JOIN `Case_Target_sql.Customers` cust
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY avg_diff_estimated_delivery ASC
LIMIT 5;
```



>>> Here these states with average diff estimated delivery shows us that how faster the order was delivered compared to the estimated day of deliver. Here we can see some states like AC & RO have where order was delivered earlier with average value of 19 days.

```
SELECT cust.customer_state,
ROUND(AVG(date_diff(order_delivered_customer_date,order_estimated_delivery_date, day)),2) as avg_diff_estimat
ed_delivery
FROM `Case_Target_sql.Orders` ord
INNER JOIN `Case_Target_sql.Customers` cust
```

```
ON cust.customer_id=ord.customer_id
GROUP BY cust.customer_state
ORDER BY avg_diff_estimated_delivery DESC
LIMIT 5;
```



>>>Here states like AL, MA & other 3 where we can say order is delivered just 1 week before.

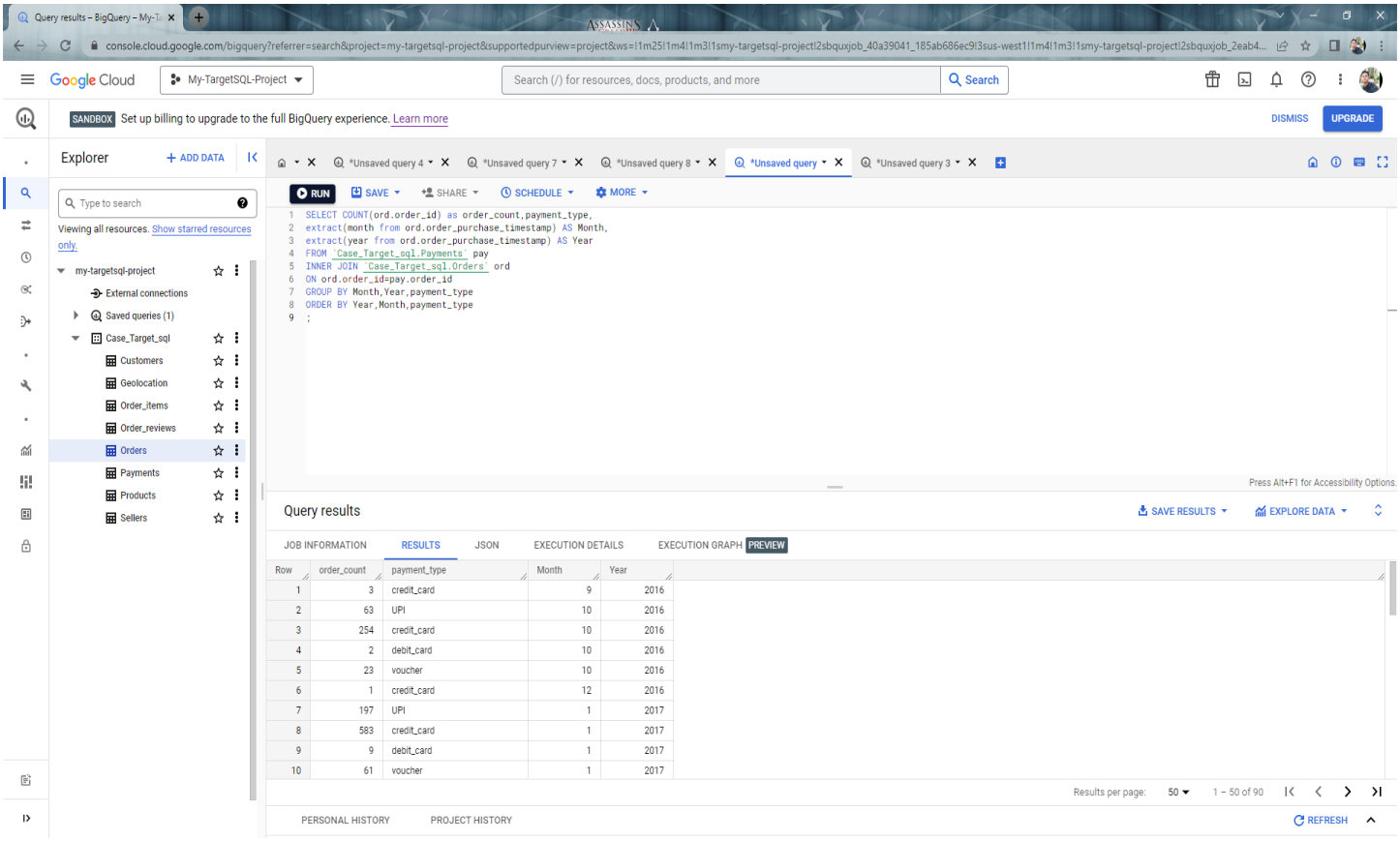
Overall delivery estimated difference is always > 7days in every states.

6. Payment type analysis:

1. Month over Month count of orders for different payment types

```
SELECT COUNT(ord.order_id) as order_count,payment_type,
extract(month from ord.order_purchase_timestamp) AS Month,
extract(year from ord.order_purchase_timestamp) AS Year
FROM `Case_Target_sql.Payments` pay
INNER JOIN `Case_Target_sql.Orders` ord
```

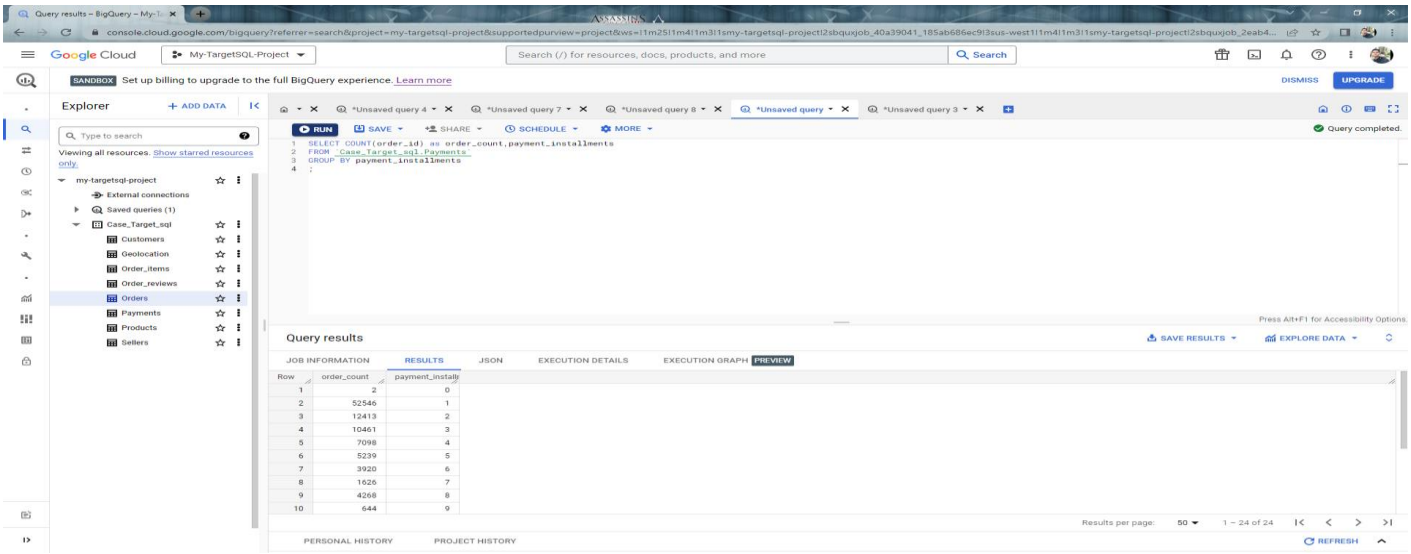
```
ON ord.order_id=pay.order_id
GROUP BY Month,Year,payment_type
ORDER BY Year,Month,payment_type
;
```



>>> Payment type Credit Card is by far the most used by customer to purchase an order.

2. Count of orders based on the no. of payment installments

```
SELECT COUNT(order_id) as order_count,payment_installments
FROM `Case_Target_sql.Payments`
GROUP BY payment_installments;
```



>>> Maximum customers opt for the payment installment number between 2 to 4 of about nearly 75% of the total payment types. While only 2 orders with full payment are observed.

Actionable Insights:	Recommendations:
Initially in October 2016 there is increase in order purchases/sales which again declines for next 2 months. However, during year 2017 there was upward trend in order purchases/sales for whole year. Also observed some sudden spike during Month of November 2017 as its month with Black Friday sale and to be considered	As we can see there is increase in the demand during the specific seasonal sale like black Friday, strategies should be defined accordingly. More count of products shall be made available. Staff shall be increased during such peak season. Inventory shall be well maintained during season. Server maintenance shall be done prior to seasonal peak.

seasonal. Also, 1st quarter of the year 2018 shows high sales which slowly declines in next quarter and sharply declines further from the end of 3rd quarter. Overall, we can say there is growing trend on e-commerce in Brazil since 2016 till 2018 as it has grown quite a lot than initial stage.	Quality check or audits shall be conducted (internal or external) before season to ensure best products and services given to customer. Overall looking at the growing trend considering percentage growth in previous years; accordingly, varied product category, services and support can be increased.
By looking at time of purchases we can say customers are active during daytime and maximum purchases are made during afternoon time following nighttime at second place. Morning time purchases are still made at commendable rate however dawn phase is the least to have activity from customers.	More support shall be provided during peak hours of purchases. Customer support shall be planned accordingly with more resources during such hours. Minor maintenance activities can be performed/downtime window when less activity of customer. E.g., Dawn phase. Staff can also get some breathing period during this time.
Out of 27 Brazilian states Sao Paulo leads with the number of maximum customers purchases of around 40% of the total orders. Rio de Jannero and Minas Gerais following Sao Paulo where together 3 of these states have customers of around 66% out of total. SP and MG are the two major states with highest number of populations.	As most populated states have a greater number of customers. Product category, inventory, customer support and quality service shall be on priority in these states. Whereas in states, where there are less customers placing orders can be focused with more marketing of the products and brand to attract new customer.
Total cost of orders between month Jan to Aug for years 2017 & 2018. Shows significant increase after first two months where it reached maximum of 18% of total cost during August 2017. For an entire year 2018 it is quite steady trend compared to year 2017 with an average 12.5% of total cost of orders.	As it is observed that there is usual growing trend/pattern during some specific months, business strategy can be aligned in way to increase the profit margin during peak time. It is also advised to encourage customer during other time through various marketing strategies when it is not season.
There are some states where average freight cost is higher than rest of the states. Where the more populated states have lower freight cost. Average Freight value in the states where it is highest like RR, PB, RO & AC is twice more than compared to states with lower freight cost.	Freight cost can be reduced in such states where it is very high compared to more popular states. Here delivery time also is major factor where along with freight cost delivery time is also high. In such states where courier companies with slow services can be dropped off and alternative can be found on to improve services or else some cost negotiation shall be done with current courier services.
Average time for the delivery around state like RR, AP, AM looks higher, also average fright value looks higher in these states.	For the delivery time in these states there can be several affecting factors which might need to be found of with further in-depth analysis and counter measures can be defined accordingly.
The states SP, PR, MG & RJ with less average freight value are one of the most populated states where most customers are located in. The state like RR, PB, RO & AC with most average freight cost is having twice more than the states with lowest freight cost.	Strategies in such states where freight cost is low can also be studied to understand how and why it is less compared to the higher ones. Similar strategies can be applied in the states where it is high to reduce fright cost.
These states are those where the average delivery time is less than compared to rest of the states. State Sao Paulo with lowest average delivery time of 1 week only.	Factors like traffic scenarios, roads, locality, number of stores shipping products & their availability can be analyzed to check how it is affecting the deliveries in these states for improvement of other states and its strategies.
States RR, AP, AM, AL & PA are the topmost states where average delivery time is highest up to 29 days.	These states can be kept as an improvement objective as in these states average delivery time is very high, and we may risk losing customer due to delay in the delivery or even long time taken for delivery.

Here these states with average diff estimated delivery shows us that how faster the order was delivered compared to the estimated day of deliver. Here we can see some states like AC & RO have where order was delivered earlier with average value of 19 days	Average diff estimated delivery can also be one of the KPI factor to be added for next year. If delivery is always faster in such states, then estimated days shall also not vary much and can be kept likewise and with less difference. It can be tested and slowly reduced to 1 week.
Here states like AL, MA & other 3 where we can say order is delivered just 1 week before. Overall delivery estimated difference is always > 7days in every states.	Here, it seems not much difference between actual delivery date and estimated date. However, it can still be kept low priority objective to reduce it to half week amount.
For Payment type maximum customers opt for the payment installment number between 2 to 4 of about nearly 75% of the total payment types. While only 2 orders with full payment are observed. Also, payment type Credit Card is by far the most used by customer to purchase an order.	For payment type, where most cost is involved then there can be more profit margin. It could be one of the potential areas where profit can be increased from by applying slightly higher interest or less discount in that range may work. Whereas to have instant cashflow we can encourage customers to go for full payment with some minor discounts.