**Project Title : Target Business Case Analysis**

**Project Description :** Target corporation is an American retailer corporation headquartered in Minneapolis, Minnesota. It is Known for its outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular project focuses on analysis and interpretation of Target retail operations in Brazil and provides insightful information about 1,00,000 orders placed between 2016 and 2018. The project analysis can be done by using a comprehensive dataset of Target, which consists of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By using dataset, it allows us to analyse valuable insights of Target operations in Brazil. It provides various business aspects such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

**Question I :**

**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.**

**Query :**

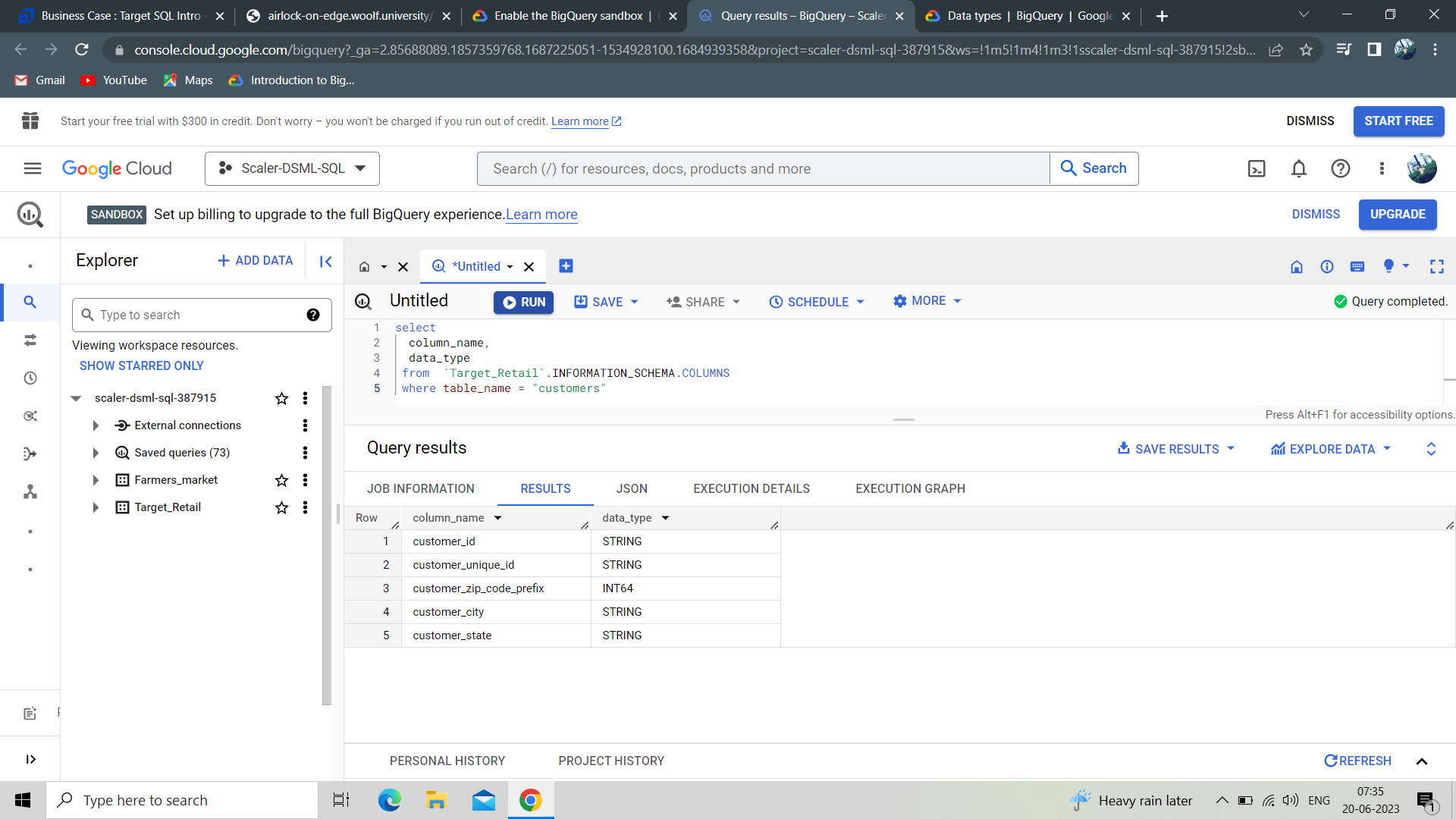
**select**

**column\_name,**

**data\_type**

**from `Target\_Retail`.INFORMATION\_SCHEMA.COLUMNS**

**where table\_name = "customers"**

**Query results :** 

**Insights :**

In this question we can see each column in the Customers table along with the Datatype. From the Query results we get to know various column data types such as customer\_id, customer\_unique\_id, customer\_city, customer\_state as ‘STRING’ data type and customer\_zip\_code\_prefix as ‘INTEGER’ datatype.

1. **Get the time range between which the orders were placed.**

**Query:**

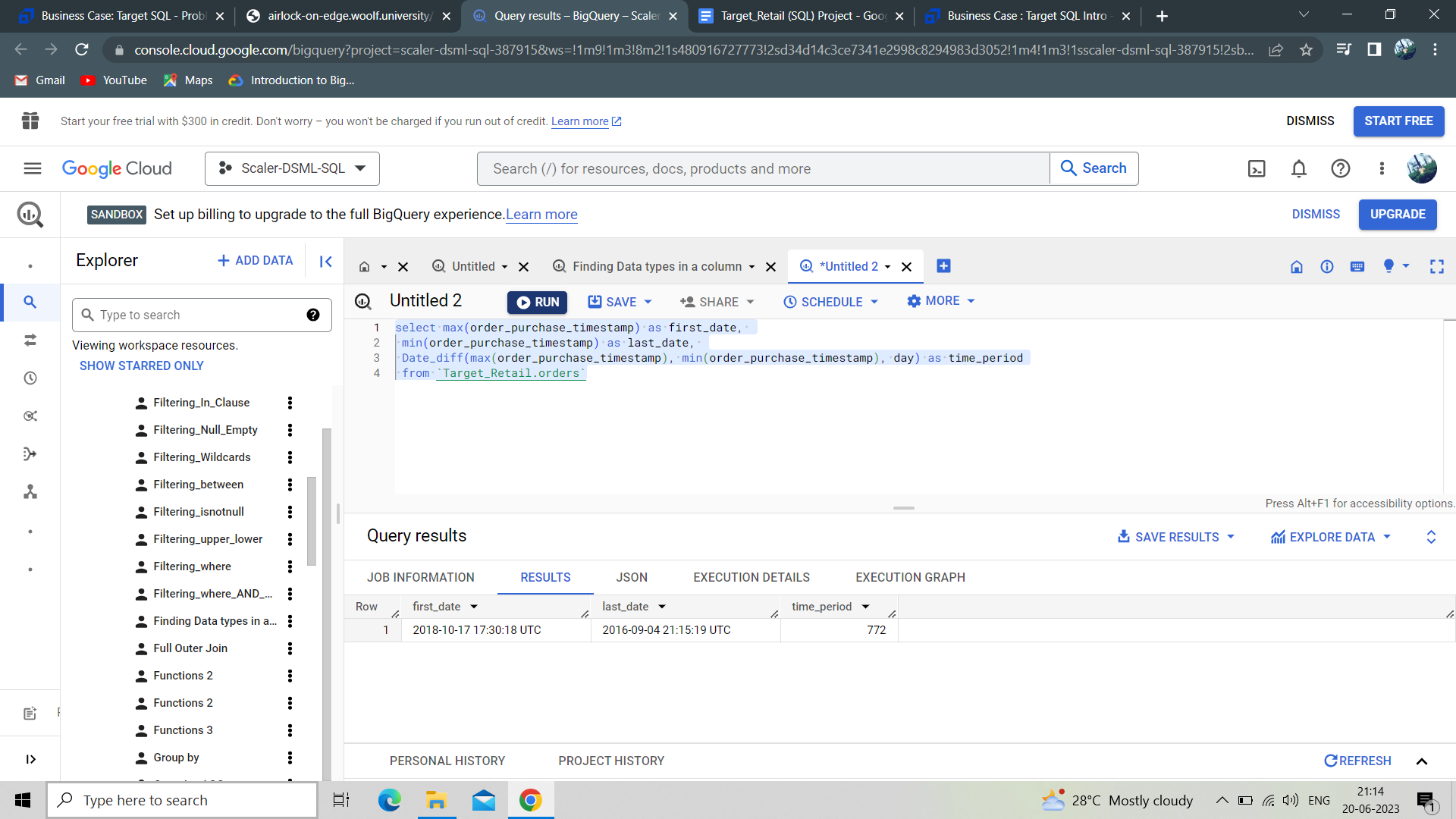
**select max(order\_purchase\_timestamp) as first\_date,**

**min(order\_purchase\_timestamp) as last\_date,**

**Date\_diff(max(order\_purchase\_timestamp), min(order\_purchase\_timestamp), day) as time\_period**

**from `Target\_Retail.orders`**

**Query results :**



**Insights :**

The results gave the details of date of order which is first placed, date of order which is last placed and the time period between the first order and the last order. We see the date of order first placed on “2018-10-17 17:30:18 UTC”, date of order last placed on “2016-09-04 21:15:19 UTC” and time period between first date and last date as 772 days.

1. **Count the number of Cities and States in our dataset.**

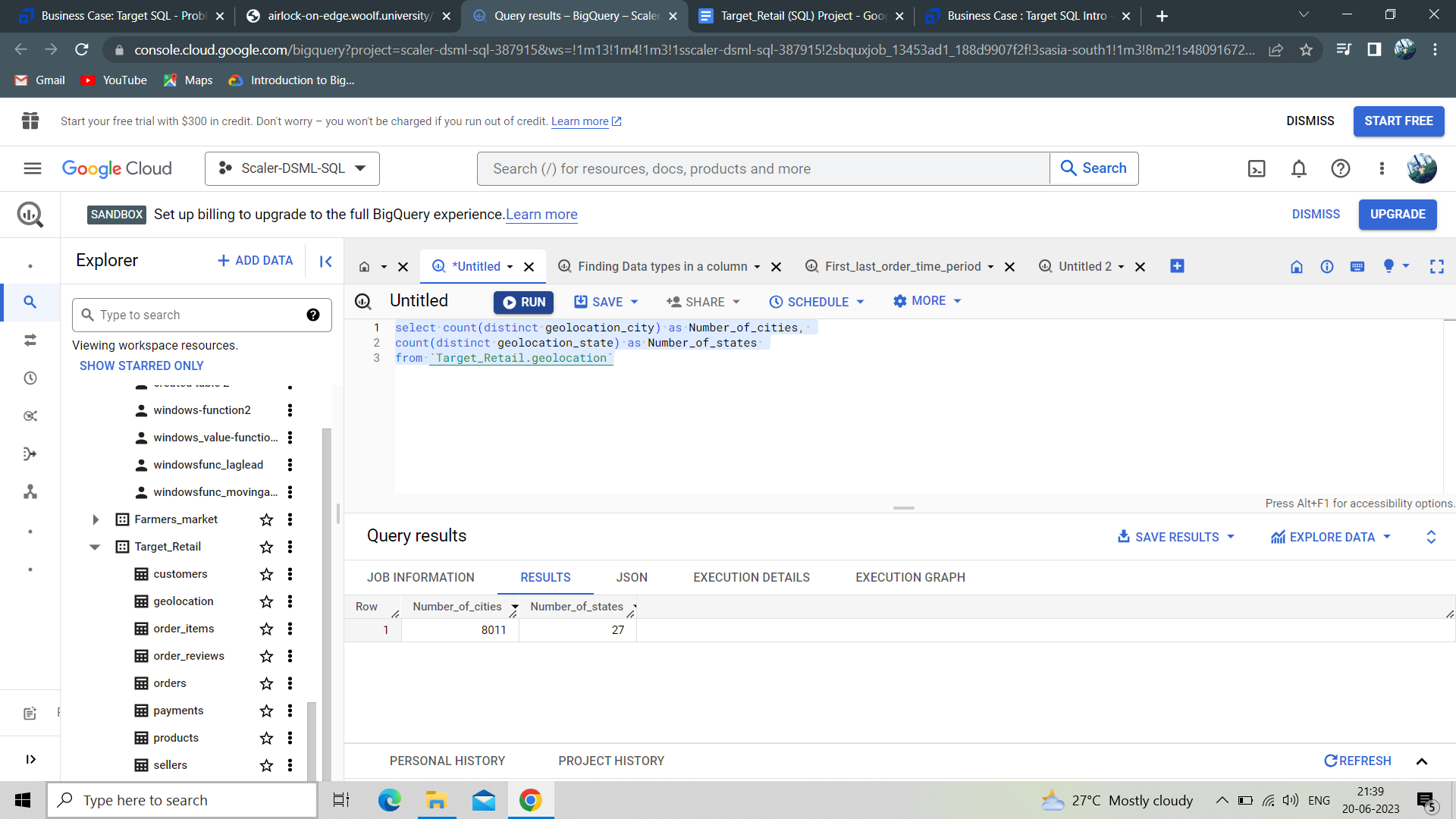
**Query :**

select count(distinct geolocation\_city) as Number\_of\_cities,

count(distinct geolocation\_state) as Number\_of\_states

from `Target\_Retail.geolocation`

**Query results :**



**Insights :**

In this question we can see the number of cities and the number of states of Brazil. The results show 8011 Cities and 27 States from Brazil, where Target carried out its retail business operations.

**Question II :**

**In-depth Exploration:**

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

**Query :**

**select Year, count(order\_id) as number\_of\_orders from**

**(**

**select order\_id,**

**extract(year from order\_purchase\_timestamp) as Year**

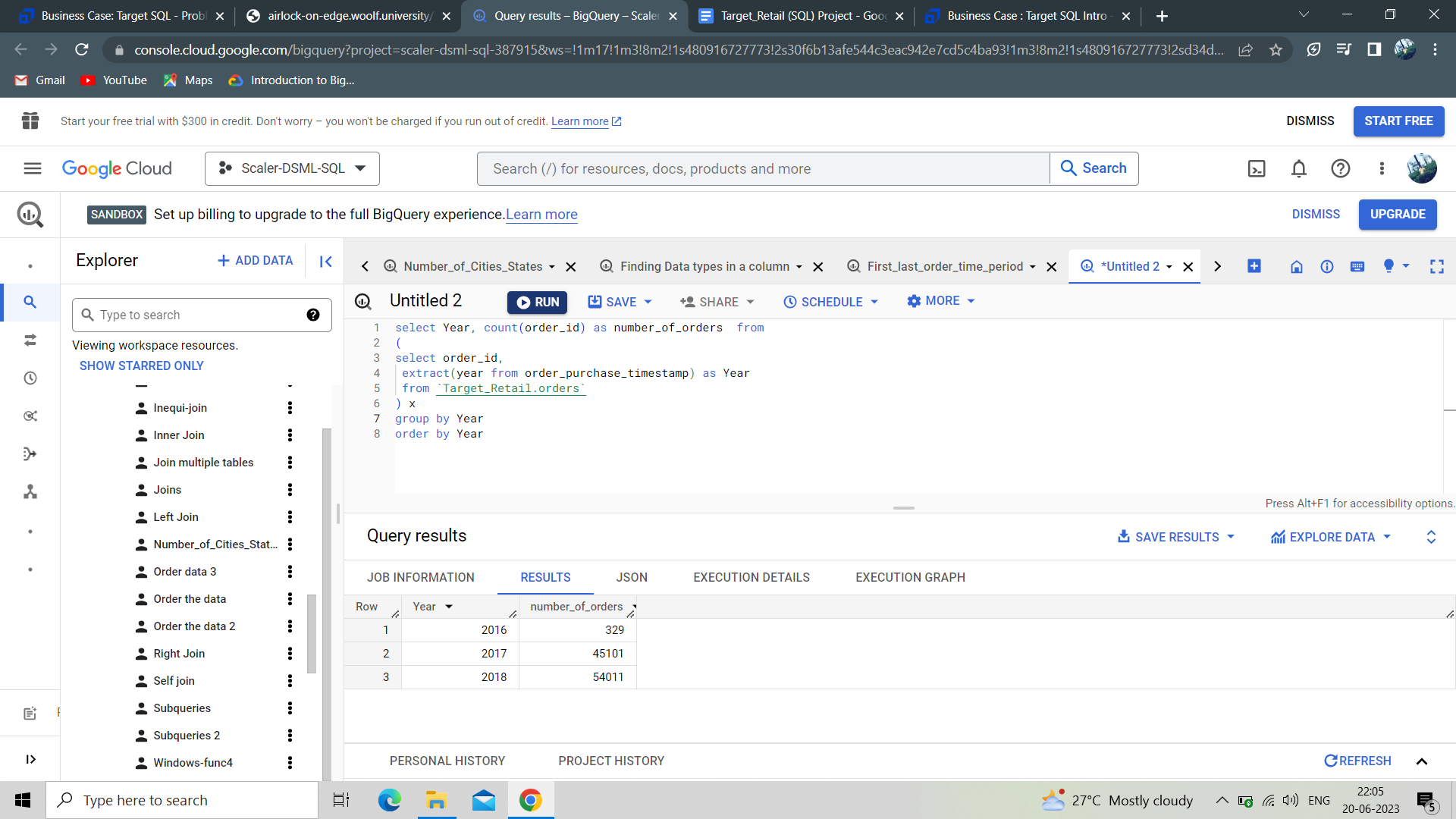
**from `Target\_Retail.orders`**

**) x**

**group by Year**

**order by Year**

**Query results:**



**Insights :**

Here we see In 2016, 329 orders were placed, In 2017 45,101 orders were placed and In 2018 54,011 orders were placed. It clearly shows that the number of orders placed increased over the year from 2016 to 2018. The number of orders placed increased over the years though better marketing and customer servicing.

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

**Query :**

**select**

**extract(month from order\_purchase\_timestamp) as month,**

**format\_datetime("%B", order\_purchase\_timestamp) as month\_name,**

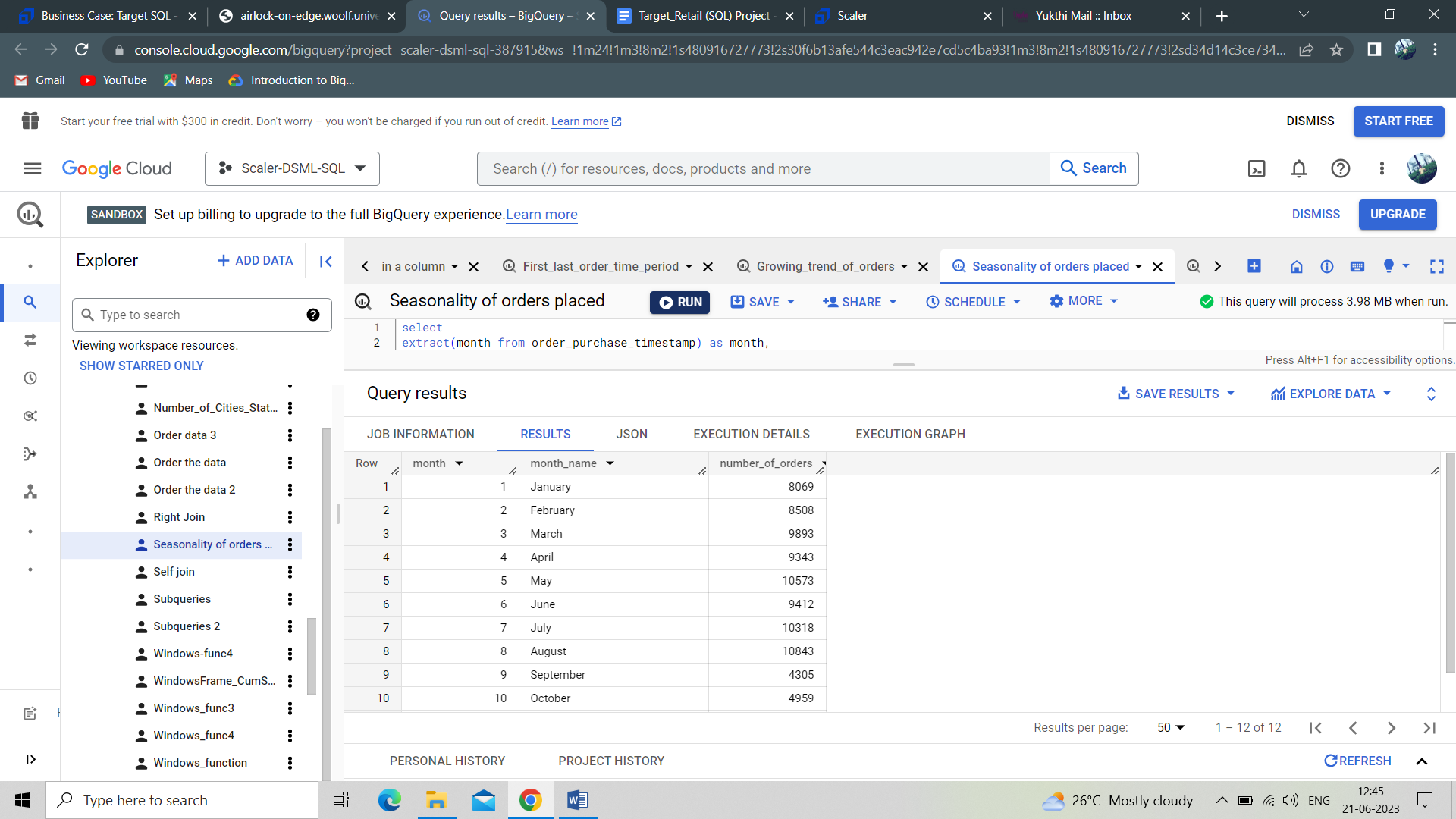
**count(order\_id) as number\_of\_orders,**

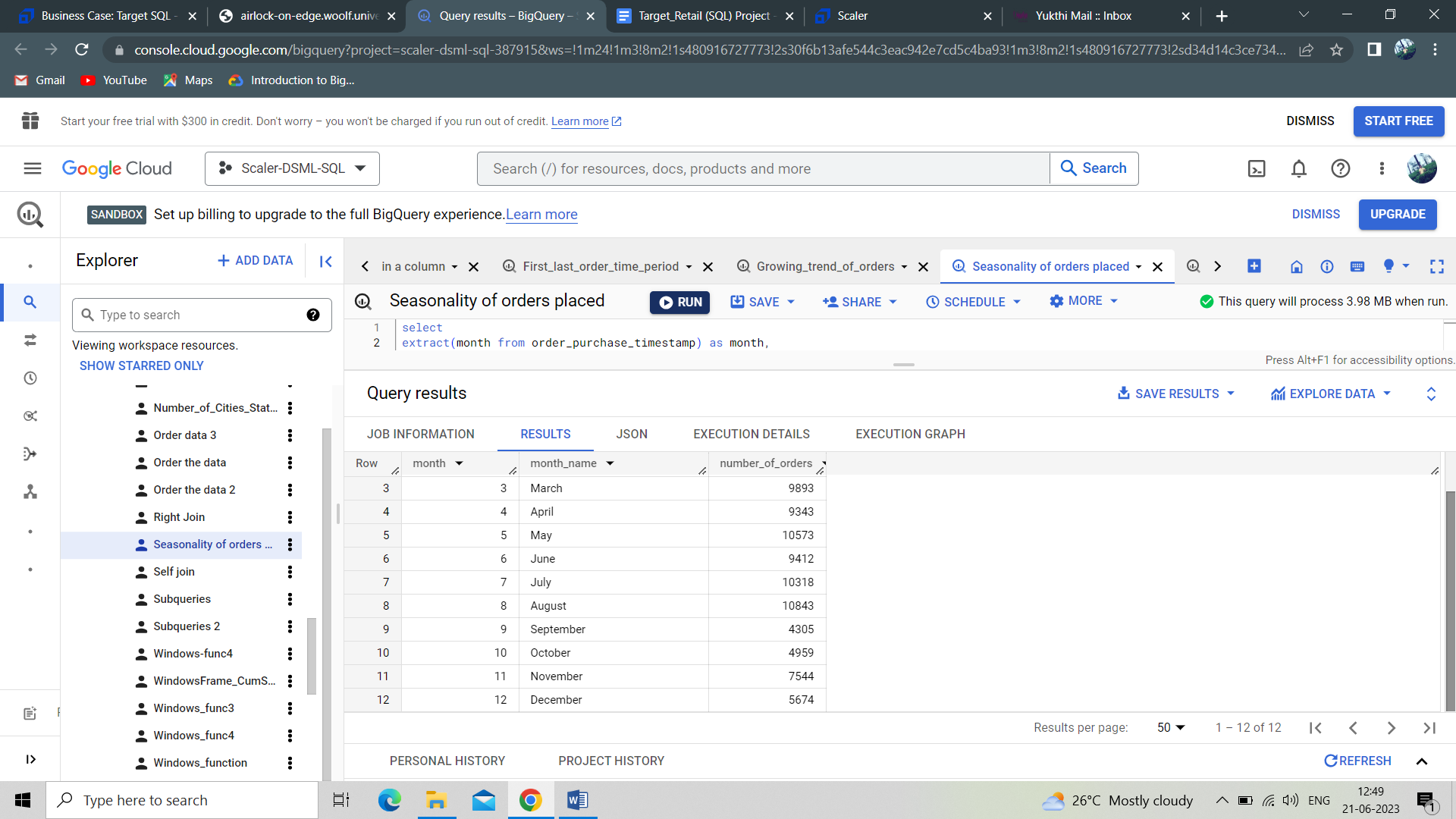
**from `Target\_Retail.orders`**

**group by month, month\_name**

**order by month**

**Query results :**





**Insights :**

The results provide seasonality of orders placed in Brazil. Where we see the highest number of orders in August with 10,843 and the lowest number of orders in September with 4305. Here, an increase in orders placed in September might be possible through better customer engagement programs and better offerings.

1. **During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)**
   1. **0-6 hrs : Dawn**
   2. **7-12 hrs : Mornings**
   3. **13-18 hrs : Afternoon**
   4. **19-23 hrs : Night**

**Query :**

**select**

**case when Time between 0 and 6 then "Dawn"**

**when Time between 7 and 12 then "Mornings"**

**when Time between 13 and 18 then "Afternoon"**

**Else "Night"**

**End as Time\_of\_the\_day,**

**count(order\_id) as number\_of\_orders\_placed**

**from (**

**select order\_id, order\_purchase\_timestamp,**

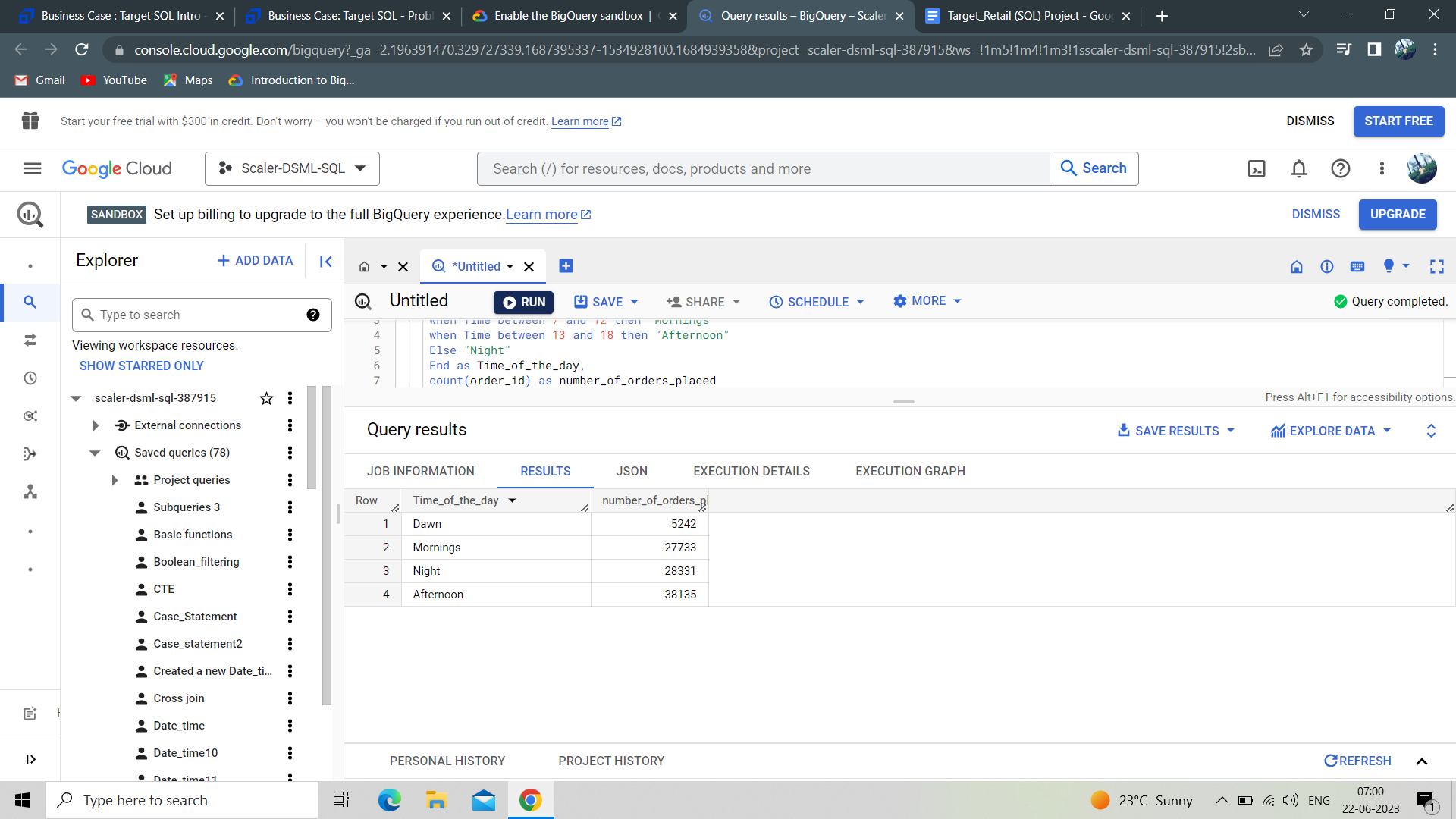
**extract(hour from order\_purchase\_timestamp) as Time**

**from `Target\_Retail.orders`**

**) x**

**group by Time\_of\_the\_day**

**order by number\_of\_orders\_placed**

**Query results:** 

**Insights:**

The results show that 5242 orders placed during Dawn, 27733 orders placed during the morning, 28331 orders placed during the Night, 38135 orders placed during the Afternoon. Overall we see more orders placed during the Afternoon session. So the orders placed in draws are less. It might increase by offering better discounts, coupons, vouchers to achieve better numbers in that particular time.

**Question III :**

**Evolution of E-commerce orders in the Brazil region**

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

**Query :**

**select customer\_state, format\_datetime("%B", order\_purchase\_timestamp) as Month,**

**count(order\_id) as number\_of\_orders,**

**from `Target\_Retail.orders` o**

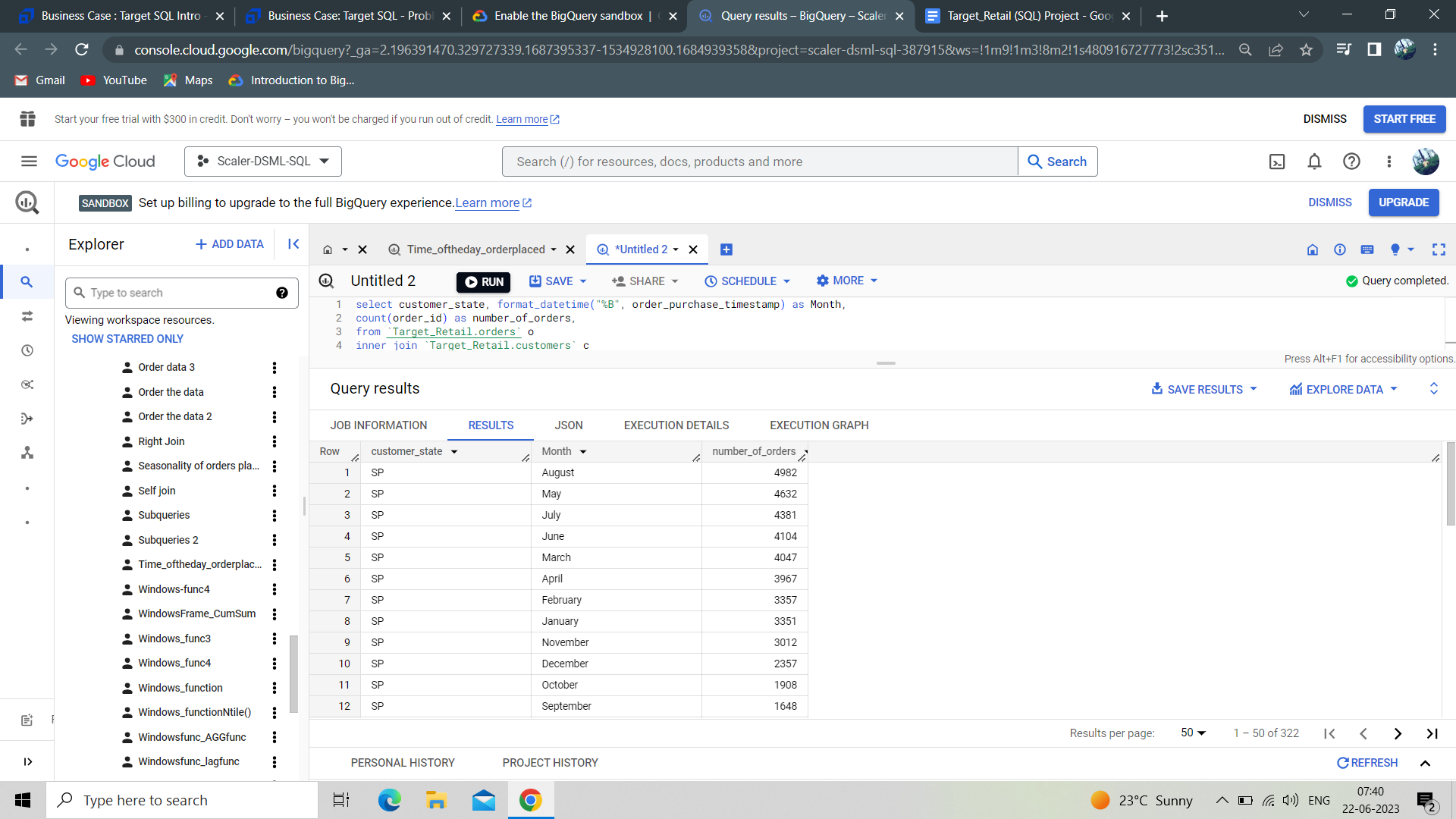
**inner join `Target\_Retail.customers` c**

**on o.customer\_id = c.customer\_id**

**group by customer\_state, Month**

**order by number\_of\_orders desc**

**Query results :**



**Insights :**

The results provide insights about the number of orders placed in SP state in each month. Here we see the highest number of orders placed in August month with 4982 orders and lowest number of orders placed in September month with 1648 orders. It might be because of the highest order placed in the previous month and less customer engagement in September. So by doing better customer engagement programs might help in the increase of orders placed.

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

**Query :**

select customer\_state,

count(distinct customer\_unique\_id) as number\_of\_customers

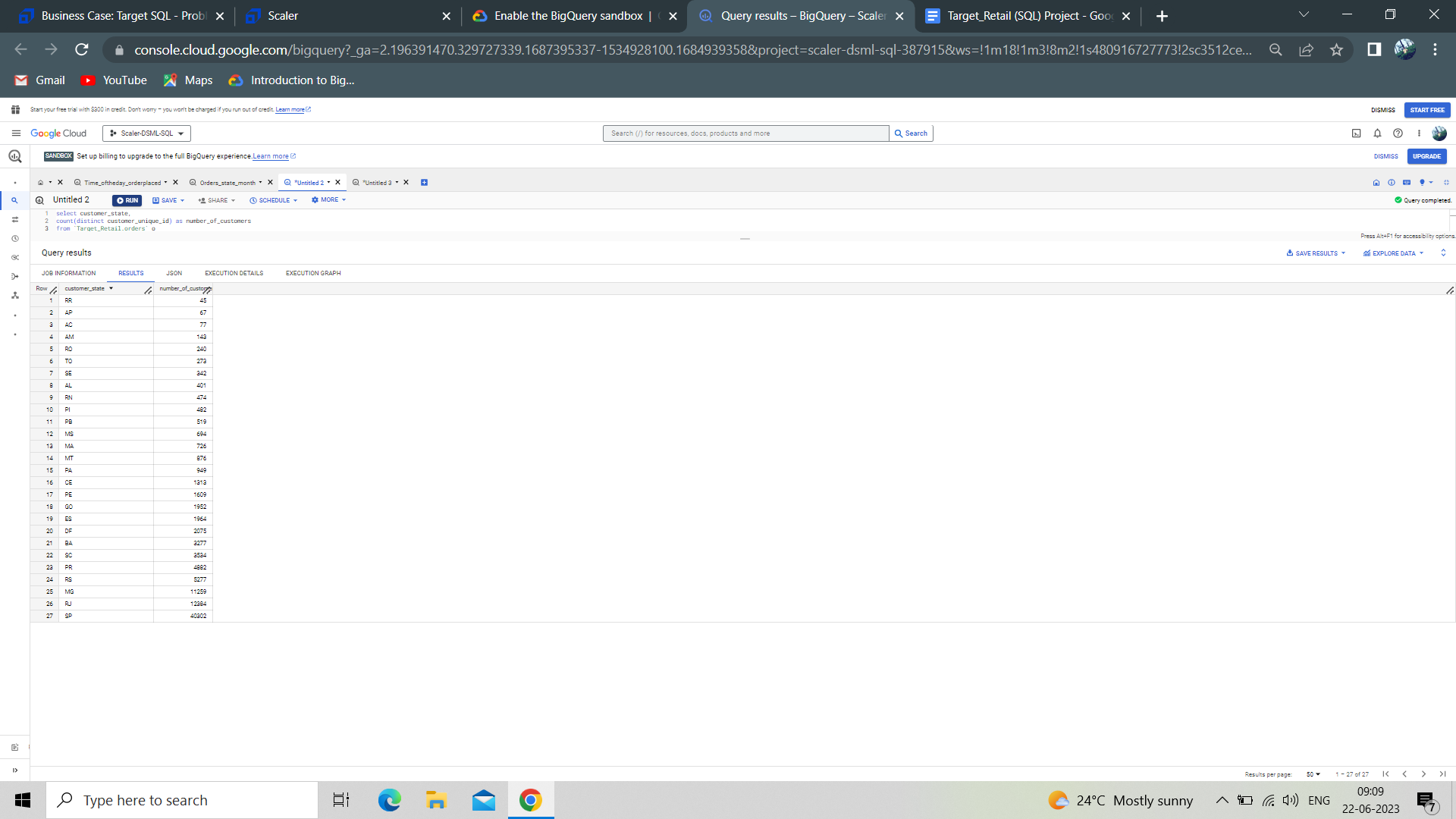
from `Target\_Retail.orders` o

left join `Target\_Retail.customers` c

on o.customer\_id = c.customer\_id

group by customer\_state

order by number\_of\_customers

**Query results :** 

**Insights :**

The result shows SP state has more number of customers with 40302 and RR state has less number of customers with 45. So through marketing and better customer servicing achieving a good number of customers in RR state might be possible.

**Question IV :**

**Impact on Economy: Analyse 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.**

**Query :**

**select year, month, month\_name, sum(cost\_of\_orders) as cost\_of\_orders,**

**lag(sum(cost\_of\_orders)) over (partition by year order by month) as orders\_cost,**

**(((sum(cost\_of\_orders) - lag(sum(cost\_of\_orders)) over (partition by year order by month))/sum(cost\_of\_orders))\*100) as percentage\_increase\_cost**

**from (**

**select extract(year from order\_purchase\_timestamp) as year,**

**extract(month from order\_purchase\_timestamp) as month,**

**format\_datetime("%B", order\_purchase\_timestamp) as month\_name,**

**(payment\_installments \* payment\_value) as cost\_of\_orders,**

**from `Target\_Retail.orders` o**

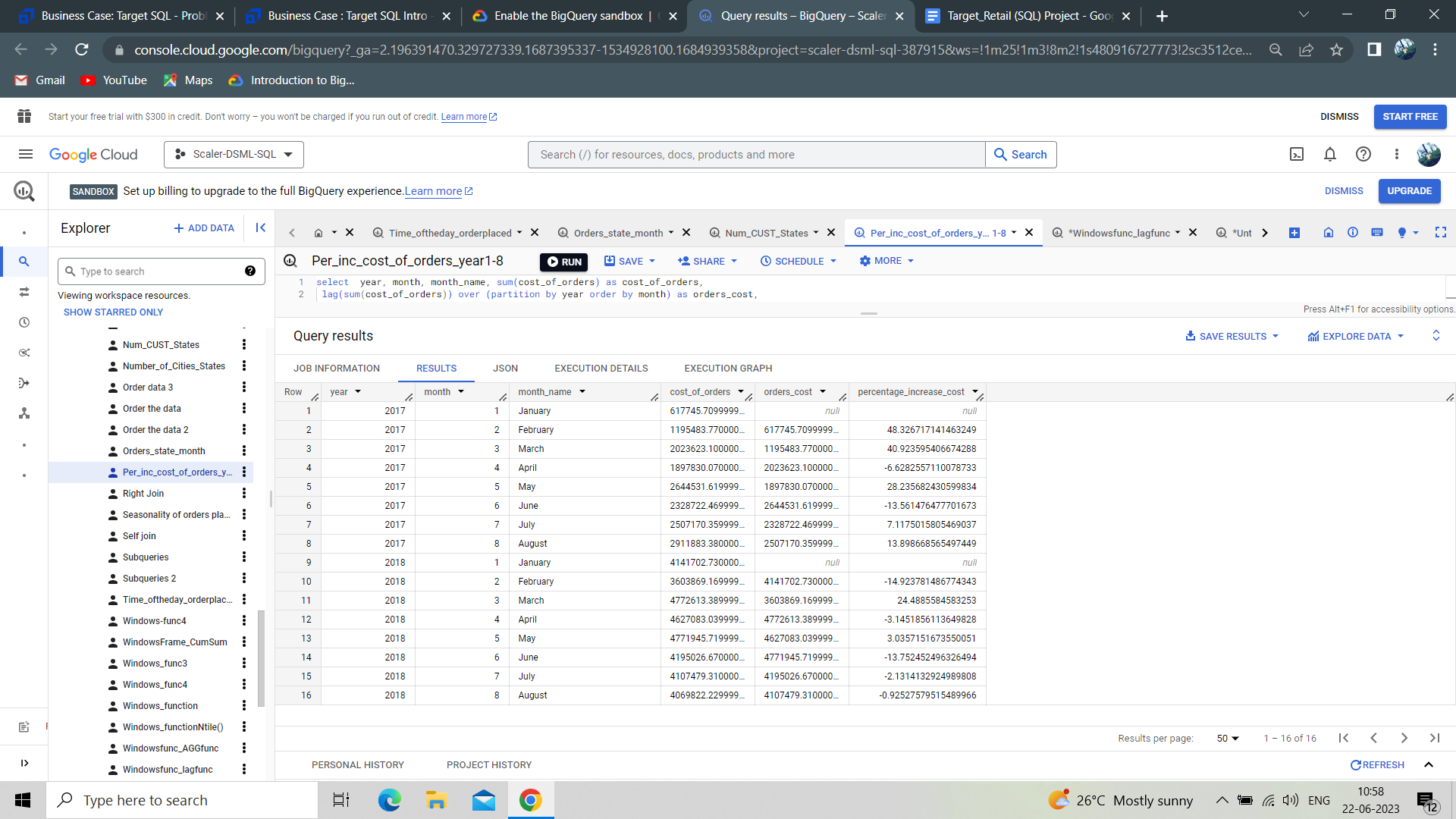
**inner join `Target\_Retail.payments` p**

**on o.order\_id = p.order\_id) x**

**where month between 1 and 8**

**group by year, month, month\_name**

**order by year**

**Query results :**

**Insights:**

The results show that in 2017 the percentage increase in cost of orders is highest in the month Feb and March and in 2018 the percentage increase in cost of orders is highest in the month March. It might happen due to the high number of orders placed in Feb and March compared to Jan, April, May and June in 2017 and 2018.

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

**Query :**

**select customer\_state, sum(price) as total\_price,**

**avg(price) as average\_price**

**from `Target\_Retail.order\_items` ot**

**inner join `Target\_Retail.orders` o**

**on ot.order\_id = o.order\_id**

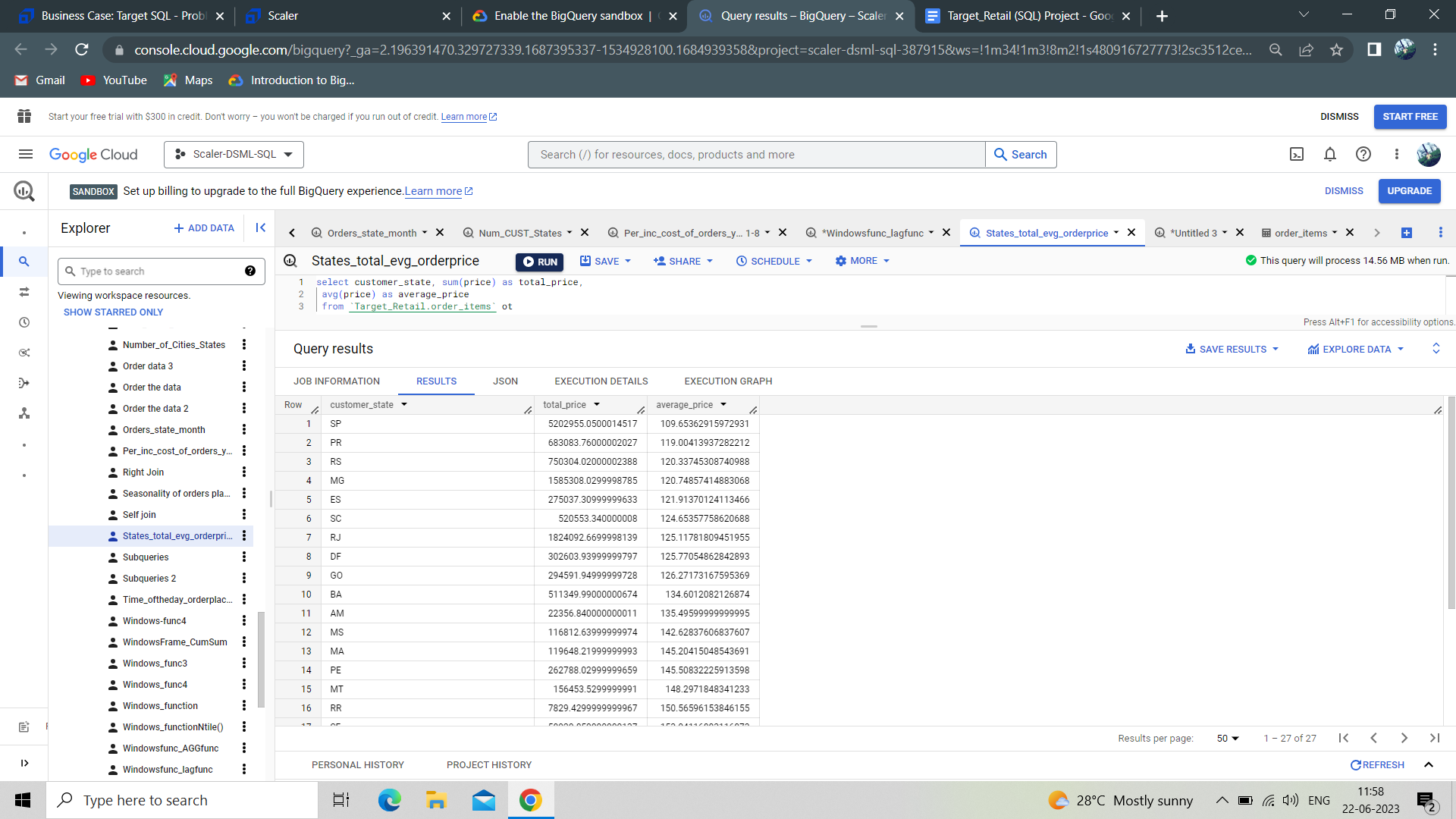
**inner join `Target\_Retail.customers` c**

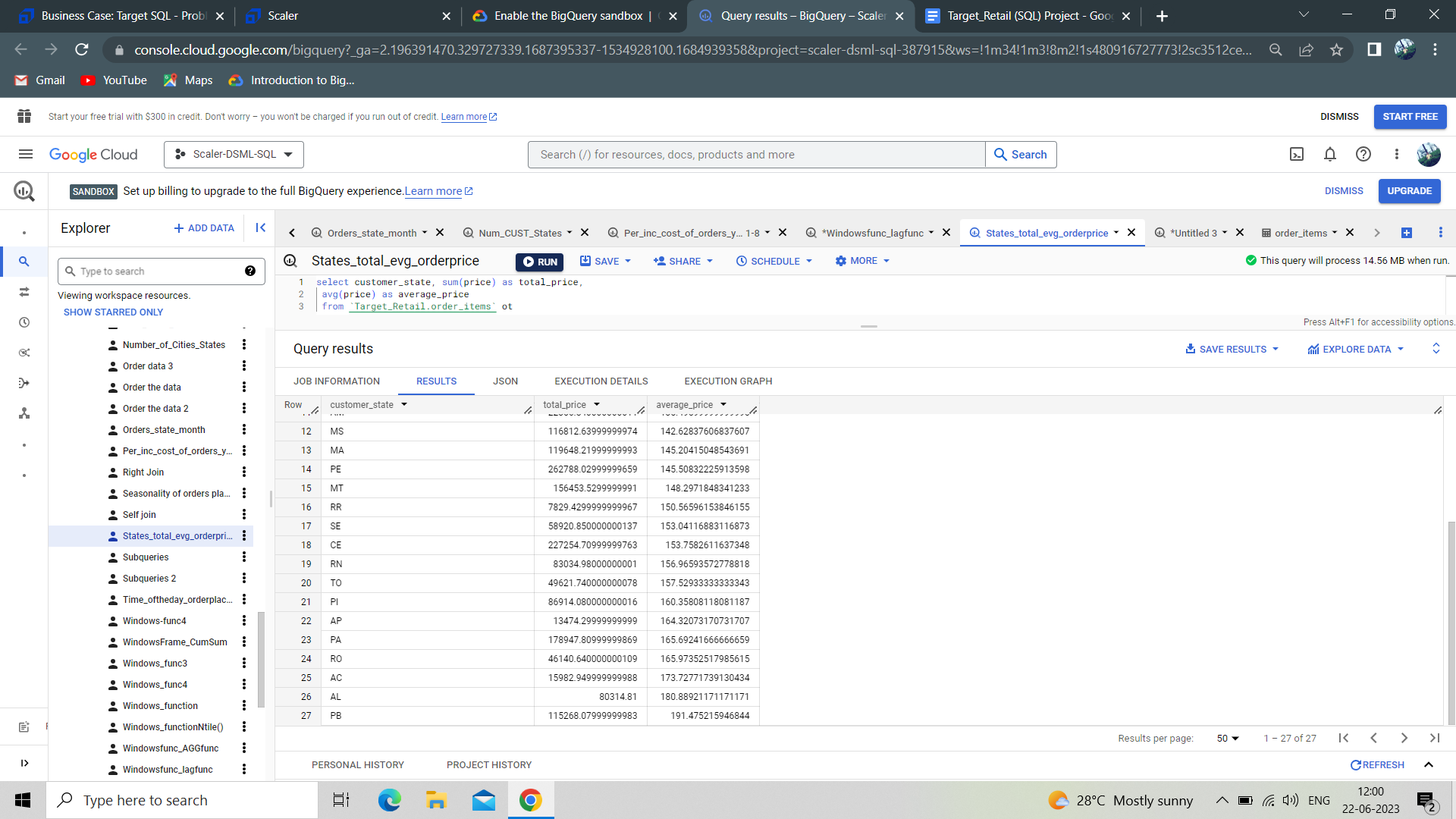
**on o.customer\_id = c.customer\_id**

**group by customer\_state**

**order by average\_price, total\_price**

**Query results:**





**Insights :**

The results show that States SP have the highest total order price and lowest average order price, RR have the lowest total order price and PB have highest average order price. In State RR we see less order price because of less orders, So through marketing and proper customer servicing increase in orders placed might increase in this State.

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

**Query :**

**select customer\_state, sum(freight\_value) as total\_fright\_value,**

**avg(freight\_value) as average\_freight\_value**

**from `Target\_Retail.order\_items` ot**

**inner join `Target\_Retail.orders` o**

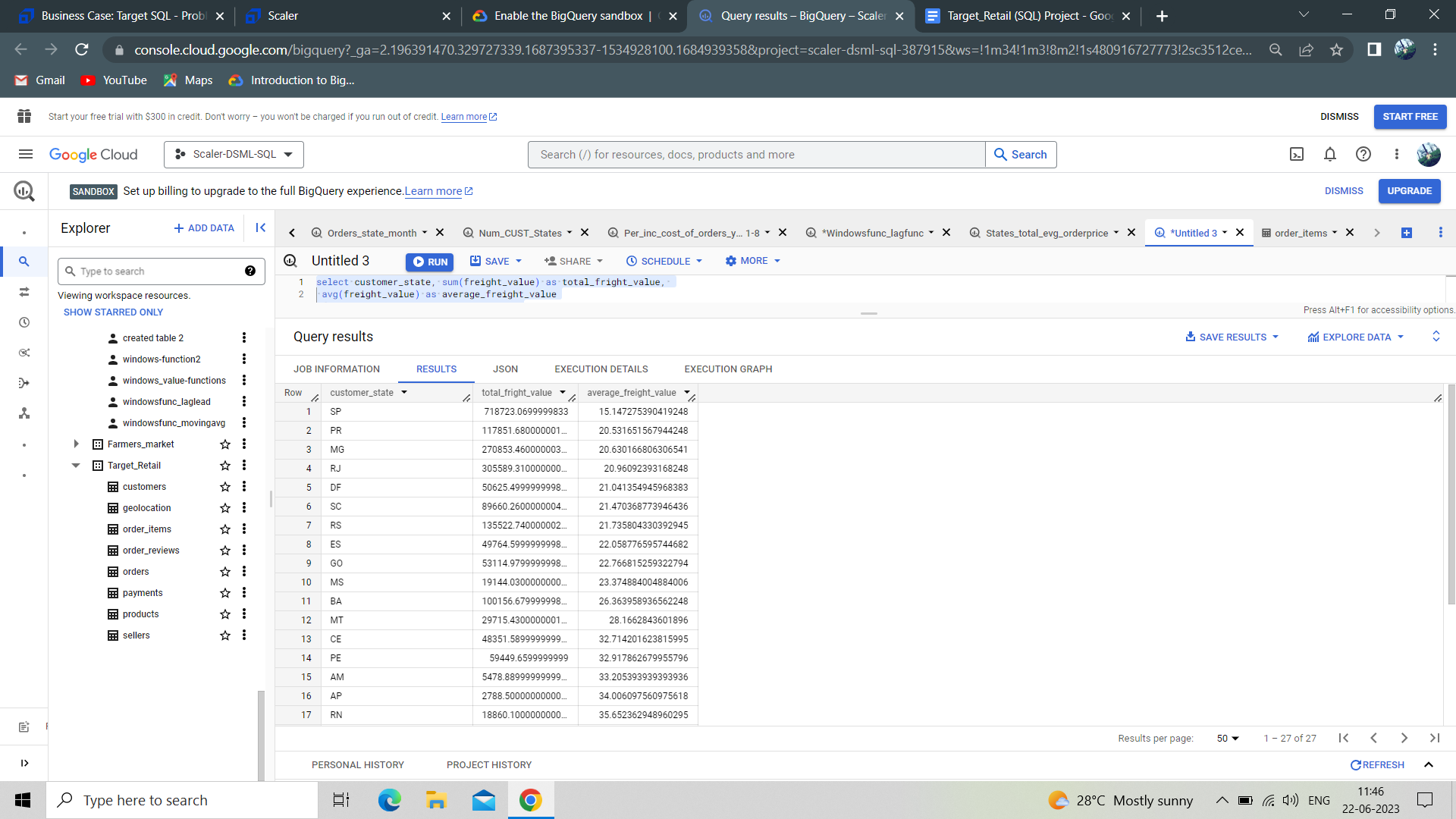
**on ot.order\_id = o.order\_id**

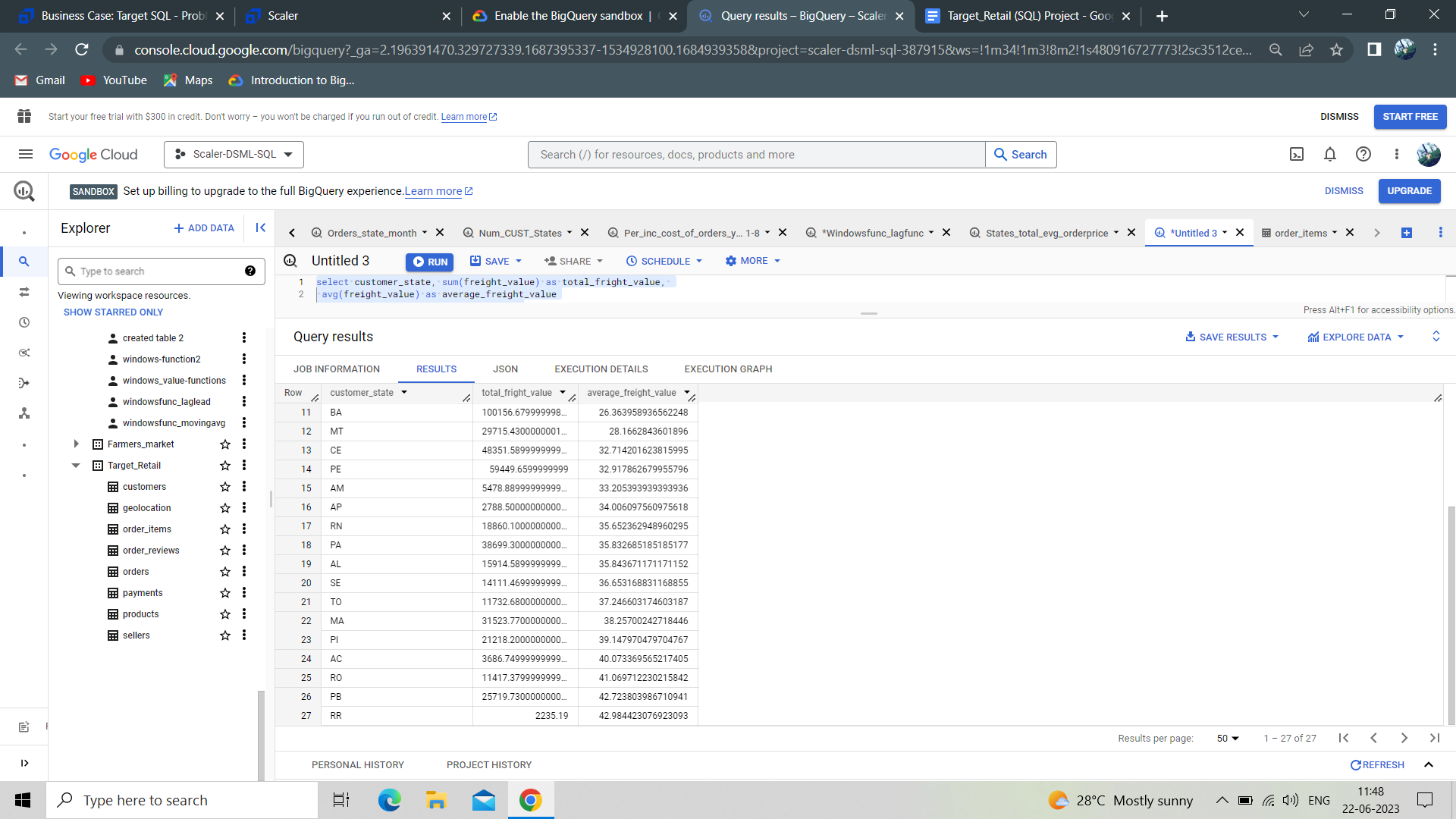
**inner join `Target\_Retail.customers` c**

**on o.customer\_id = c.customer\_id**

**group by customer\_state**

**order by average\_freight\_value, total\_fright\_value**

**Query results:**



**Insights :**

Here we see state SP is highest in total freight value and lowest in average freight value, RR is lowest in total freight value and highest in average freight value. By reducing the cost of transportation and confirming the orders at initial stage to avoid rejections might help decrease in Freight value.

**Question V :**

**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.**

**Query :**

**select order\_id,**

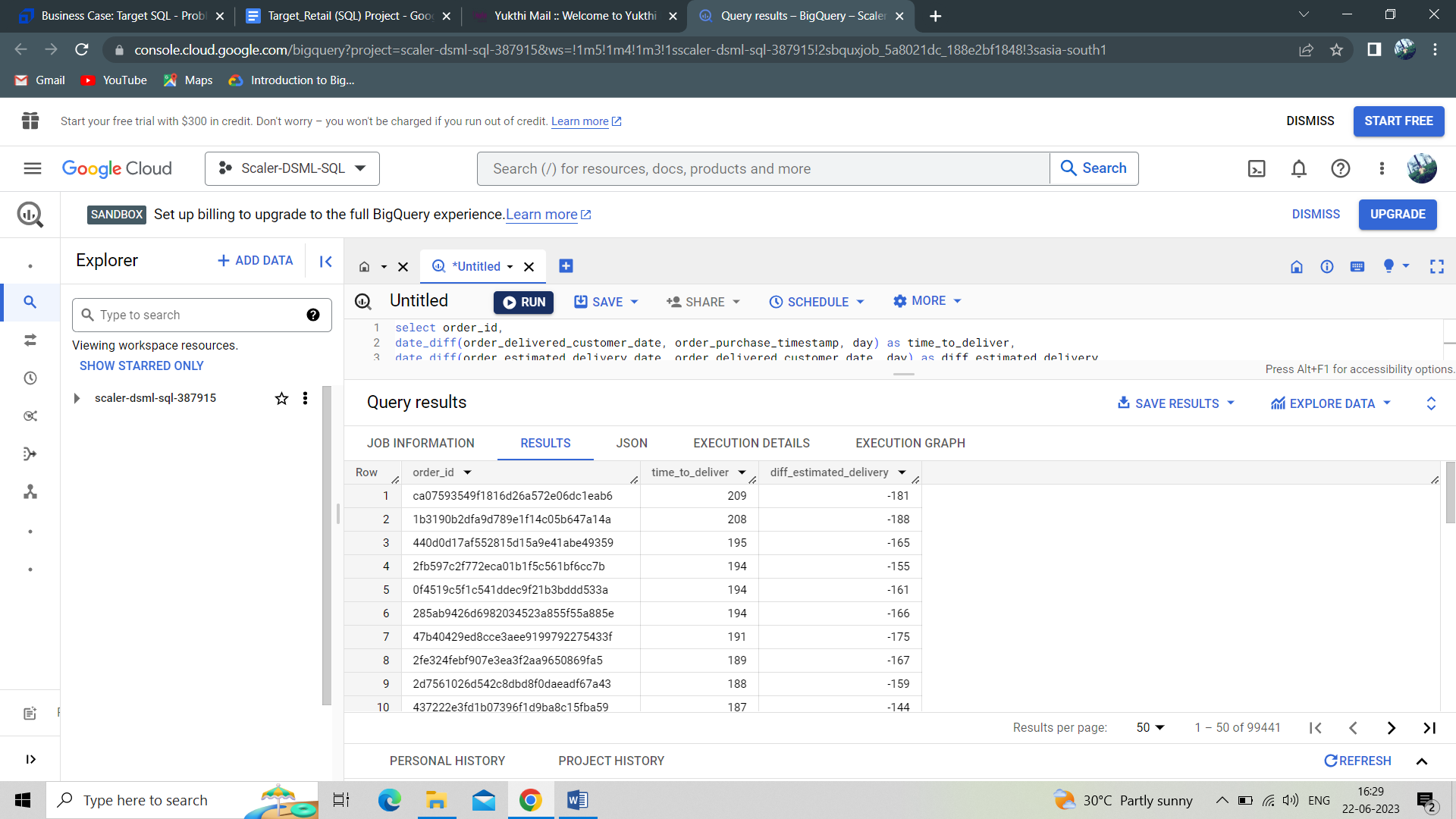
**date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day) as time\_to\_deliver,**

**date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day) as diff\_estimated\_delivery**

**from `Target\_Retail.orders`**

**order by time\_to\_deliver desc, diff\_estimated\_delivery desc**

**Query results :**



**Insights :**

Here we see more days taken by an order delivered to the customer from the date of order placed is 209 days and more number where an order delayed to deliver to the customer from the estimated delivery date is 188 days. By increasing delivery agents where the delivery is delayed might decrease in delay in delivery of orders.

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

**Query :**

**SELECT customer\_state, Top\_Bottom\_5\_Freight\_value**

**FROM (**

**SELECT customer\_state, AVG(freight\_value) AS Top\_Bottom\_5\_Freight\_value**

**FROM `Target\_Retail.order\_items` ot**

**INNER JOIN `Target\_Retail.orders` o ON ot.order\_id = o.order\_id**

**INNER JOIN `Target\_Retail.customers` c ON o.customer\_id = c.customer\_id**

**GROUP BY customer\_state**

**ORDER BY Top\_Bottom\_5\_Freight\_value DESC**

**LIMIT 5**

**) AS top\_5**

**UNION ALL**

**SELECT customer\_state, Bottom\_5\_Freight\_value**

**FROM (**

**SELECT customer\_state, AVG(freight\_value) AS Bottom\_5\_Freight\_value**

**FROM `Target\_Retail.order\_items` ot**

**INNER JOIN `Target\_Retail.orders` o ON ot.order\_id = o.order\_id**

**INNER JOIN `Target\_Retail.customers` c ON o.customer\_id = c.customer\_id**

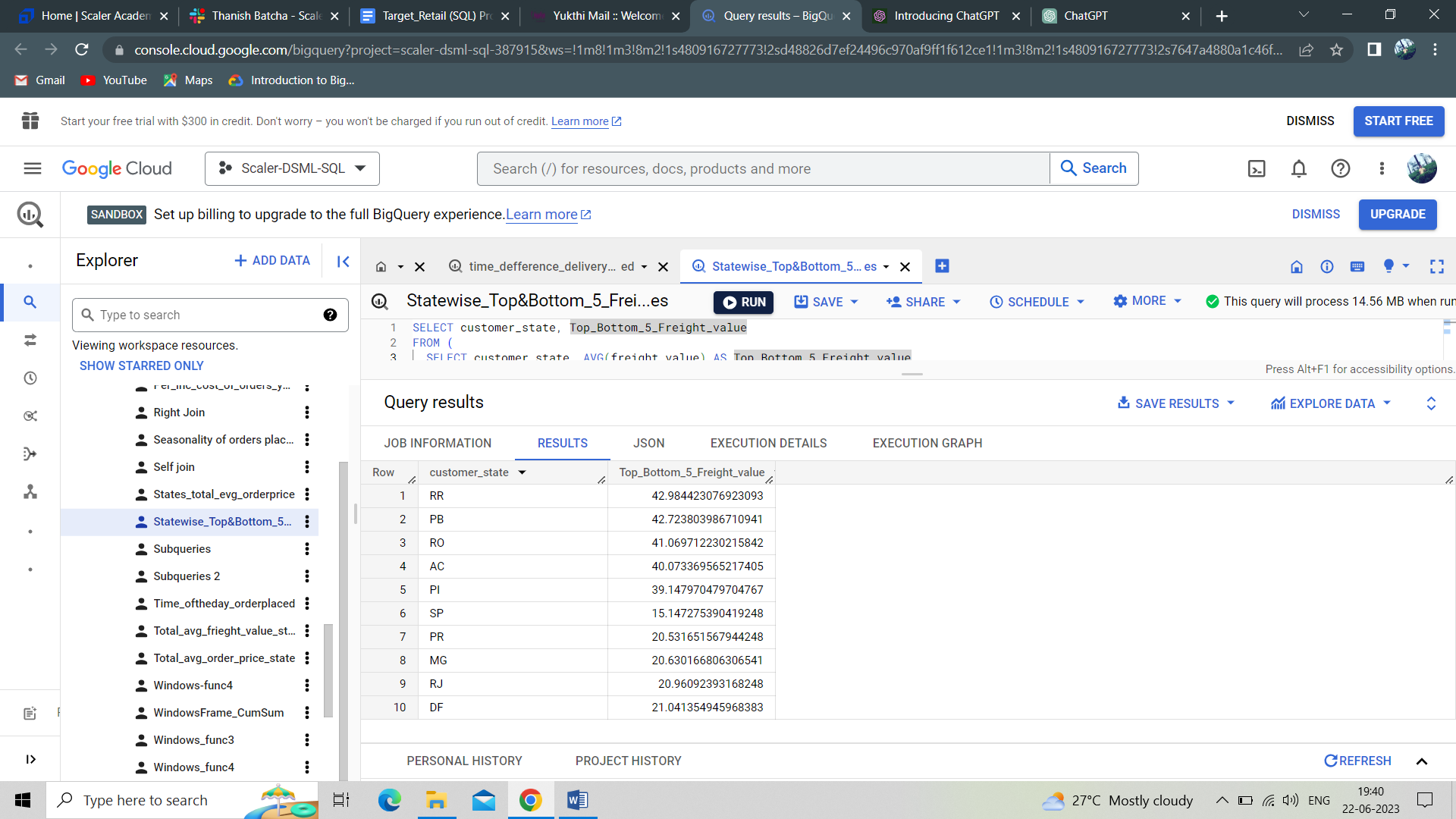
**GROUP BY customer\_state**

**ORDER BY Bottom\_5\_Freight\_value**

**LIMIT 5**

**) AS bottom\_5**

**Query results:**



**Insights :**

The results show that RR, PB, RO, AC, PI as Top 5 states which have highest average freight value and SP, PR, MG, RJ, DF as Bottom 5 states which have lowest average freight value. Decrease in freight value is possible based on less costly mode of transportation and order confirmations to avoid rejections at the end.

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

**Query :**

**with table\_1 as (select customer\_state,**

**avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day)) as Avg\_delivery\_time**

**from `Target\_Retail.orders` o**

**left join `Target\_Retail.customers` c**

**on o.customer\_id = c.customer\_id**

**group by customer\_state**

**order by Avg\_delivery\_time),**

**table\_2 as (select customer\_state, Avg\_delivery\_time,**

**row\_number() over (order by Avg\_delivery\_time desc) as Top\_Avg\_delivery\_time,**

**row\_number() over (order by Avg\_delivery\_time asc) as Bottom\_avg\_delivery\_time**

**from table\_1),**

**table\_3 as (select \* from table\_2**

**where Top\_Avg\_delivery\_time <= 5 or Bottom\_avg\_delivery\_time <= 5 )**

**select \* from table\_3**

**Query results:**



**Insights:**

The above shows that states RR, AP, AM, AL, PA are the top 5 states with highest average delivery time and SP, PR, MG, DF, SC are the Top 5 states with lowest average delivery time. So an increase in delivery agents in these states might increase average delivery time.

1. **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.**

**Query :**

**with table\_1 as (select customer\_state,**

**avg(date\_diff(order\_delivered\_customer\_date, order\_estimated\_delivery\_date, day)) as Avg\_delivery\_time\_difference**

**from `Target\_Retail.orders` o**

**left join `Target\_Retail.customers` c**

**on o.customer\_id = c.customer\_id**

**group by customer\_state),**

**table\_2 as (select customer\_state, Avg\_delivery\_time\_difference,**

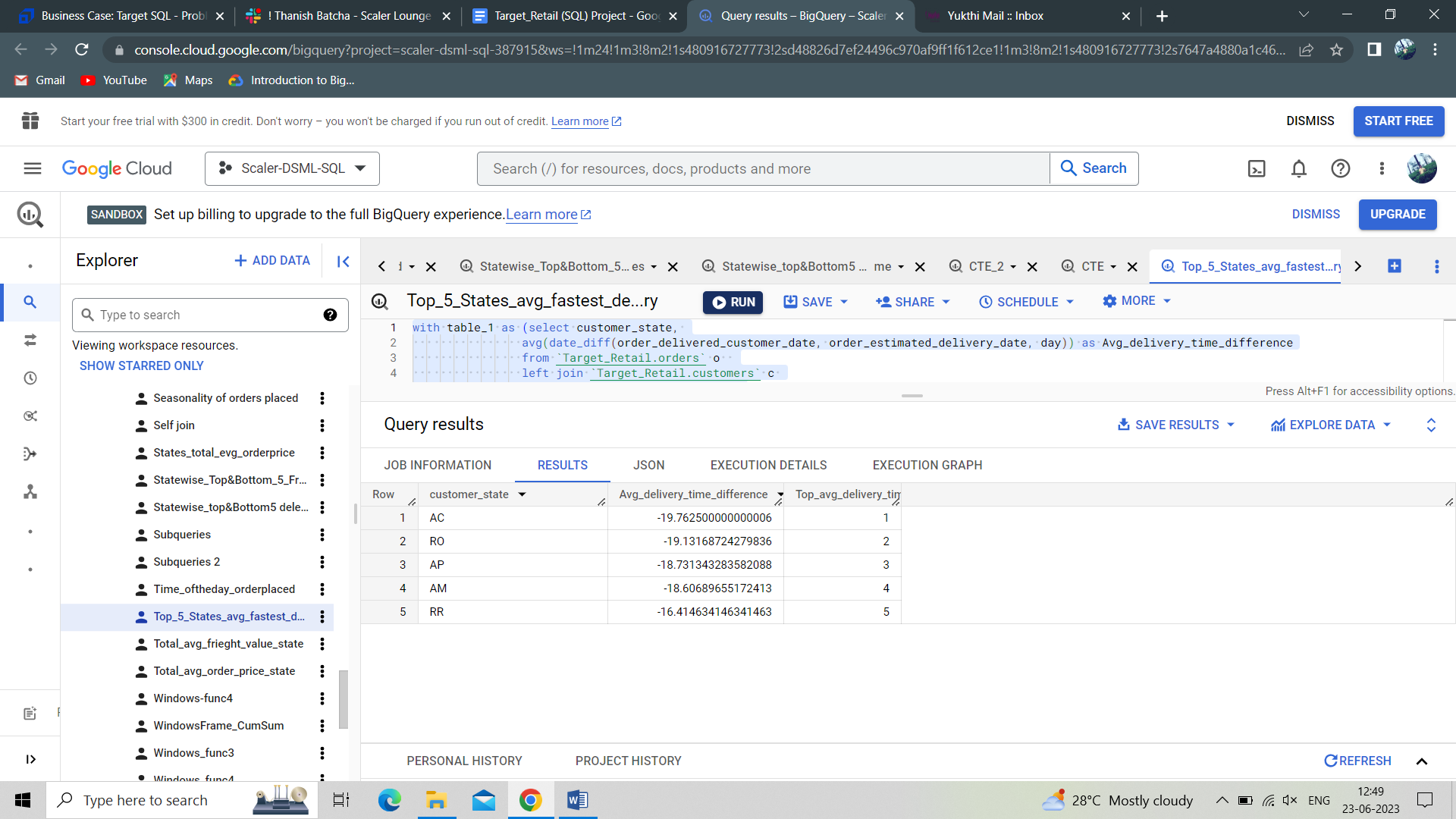
**row\_number() over (order by Avg\_delivery\_time\_difference) as Top\_avg\_delivery\_time\_diffrence**

**from table\_1)**

**select \* from table\_2**

**where Top\_avg\_delivery\_time\_diffrence <= 5**

**Query results:**



**Insights :**

The results show that AC, RO, AP, AM, RR are the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

**Question VI :**

**Analysis based on the payments :**

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

**Query :**

**select extract(month from order\_purchase\_timestamp) as month,**

**format\_datetime("%B", order\_purchase\_timestamp) as month\_name,**

**payment\_type,**

**count(o.order\_id) as num\_of\_order\_id**

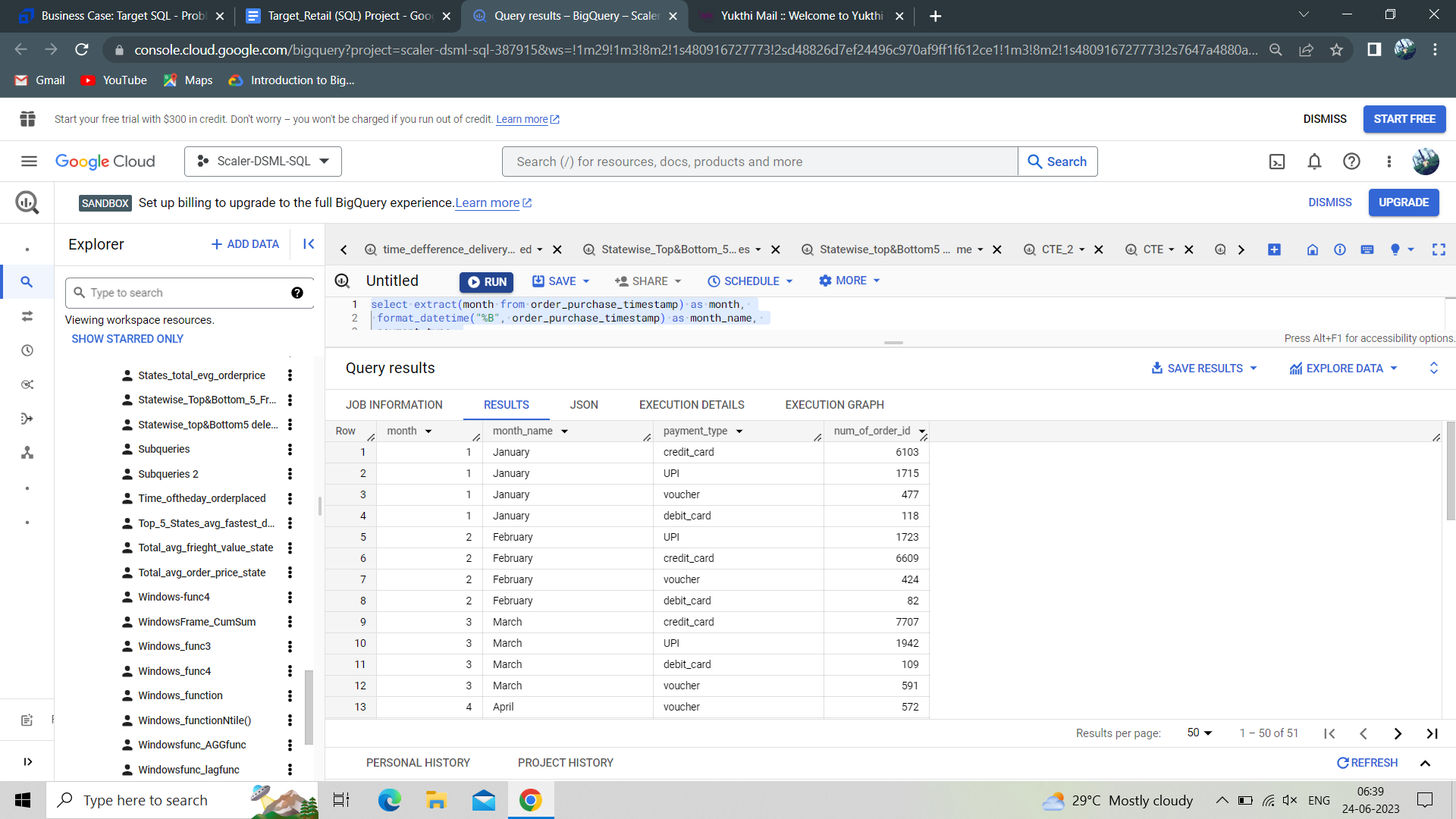
**from `Target\_Retail.orders` o**

**left join `Target\_Retail.payments` p**

**on p.order\_id = o.order\_id**

**group by month, month\_name, payment\_type**

**order by month**

**Query results :** 

**Insights :**

From the results we come to know that the number of orders placed through credit card payments in each month over the years is high as compared to other payment types and debit card payments are less in each month over the years. So the debit card payment might be less because of bank rejections and server issues.

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

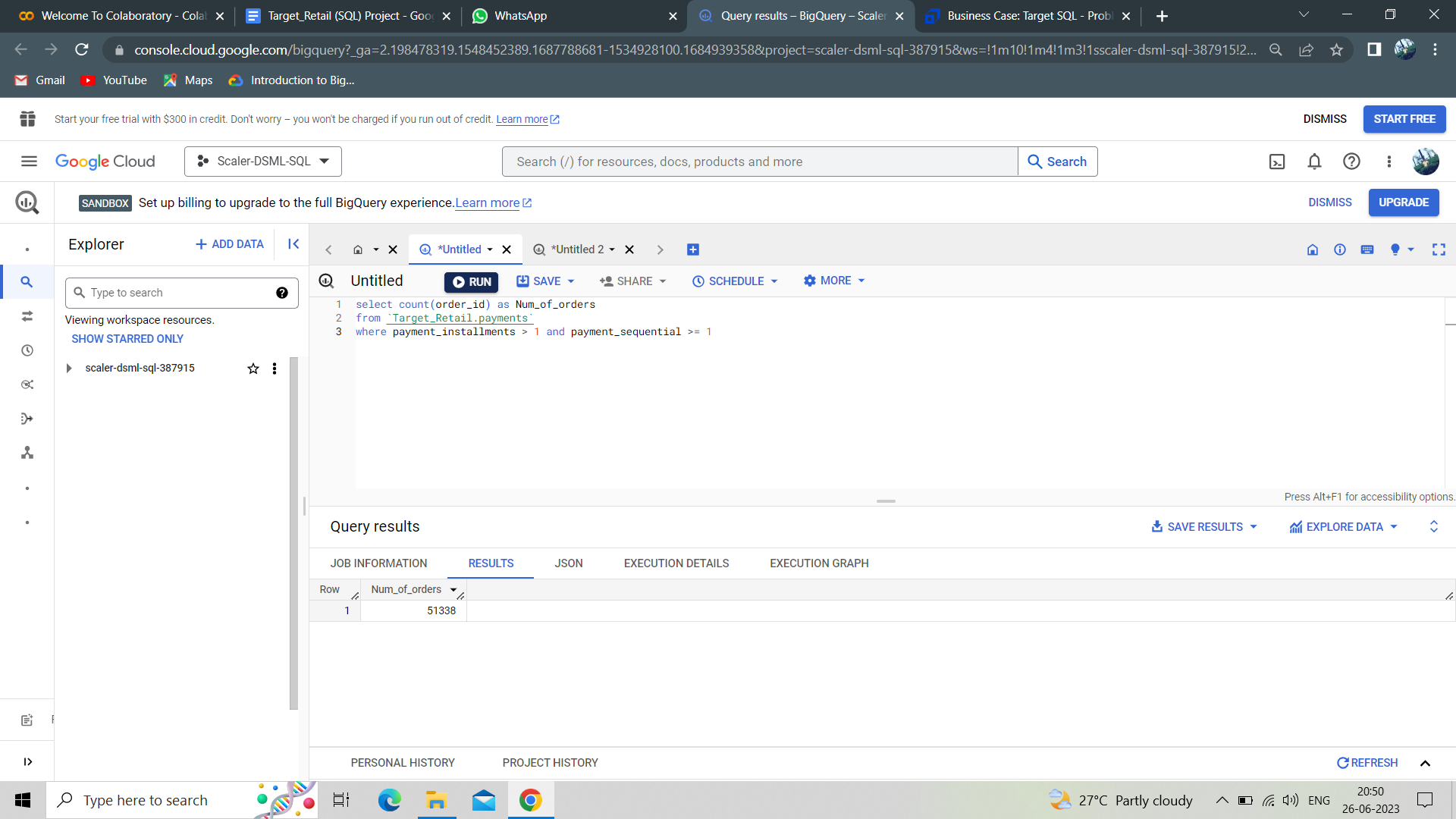
**Query :**

**select count(order\_id) as Num\_of\_orders**

**from `Target\_Retail.payments`**

**where payment\_installments > 1 and payment\_sequential >= 1**

**Query results:**



**Insights:**

Out of 103886 orders 51338 are paid under instalment basis. So more than 50 % of orders are paid under instalment basis and it is relatively greater than orders which got the entire payment at one time.

**Recommendations :**

* During the Dawn period offering better discounts, coupons, vouchers might increase in orders placed and to avoid the cancellations of delivery, order confirmation and also reducing the delivery time might help in increasing the orders placed.
* In states like RR, increasing delivery agents helps in reducing delivery time, launching new retail outlets, offering better discounts, coupons, vouchers, conducting customer engagement programs and better customer serving also helps in increasing the orders placed.
* In September the orders placed were less compared to other months because of the highest orders placed in the previous month or low customer serving and customer benefits. So offering better discounts, vouchers, coupons on orders and also through better customer servicing might increase the number of orders placed in september.
* We see total freight value is high in case RR state because of lesser orders and SP have highest average freight value because of mode of transportation and order cancellations. So to avoid higher freight charges reducing the transportation cost, reducing order cancellations and increasing orders placed in RR state with good customer servicing and benefits might help.
* Debit card payments are very less while placing the orders because of bank server issues and rejections by the bank during the payment. It is the same in case of UPI payments as well. In order to increase the Debit and UPI transactions, optimising the payment gateway might be helpful, since most of the customers use these payment modes for placing the orders.
* State RR has the lowest order price because of less orders. So increasing the retail stores, delivery agents, offering better discounts, coupons, vouchers, customer services and through marketing. The Target can increase the orders placed in the state.
* Irrespective of considerable orders placed. State PB has highest freight values because of poor mode of transport, less outlets and delivery agents. So Increase in outlet stores, delivery agents with better transport might help in reducing the freight charges.