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

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

customer\_id=String-VARCHAR

customer\_unique\_id=String-VARCHAR

customer\_zip\_code\_prefix=Integer

Customer\_city= String -CHAR(3)

customer\_state= String -CHAR (2)

**Insight/Recommendation** : Datatype String with VARCHAR,CHAR is used along with Integer

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

SELECT

concat(x.start\_date," ",x.start\_time) as First\_order,

concat(x.End\_date," ",x.end\_time) as last\_order,

from (

select

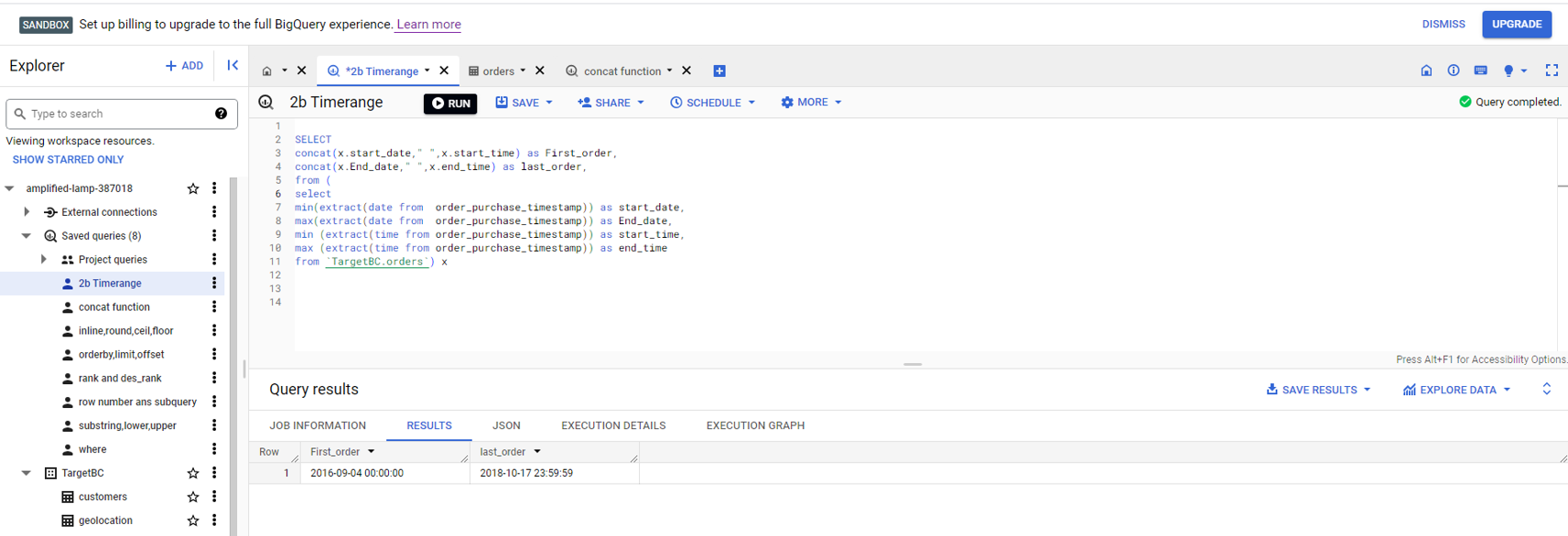
min(extract(date from  order\_purchase\_timestamp)) as start\_date,

max(extract(date from  order\_purchase\_timestamp)) as End\_date,

min (extract(time from order\_purchase\_timestamp)) as start\_time,

max (extract(time from order\_purchase\_timestamp)) as end\_time

from `TargetBC.orders`) x



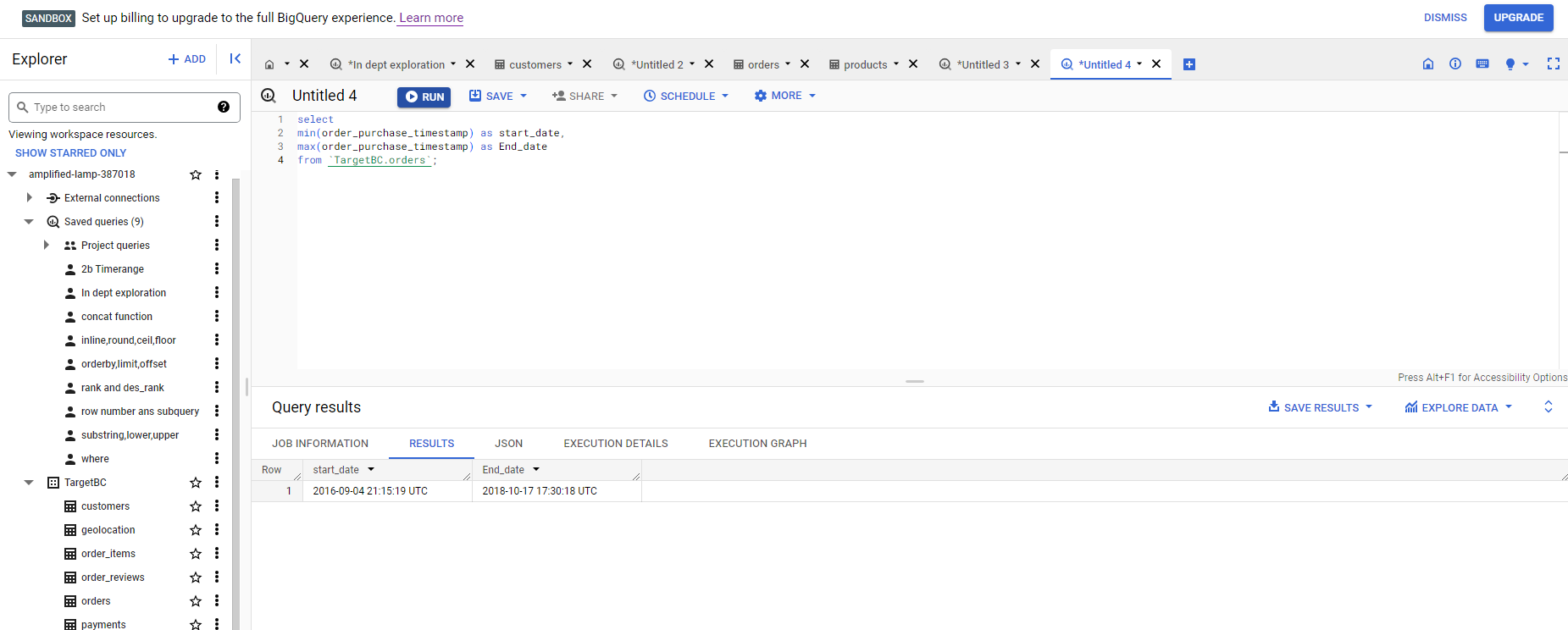
Or

select

min(order\_purchase\_timestamp) as start\_date,

max(order\_purchase\_timestamp) as End\_date

from `TargetBC.orders`;



**Insight : Time range for give dataset found to be in between Sept-2016 to Oct-2018**

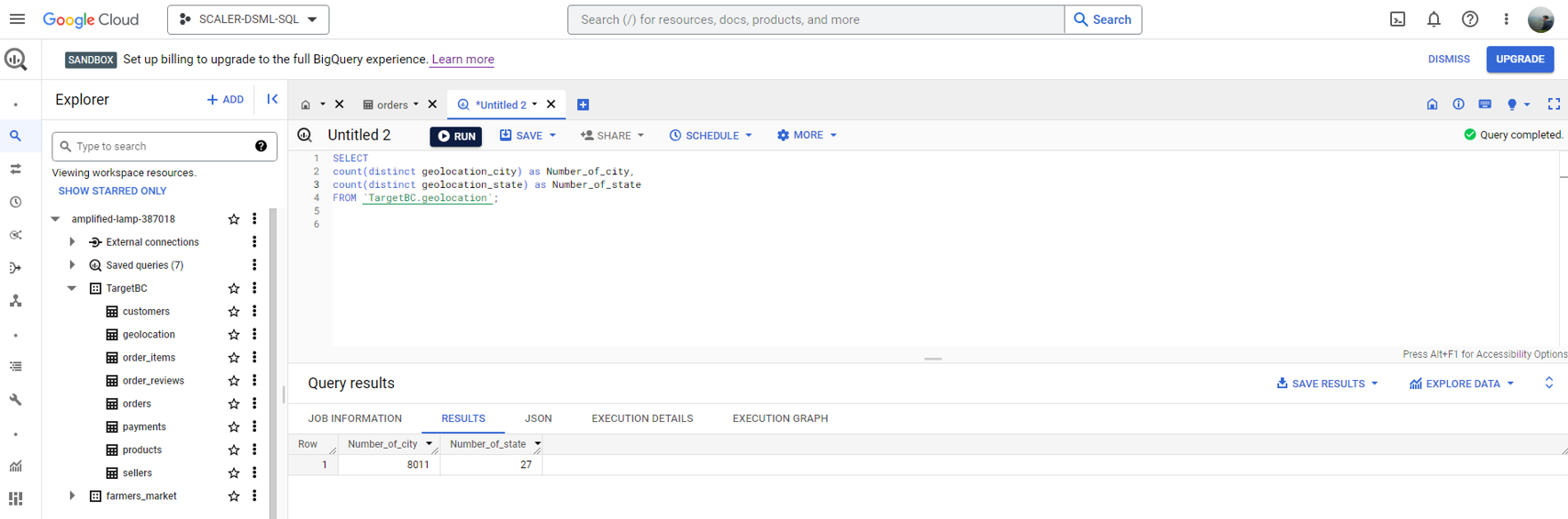
c. Count the number of Cities and States in our dataset.

SELECT

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

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

FROM `TargetBC.geolocation`;



**Insight/Recommendation: Target has spectrum of order from across all state 27 states and 8011 cities covering all region within brazil.**

**2. In-depth Exploration:**

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

For Year :

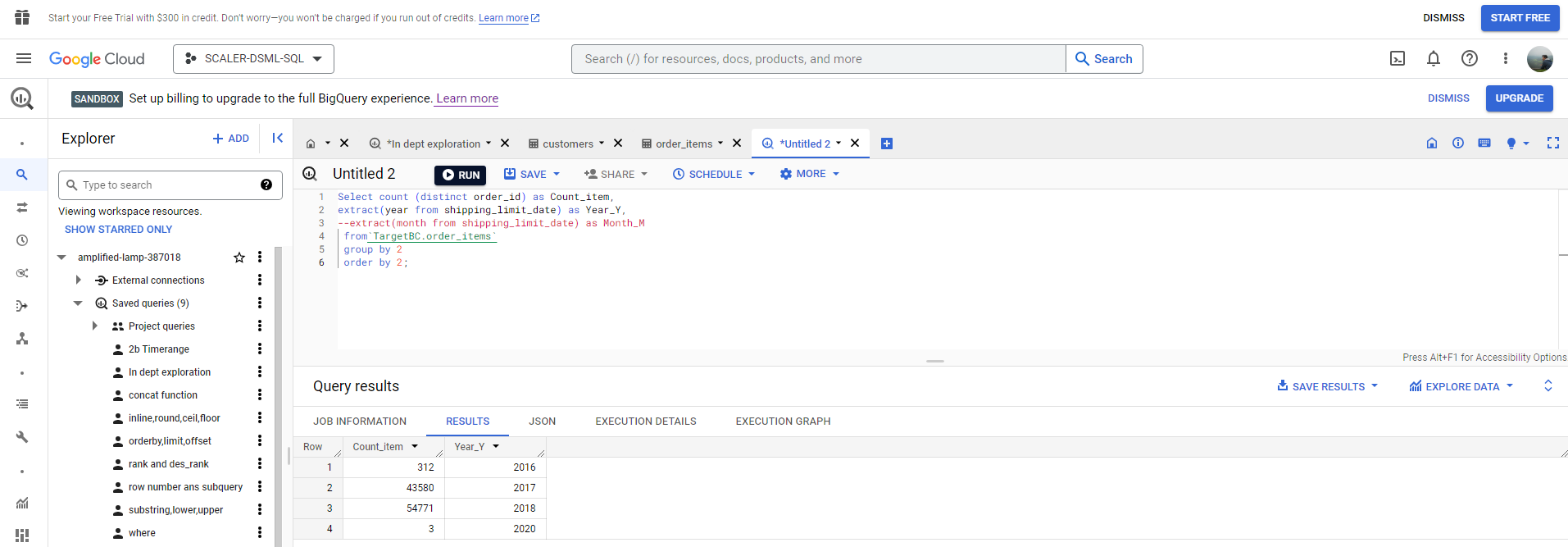
Select count (distinct order\_id) as Count\_item,

extract(year from shipping\_limit\_date) as Year\_Y,

from`TargetBC.order\_items`

 group by 2

 order by 2;



**Insight/Recommendation : In past year 2017,2018 was significant growth but in recent year 2019 onwards it has sharp downfall in number of orders.**

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

Select count (distinct order\_id) as Count\_item,

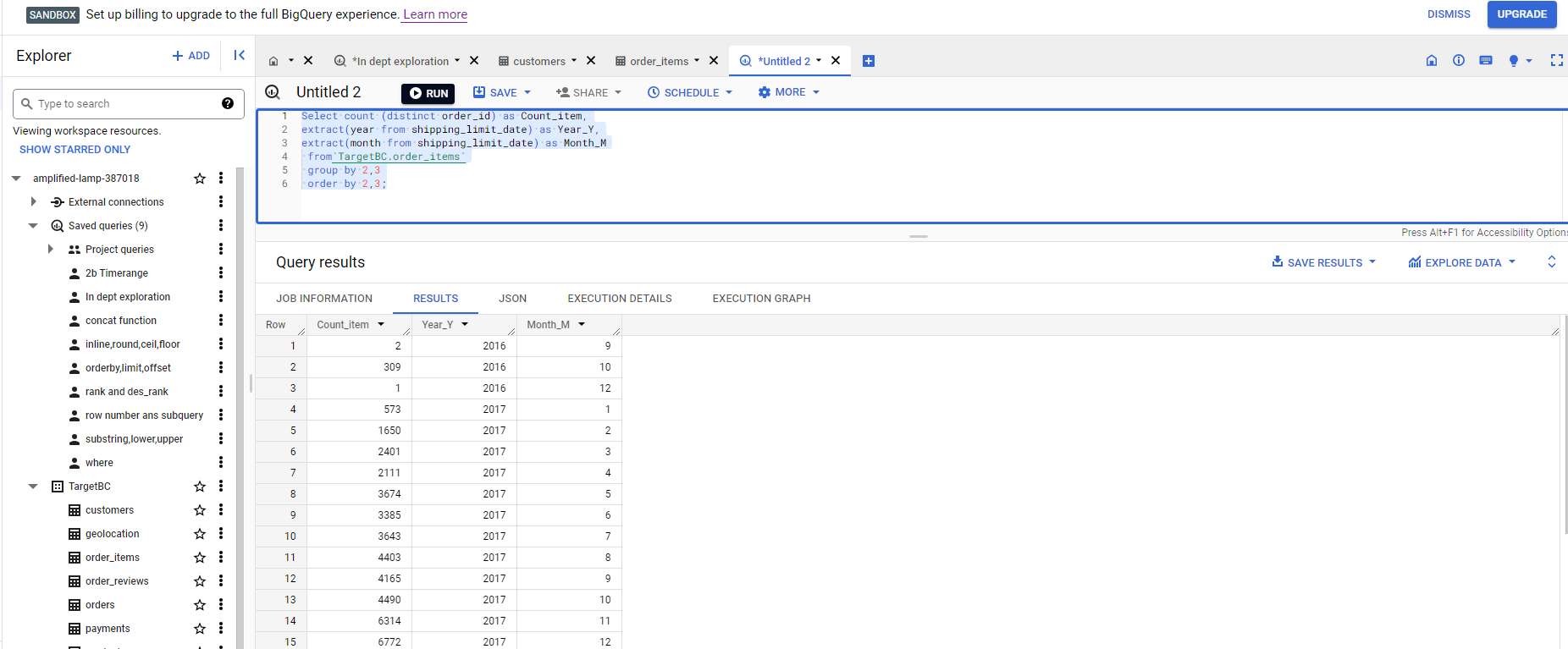
extract(year from shipping\_limit\_date) as Year\_Y,

extract(month from shipping\_limit\_date) as Month\_M

 from`TargetBC.order\_items`

 group by 2,3

 order by 2,3;



And

select max (Count\_item)as Max\_order,x.Month\_M,

from (

Select count (distinct order\_id) as Count\_item,

extract(year from shipping\_limit\_date) as Year\_Y,

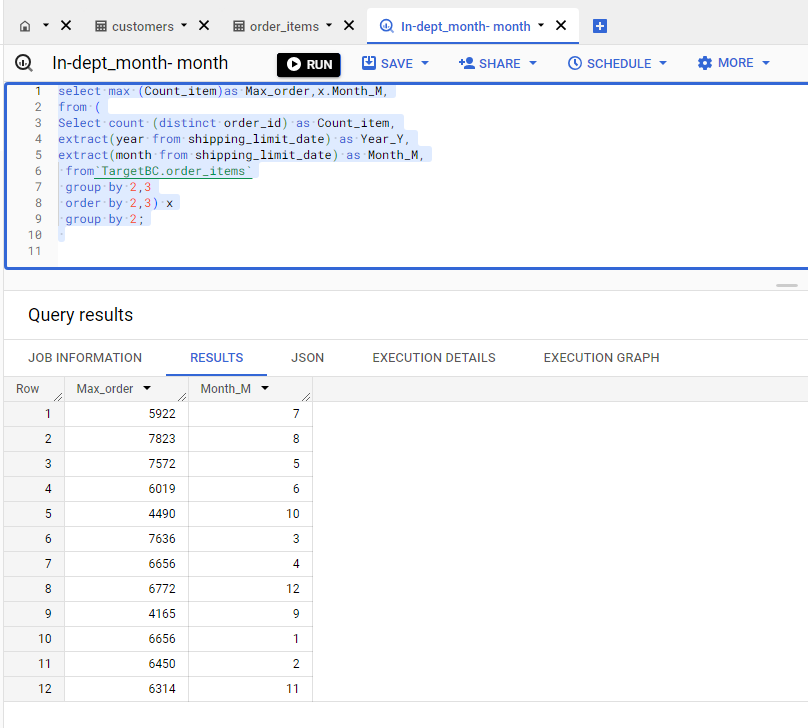
extract(month from shipping\_limit\_date) as Month\_M,

 from`TargetBC.order\_items`

 group by 2,3

 order by 2,3) x

 group by 2;



**Insight/Recommendation : Max order trend is shown in month summer start from Dec to March,and then in winter session June to Aug.**

c. 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

Select order\_purchase\_timestamp,

extract(hour from order\_purchase\_timestamp) Time\_hour,

case

when extract(hour from order\_purchase\_timestamp) between 0 and 6 then "Dawn"

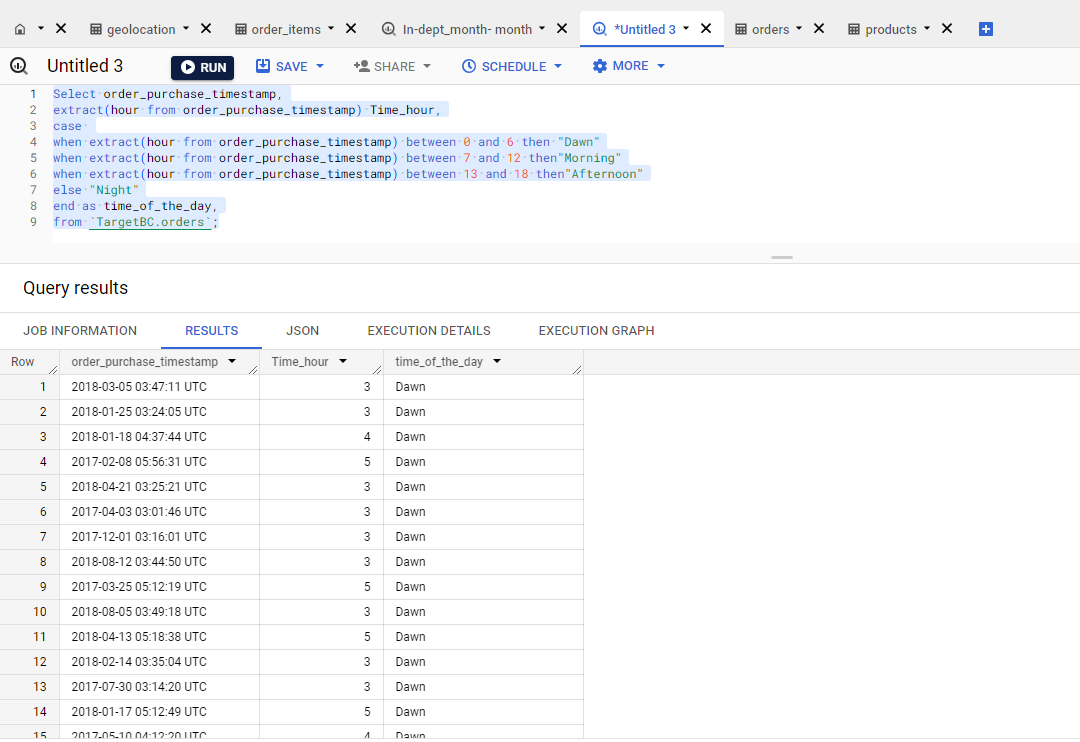
when extract(hour from order\_purchase\_timestamp) between 7 and 12 then"Morning"

when extract(hour from order\_purchase\_timestamp) between 13 and 18 then"Afternoon"

else "Night"

end as time\_of\_the\_day,

from `TargetBC.orders`;



**Insight/Recommendation :** Brazilian customers mostly place their orders in Night.

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

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

select

c.customer\_state,

extract(month from o.order\_purchase\_timestamp) as Month\_M,

count(distinct order\_id) count\_row,

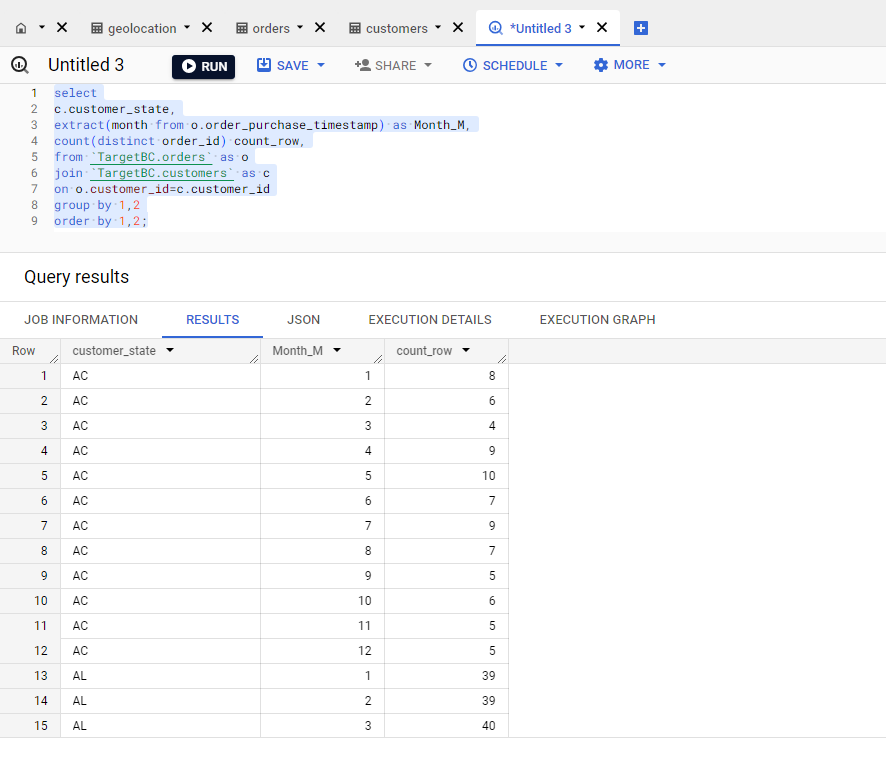
from `TargetBC.orders` as o

join `TargetBC.customers` as c

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

group by 1,2

order by 1,2;



**Insight/Recommendation :** Brazilline state Bahia (BA) got highest number of Month-on Month order compared to all state.

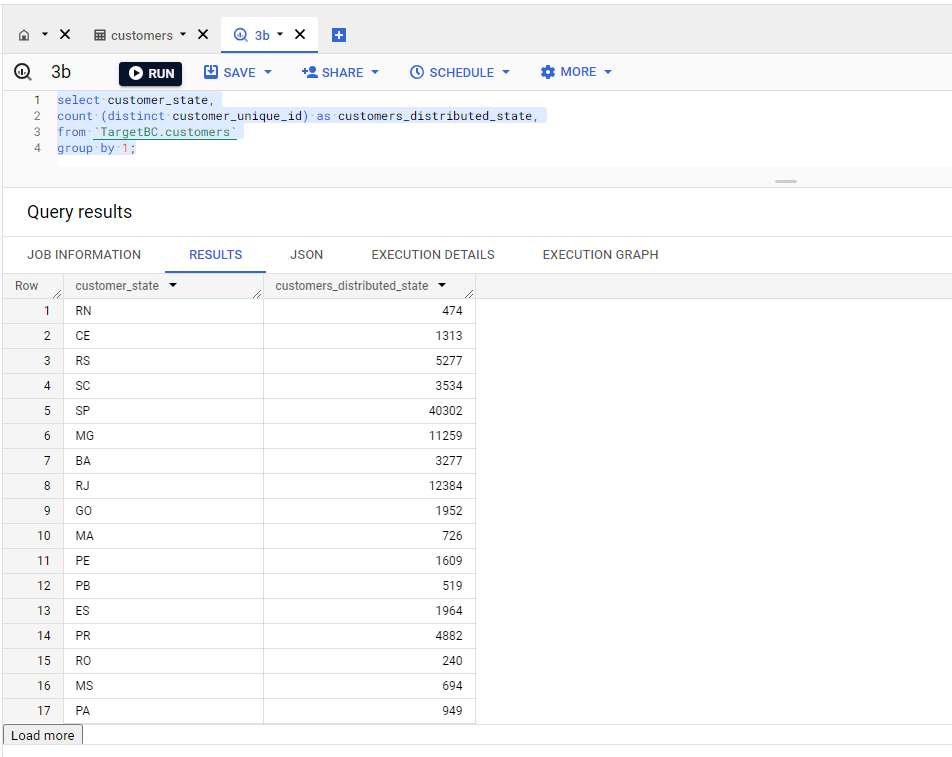
b. How are the customers distributed across all the states?

select customer\_state,

count (distinct customer\_unique\_id) as customers\_distributed\_state,

from `TargetBC.customers`

group by 1;



**Insight/Recommendation :** Customer distribution is more in North-east,south,South-east region state.

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

a. 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.

select x.Sum\_payment,

lag (x.Sum\_payment,1) over (partition by x.Year\_y order by x.Month\_m) as prev\_month,

x.Year\_y,x.Month\_m,

round((((x.Sum\_payment- lag (x.Sum\_payment,1) over (partition by x.Year\_y order by x.Month\_m))/lag (x.Sum\_payment,1) over (partition by x.Year\_y order by x.Month\_m))\*100),2) as Increase\_year,

from (

Select

sum(payment\_value) as Sum\_payment,

extract(year from shipping\_limit\_date) as Year\_y,

extract(month from shipping\_limit\_date) as Month\_m,

from `TargetBC.payments` as P

join `TargetBC.order\_items` O

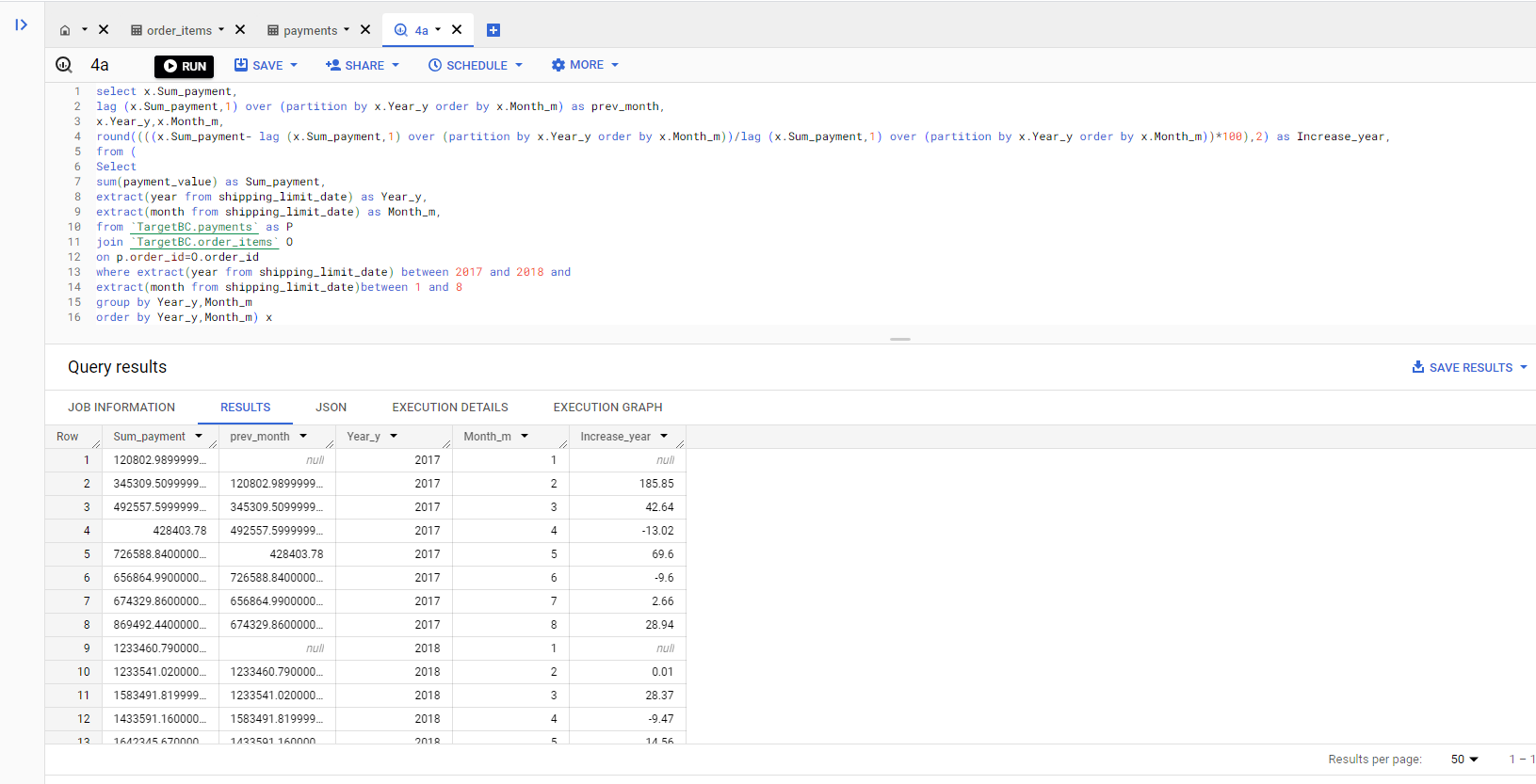
on p.order\_id=O.order\_id

where extract(year from shipping\_limit\_date) between 2017 and 2018 and

extract(month from shipping\_limit\_date)between 1 and 8

group by Year\_y,Month\_m

order by Year\_y,Month\_m) x



**Insight/Recommendation :**Growth inSales are not consistent (Jan to Aug) for period of 2017 and 2018

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

select C.customer\_state,p.payment\_value,

Sum(P.payment\_value) over (partition by C.customer\_state) as Total\_order\_price,

round(avg(P.payment\_value) over (partition by C.customer\_state),2) as Avg\_value,

from `TargetBC.payments` as P

join `TargetBC.orders` as O

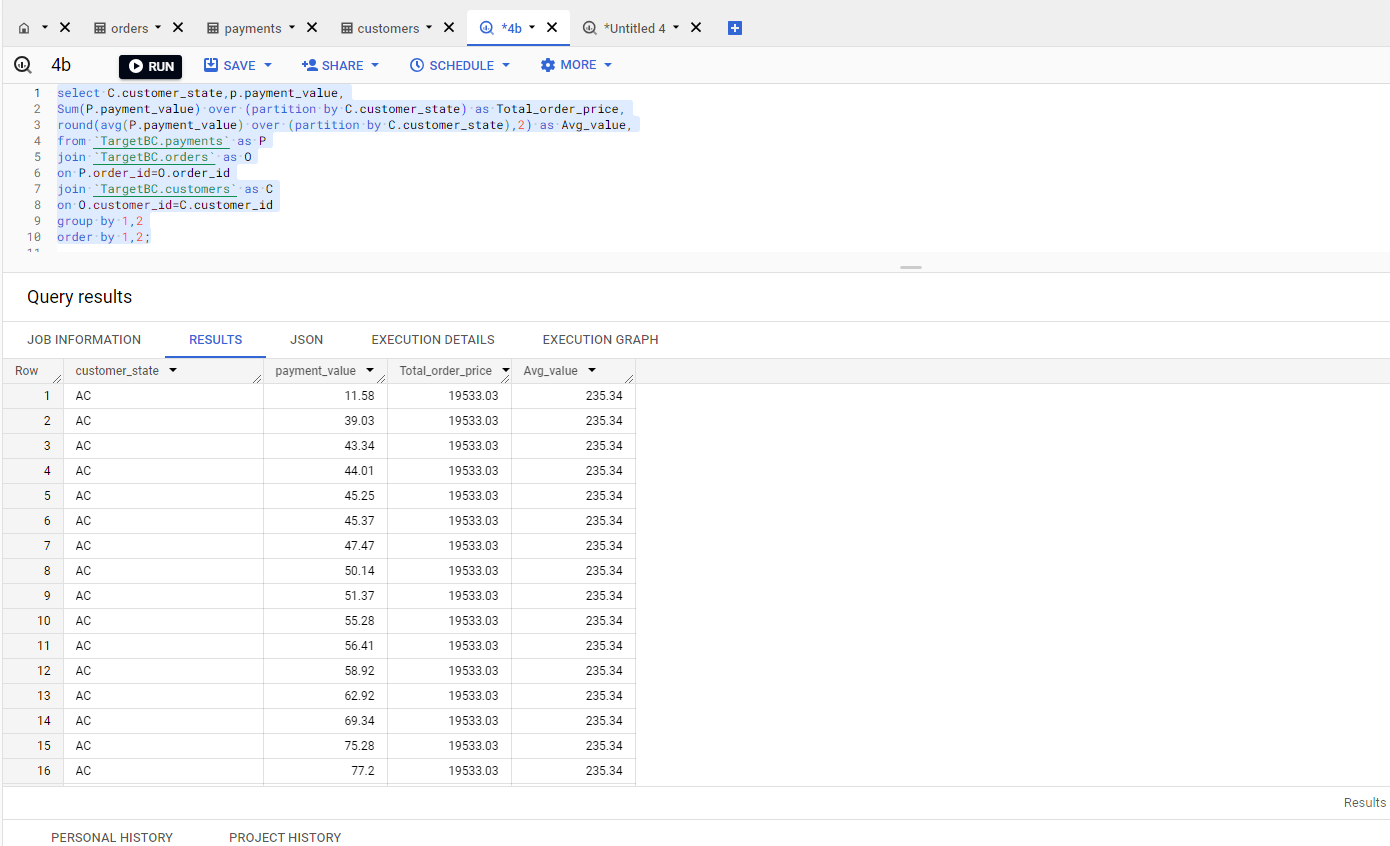
on P.order\_id=O.order\_id

join `TargetBC.customers` as C

on O.customer\_id=C.customer\_id

group by 1,2

order by 1,2;



**Insight/Recommendation :**  Average ticket size across all state is more than 200 but need to work to increase Average ticket size.

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

select C.customer\_state,OI.freight\_value,

sum(OI.freight\_value) over (partition by C.customer\_state) as Total\_freight,

avg(OI.freight\_value) over (partition by C.customer\_state) as Avg\_freight,

from `TargetBC.order\_items` as OI

join `TargetBC.orders` as O

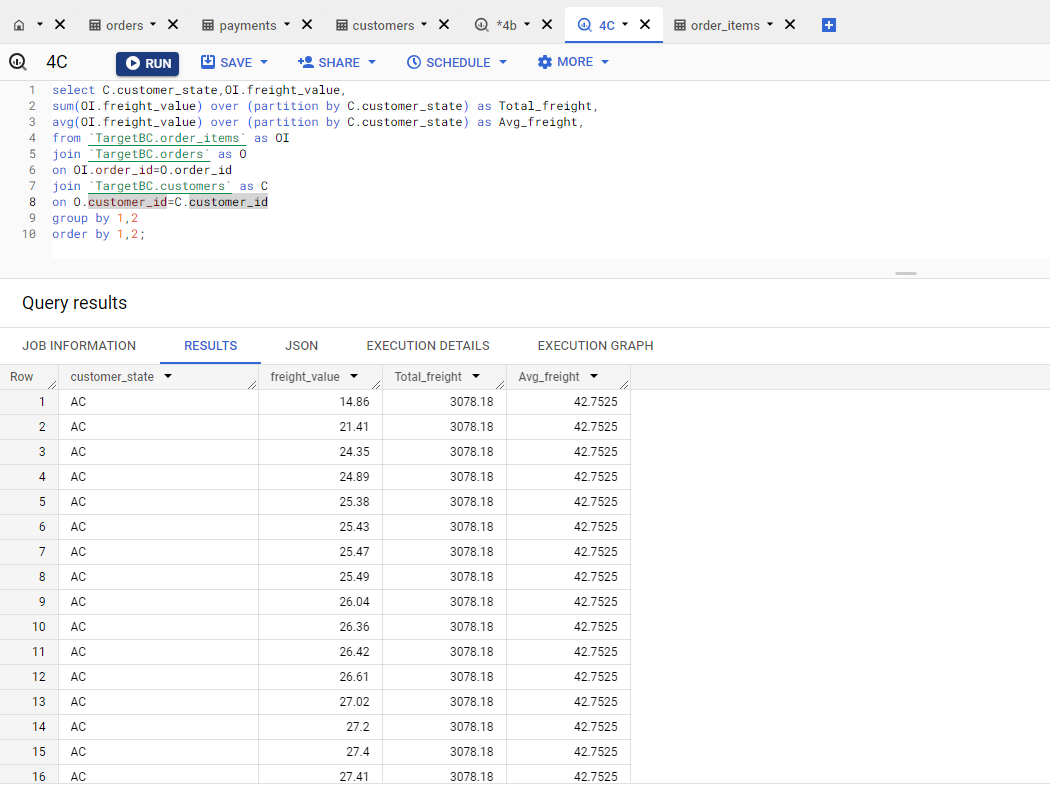
on OI.order\_id=O.order\_id

join `TargetBC.customers` as C

on O.customer\_id=C.customer\_id

group by 1,2

order by 1,2;



**Insight/Recommendation :** Total and Average freight in North state is quite high~ 40 compared to south region state ~29, Need to setup sub-warehouse to lower down freight costing.

**5. Analysis based on sales, freight and delivery time**.

a. 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

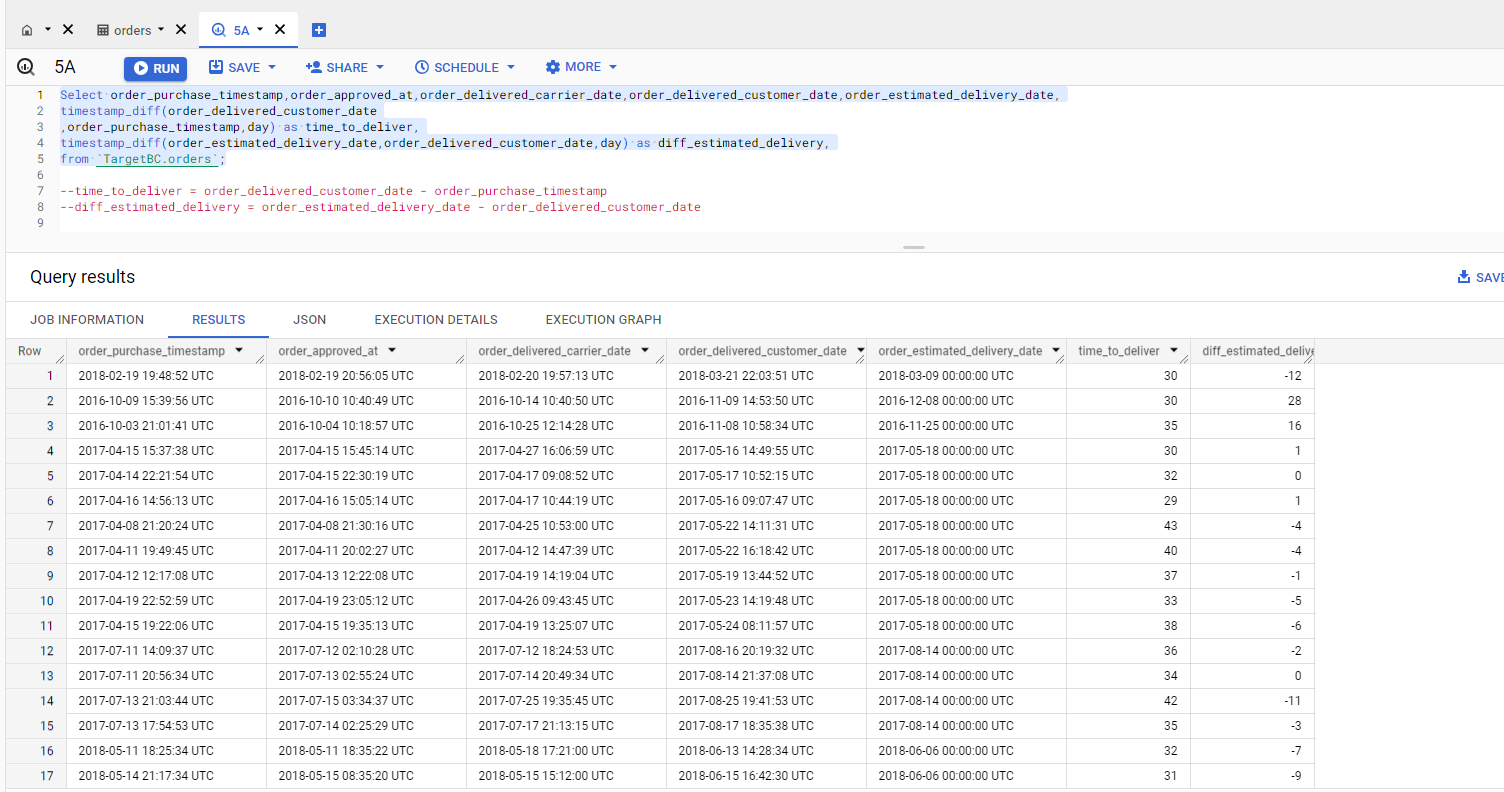
Select order\_purchase\_timestamp,order\_approved\_at,order\_delivered\_carrier\_date,order\_delivered\_customer\_date,order\_estimated\_delivery\_date,

timestamp\_diff(order\_delivered\_customer\_date

,order\_purchase\_timestamp,day) as time\_to\_deliver,

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

from `TargetBC.orders`;



**Insight/Recommendation :**  Time taken to deliver order is almost near to 30 day and every 2nd order is missing estimated delivery date.

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

Top 5 Highest states :

select C.customer\_state,

round (avg(OI.freight\_value),2) as average\_freight\_value\_asc,

from `TargetBC.order\_items` as OI

join `TargetBC.orders` as O

on OI.order\_id=O.order\_id

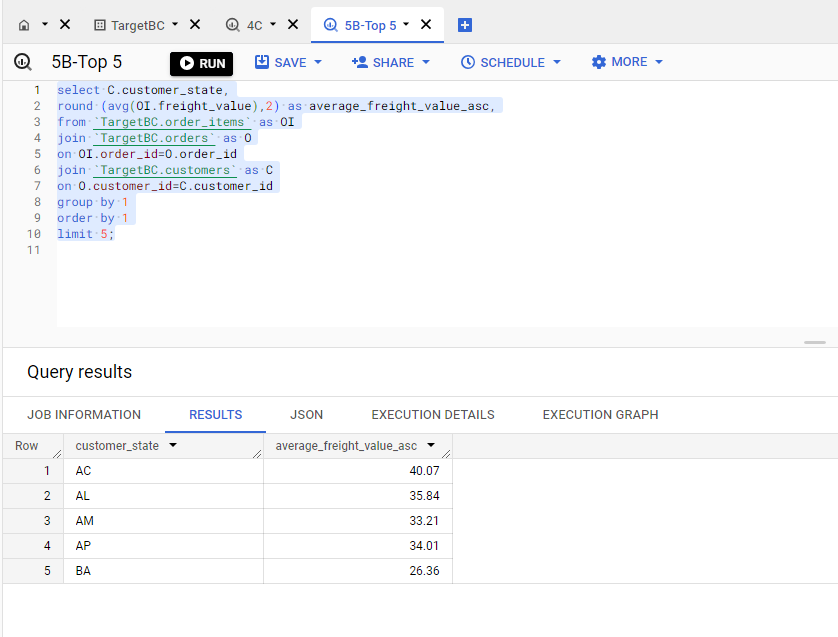
join `TargetBC.customers` as C

on O.customer\_id=C.customer\_id

group by 1

order by 1

limit 5;



**Insight/Recommendation :** All these 5 state are for North region of brazil where average freight is higher compared to south,south-east region.

Top 5 Lowest states :

select C.customer\_state,

round (avg(OI.freight\_value),2) as average\_freight\_value\_desc,

from `TargetBC.order\_items` as OI

join `TargetBC.orders` as O

on OI.order\_id=O.order\_id

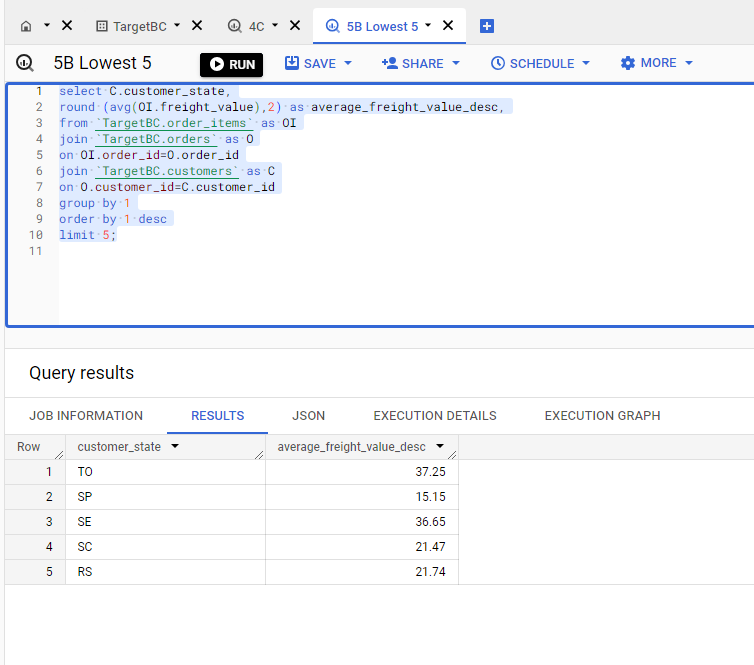
join `TargetBC.customers` as C

on O.customer\_id=C.customer\_id

group by 1

order by 1 desc

limit 5;



**Insight/Recommendation :** State from South,South-east region has lowest Avg\_freight\_value

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

Top 5 state Highest average delivery time :

Select C.customer\_state,

round (AVG (timestamp\_diff(order\_delivered\_customer\_date

,order\_purchase\_timestamp,day)),2) as Avg\_time\_to\_deliver

from `TargetBC.customers` C

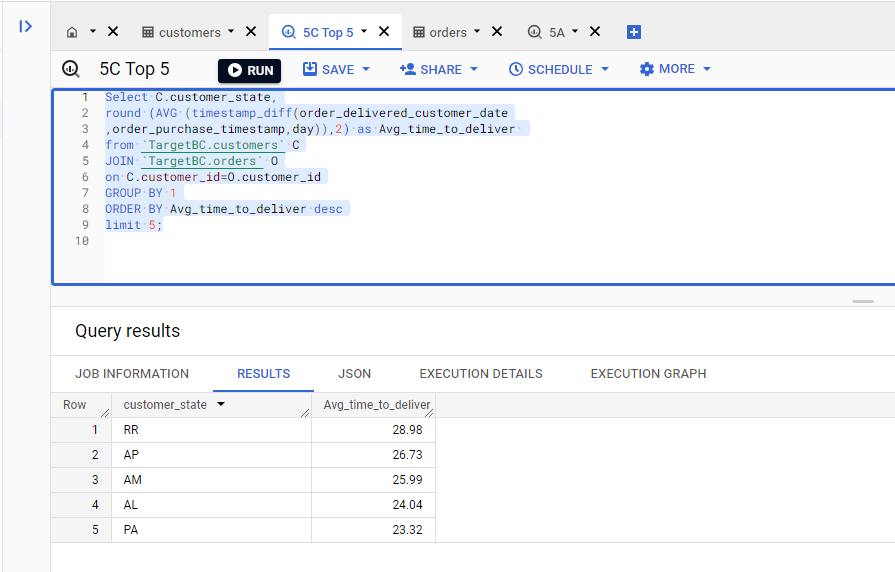
JOIN `TargetBC.orders` O

on C.customer\_id=O.customer\_id

GROUP BY 1

ORDER BY Avg\_time\_to\_deliver desc

limit 5;



**Insight/Recommendation :** Delivery time for North state is quite high compared to south,southest state.

Top 5 state lowest average delivery time:

Select C.customer\_state,

round (AVG (timestamp\_diff(order\_delivered\_customer\_date

,order\_purchase\_timestamp,day)),2) as Avg\_time\_to\_deliver

from `TargetBC.customers` C

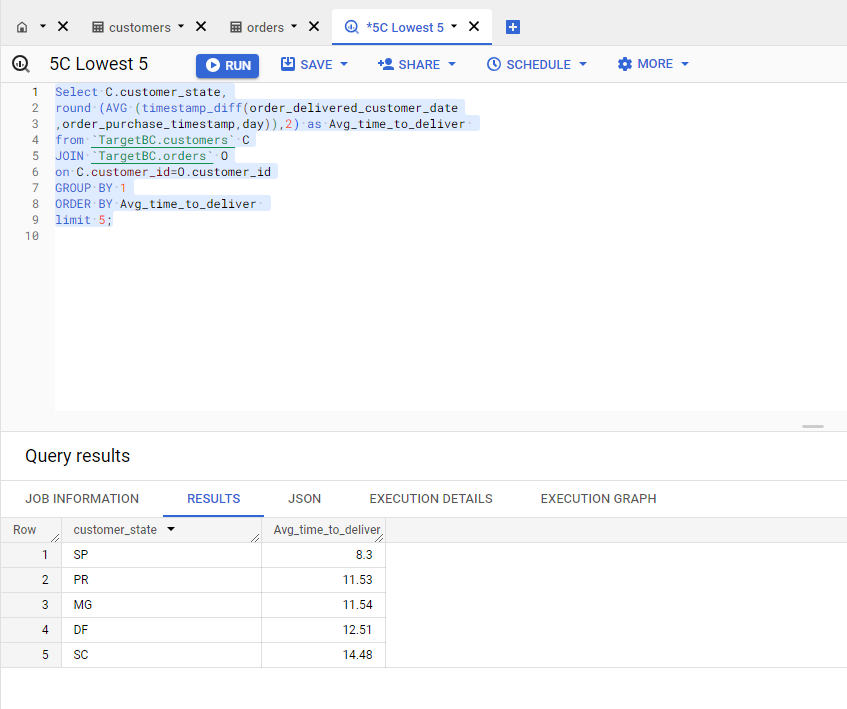
JOIN `TargetBC.orders` O

on C.customer\_id=O.customer\_id

GROUP BY 1

ORDER BY Avg\_time\_to\_deliver

limit 5;



**Insight/Recommendation :** Delivery time for to south,southest state are good compared to north state.

d. 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

Select C.customer\_state,

round (avg(timestamp\_diff(O.order\_delivered\_customer\_date

,O.order\_estimated\_delivery\_date

,day)),2) as delivery\_difference\_day

from `TargetBC.customers` C

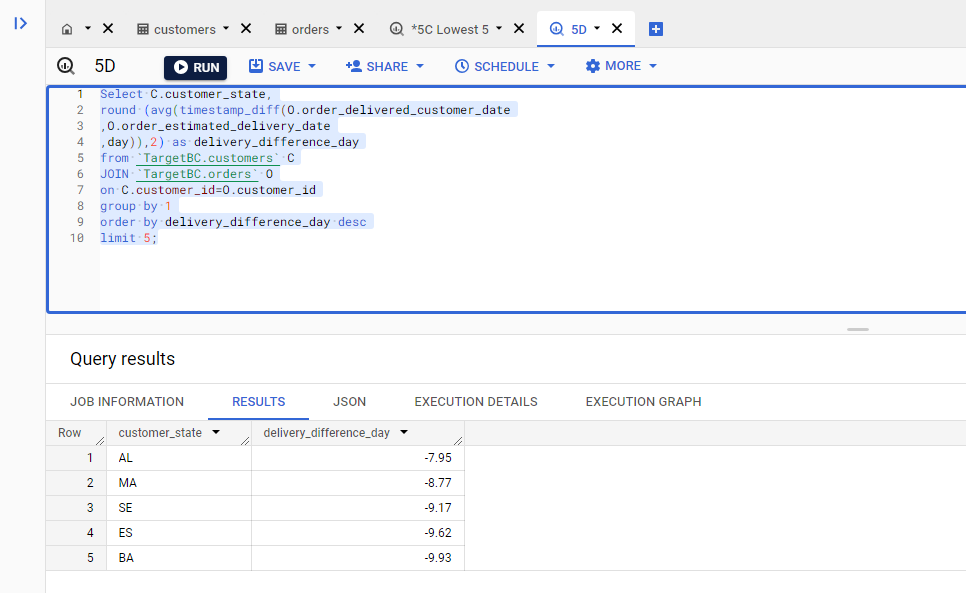
JOIN `TargetBC.orders` O

on C.customer\_id=O.customer\_id

group by 1

order by delivery\_difference\_day desc

limit 5;



6. **Analysis based on the payments:**

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

Select C.customer\_state,P.payment\_type,O.order\_purchase\_timestamp,

extract (month from O.order\_purchase\_timestamp) as Month\_,

extract (year from O.order\_purchase\_timestamp) as year\_,

count(P.payment\_type) over (partition by P.payment\_type order by C.customer\_state) as Number\_payment\_type

from `TargetBC.payments` as P

join `TargetBC.orders` as O

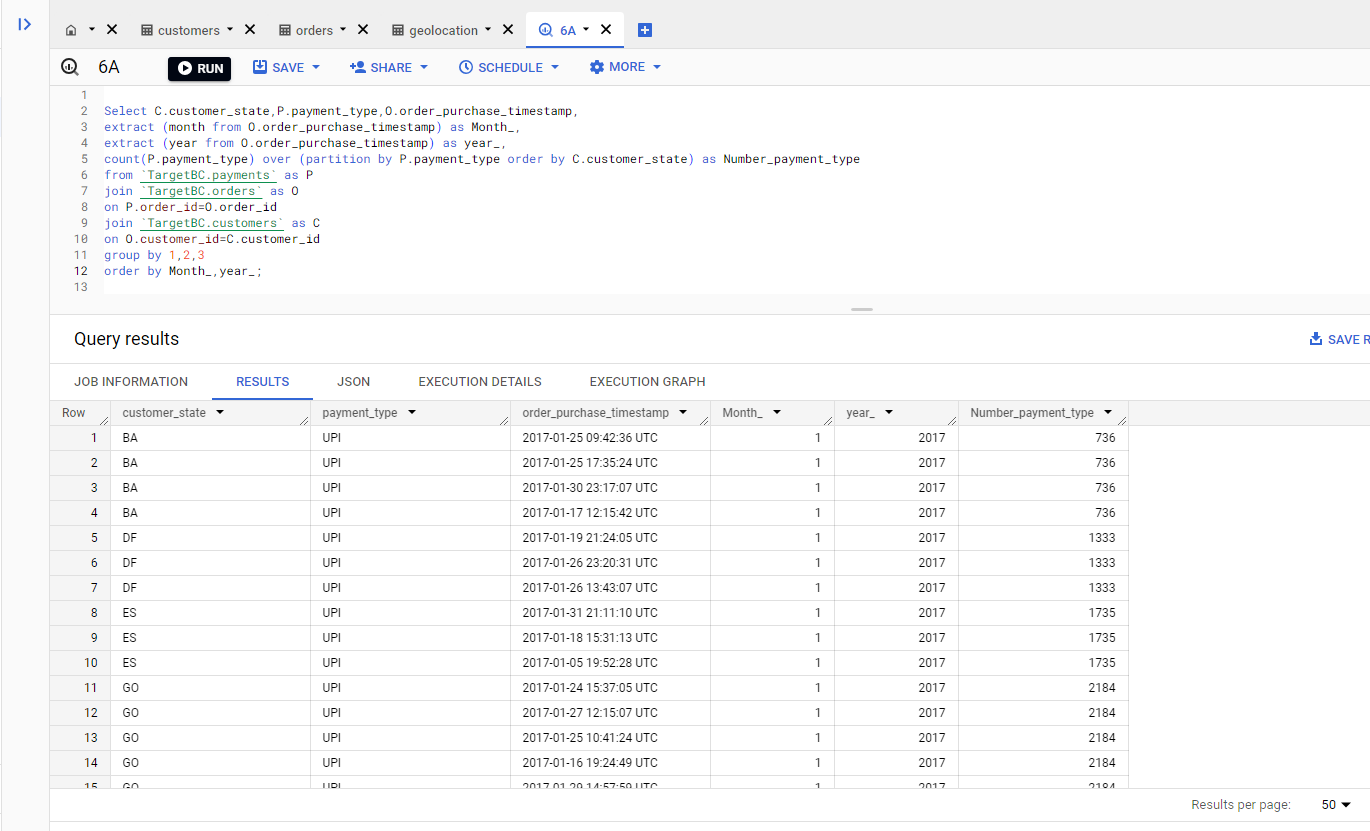
on P.order\_id=O.order\_id

join `TargetBC.customers` as C

on O.customer\_id=C.customer\_id

group by 1,2,3

order by Month\_,year\_;



**Insight/Recommendation :** Customer mostly user credit card payment mode compared to debit,UPI this trend is shown across all state.

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

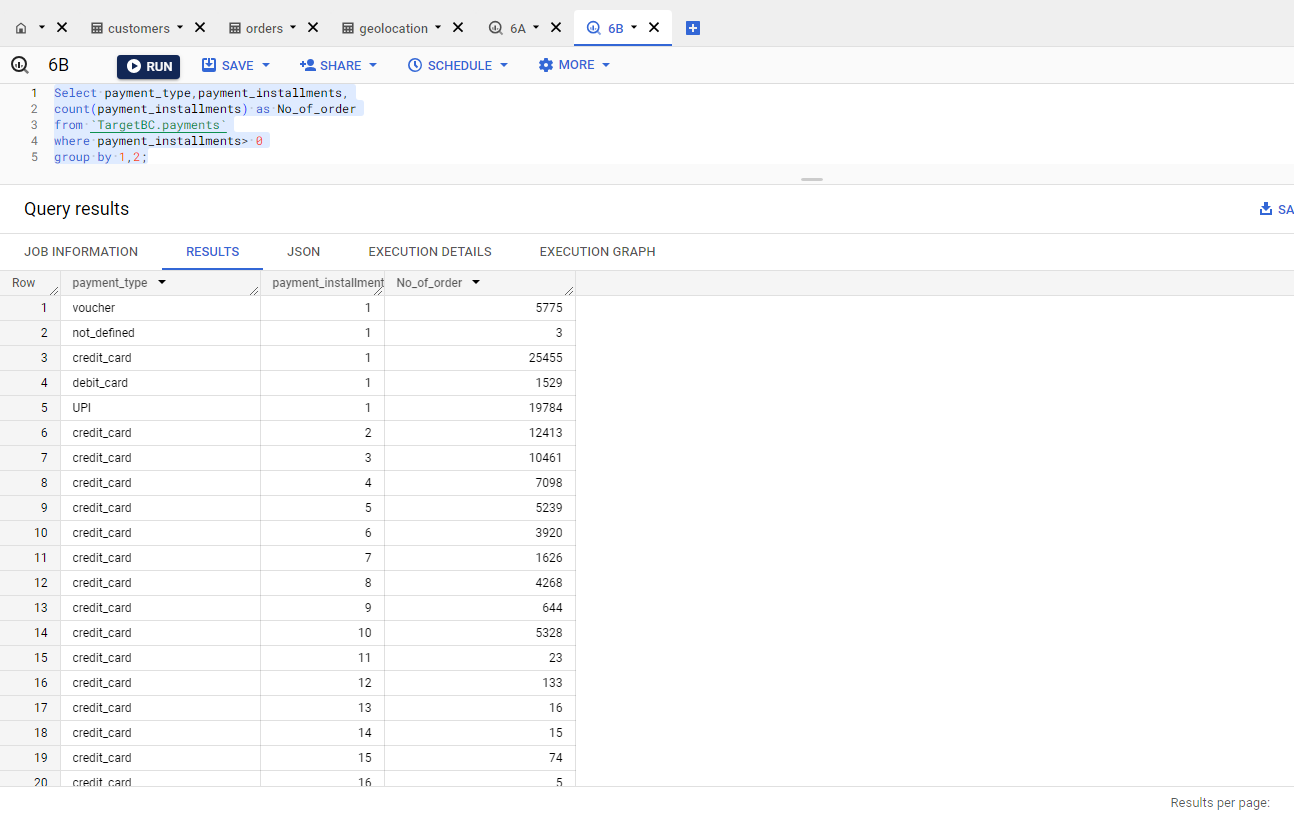
Select payment\_type,payment\_installments,

count(payment\_installments) as No\_of\_order

from `TargetBC.payments`

where payment\_installments> 0

group by 1,2;



**Insight/Recommendation :** Customer mostly take short period installment form credit cards.