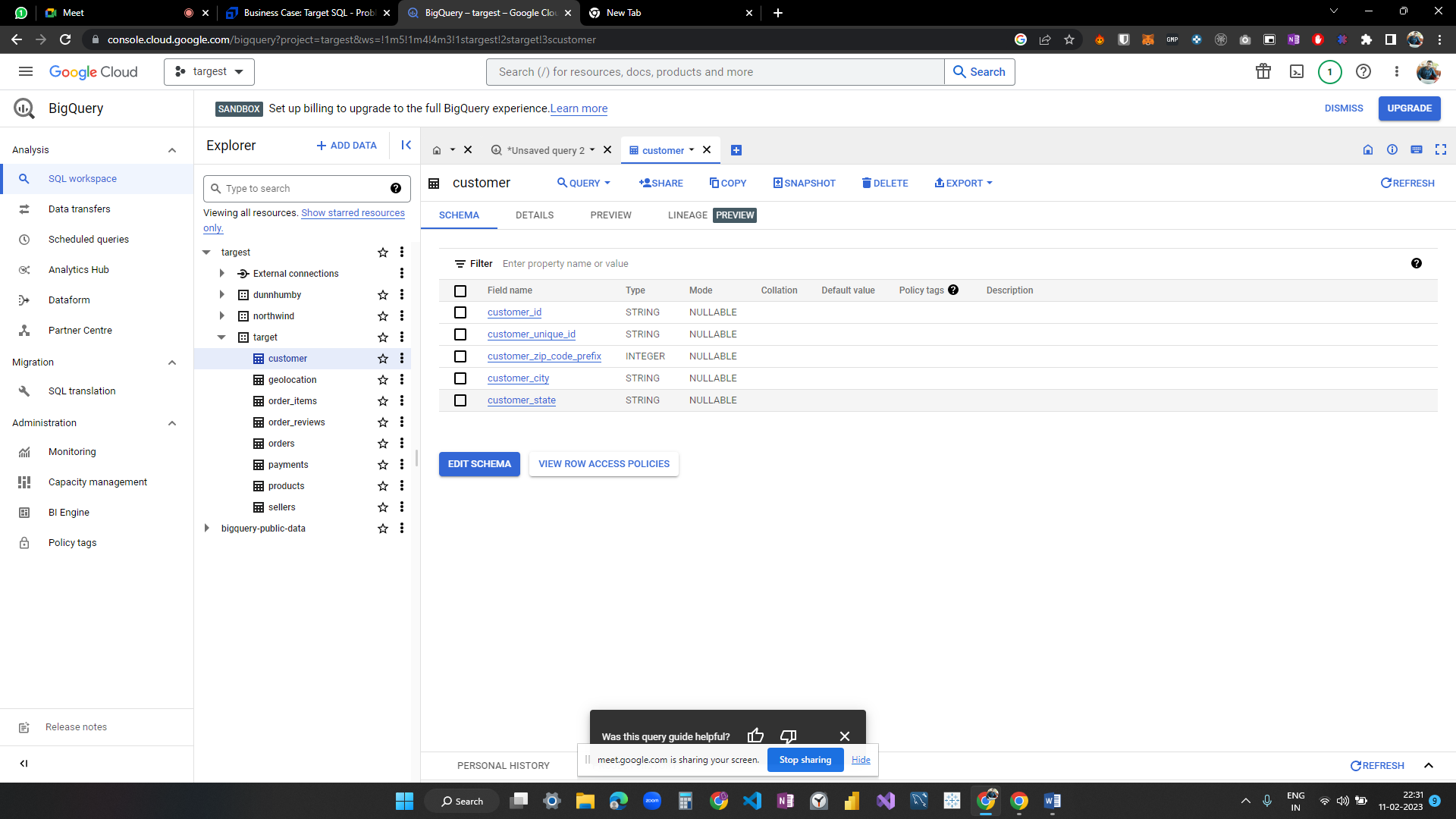
Data type informations

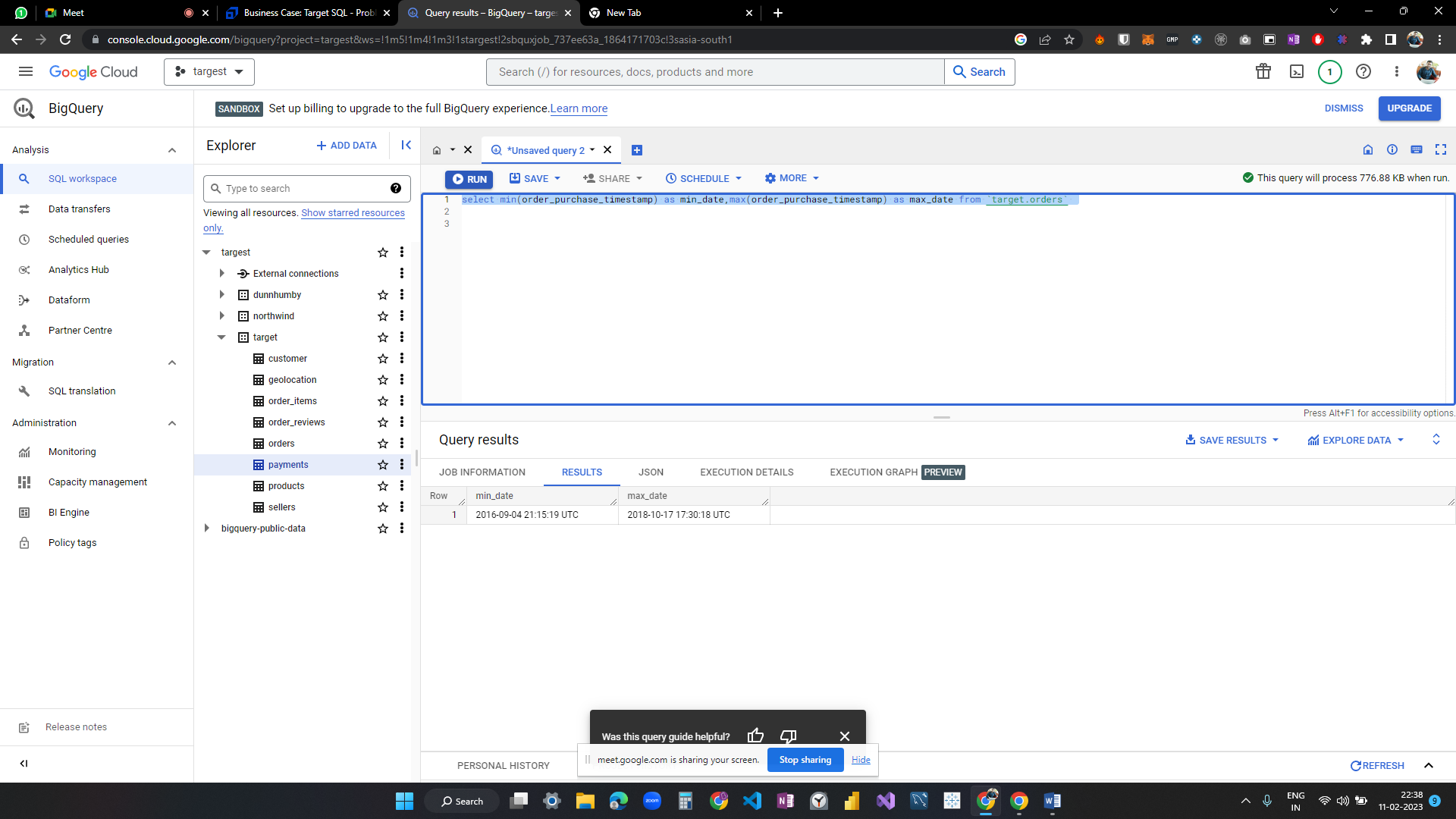


Time period for which the data is given

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

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

from `target.orders`

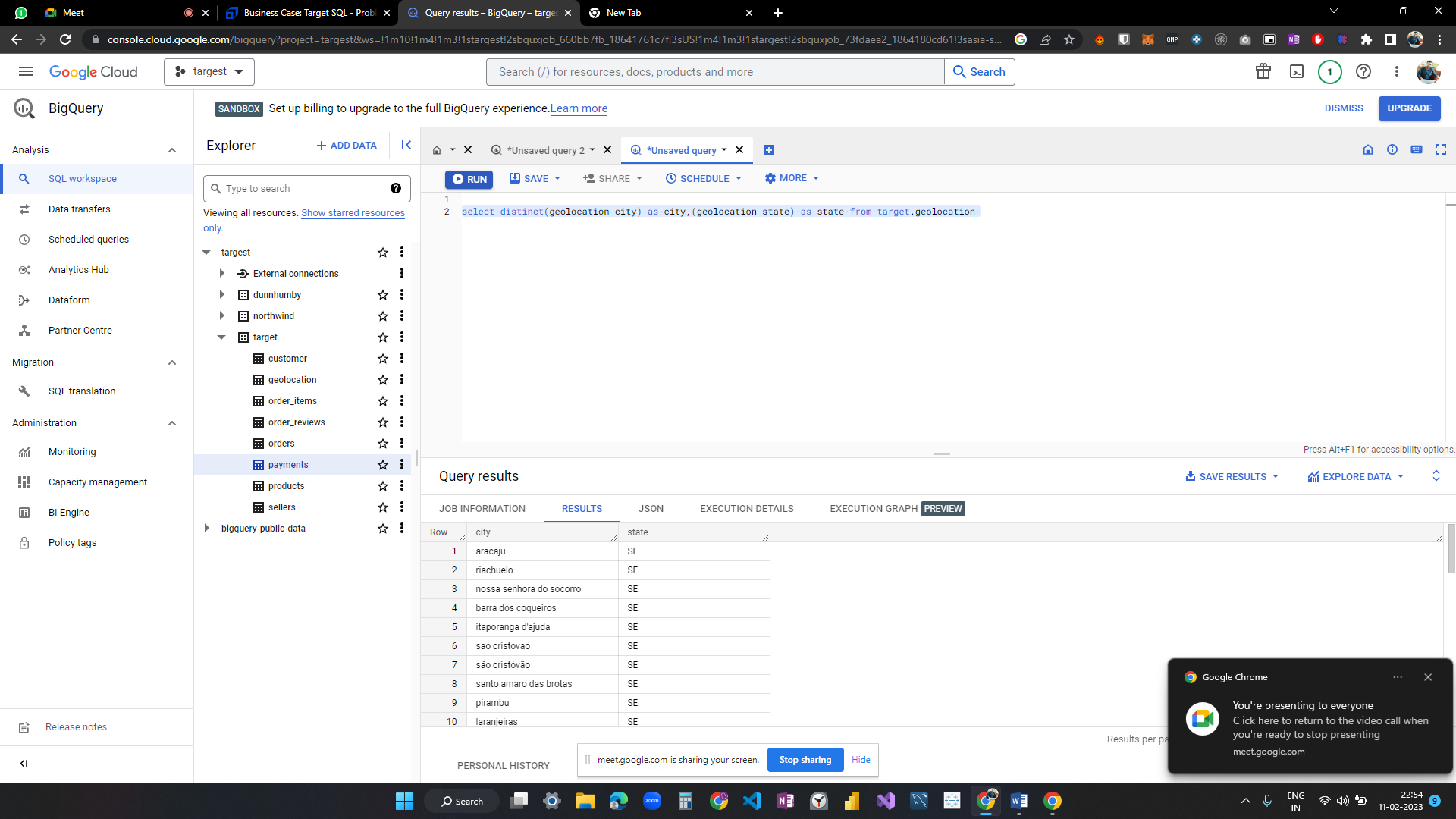


Cities and States of customers ordered during the given period

select distinct(geolocation\_city) as city,

(geolocation\_state) as state

from target.geolocation



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?

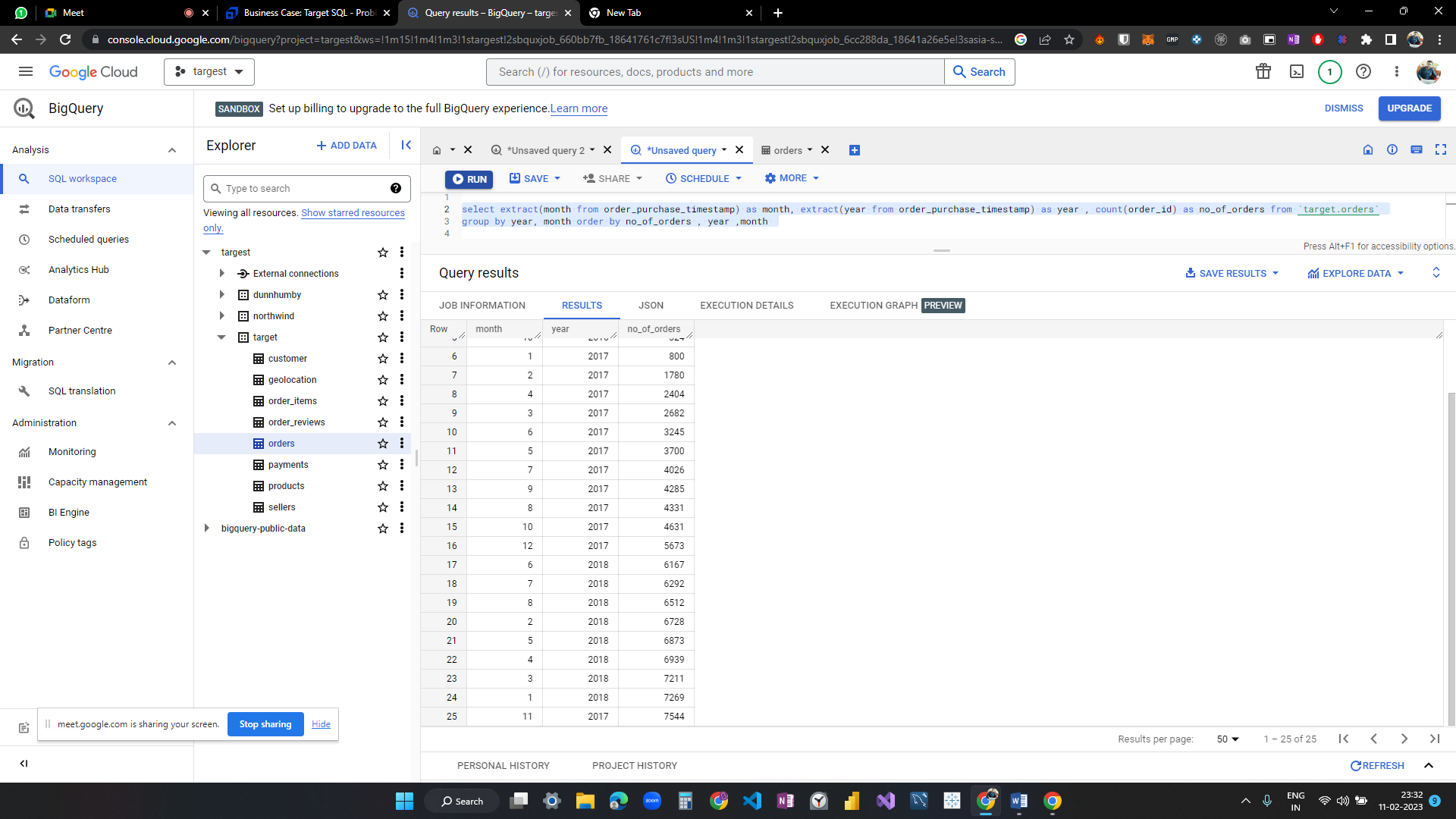
select extract(month from order\_purchase\_timestamp) as month,

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

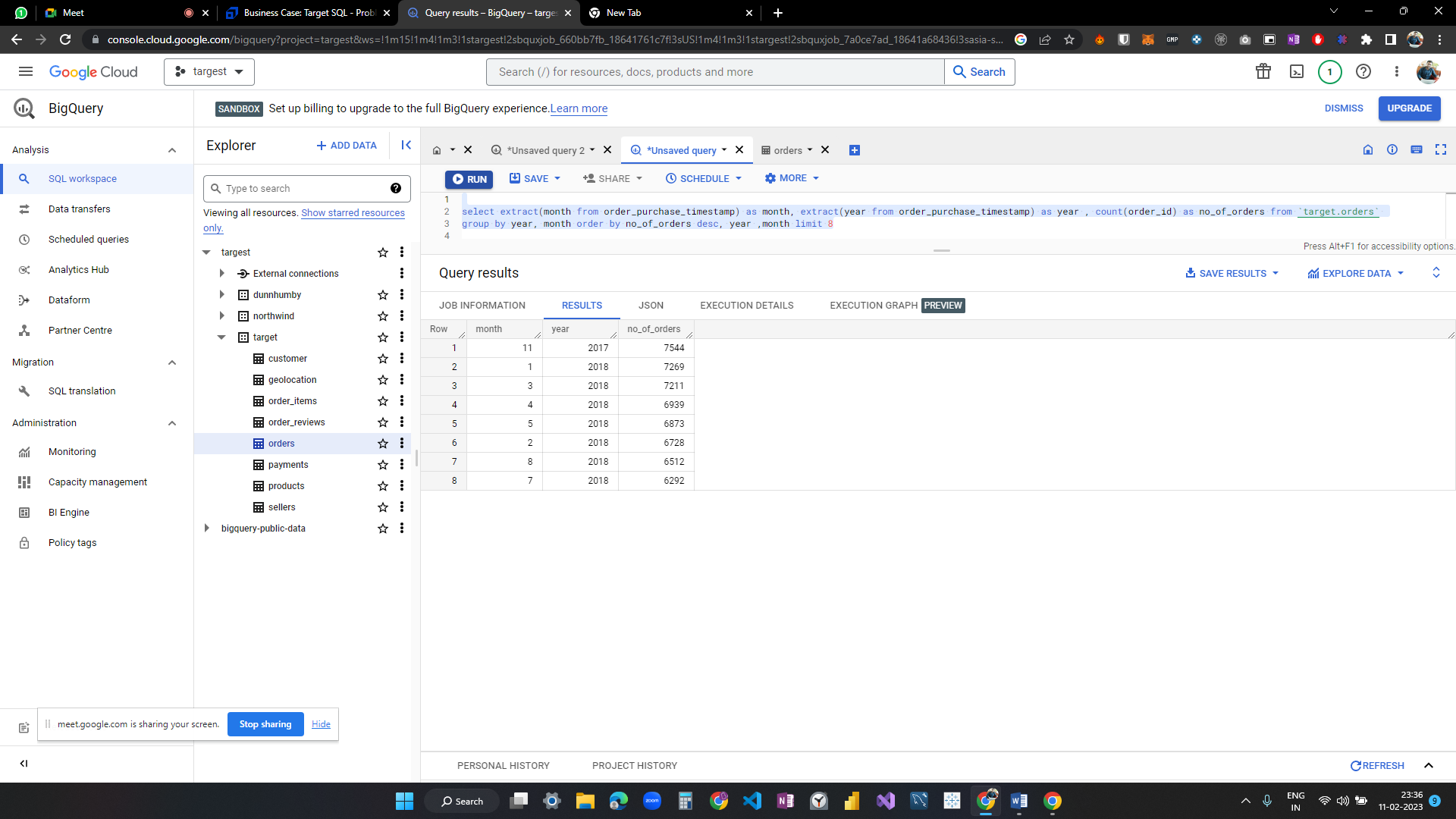
count(order\_id) as no\_of\_orders

from `target.orders` group by year, month

order by no\_of\_orders , year ,month



As we can observe no of orders are increasing almost on monthly basis with some slight dip oin no of orders but overall it is increasing



As we cannot observe any outlier in any month, we fail to accept that there is any peak sale due to season.

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

select case when extract(hour from order\_purchase\_timestamp) between 4 and 5 then "dawn"

when extract(hour from order\_purchase\_timestamp) between 5 and 12 then "morning"

when extract(hour from order\_purchase\_timestamp) between 12 and 18 then "afternoon"

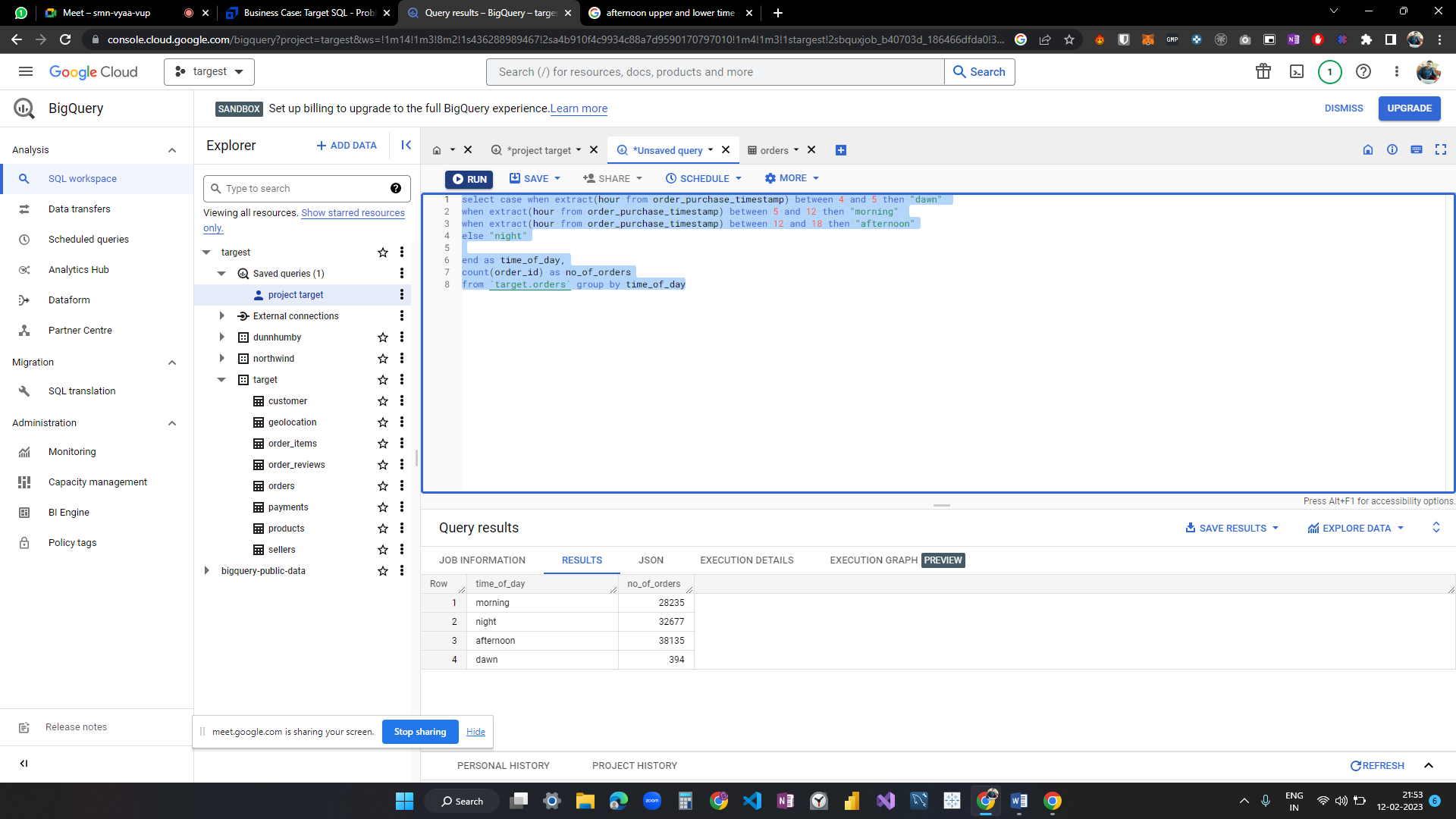
else "night"

end as time\_of\_day,

count(order\_id) as no\_of\_orders

from `target.orders`

group by time\_of\_day



Most brazillian do shopping in afternoon i.e from 12 pm to 6 pm so if we want to run some offer that will be perfect time and we can consider this time as feasible for reaching out to customers

Get month on month orders by states

select  c.customer\_state,

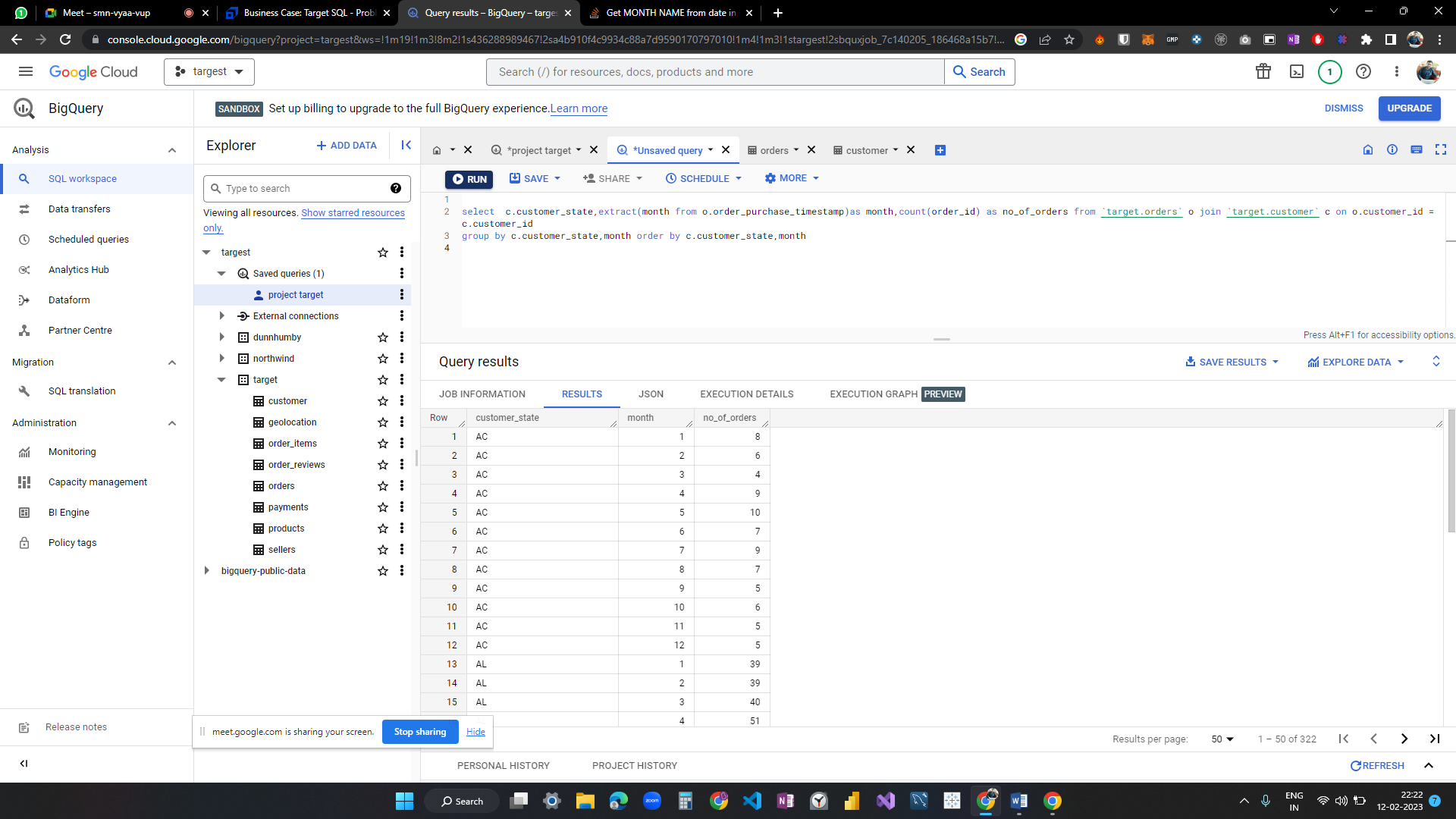
extract(month from o.order\_purchase\_timestamp)as month,

count(order\_id) as no\_of\_orders

from `target.orders` o join `target.customer` c on o.customer\_id = c.customer\_id

group by c.customer\_state,month

order by c.customer\_state,month



Trying month name

select  c.customer\_state,

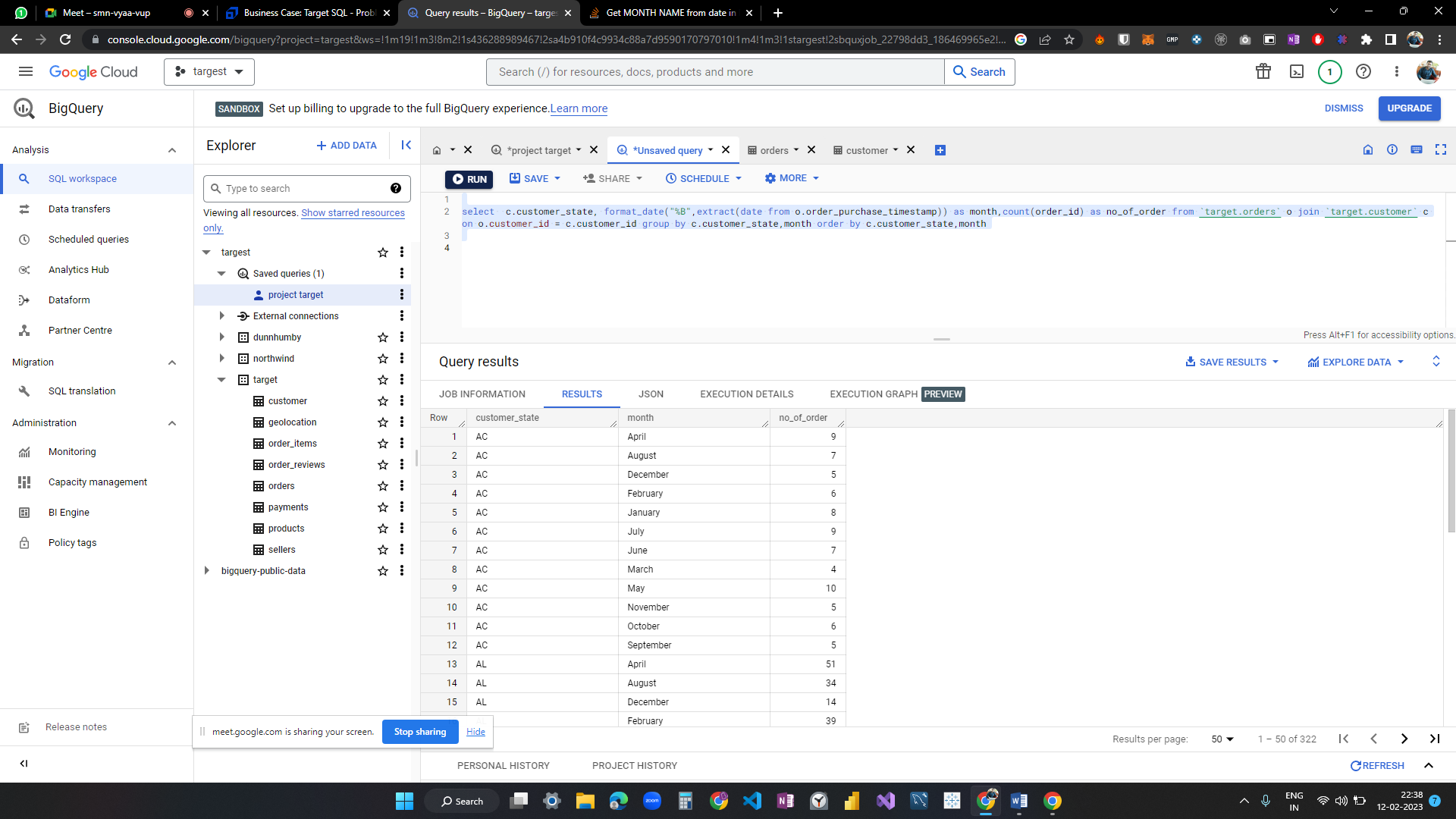
 format\_date("%B",extract(date from o.order\_purchase\_timestamp)) as month,

count(order\_id) as no\_of\_order from `target.orders` o

join `target.customer` c on o.customer\_id = c.customer\_id

group by c.customer\_state,month

order by c.customer\_state,month



Distribution of customers across the states in Brazil

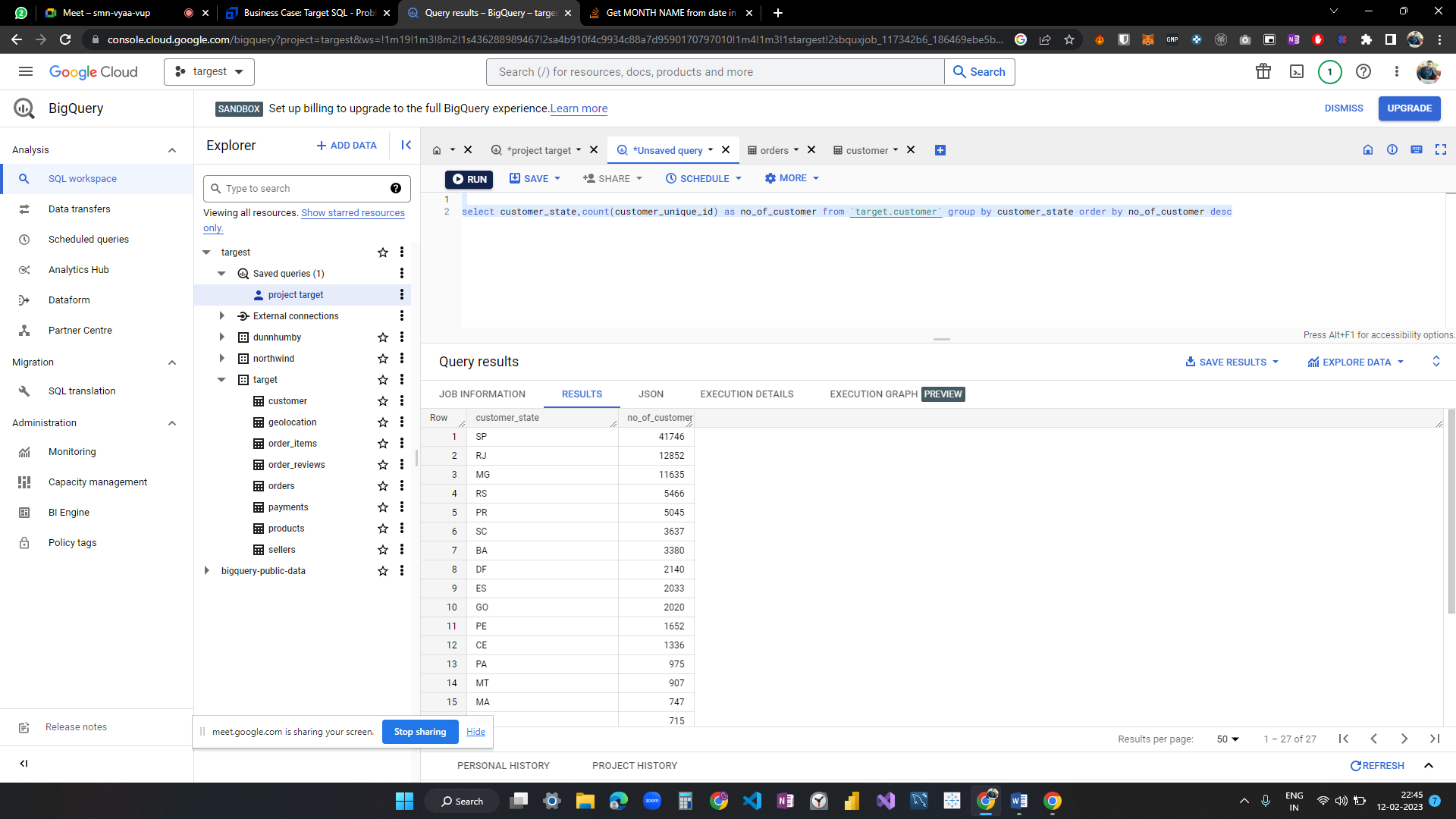
select customer\_state,

count(customer\_unique\_id) as no\_of\_customer

from `target.customer`

group by customer\_state

order by no\_of\_customer desc



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

with base as(

select extract(year from

order\_purchase\_timestamp ) as year,

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

p.payment\_value total from `target.payments` p join `target.orders` o on p.order\_id = o.order\_id

)

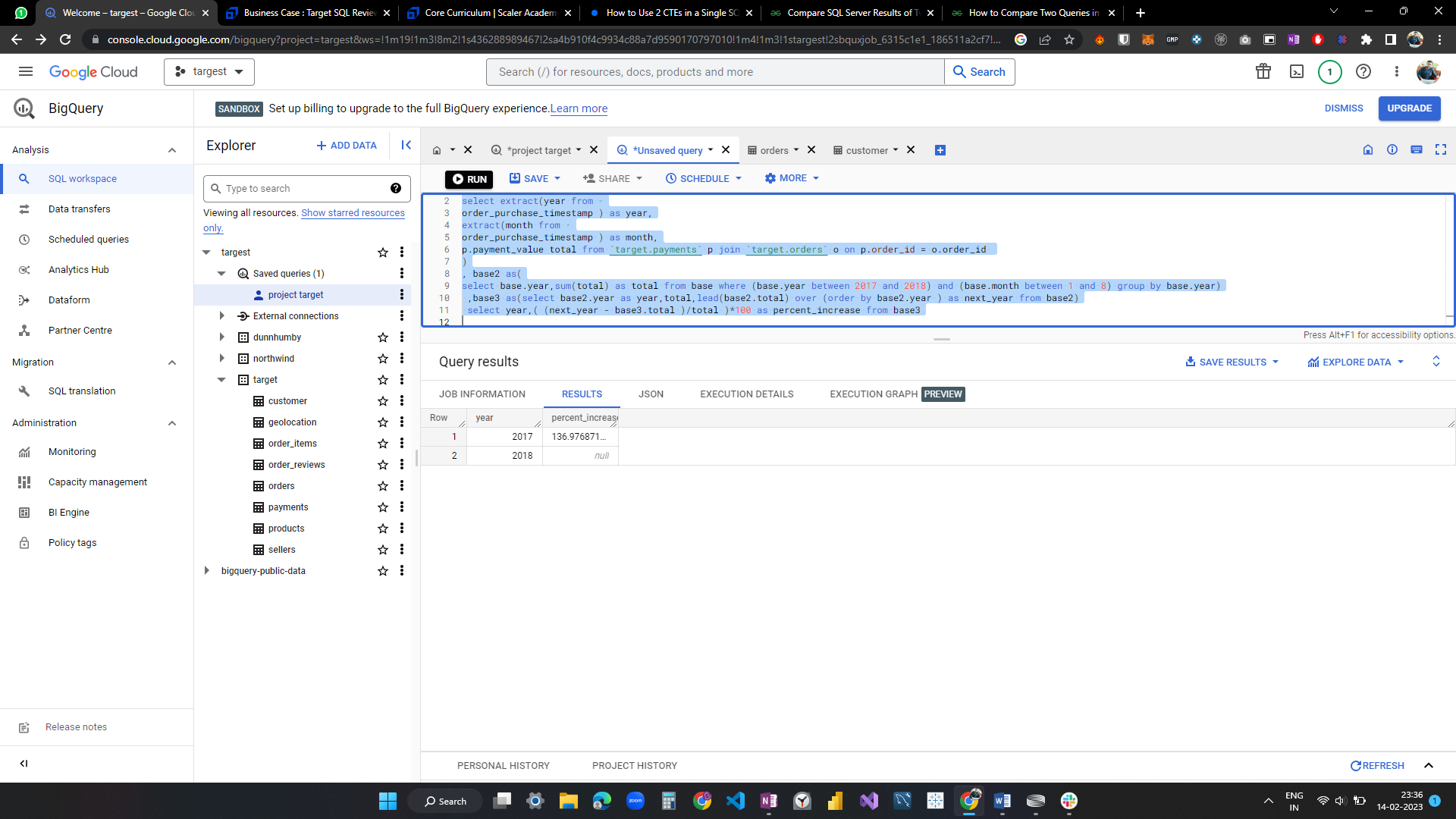
, base2 as(

select base.year,sum(total) as total from base where (base.year between 2017 and 2018) and (base.month between 1 and 8) group by base.year)

,base3 as(

select base2.year as year,total,lead(base2.total) over (order by base2.year ) as next\_year from base2)

 select year,( (next\_year - base3.total )/total )\*100 as percent\_increase from base3



Mean & Sum of price and freight value by customer state

select customer\_state,

avg(price) as avg\_price,

avg(freight\_value) as avg\_freight\_value,

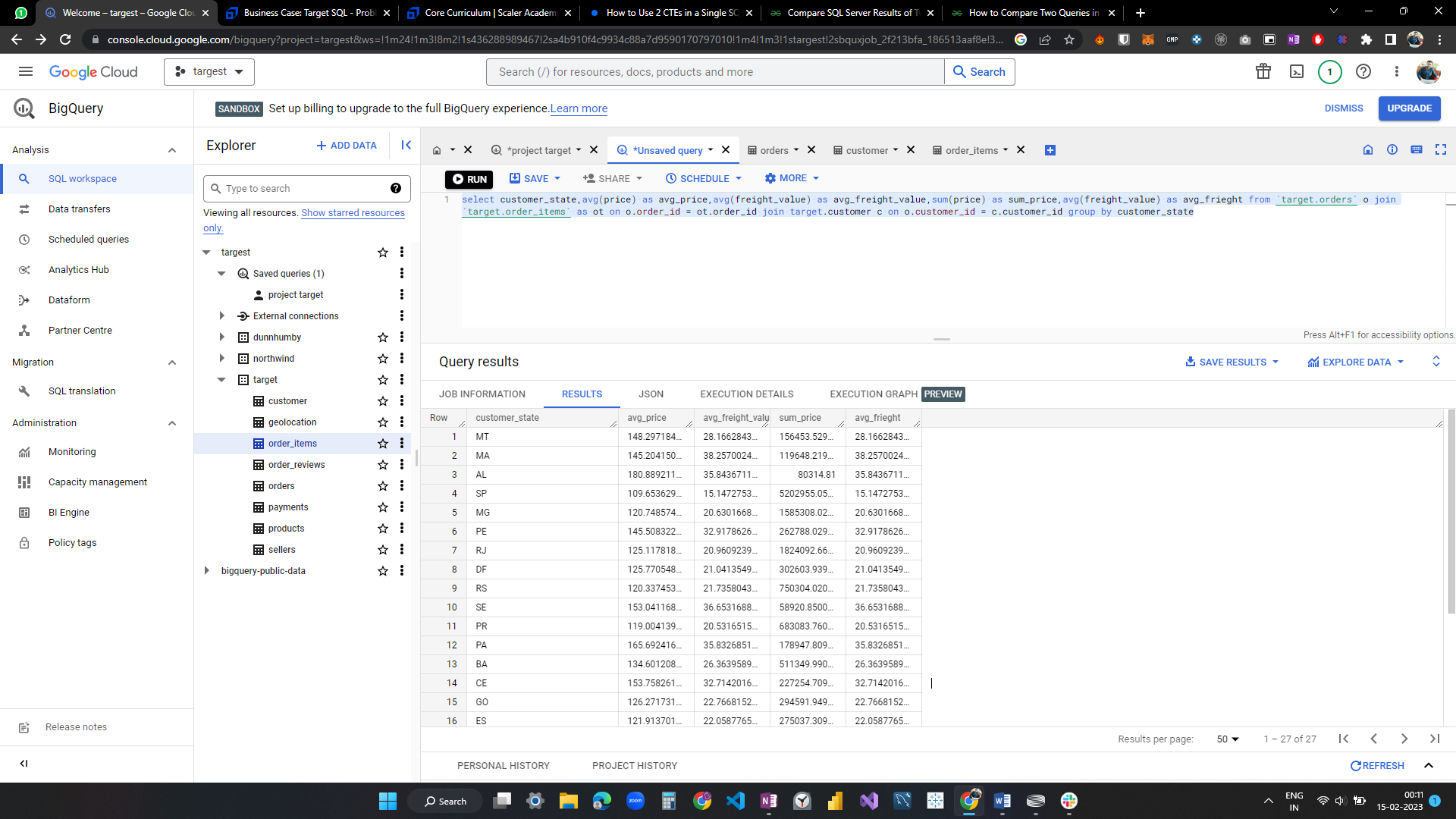
sum(price) as sum\_price,

avg(freight\_value) as avg\_frieght

from `target.orders` o join `target.order\_items` as ot on o.order\_id = ot.order\_id

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

group by customer\_state



Calculate days between purchasing, delivering and estimated delivery

select order\_id,

date\_diff(order\_estimated\_delivery\_date, order\_purchase\_timestamp, day)

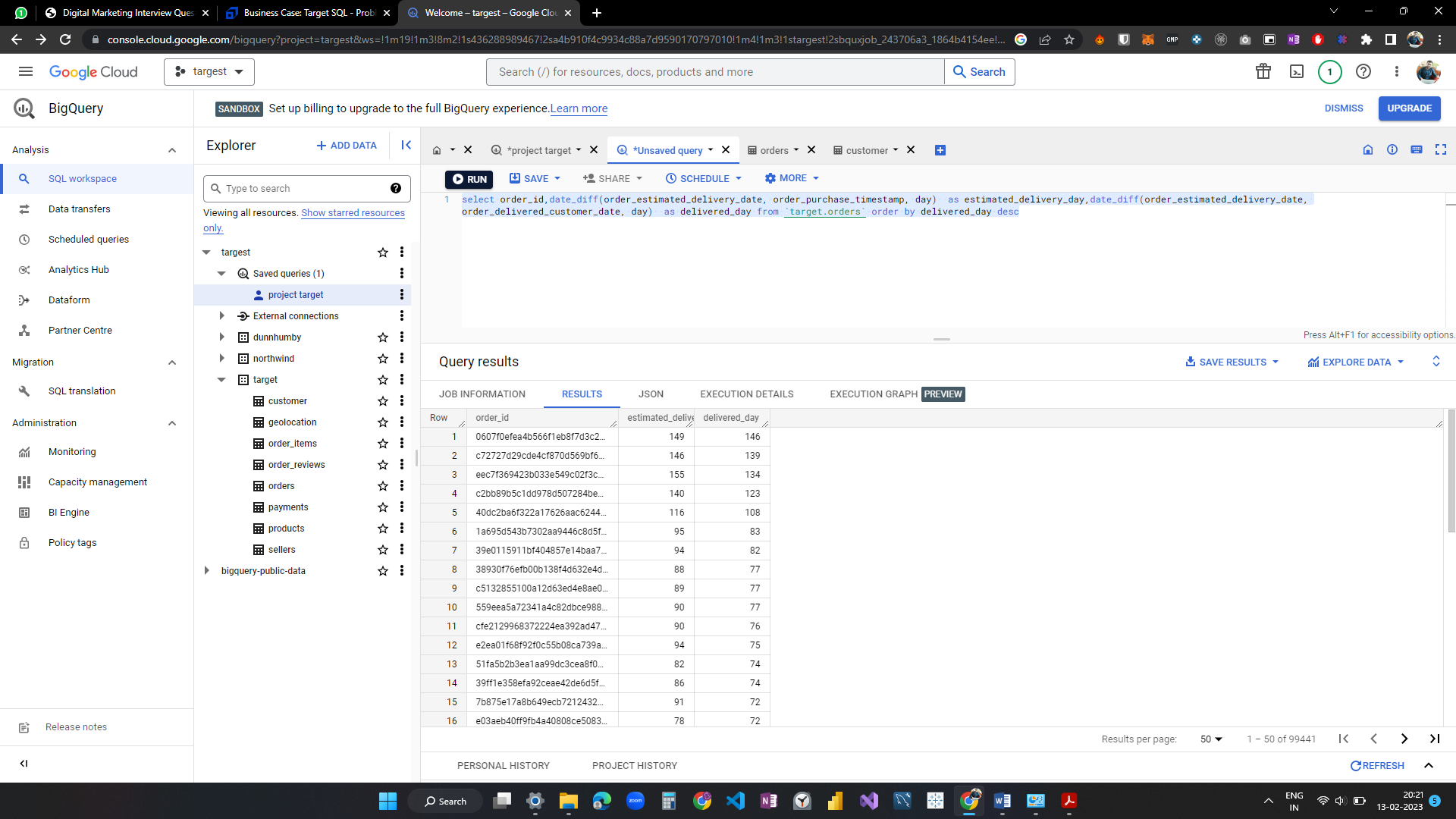
as estimated\_delivery\_day,

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

as delivered\_day

from `target.orders`

order by delivered\_day desc



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 order\_id,

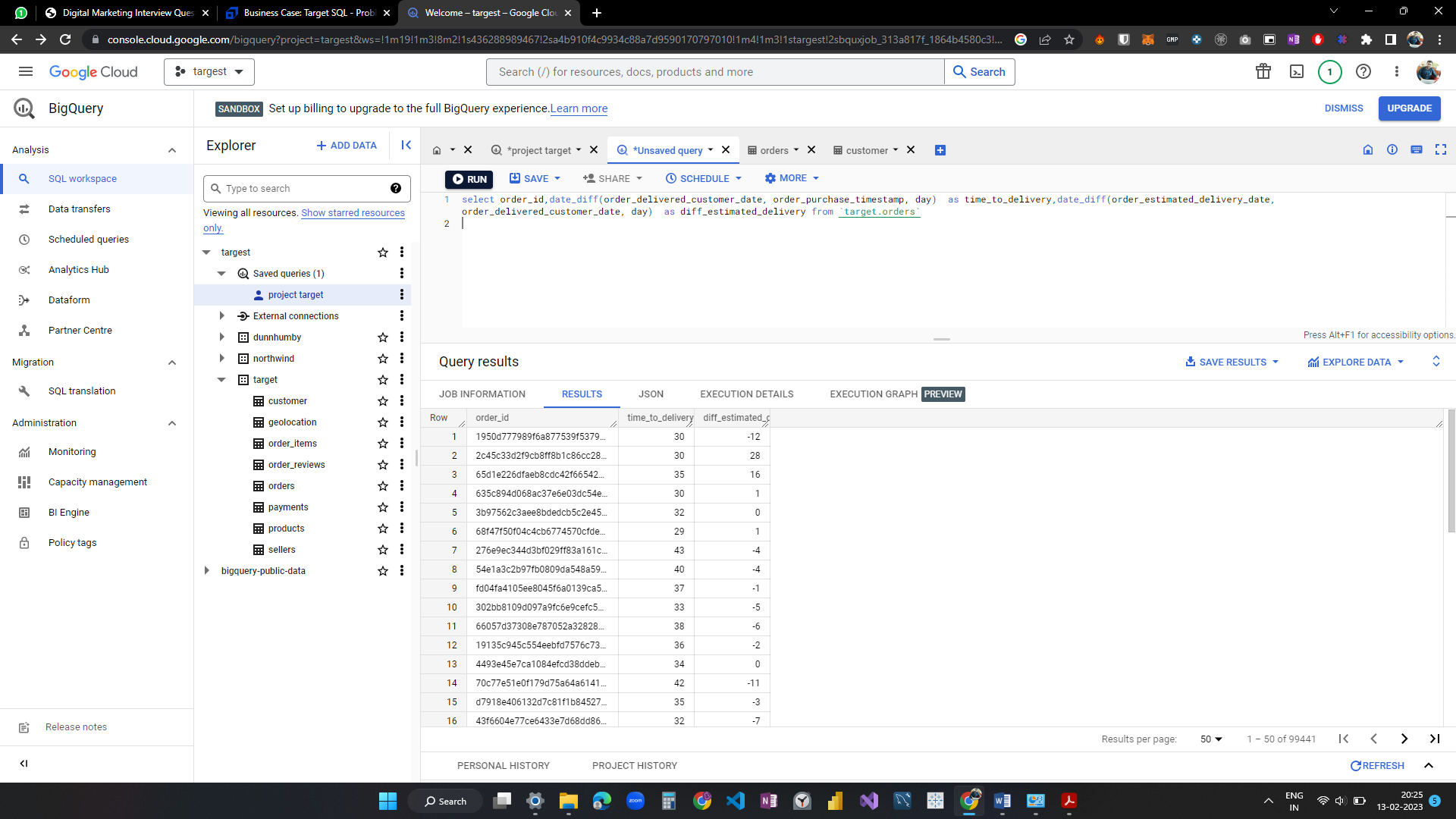
date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day)

 as time\_to\_delivery,

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

as diff\_estimated\_delivery

from `target.orders`



Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

select c.customer\_state,

avg(ot.freight\_value) as freight\_value,

avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day))

as time\_to\_delivery ,

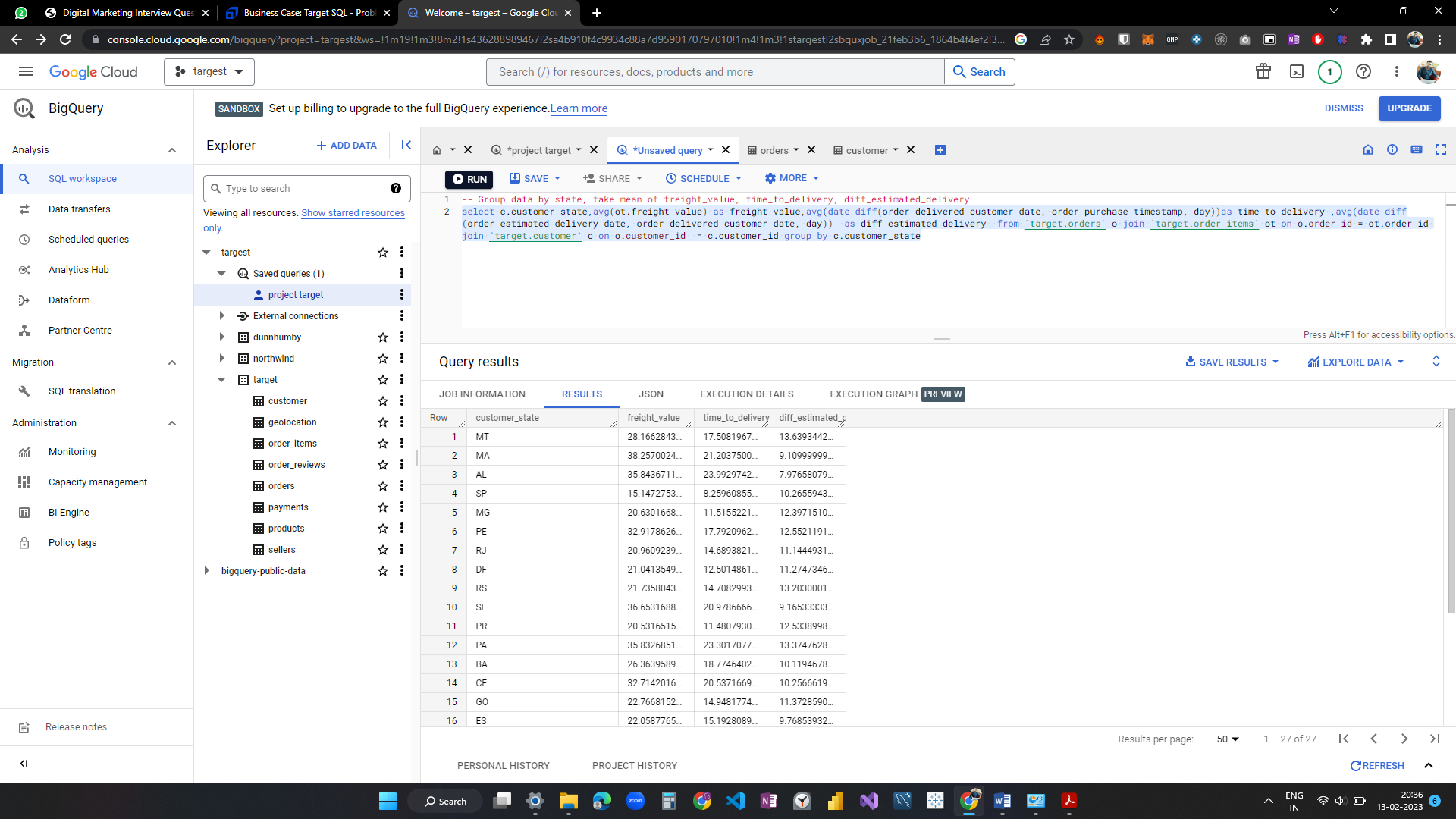
avg(date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day))

as diff\_estimated\_delivery

from `target.orders` o join `target.order\_items` ot on o.order\_id = ot.order\_id

join `target.customer` c on o.customer\_id  = c.customer\_id

group by c.customer\_state



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

select c.customer\_state,

avg(ot.freight\_value) as freight\_value,

avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day))

as time\_to\_delivery ,

avg(date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day))

as diff\_estimated\_delivery

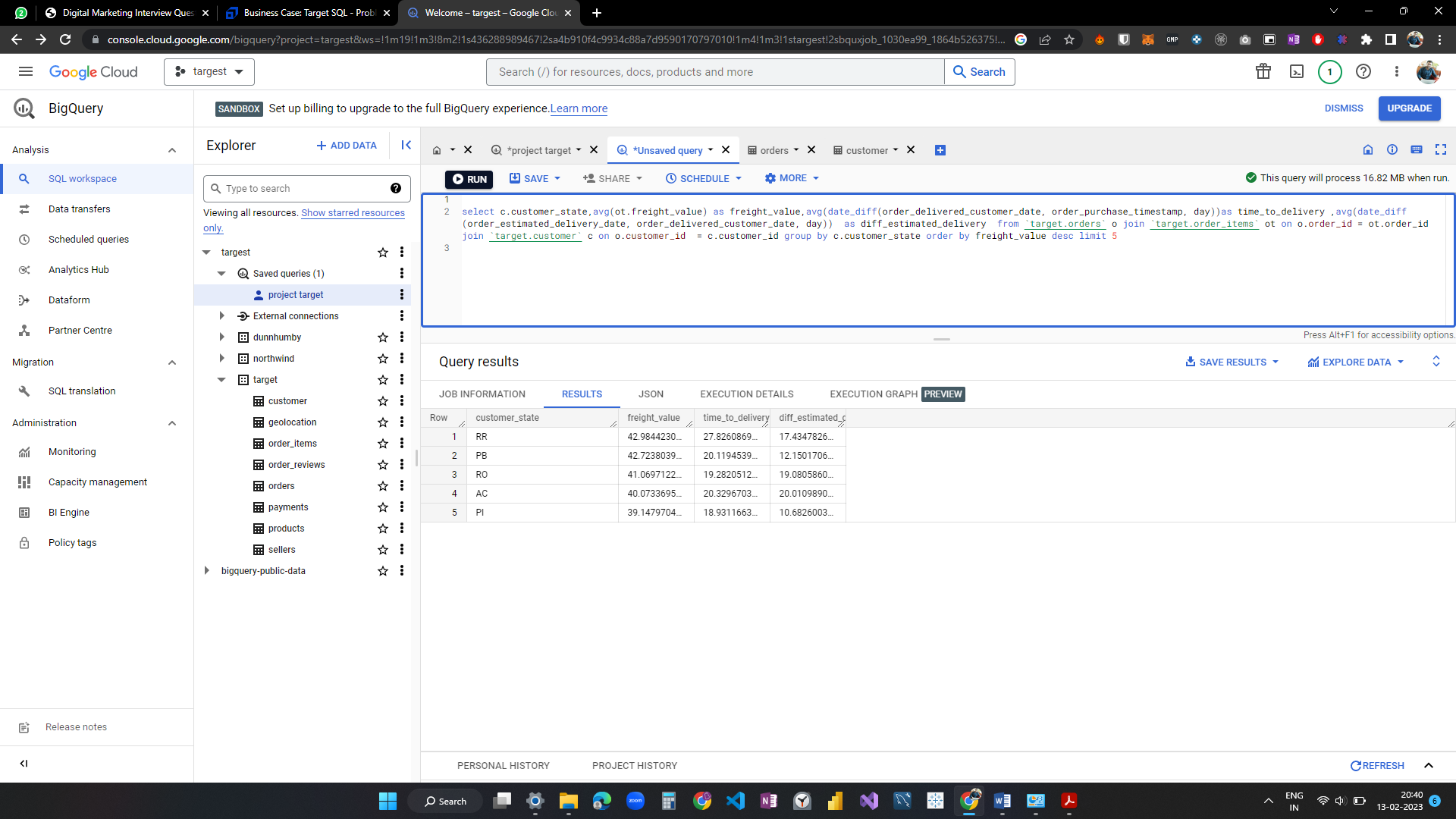
from `target.orders` o join `target.order\_items` ot on o.order\_id = ot.order\_id

join `target.customer` c on o.customer\_id  = c.customer\_id

group by c.customer\_state

order by freight\_value desc

limit 5



Top 5 states with highest/lowest average time to delivery

select c.customer\_state,

avg(ot.freight\_value) as freight\_value,

avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day))

as time\_to\_delivery ,

avg(date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day))

as diff\_estimated\_delivery

from `target.orders` o join `target.order\_items` ot on o.order\_id = ot.order\_id

join `target.customer` c on o.customer\_id  = c.customer\_id

group by c.customer\_state

order by time\_to\_delivery desc

limit 5



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

select c.customer\_state,

avg(ot.freight\_value) as freight\_value,

avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day))

as time\_to\_delivery ,

avg(date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day))

as diff\_estimated\_delivery

from `target.orders` o join `target.order\_items` ot on o.order\_id = ot.order\_id

join `target.customer` c on o.customer\_id  = c.customer\_id

group by c.customer\_state

order by diff\_estimated\_delivery desc

limit 5

Month over Month count of orders for different payment types

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

format\_date("%B",extract(date from o.order\_purchase\_timestamp)) as month\_name,

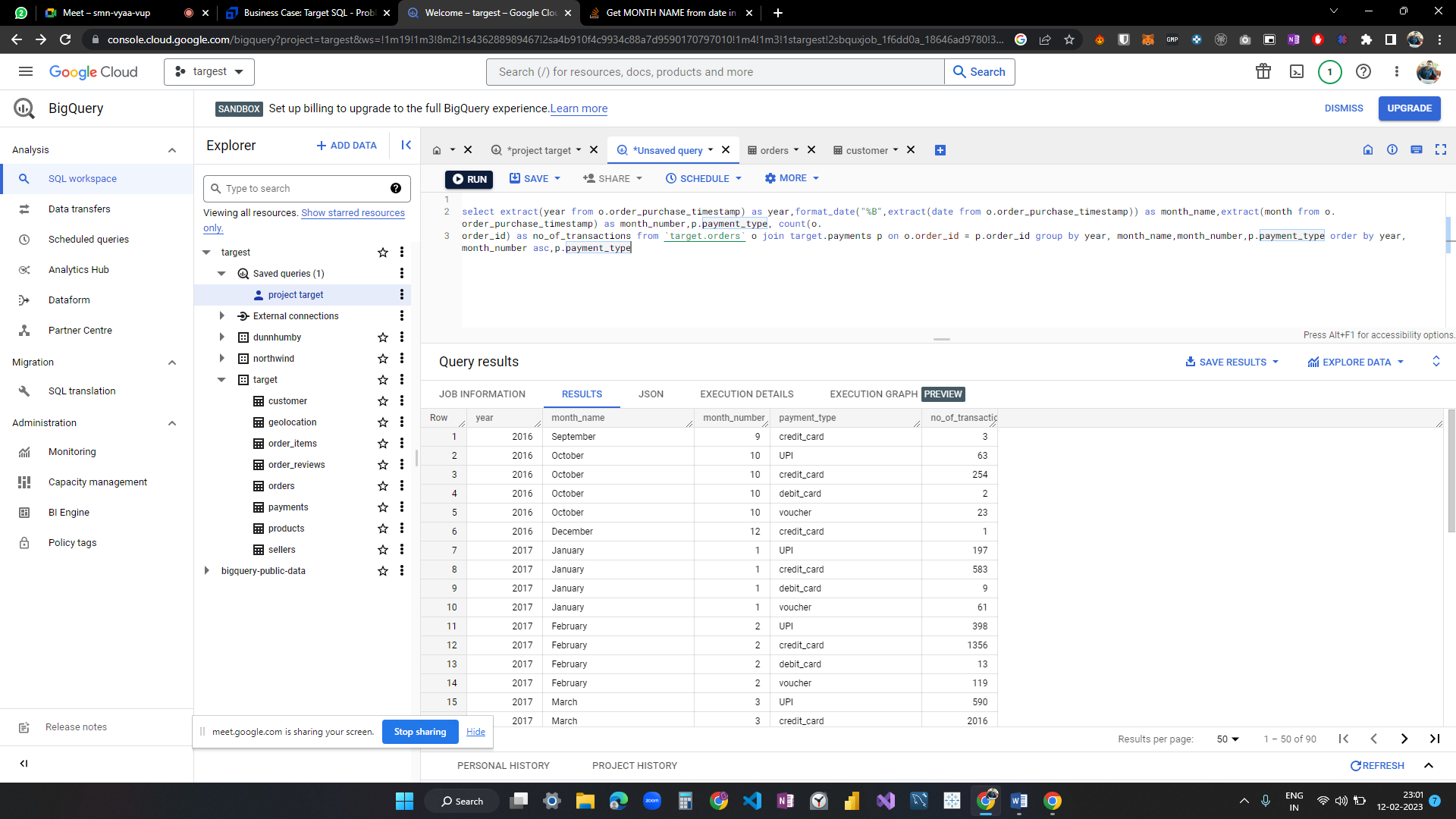
extract(month from o.order\_purchase\_timestamp) as month\_number,

p.payment\_type,

count(o.order\_id) as no\_of\_transactions from `target.orders` o

join target.payments p on o.order\_id = p.order\_id   
group by year, month\_name,month\_number,p.payment\_type

order by year,month\_number asc,p.payment\_type



Count of orders based on the no. of payment installments

select payment\_installments,

count(order\_id) as no\_of\_payments

from `target.payments`

group by payment\_installments

