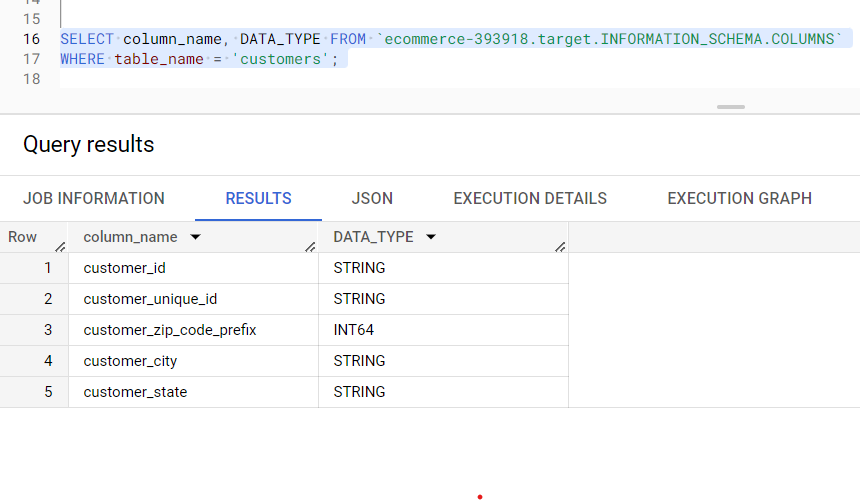
**1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.**

1) Data type of all columns in the “customers” table.

SELECT column\_name, DATA\_TYPE FROM `ecommerce-393918.target.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name = 'customers';



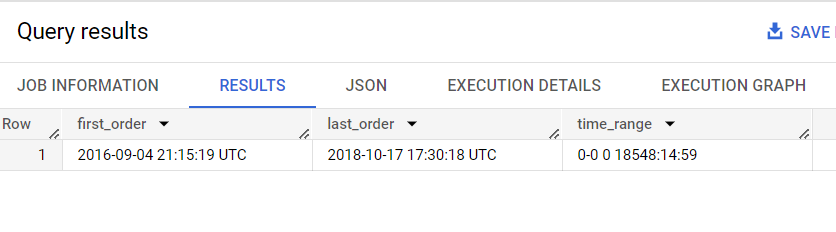
2) Get the time range between which the orders were placed.

select min(order\_purchase\_timestamp) as first\_order,

max(order\_purchase\_timestamp) as last\_order ,

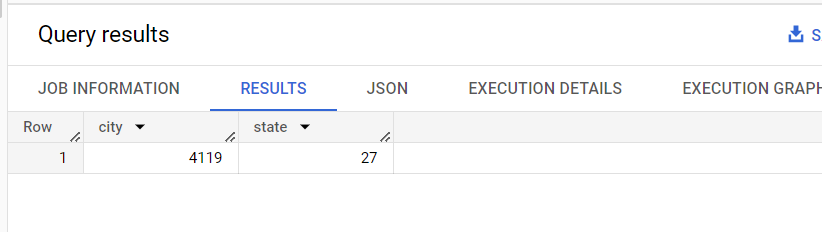
max(order\_purchase\_timestamp)- min(order\_purchase\_timestamp) as time\_range

from target.orders



3) Count the number of Cities and States in our dataset.

select count(distinct customer\_city) as city, count(distinct customer\_state)  as state from `target.customers`;



**2)**

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

SELECT

  EXTRACT(YEAR FROM order\_purchase\_timestamp) AS order\_year,

  EXTRACT(MONTH FROM order\_purchase\_timestamp) AS order\_month,

  COUNT(\*) AS order\_count

FROM target.orders

GROUP BY

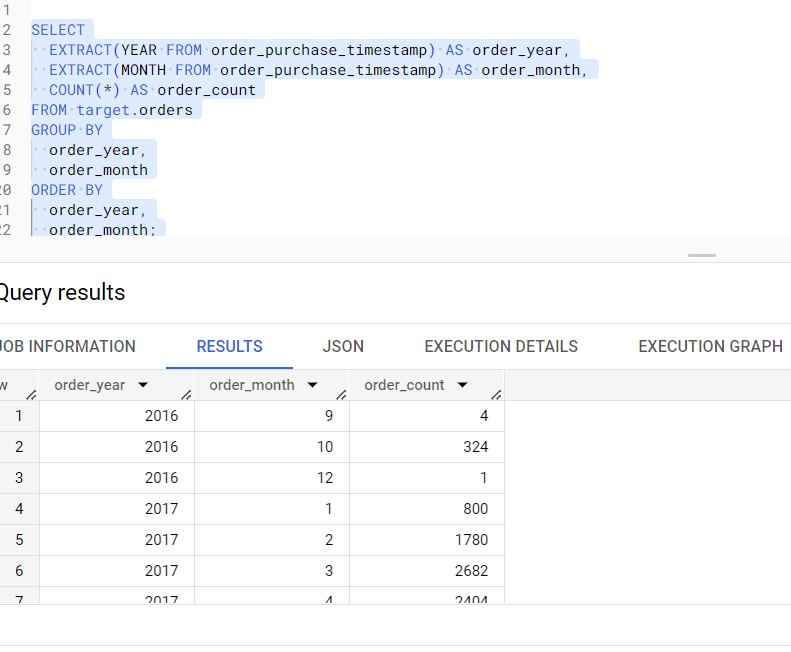
  order\_year,

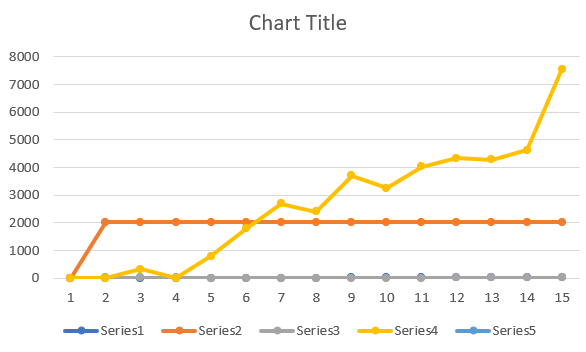
  order\_month

ORDER BY

  order\_year,

  order\_month;



****

**Analysis -** There is a gradual increase in the count of order in comparison with 2016 data and regular trends of increase of counts from 2016 to 2017 to, there is sudden downward trend in 2018 in the month of September and October but this downfall gradually started from April, 2018 by decrease in count by some hundreds.

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

SELECT

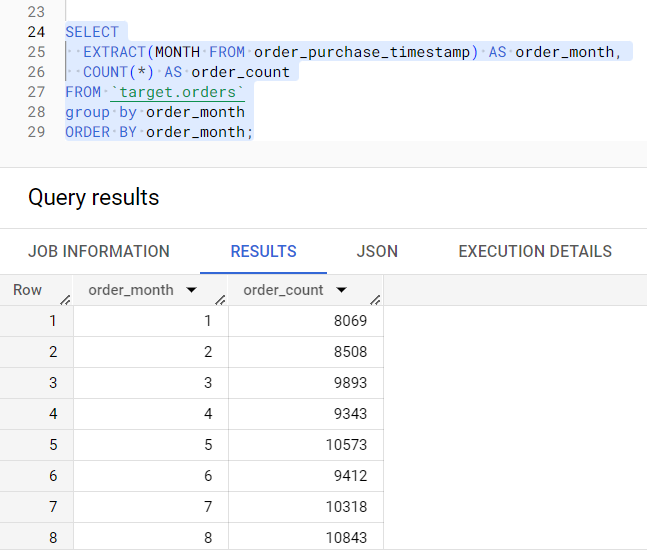
  EXTRACT(MONTH FROM order\_purchase\_timestamp) AS order\_month,

  COUNT(\*) AS order\_count

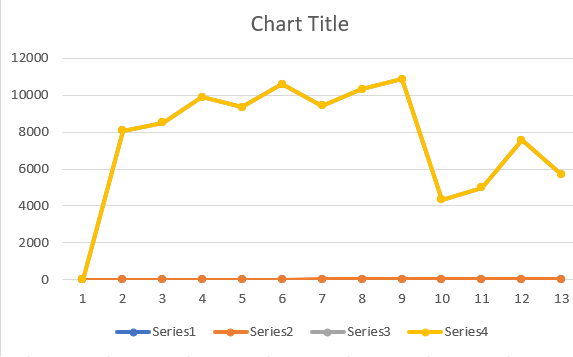
FROM `target.orders`

group by order\_month

ORDER BY order\_month;



**Analysis -** Yes, there is a monthly seasonality while ordering the orders. There are maximum orders placed in the month of May, July and August.



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

SELECT

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'

        WHEN EXTRACT(HOUR FROM order\_purchase\_timestamp) BETWEEN 19 AND 23 THEN 'Night'

    END AS time\_interval,

    COUNT(\*) AS order\_count

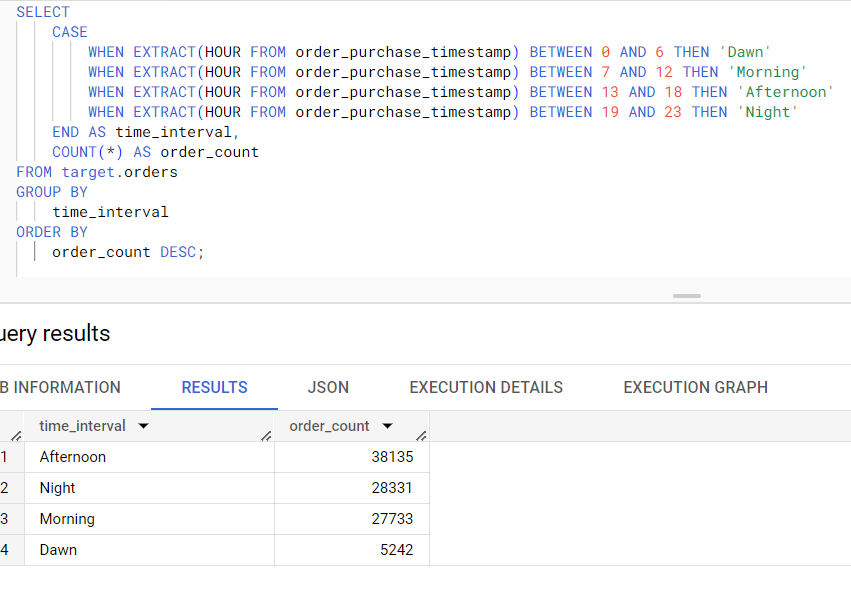
FROM target.orders

GROUP BY

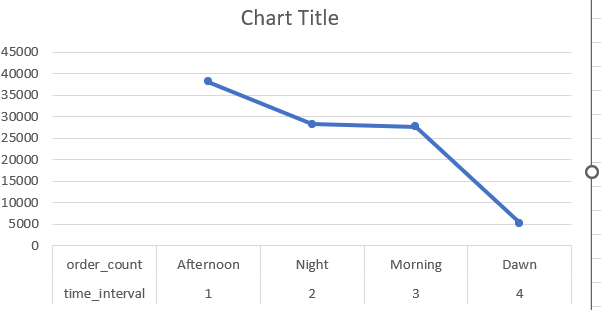
    time\_interval

ORDER BY

    order\_count DESC;



**Analysis -** There are maximum orders placed in the afternoon hour from 13 to 18.



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

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

SELECT

    customer\_state,

    EXTRACT(month FROM order\_purchase\_timestamp) AS mth,

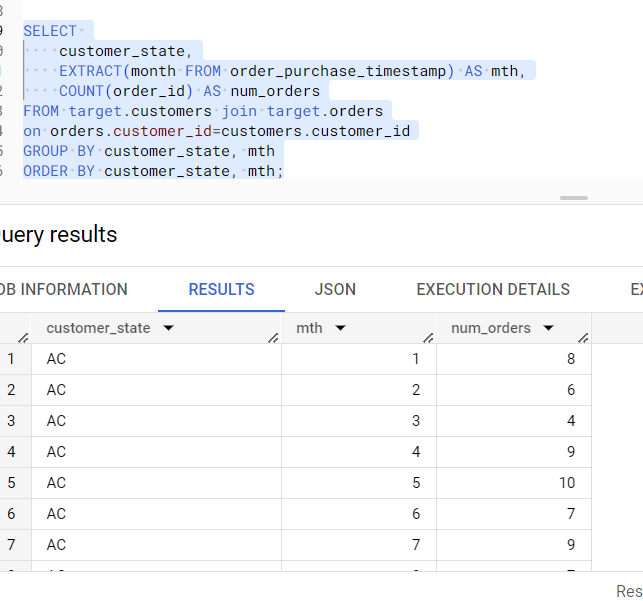
    COUNT(order\_id) AS num\_orders

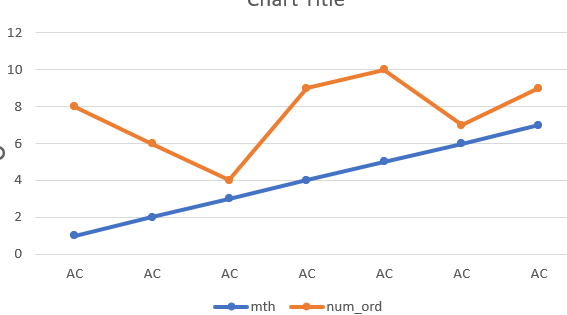
FROM target.customers join target.orders

on orders.customer\_id=customers.customer\_id

GROUP BY customer\_state, mth

ORDER BY customer\_state, mth;





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

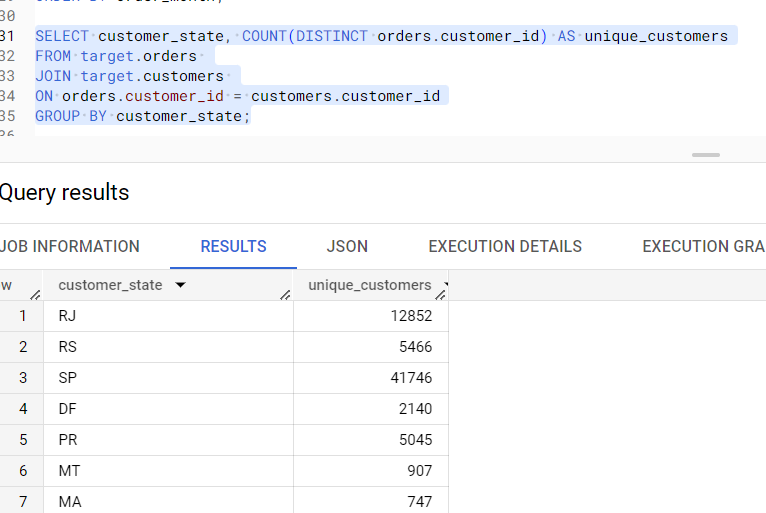
SELECT customer\_state, COUNT(DISTINCT orders.customer\_id) AS unique\_customers

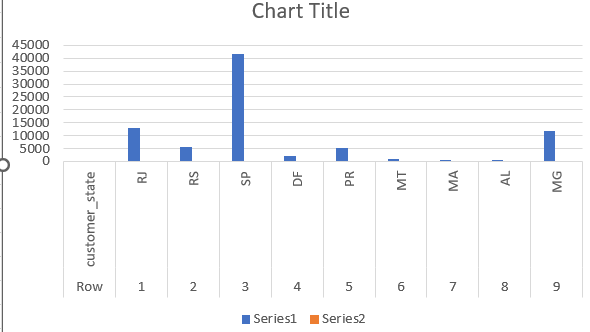
FROM target.orders

JOIN target.customers

ON orders.customer\_id = customers.customer\_id

GROUP BY customer\_state;





**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)

SELECT ((sum(

          Case

          when extract(year from order\_purchase\_timestamp)=2018 and extract(month from order\_purchase\_timestamp)<=8

          then payment\_value

          else 0 end)- sum(

          Case

          when extract(year from order\_purchase\_timestamp)=2017

          then payment\_value

          else 0 end))/sum(

          Case

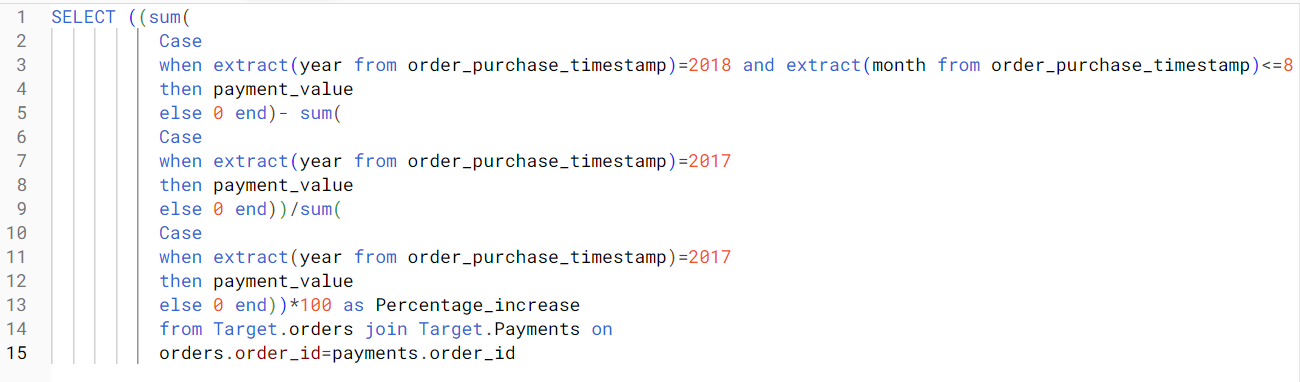
          when extract(year from order\_purchase\_timestamp)=2017

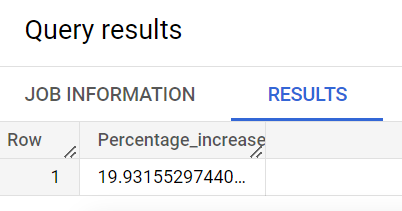
          then payment\_value

          else 0 end))\*100 as Percentage\_increase

          from Target.orders join Target.Payments on

          orders.order\_id=payments.order\_id





ANALYSIS : There has been an increase of ~20 % in cost in year 2018 compared to year 2017

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

SELECT customer\_state,round(sum(payment\_value),2)as total\_order\_price,

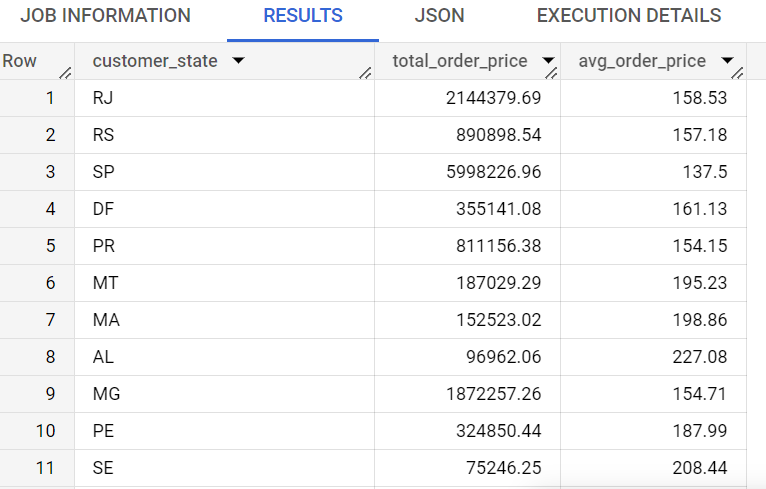
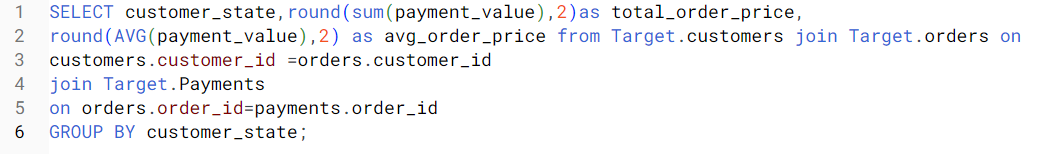
round(AVG(payment\_value),2) as avg\_order\_price from Target.customers join Target.orders on

customers.customer\_id =orders.customer\_id

join Target.Payments

on orders.order\_id=payments.order\_id

GROUP BY customer\_state;



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

select customer\_state,

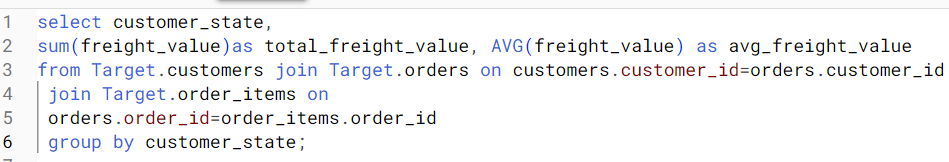
sum(freight\_value)as total\_freight\_value, AVG(freight\_value) as avg\_freight\_value

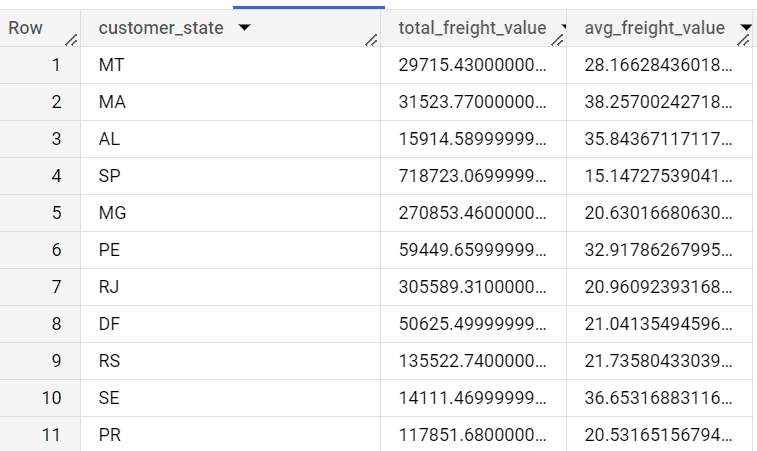
from Target.customers join Target.orders on customers.customer\_id=orders.customer\_id

 join Target.order\_items on

 orders.order\_id=order\_items.order\_id

 group by customer\_state;





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

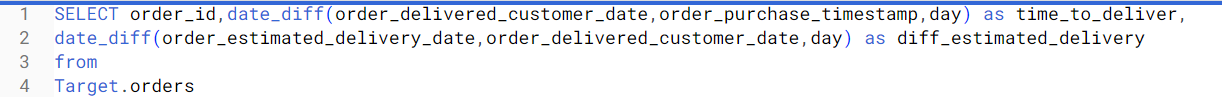
1. Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.  
   Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

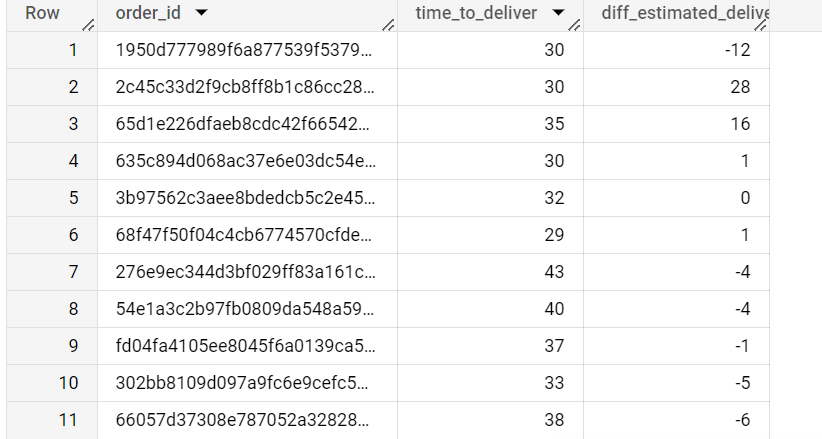
SELECT order\_id,date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) as time\_to\_deliver,

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

from

Target.orders







ANALYSIS : As per the data Delay in orders was maximum 188 days where as in case of fast delivery it reached only 55 days prior to estimated delivery date

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

SELECT customer\_state,

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

from

Target.customers

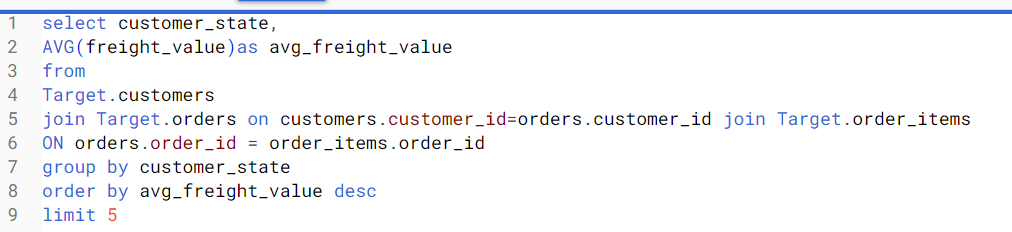
join Target.orders on customers.customer\_id=orders.customer\_id join Target.order\_items

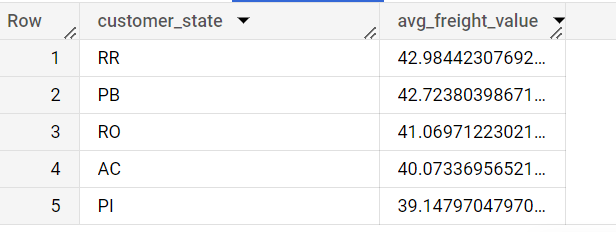
ON orders.order\_id = order\_items.order\_id

group by customer\_state

order by avg\_freight\_value desc

limit 5





SELECT customer\_state,

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

from

Target.customers

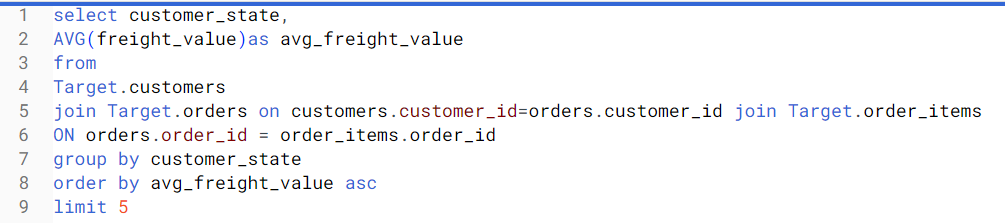
join Target.orders on customers.customer\_id=orders.customer\_id join Target.order\_items

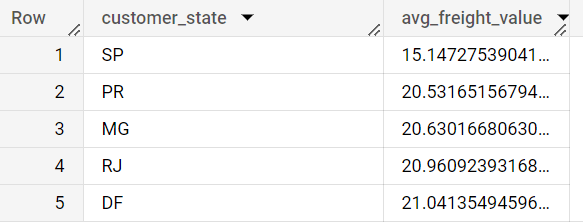
ON orders.order\_id = order\_items.order\_id

group by customer\_state

order by avg\_freight\_value asc

limit 5





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

SELECT customer\_state,

AVG(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)) as Avg\_Delivery\_Time

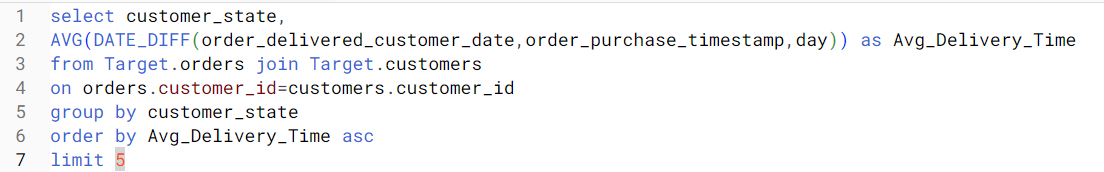
from Target.orders join Target.customers

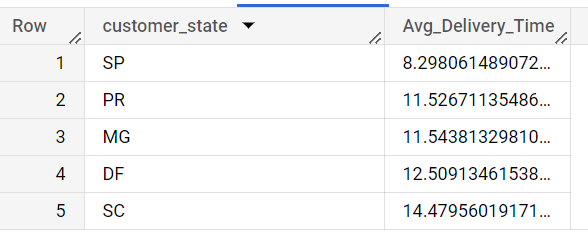
on orders.customer\_id=customers.customer\_id

group by customer\_state

order by Avg\_Delivery\_Time asc

limit 5





SELECT customer\_state,AVG(DATE\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,day)) as Avg\_Delivery\_time

from

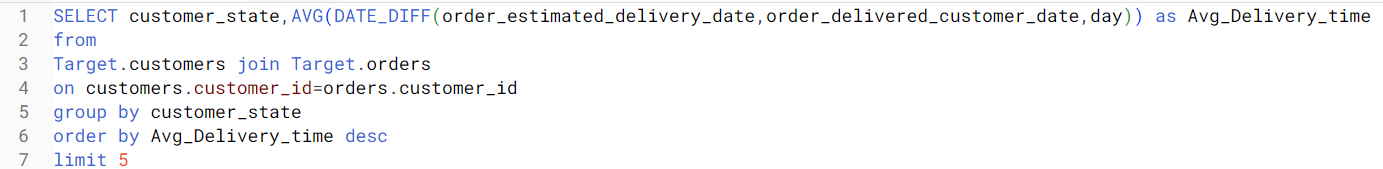
Target.customers join Target.orders

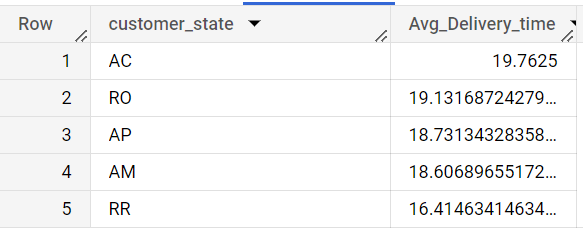
on customers.customer\_id=orders.customer\_id

group by customer\_state

order by Avg\_Delivery\_time desc

limit 5





4.Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

SELECT customer\_state,

AVG(DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)) as Avg\_Delivery\_Time

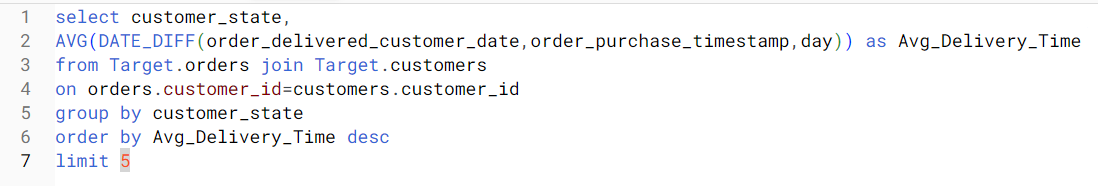
from Target.orders join Target.customers

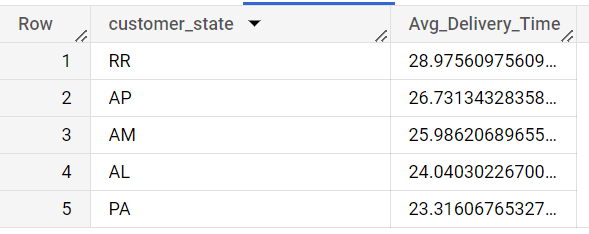
on orders.customer\_id=customers.customer\_id

group by customer\_state

order by Avg\_Delivery\_Time desc

limit 5





**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 payment\_month,

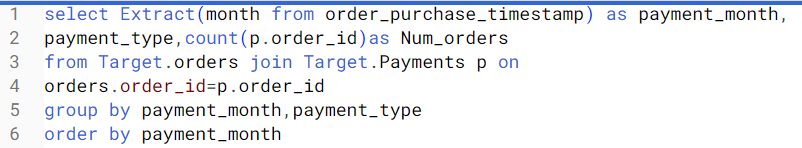
payment\_type,count(p.order\_id)as Num\_orders

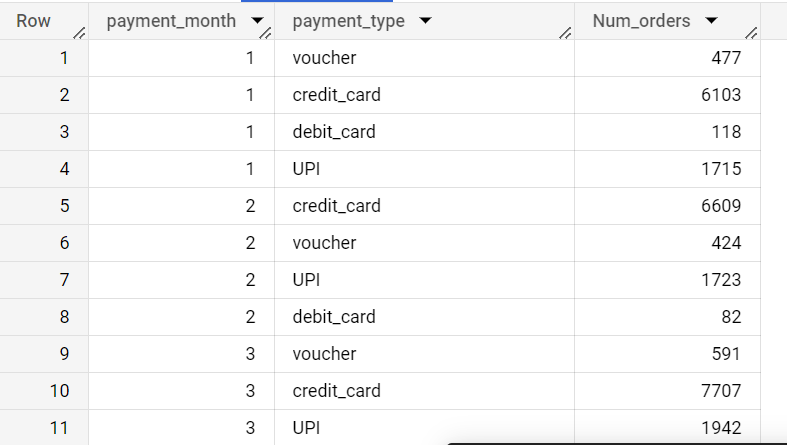
from Target.orders join Target.Payments p on

orders.order\_id=p.order\_id

group by payment\_month,payment\_type

order by payment\_month





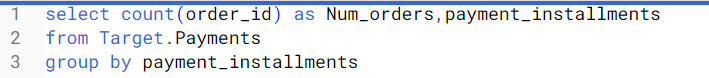
ANALYSIS : With the given data we can find that maximum purchases are made by credit cars followed by UPI , Voucher and Debit Card respectively.

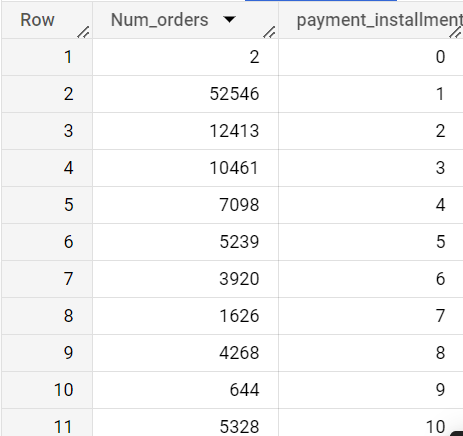
2.Find the no. of orders placed on the basis of the payment installments that have been paid.

select count(order\_id) as Num\_orders,payment\_installments

from Target.Payments

group by payment\_installments





ANALYSIS: As we can observe orders with maximum payment installment are 1,2 & 3 which tells most of the orders were placed in last month.