**Case\_study\_1**

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

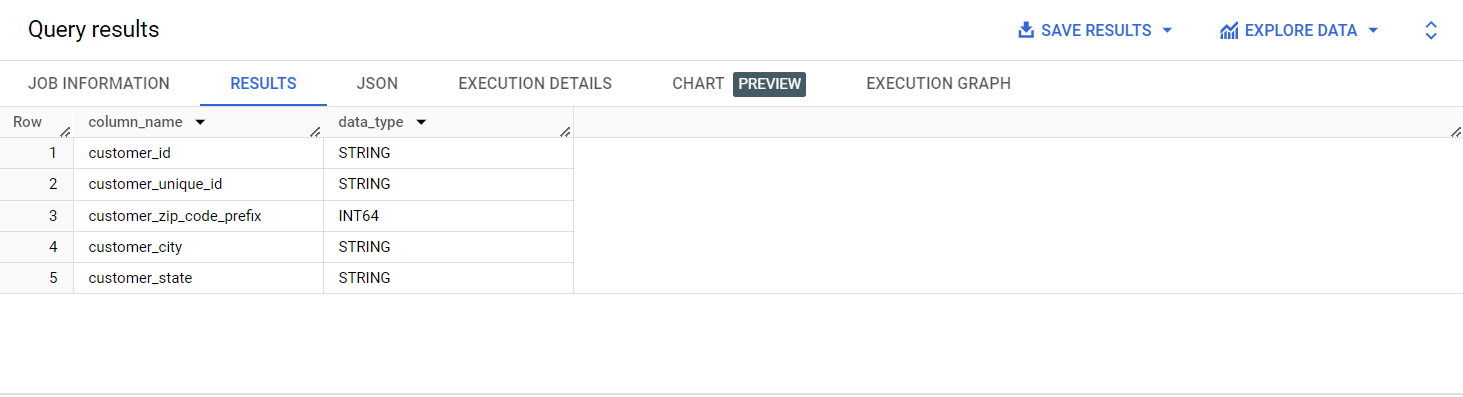
**1.Data type of all columns in the "customers" table**

select column\_name, data\_type

from `sinuous-anvil-396816.case\_study\_1.INFORMATION\_SCHEMA.COLUMNS`

where table\_name = 'customers'

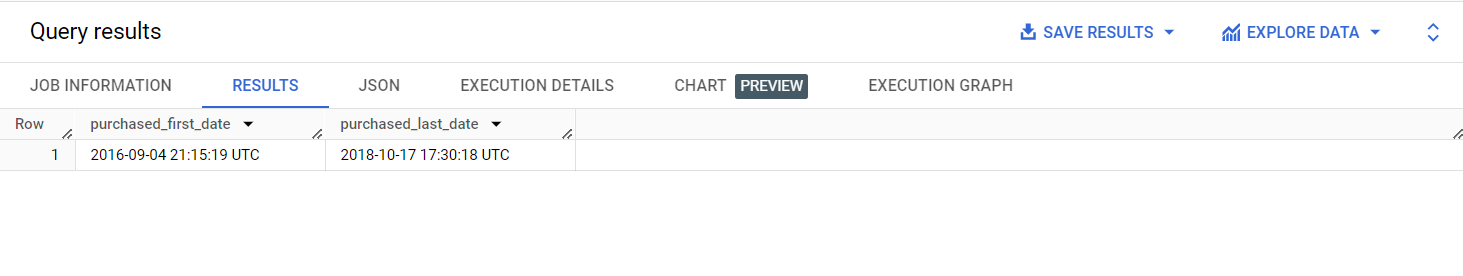
order by ordinal\_position

**2.Get the time range between which the orders were place.**

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

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

from `case\_study\_1.orders`



**Additional evaluation:-**

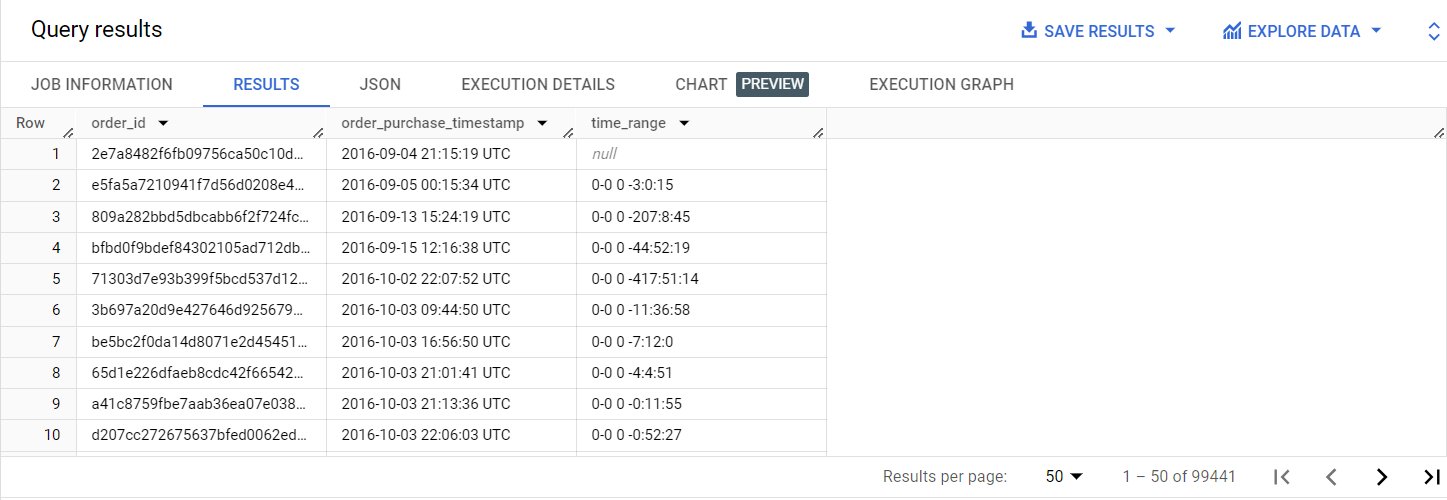
**Time diffrence between consecutive purchases happenedand to find maximum and minimum intervals.**

select order\_id,order\_purchase\_timestamp,

(lag(order\_purchase\_timestamp)over(order by order\_purchase\_timestamp)-order\_purchase\_timestamp) as time\_range

from `case\_study\_1.orders`

order by order\_purchase\_timestamp



**Maximum and minimum time interval between purchases**

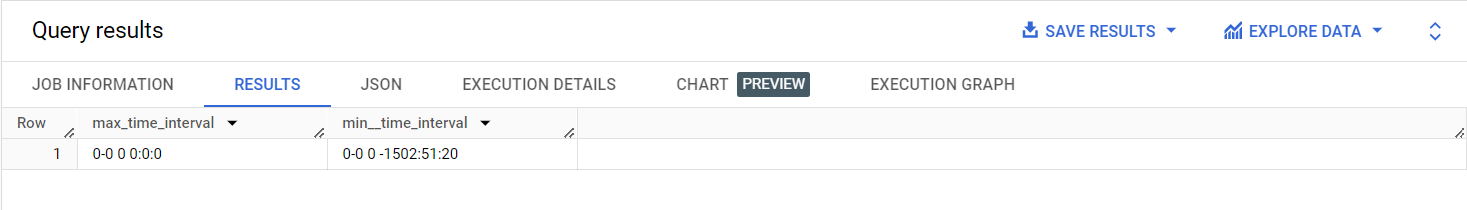
select max(time\_range) as max\_time\_interval,min(time\_range) as min\_\_time\_interval from (

select \*, (lag(order\_purchase\_timestamp)over(order by order\_purchase\_timestamp)-order\_purchase\_timestamp) as time\_range

from `case\_study\_1.orders`

order by order\_purchase\_timestamp

)

  
**3.Count the Cities & States of customers who ordered during the given period.**

with cte\_1b as(

select c.customer\_id,

c.customer\_zip\_code\_prefix,

o.order\_purchase\_timestamp,

g.geolocation\_zip\_code\_prefix,

g.geolocation\_city,g.geolocation\_state

from `case\_study\_1.customers` c inner join `case\_study\_1.orders` o on c.customer\_id=o.customer\_id

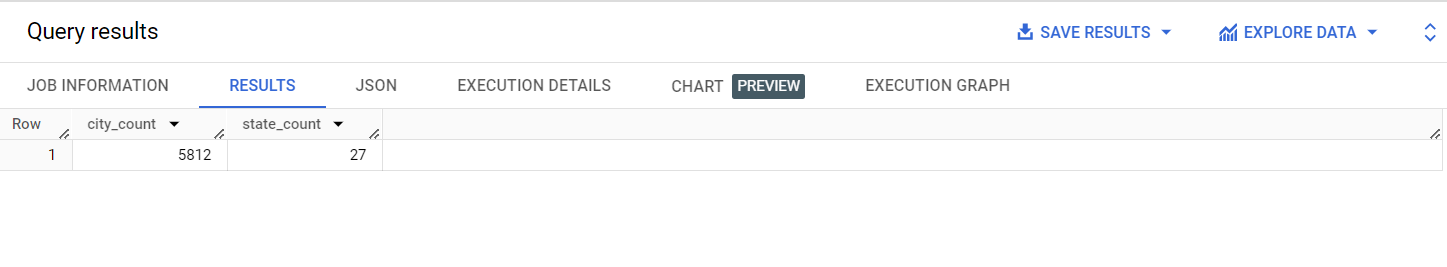
inner join `case\_study\_1.geolocation` g on c.customer\_zip\_code\_prefix=g.geolocation\_zip\_code\_prefix

)

select count(distinct geolocation\_city) as city\_count,

count(distinct geolocation\_state) as state\_count

from cte\_1b



**2. In-depth Exploration:**

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

select count(\*) as no\_of\_purchases,year

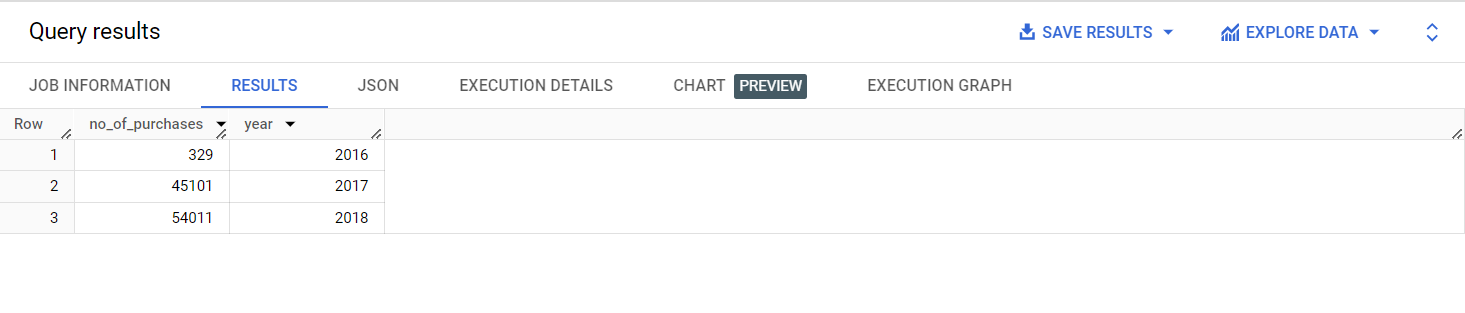
from (

select \*, EXTRACT(year FROM order\_purchase\_timestamp) AS year

from `case\_study\_1.orders`)t

group by t.year

order by t.year



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

with cte\_q2 as(

select \*, EXTRACT(year FROM order\_purchase\_timestamp) AS year,

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

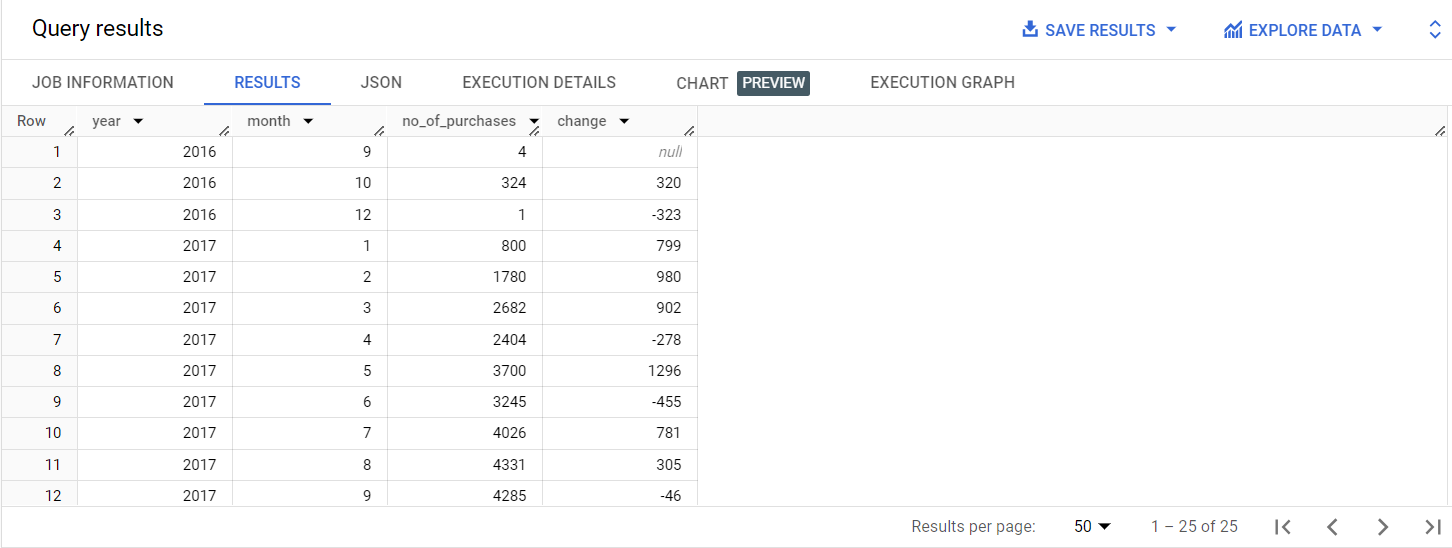
from `case\_study\_1.orders`

)

select year,month,count(\*) as no\_of\_purchases,(count(\*)-lag(count(\*))over(order by year,month)) as change from cte\_q2

group by year,month

order by year,month



**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**

with cte\_q3 as(

select c.customer\_id,o.order\_purchase\_timestamp,g.geolocation\_state,

format\_timestamp('%H',o.order\_purchase\_timestamp,"UTC") as time\_range

from `case\_study\_1.customers` c inner join `case\_study\_1.orders` o on c.customer\_id=o.customer\_id

inner join `case\_study\_1.geolocation` g on c.customer\_zip\_code\_prefix=g.geolocation\_zip\_code\_prefix

where g.geolocation\_state='BA'

)

select count(\*) as num\_of\_purchases,

case

when time\_range>='00' and time\_range<='06' then 'Dawn'

when time\_range>='07' and time\_range<='12' then 'Mornings'

when time\_range>='13' and time\_range<='18' then 'Afternoon'

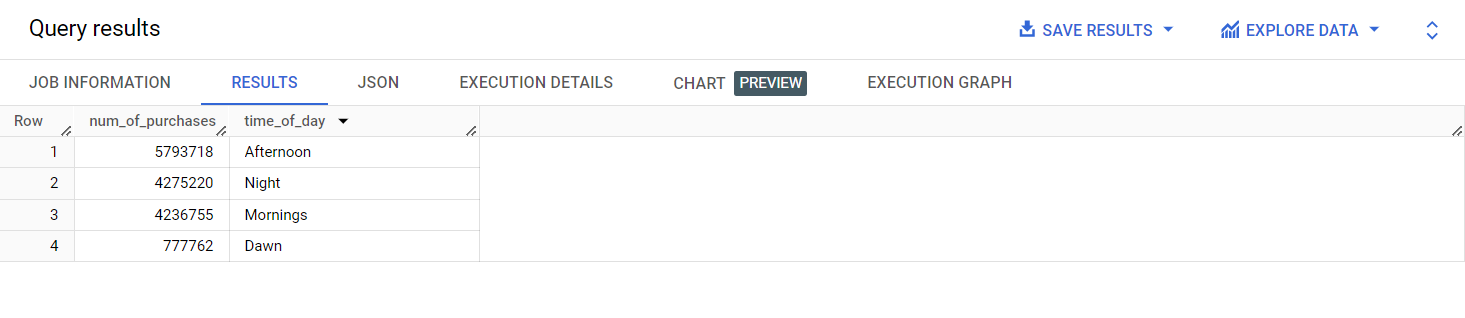
else 'Night'

end as time\_of\_day

from cte\_q3

group by time\_of\_day

order by num\_of\_purchases desc

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

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

with cte\_3a as (

select g.geolocation\_state as STATE,

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

o.order\_purchase\_timestamp

from `case\_study\_1.customers` c inner join `case\_study\_1.orders` o on c.customer\_id=o.customer\_id

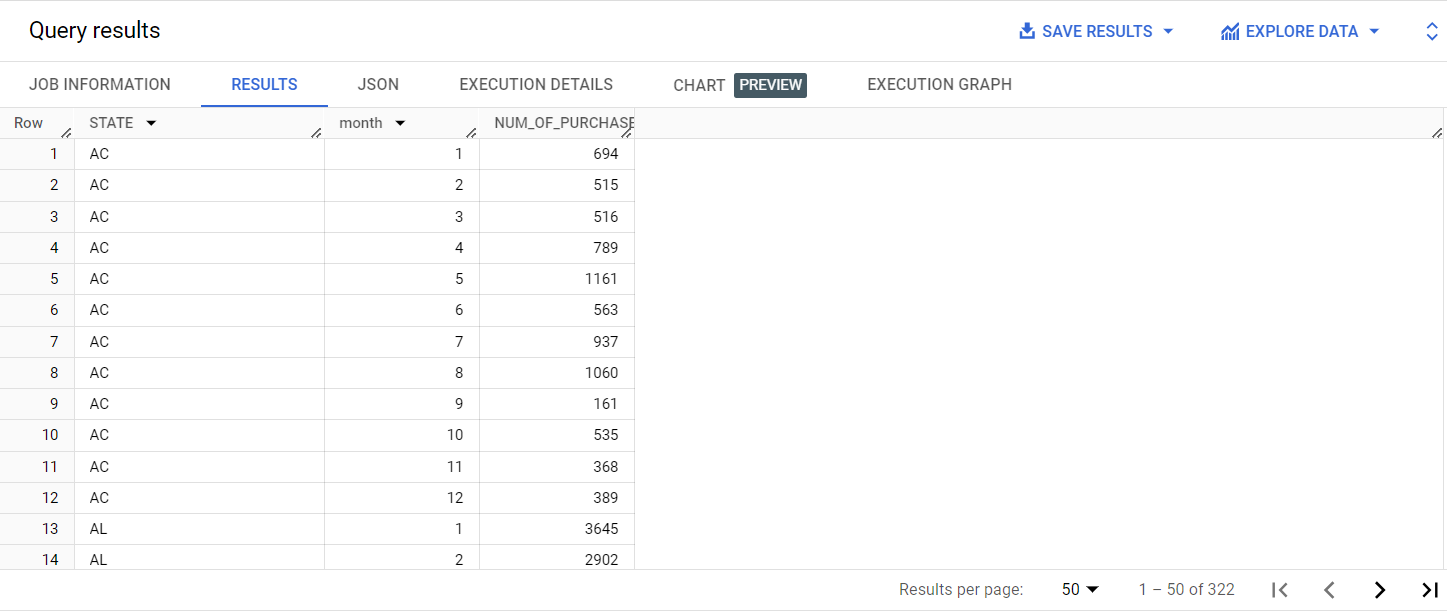
inner join `case\_study\_1.geolocation` g on c.customer\_zip\_code\_prefix=g.geolocation\_zip\_code\_prefix)

select DISTINCT STATE,month,COUNT(\*) AS NUM\_OF\_PURCHASES

from cte\_3a

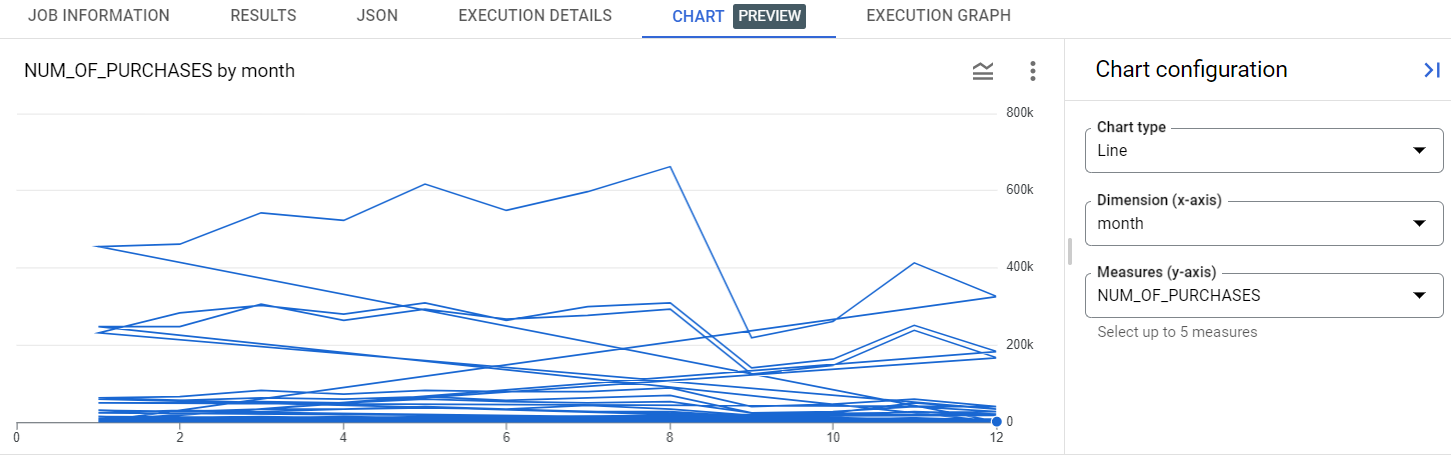
group by STATE,month

ORDER BY STATE,month



**Additional :-**

**To observe the pattern in the number of purchases with month for different states**



**2.How are the customers distributed across all the state**

select g.geolocation\_state as STATE,

count(distinct c.customer\_id) AS UNIQUE\_CUSTOMERS

from `case\_study\_1.customers` c inner join `case\_study\_1.geolocation` g

on c.customer\_zip\_code\_prefix=g.geolocation\_zip\_code\_prefix

group by STATE

ORDER BY STATE

**Additional:-**

**Arranging the above result to find states with highest number of unique customers.**

select g.geolocation\_state as STATE,

count(distinct c.customer\_id) AS UNIQUE\_CUSTOMERS

from `case\_study\_1.customers` c inner join `case\_study\_1.geolocation` g

on c.customer\_zip\_code\_prefix=g.geolocation\_zip\_code\_prefix

group by STATE

ORDER BY UNIQUE\_CUSTOMERS desc



**4. 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).You can use the "payment\_value" column in the payments table to get the cost of orders**

with cte\_go as (

select sum(p.payment\_value) as tot\_val,

EXTRACT(year FROM o.order\_purchase\_timestamp) as year

from `case\_study\_1.payments` p inner join `case\_study\_1.orders` o using(order\_id)

where EXTRACT(month FROM o.order\_purchase\_timestamp)>1 and EXTRACT(month FROM o.order\_purchase\_timestamp)<8

group by year

order by year)

select round(((tot\_val-lag\_tot\_val)/lag\_tot\_val)\*100,2) as per\_increse\_cost

from(

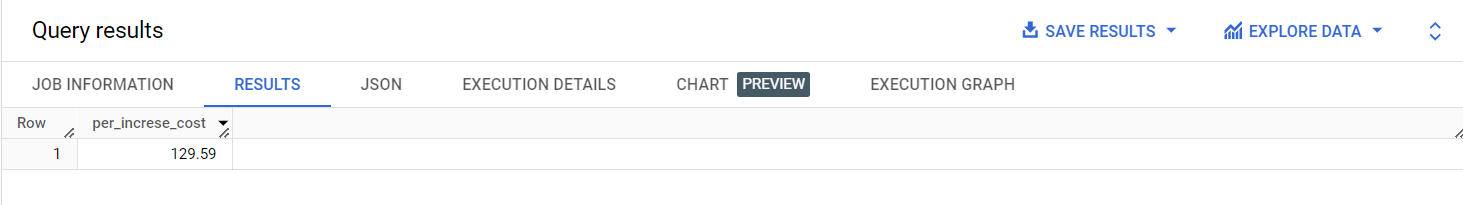
select tot\_val,year,lag(tot\_val) over(order by year) as lag\_tot\_val

from cte\_go

order by year

)t

where year=2018



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

select c.customer\_state as State,

round(sum(i.price),2) as Total\_order\_value,

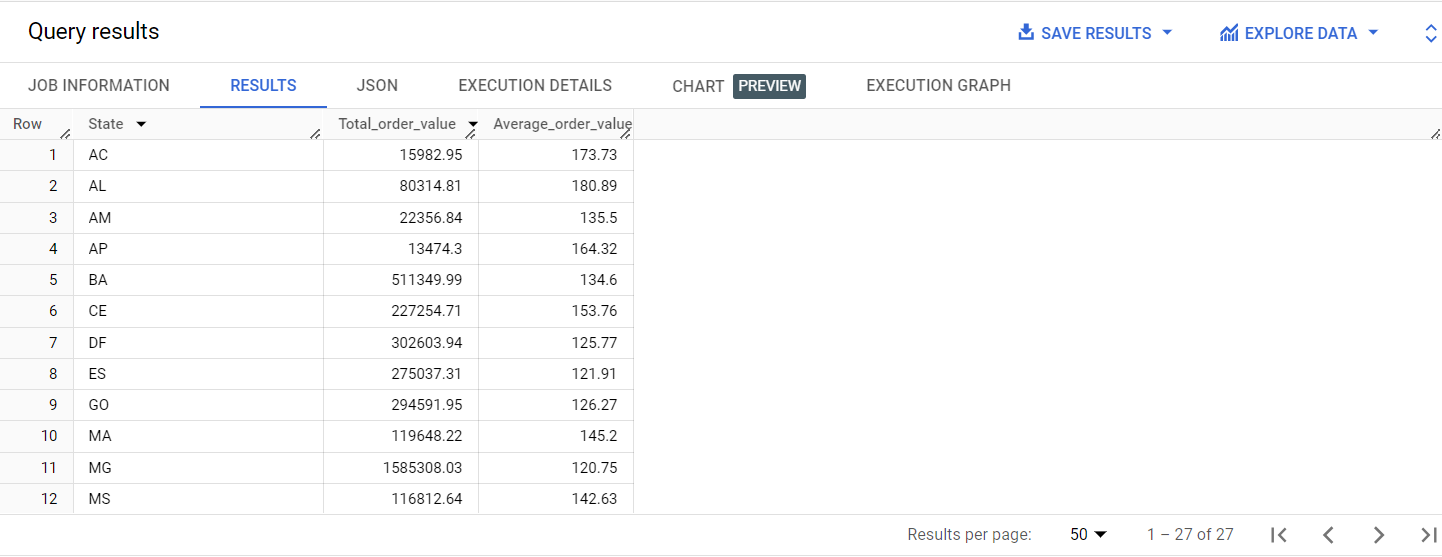
round(avg(i.price),2) as Average\_order\_value

from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state

order by customer\_state



**Additional:-**

**Top 5 states with highest average order value**

select State from (select c.customer\_state as State,

round(sum(i.price),2) as Total\_order\_value,

round(avg(i.price),2) as Average\_order\_value

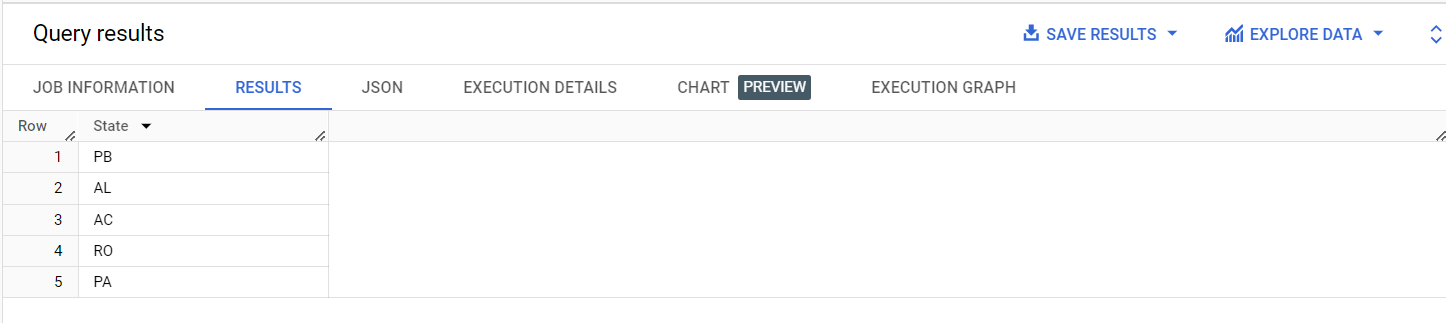
from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state)t

order by Average\_order\_value desc

limit 5



**Top 5 states with lowest average order value**

select State from (select c.customer\_state as State,

round(sum(i.price),2) as Total\_order\_value,

round(avg(i.price),2) as Average\_order\_value

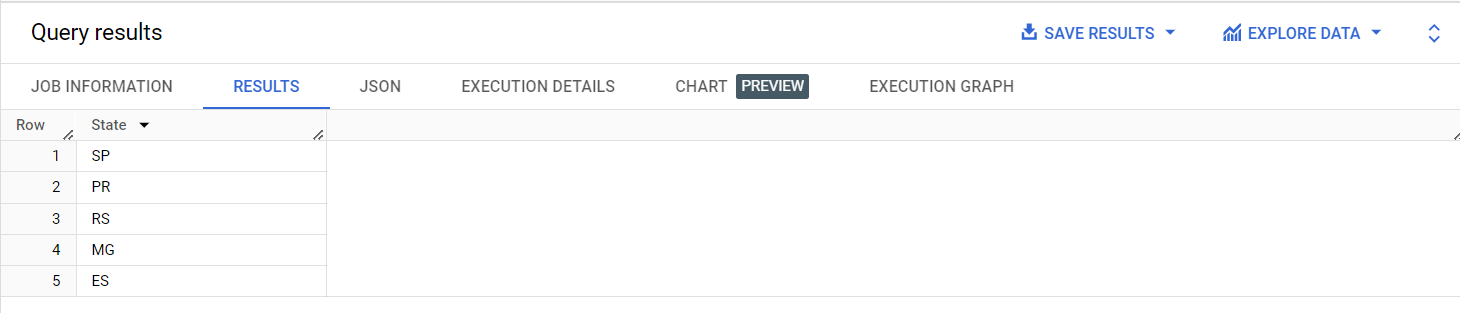
from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state)t

order by Average\_order\_value asc

limit 5



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

select c.customer\_state as State,

round(sum(i.freight\_value),2) as Total\_freight\_value,

round(avg(i.freight\_value),2) as Average\_freight\_value

from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state

order by customer\_state



**5. 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.**  
**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:

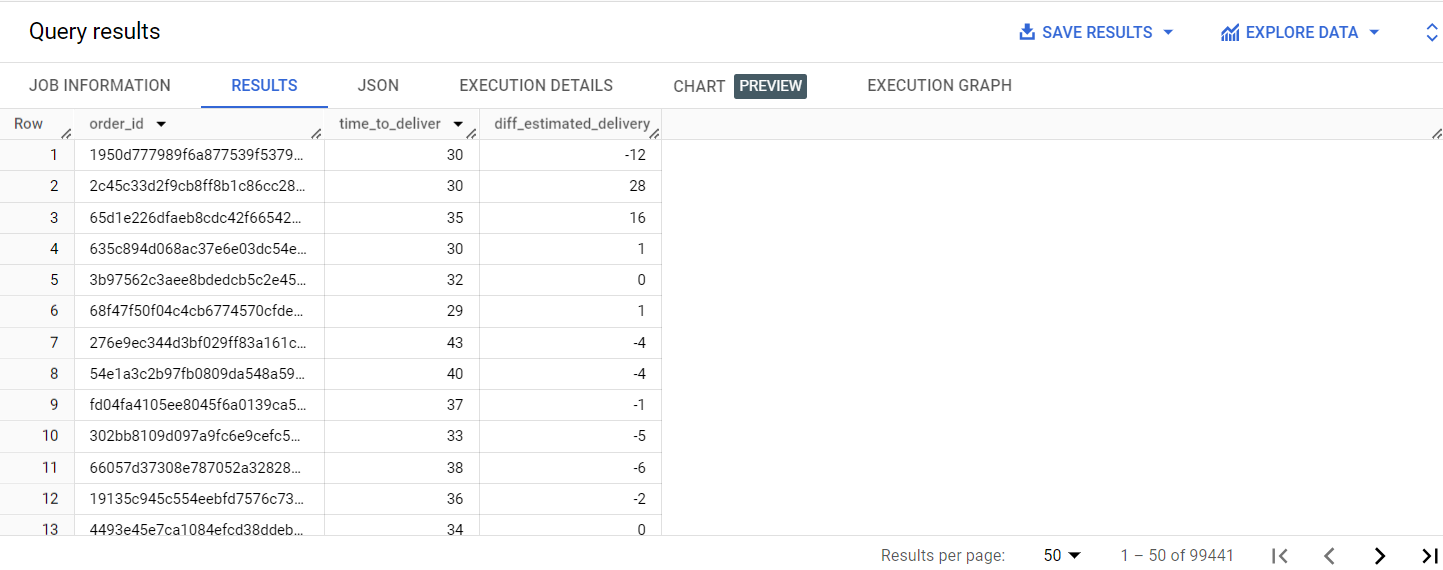
* time\_to\_deliver = order\_delivered\_customer\_date - order\_purchase\_timestamp
* diff\_estimated\_delivery = order\_estimated\_delivery\_date - order\_delivered\_customer\_date

select order\_id,

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 `case\_study\_1.orders`



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

**Heighest:-**

select State from (

select c.customer\_state as State,

round(sum(i.freight\_value),2) as Total\_freight\_value,

round(avg(i.freight\_value),2) as Average\_freight\_value

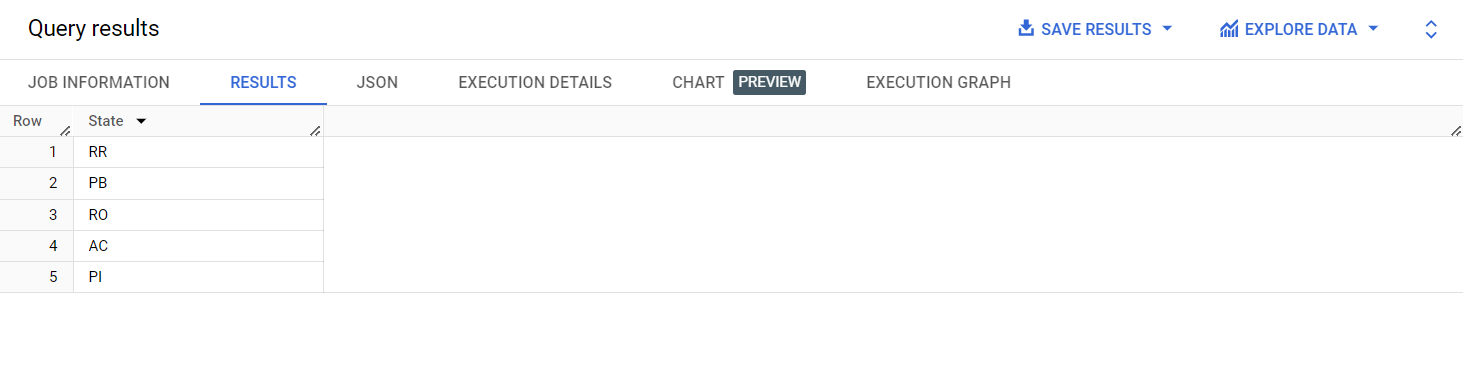
from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state)t

order by Average\_freight\_value DESC

limit 5



**Lowest:-**

select State from (

select c.customer\_state as State,

round(sum(i.freight\_value),2) as Total\_freight\_value,

round(avg(i.freight\_value),2) as Average\_freight\_value

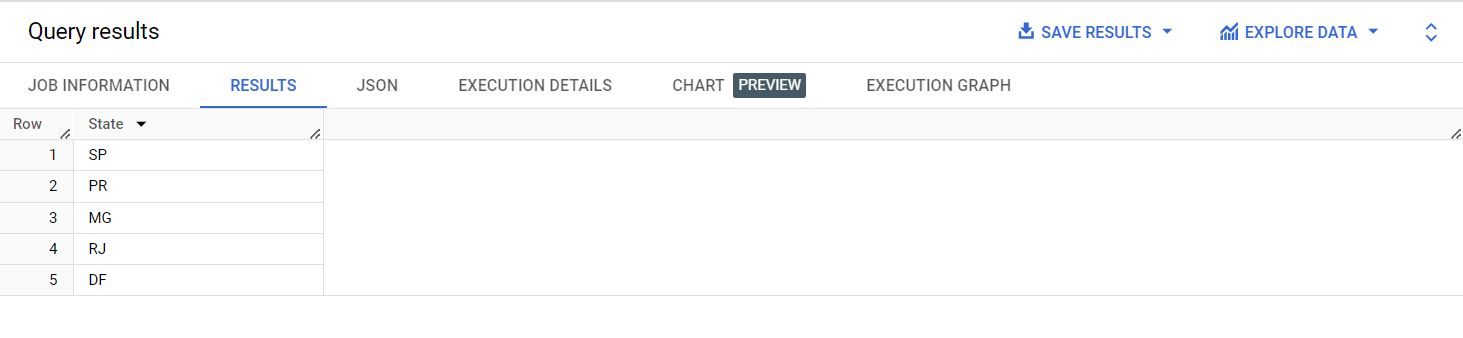
from `case\_study\_1.order\_items` i inner join `case\_study\_1.orders` using(order\_id)

inner join `case\_study\_1.customers`c using(customer\_id)

group by customer\_state )t

order by Average\_freight\_value asc

limit 5

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

**Heighest:-**

select State from(

select distinct c.customer\_state as State,

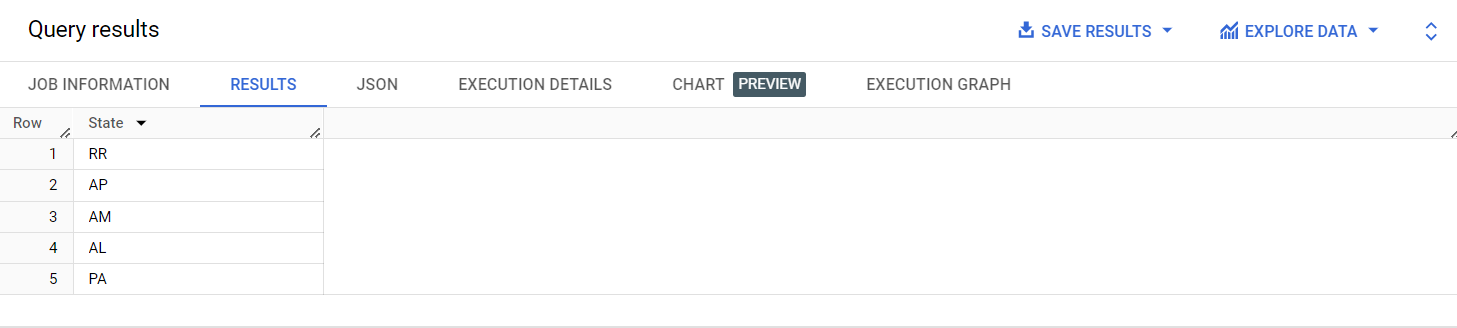
round(avg(TIMESTAMP\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) over(partition by c.customer\_state),2)

as avg\_time\_to\_deliver

from `case\_study\_1.orders`o inner join `case\_study\_1.customers`c using(customer\_id))t

order by avg\_time\_to\_deliver desc

limit 5



**Lowest:-**

select State from (

select distinct c.customer\_state as State,

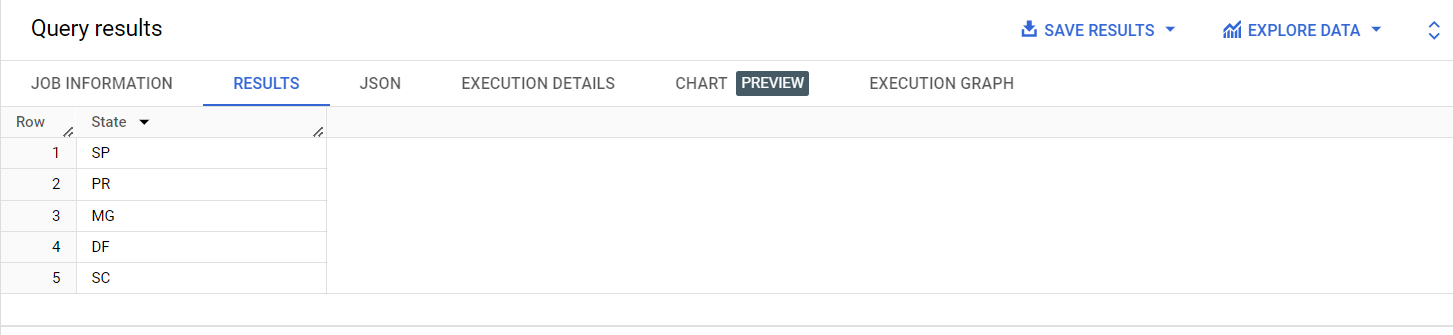
round(avg(TIMESTAMP\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) over(partition by c.customer\_state),2)

as avg\_time\_to\_deliver

from `case\_study\_1.orders`o inner join `case\_study\_1.customers`c using(customer\_id))t

order by avg\_time\_to\_deliver asc

limit 5



4.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.

with cte\_5D AS(

select DISTINCT c.customer\_state as State,

avg(TIMESTAMP\_DIFF(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) over(partition by c.customer\_state)

as avg\_time\_to\_deliver,

avg(TIMESTAMP\_DIFF(o.order\_estimated\_delivery\_date, o.order\_delivered\_customer\_date, day)) over(partition by c.customer\_state)

as avg\_diff\_estimated\_delivery

from `case\_study\_1.orders`o inner join `case\_study\_1.customers`c using(customer\_id)

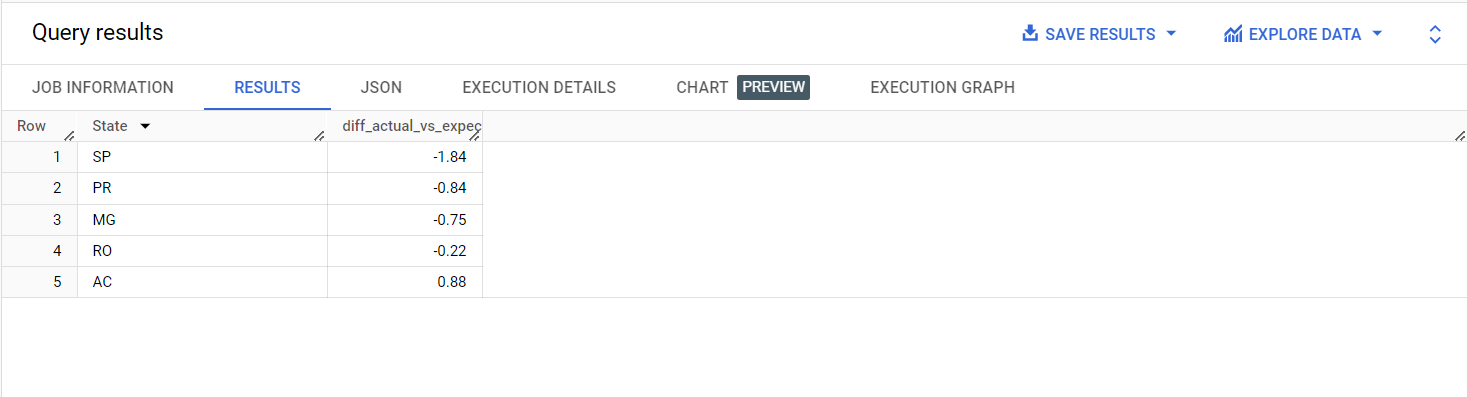
)

SELECT State,

round((avg\_time\_to\_deliver-avg\_diff\_estimated\_delivery),2) as diff\_actual\_vs\_expected

from cte\_5D

ORDER BY diff\_actual\_vs\_expected

limit 5  


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

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

with cte\_6A as (

select p.payment\_type,FORMAT\_DATE('%B',o.order\_purchase\_timestamp) AS MONTH,

EXTRACT(month FROM o.order\_purchase\_timestamp) AS month\_of\_year,o.order\_purchase\_timestamp

from `case\_study\_1.orders` o inner join `case\_study\_1.payments` p on p.order\_id=o.order\_id

)

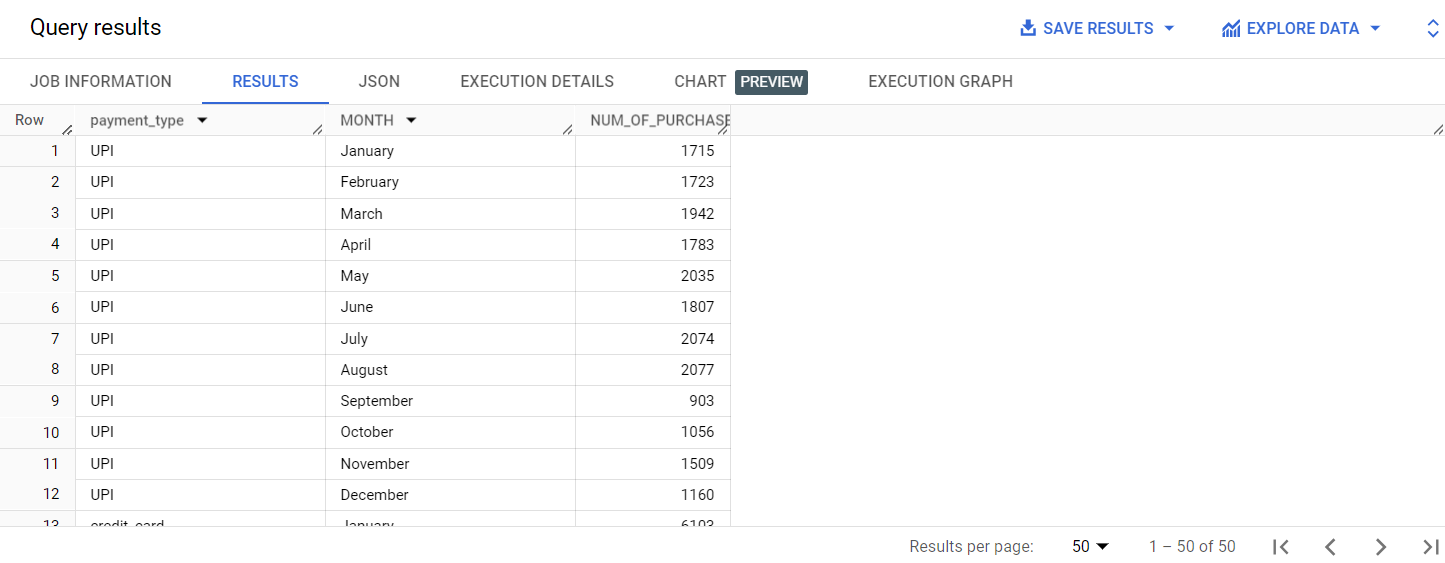
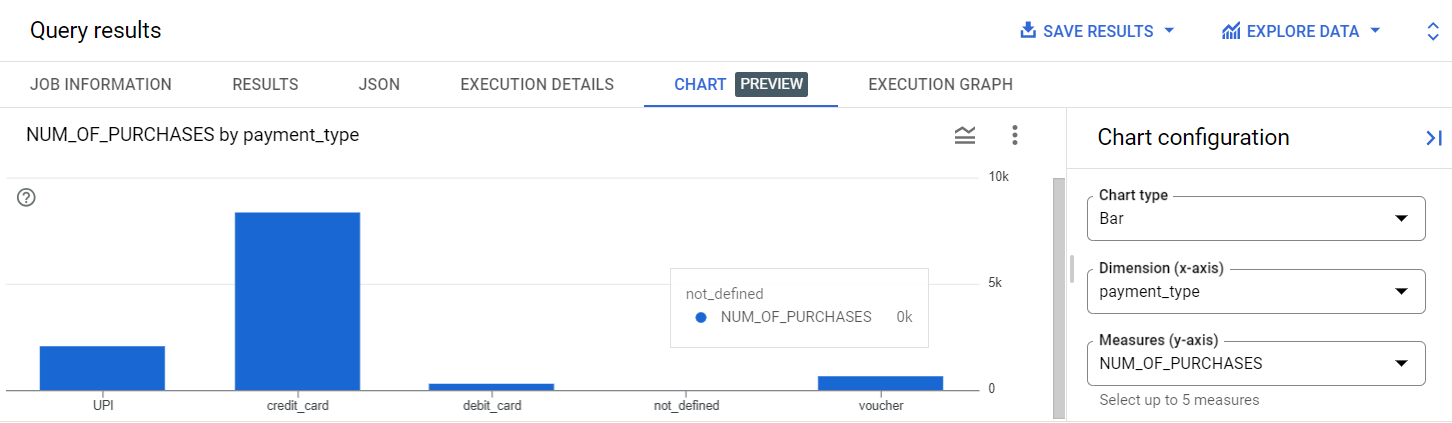
select payment\_type,MONTH,NUM\_OF\_PURCHASES from

(select DISTINCT payment\_type,MONTH,month\_of\_year,COUNT(\*) AS NUM\_OF\_PURCHASES

from cte\_6A AS NUM\_OF\_PURCHASES

group by payment\_type,MONTH,month\_of\_year

ORDER BY payment\_type,month\_of\_year)

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

with cte\_6B as (

select p.payment\_installments,o.order\_purchase\_timestamp

from `case\_study\_1.orders` o inner join `case\_study\_1.payments` p on p.order\_id=o.order\_id

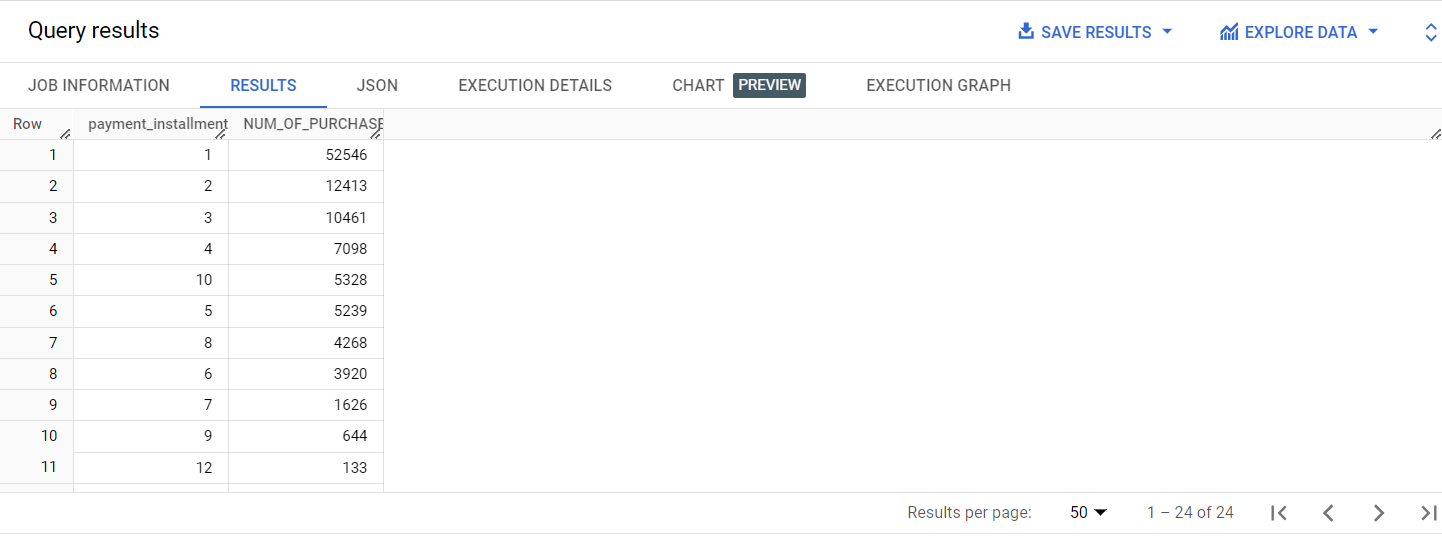
)

select DISTINCT payment\_installments,

COUNT(\*) AS NUM\_OF\_PURCHASES from cte\_6B

group by payment\_installments

ORDER BY NUM\_OF\_PURCHASES desc



**7. Actionable Insights & Recommendations**

This is an analysis carried out on purchases made from 2016 september to 2018 october by customers distributed in 5812 cities of 27 states .

1. Over the years the purchases has increased from 2017 to 2018 (data of all months not available in year 2016).
2. Customers are more likely to make purchases on the afternoon time of the day and least active at Dawn.
3. Based on the month on month analysis of of all states.It is found that there exist a seasonal pattern in purchase for many states with nearly steady purchases on initial period of year and then rapidly decline in the month of september after that slowly increases.
4. In brazil the most customers are located in the states of SP,RJ,MG and least in RR and AP.
5. There is nearly 130% increase in cost price over the year 2017 to2018 considering only months between January and August.
6. State SP has the maximun total order value and least average order value.
7. State RR have heighest average freight value and SP has the least.
8. Most fastest delivery is happening in the sates of SP,PR,MG and delivery is longest in RR,AP,AM.
9. SP,PR,MG have fastest delivery when compared to estimated delivery time.
10. Credit card is most preffered payment then comes UPI and debit card is least opted.