**Business Case study**

**Big Retail in Brazil**

**(2016-2018)**

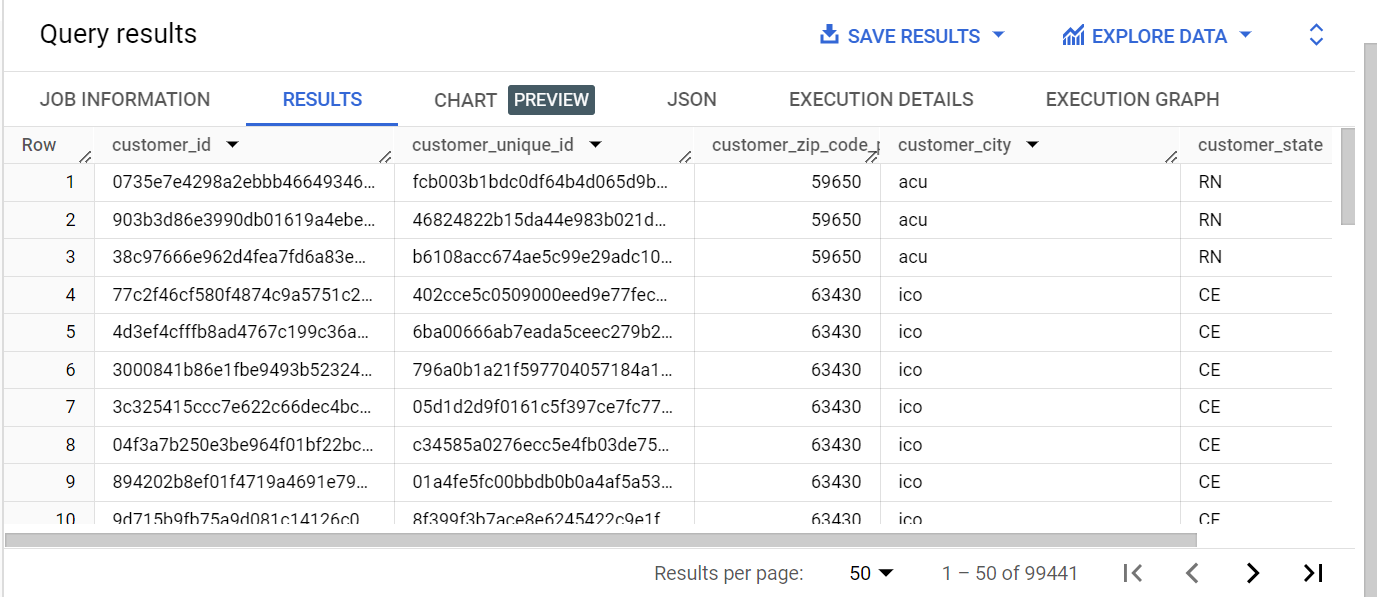
**1.1.Data type of all columns in the "customers" table.**

select \* from Retail.customers;

correct Answer:

select column\_name,data\_type from verdant-tempest-399415.Retail.INFORMATION\_SCHEMA.COLUMNS

WHERE table\_name = 'customers'

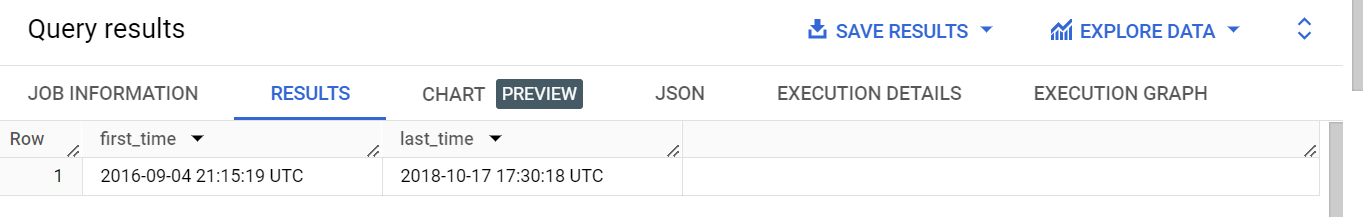


-------------------------------------------------------------------------------------------------------------------

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

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

max(order\_purchase\_timestamp) as last\_time from Retail.orders;



-------------------------------------------------------------------------------------------------------------------

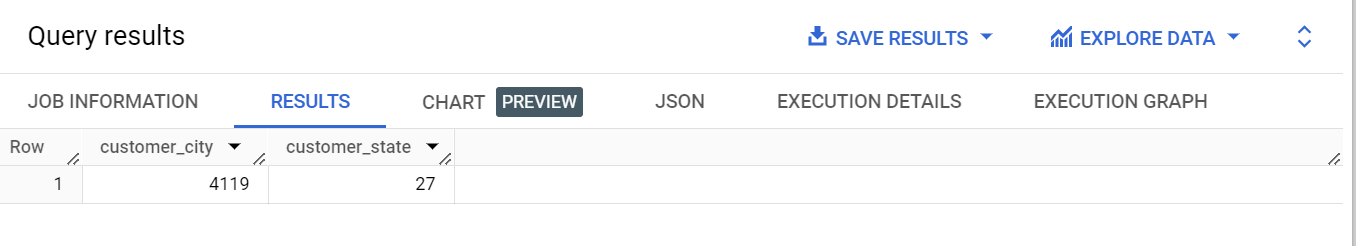
**1.3.Count the Cities & States of customers who ordered during the given period.**

select

count(distinct customer\_city) as customer\_city,

count(distinct customer\_state) as customer\_state

from Retail.customers;



-------------------------------------------------------------------------------------------------------------------

**2.In-depth Exploration:**

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

select

FORMAT\_TIMESTAMP('%Y-%m',order\_purchase\_timestamp) as month,

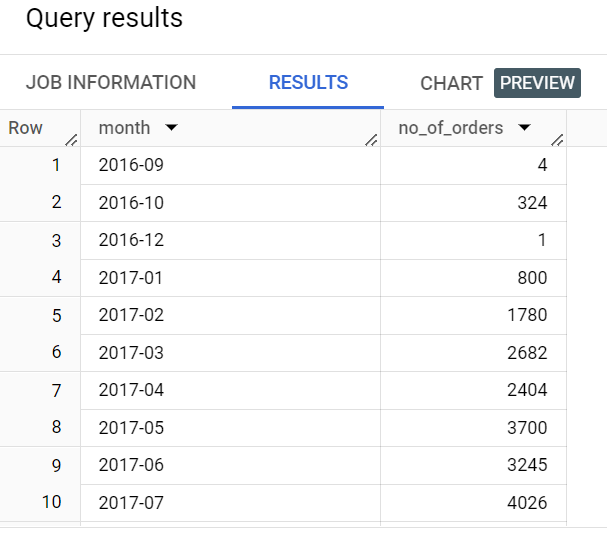
count(FORMAT\_TIMESTAMP('%Y-%m',order\_purchase\_timestamp)) as no\_of\_orders

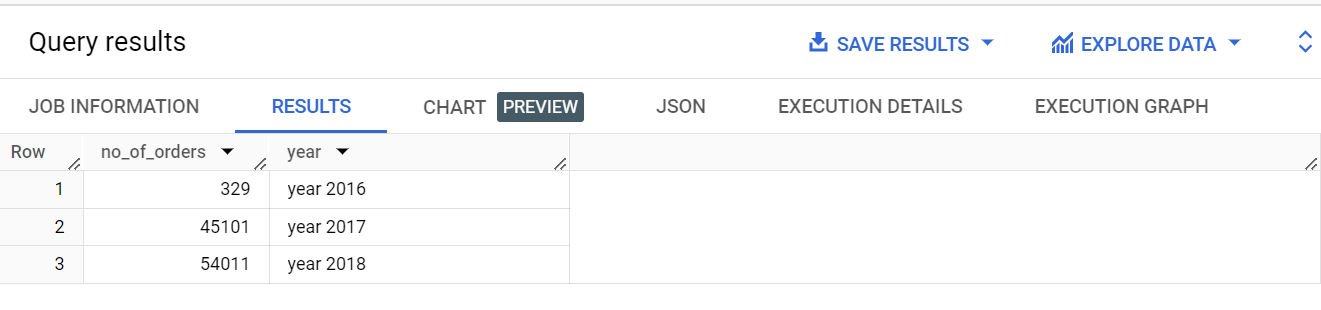
from Retail.orders

GROUP BY 1

ORDER BY 1

Ans: yes the orders are growing for the past years





-------------------------------------------------------------------------------------------------------------------

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

select

t.month,

t.no\_of\_orders

 from

(

select

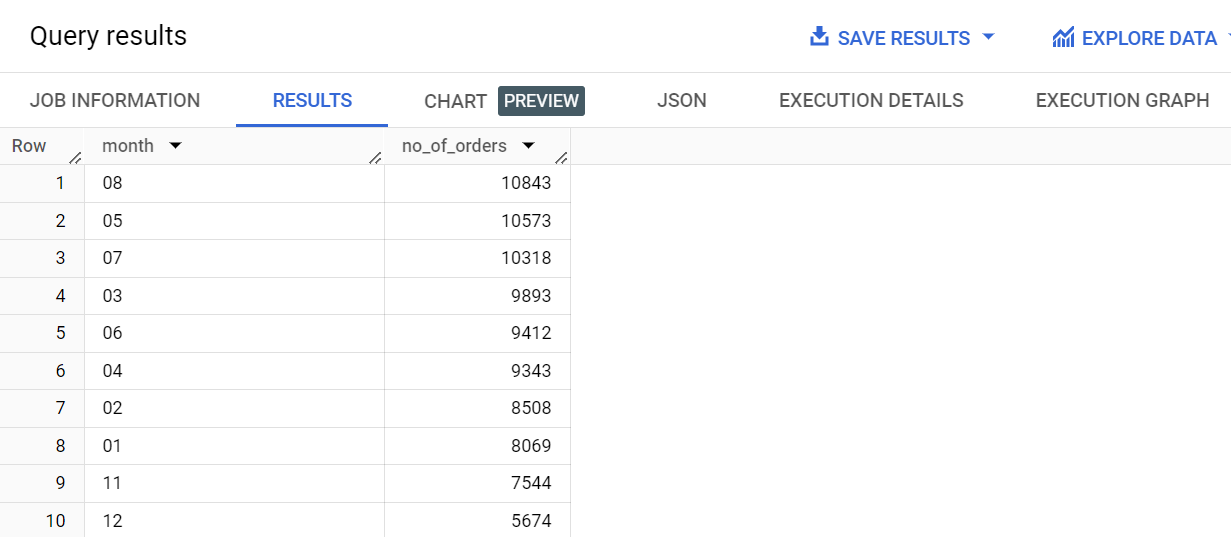
FORMAT\_TIMESTAMP('%m',order\_purchase\_timestamp) as month,

count(FORMAT\_TIMESTAMP('%m',order\_purchase\_timestamp)) as no\_of\_orders

from Retail.orders

group by FORMAT\_TIMESTAMP('%m',order\_purchase\_timestamp)) t

order by t.no\_of\_orders desc;



In month of “August,may,July” the purchase rate is high

--------------------------------------------------------------------------------------------------------------

**2.3 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 FORMAT\_TIMESTAMP('%H',order\_purchase\_timestamp) between '00' and '06' then 'Dawn'

   when FORMAT\_TIMESTAMP('%H',order\_purchase\_timestamp) between '07' and '12' then 'Mornings'

   when FORMAT\_TIMESTAMP('%H',order\_purchase\_timestamp) between '13' and '18' then 'Afternoon'

   when FORMAT\_TIMESTAMP('%H',order\_purchase\_timestamp) between '19' and '23' then 'Night'

end as Time\_of\_purchase,

count(\*) as no\_of\_orders

from Retail.orders

group by 1

order by no\_of\_orders;

A screenshot of a computer

Description automatically generated

Brazilian customers mostly place their orders in “Afternoon”

--------------------------------------------------------------------------------------------------------------

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

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

select

t.customer\_state,

t.month,

t.no\_of\_orders from (

select

c.customer\_state,

FORMAT\_TIMESTAMP('%Y-%m',o.order\_purchase\_timestamp) as month,

count(FORMAT\_TIMESTAMP('%Y-%m',order\_purchase\_timestamp)) as no\_of\_orders

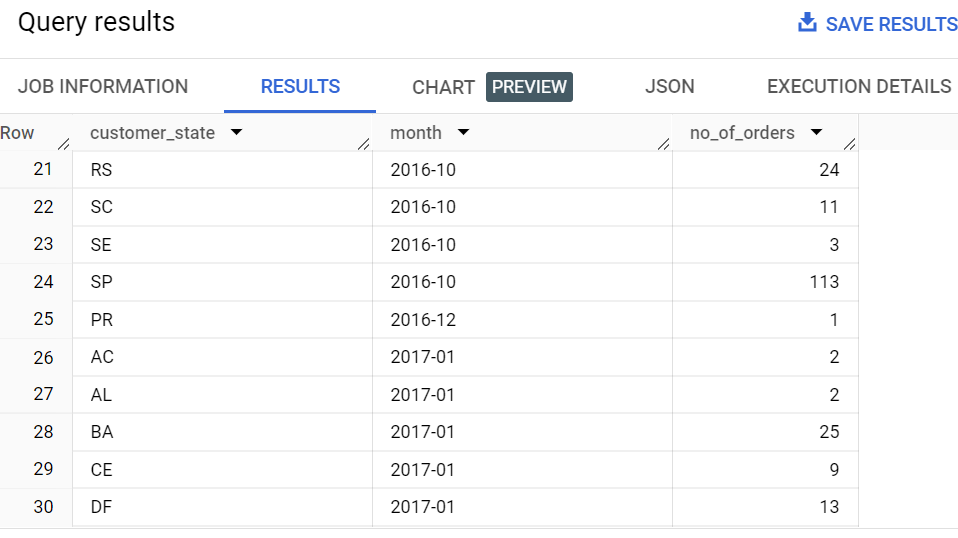
from Retail.customers c

left join Retail.orders o

using(customer\_id)

group by FORMAT\_TIMESTAMP('%Y-%m',order\_purchase\_timestamp),c.customer\_state)t

order by 2,1;



--------------------------------------------------------------------------------------------------------------

**3.2How are the customers distributed across all the states?**

select

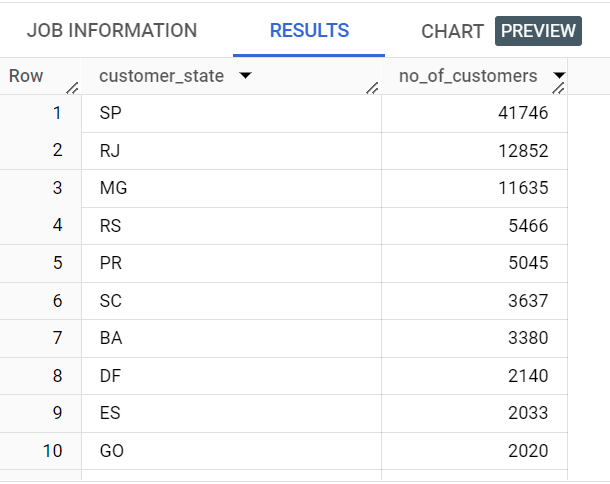
distinct customer\_state,

count(\*) as no\_of\_customers

from Retail.customers

group by customer\_state

order by no\_of\_customers desc;



Retail in Brazil have there customers more in ‘SP’,’RJ’,’MG’

--------------------------------------------------------------------------------------------------------------

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

**4.1Get 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.**

create view Retail.cte1 as

(select

year,

month,

payment\_value from

(select

FORMAT\_TIMESTAMP('%Y',o.order\_purchase\_timestamp) as year,

FORMAT\_TIMESTAMP('%m',o.order\_purchase\_timestamp) as month,

round(sum(p.payment\_value),2) as payment\_value

from Retail.payments p

join Retail.orders o

using(order\_id)

group by 1,2)t

where t.year between '2017' and '2018' and

t.month between '01' and '08'

order by t.month);

select year,month,percenatge\_increase from

(select

year,

month,

payment\_value,

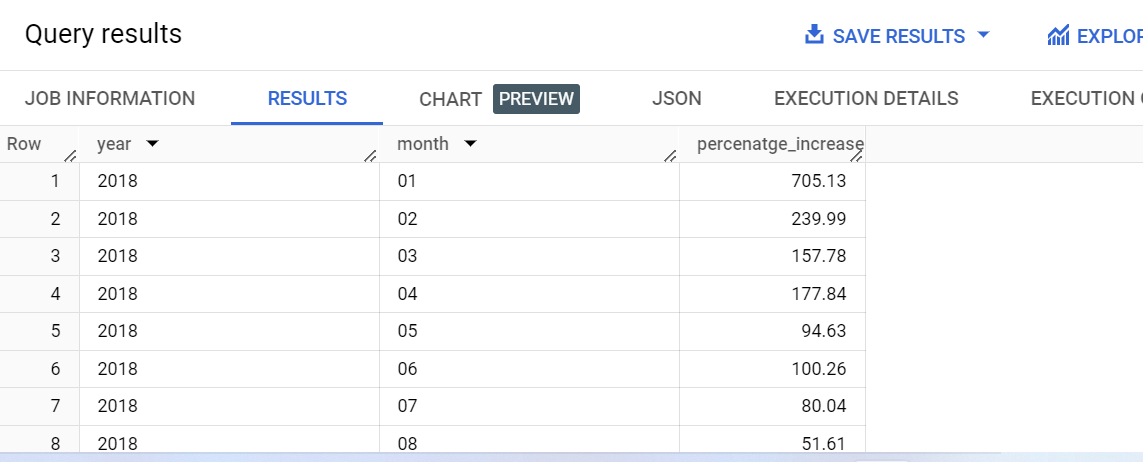
round((payment\_value - lead(payment\_value,1) over(partition by month order by payment\_value desc))

/(lead(payment\_value,1) over(partition by month order by payment\_value desc) \* .01),2) as percenatge\_increase

from Retail.cte1)t

where t.percenatge\_increase is not null

order by month;



--------------------------------------------------------------------------------------------------------------

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

select

c.customer\_state,

round(Sum(ot.price),2) as Total,

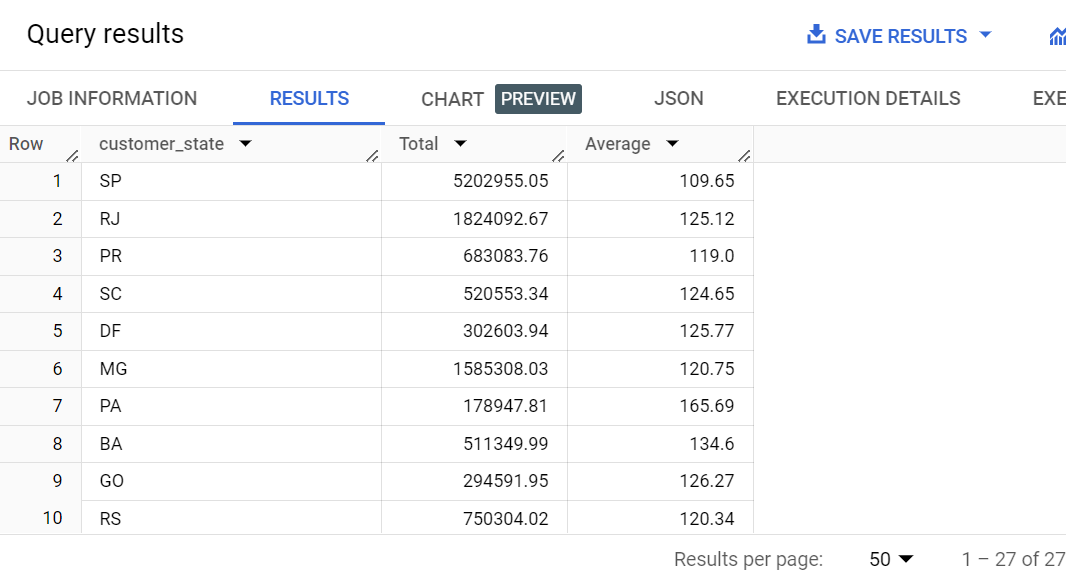
round(Sum(distinct ot.price)/count(distinct ot.price),2) as Average,

 from Retail.orders o

join Retail.order\_items ot using(order\_id)

left join Retail.customers c using (customer\_id)

group by c.customer\_state;



--------------------------------------------------------------------------------------------------------------

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

select

c.customer\_state,

round(Sum(ot.freight\_value),2) as Total\_freight,

round(Avg(ot.freight\_value),2) as Average\_freight,

 from Retail.orders o

join Retail.order\_items ot using(order\_id)

left join Retail.customers c using (customer\_id)

group by c.customer\_state;

A screenshot of a data

Description automatically generated

--------------------------------------------------------------------------------------------------------------

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

**5.1 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

distinct order\_id,

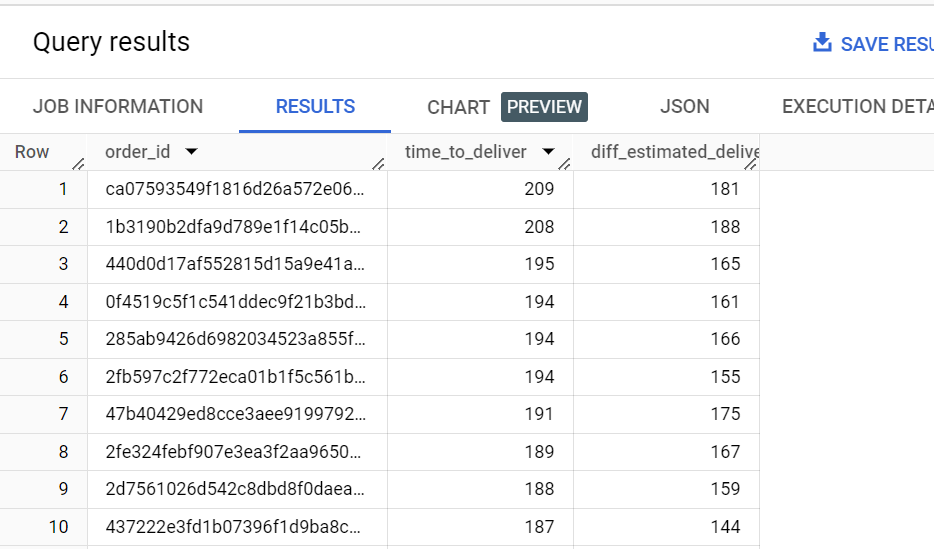
DATETIME\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) as time\_to\_deliver,

DATETIME\_DIFF(order\_delivered\_customer\_date,order\_estimated\_delivery\_date,DAY) as diff\_estimated\_delivery

from Retail.orders

where order\_status = 'delivered'

ORDER BY time\_to\_deliver DESC;



--------------------------------------------------------------------------------------------------------------

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

(select \* from (

select

c.customer\_state,

round(sum(freight\_value)/count(freight\_value),2) as Avg\_freight\_value,

dense\_rank() over(order by round(sum(freight\_value)/count(freight\_value),2) desc ) as rank

from Retail.orders o

join Retail.order\_items ot using(order\_id)

join Retail.customers c using(customer\_id)

group by c.customer\_state

order by Avg\_freight\_value desc) t

where t.rank <= 5)

union all

(select \* from

(select

c.customer\_state,

round(sum(freight\_value)/count(freight\_value),2) as Avg\_freight\_value,

dense\_rank() over(order by round(sum(freight\_value)/count(freight\_value),2)) as rank

from Retail.orders o

join Retail.order\_items ot using(order\_id)

join Retail.customers c using(customer\_id)

group by c.customer\_state

order by Avg\_freight\_value) t

where t.rank <=5)

order by Avg\_freight\_value;

**A screenshot of a data

Description automatically generated**

--------------------------------------------------------------------------------------------------------------

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

(select \* from

(select

c.customer\_state,

round(avg(DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_purchase\_timestamp,DAY)),2) as Avg\_delivery\_time,

dense\_rank() over(order by round(avg(DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_purchase\_timestamp,DAY)),2) desc ) as rank

from Retail.customers c

join Retail.orders o

using(customer\_id)

group by c.customer\_state

order by Avg\_delivery\_time desc) t

where t.rank <= 5)

union all

(select \* from

(select

c.customer\_state,

round(avg(DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_purchase\_timestamp,DAY)),2) as Avg\_delivery\_time,

dense\_rank() over(order by round(avg(DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_purchase\_timestamp,DAY)),2)) as rank

from Retail.customers c

join Retail.orders o

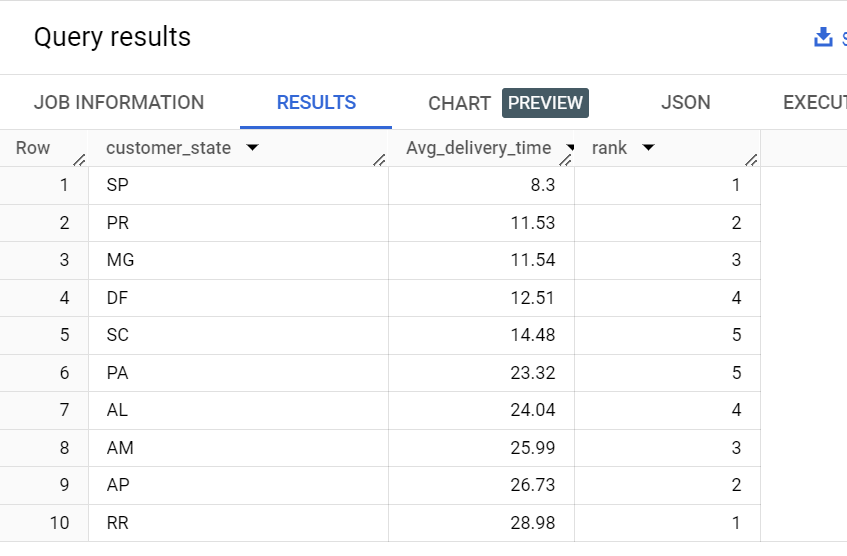
using(customer\_id)

group by c.customer\_state

order by Avg\_delivery\_time) t

where t.rank <= 5)

order by Avg\_delivery\_time;

****

--------------------------------------------------------------------------------------------------------------

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

select \* from

(select

c.customer\_state,

DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_estimated\_delivery\_date,DAY) as delivery\_speed,

dense\_rank() over(order by DATETIME\_DIFF(o.order\_delivered\_customer\_date,o.order\_estimated\_delivery\_date,DAY)) as rank

from Retail.customers c

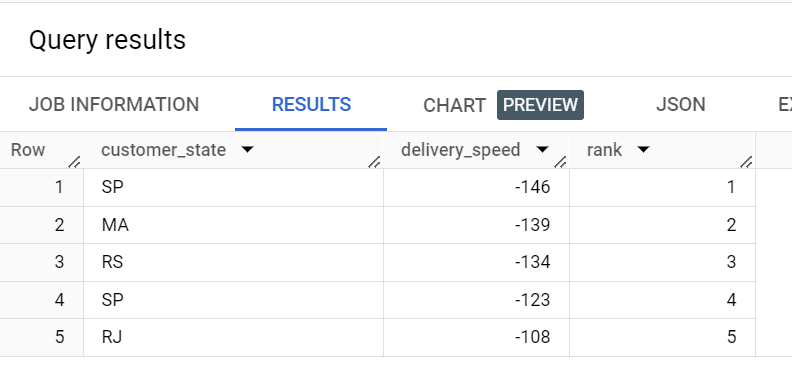
join Retail.orders o

using(customer\_id)

order by delivery\_speed) t

where t.delivery\_speed is not null

and t.rank<=5;



--------------------------------------------------------------------------------------------------------------

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

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

select

FORMAT\_TIMESTAMP('%m',o.order\_purchase\_timestamp) as month,

p.payment\_type,

count(distinct o.order\_id) as no\_of\_orders

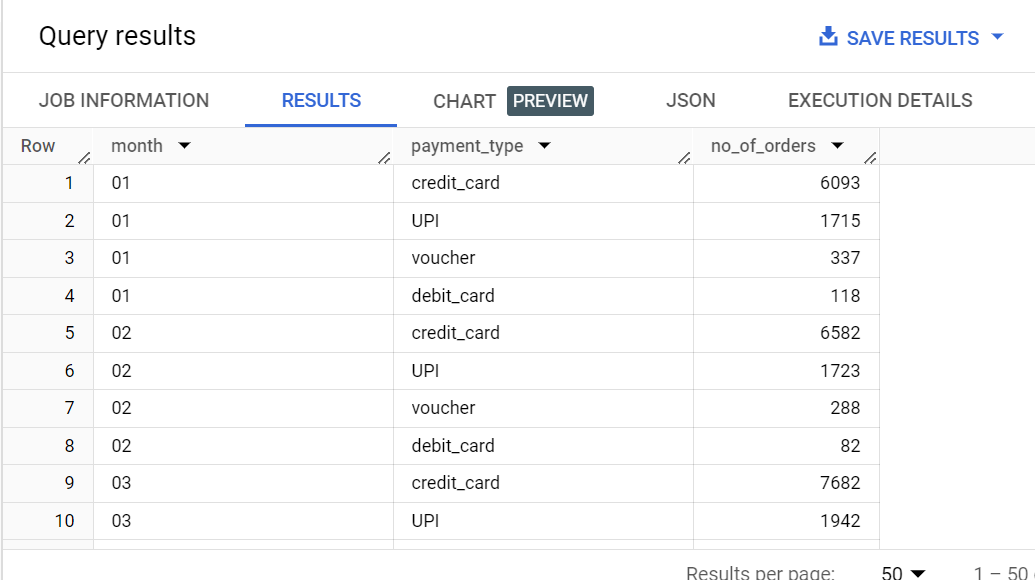
from Retail.payments p

join Retail.orders o

using(order\_id)

group by 2,1

order by month,no\_of\_orders desc;

****

--------------------------------------------------------------------------------------------------------------

**6.2Find the no. of orders placed on the basis of the payment installments that have been paid.**

select

payment\_installments,

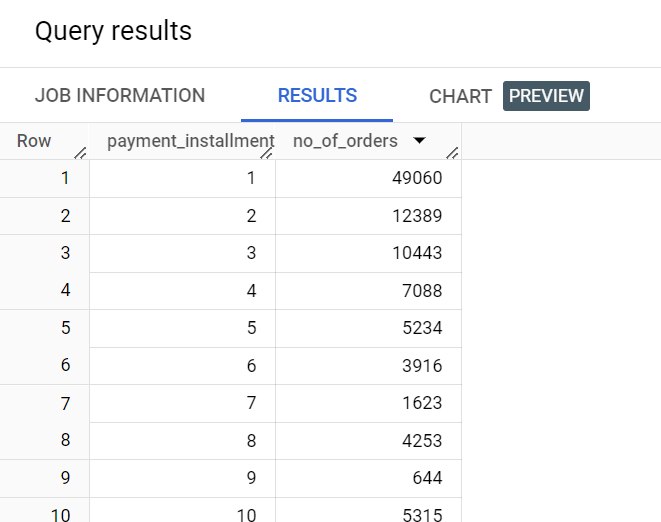
count(distinct order\_id) as no\_of\_orders

from Retail.payments

where payment\_installments > 0 and

payment\_sequential > 0

group by payment\_installments;

****

**Insights from the data**

* [São Paulo](https://en.wikipedia.org/wiki/S%C3%A3o_Paulo_(state)), Gerais, Rio have the more number of orders placed by the customer, So the keep increasing the outlet in these areas can able to attract customers, boost sales, and enhance the brand image.
* Promote with more activities and attract customers with more privilege at afternoon.
* Retails can establish more discount and best price over the month period from May to August every year.
* Freight insurance policy can implement in the high revenue states and continued to all other secondary revenue states.

### PA,AL,AM,AP,RR are the state having high delivery time, so we can keep local warehouse and Inventory Management

* for these regions to reduce the delivery time.
* Since more customers chose credit card as their mode of payment, we can offer incentives such as airline miles, hotel room rentals, gift certificates and cash back on purchases.
* And we can keep a credit card outlet to get new credit cards for customer who have not use credit card as their mode of payment.

**End of Case study**