**E Context:**

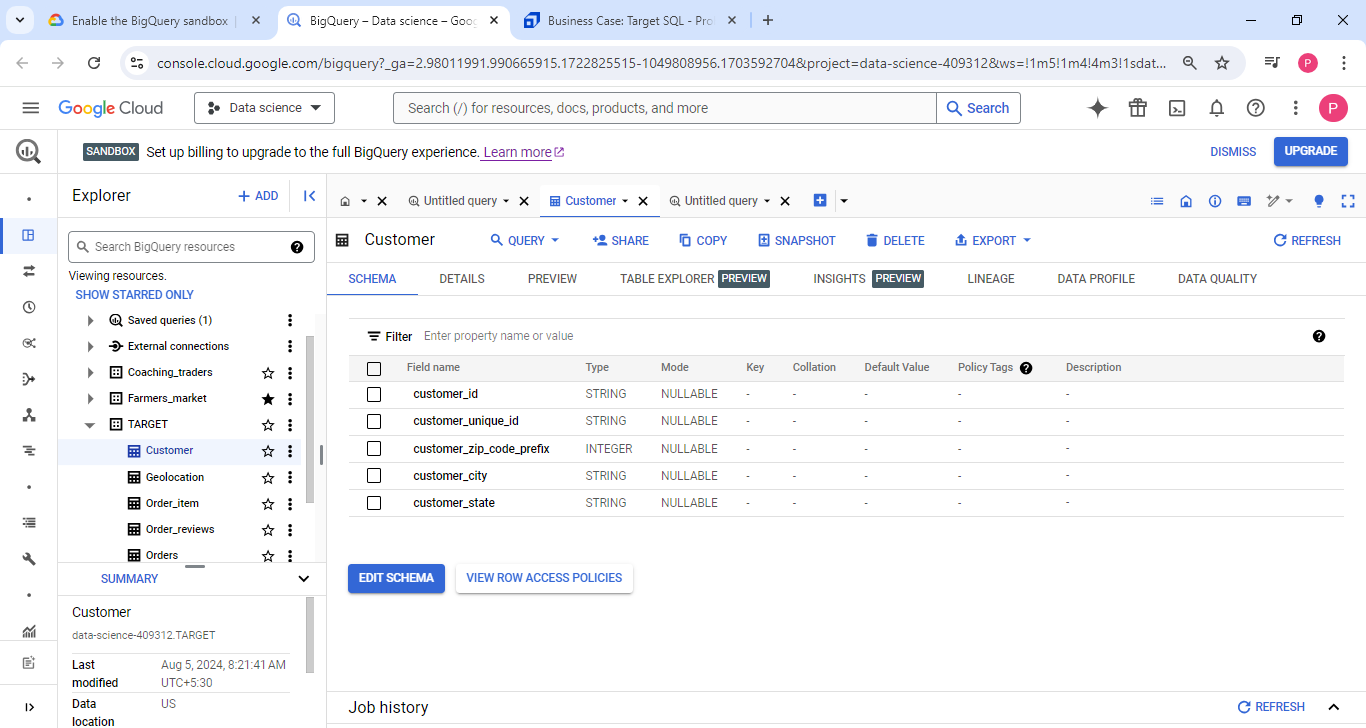
**Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.**

**This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.**

**By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels**

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

**Q1.** **Data type of all columns in the "customers" table :**

****

**INFERENCE:** As we can see from the above schema of “customers” table there are 5 columns present & there

datatypes are-

**“customer\_zip\_code\_prefix”-** It is the only column which contain **Integer** data type values out

of 5 columns.

**“customer\_id”,”customer\_unique\_id”,”customer\_city”,”customer\_state”**- All these 4 columns are having

same data type that is **String.**

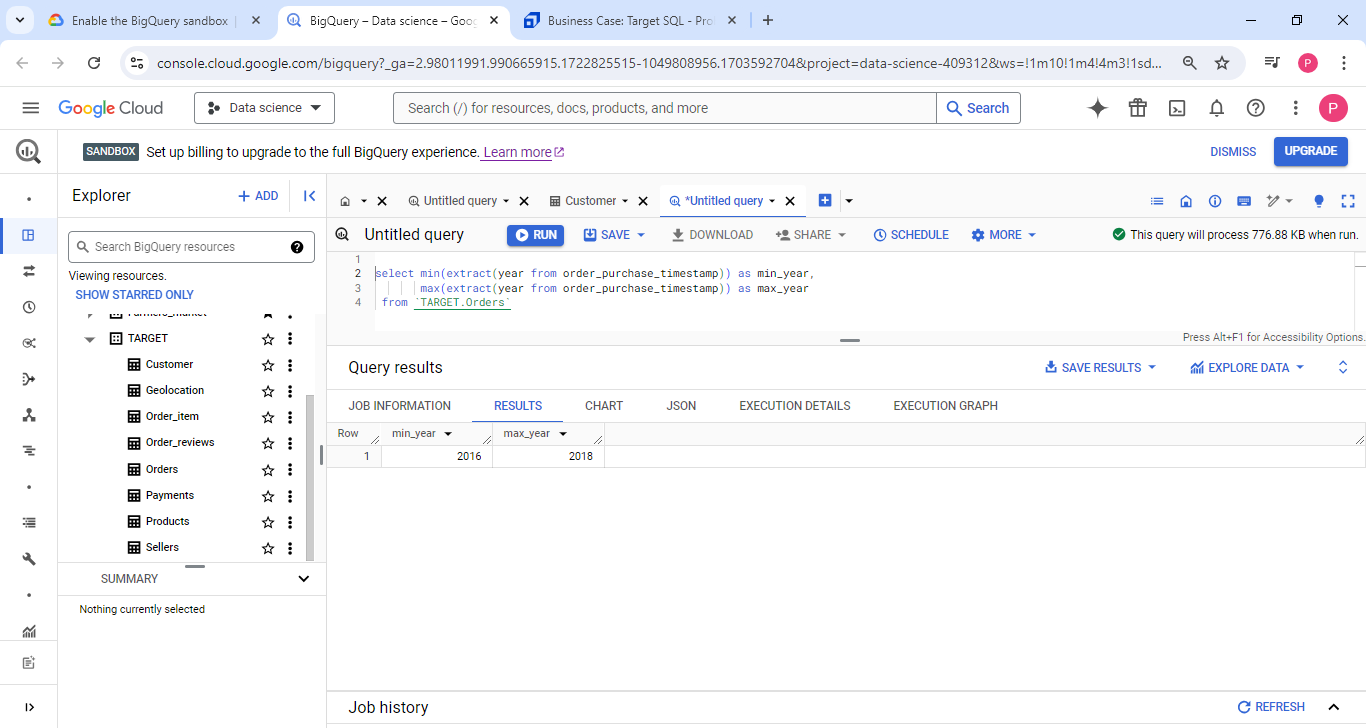
**Q2. Get the time range between which the orders were placed.**

**Query-**

select min(extract(year from order\_purchase\_timestamp)) as min\_year,

       max(extract(year from order\_purchase\_timestamp)) as max\_year

 from `TARGET.Orders`



**Inference:** From the above query we can cleary say that the orders placed in the time range between 2016 & 2018 & also in our given context it has been clearly mentioned that the orders placed between 2016 & 2018.

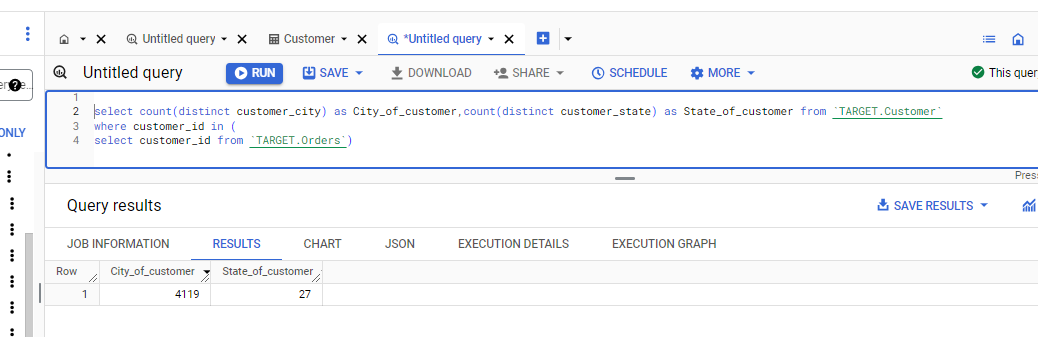
**Q3.** **Count the Cities & States of customers who ordered during the given period.**

**Query-**

select count(distinct customer\_city) as City\_of\_customer,count(distinct customer\_state) as State\_of\_customer from `TARGET.Customer`

where customer\_id in (

select customer\_id from `TARGET.Orders`)



**Inference:** From the above query we can observe that there are total **4119** nos of cities & **27** nos of states from where the Customers made the order during the time range of 2016 to 2018.

**2. In-depth Exploration:**

**Q1.Is there a growing trend in the no. of orders placed over the past years?**

**Query:**

with cte as (

select extract(year from  order\_purchase\_timestamp) as year, extract(month from  order\_purchase\_timestamp) as month,count(order\_id) count\_of\_ordermonthwise

 from `TARGET.Orders`

 group by 1,2

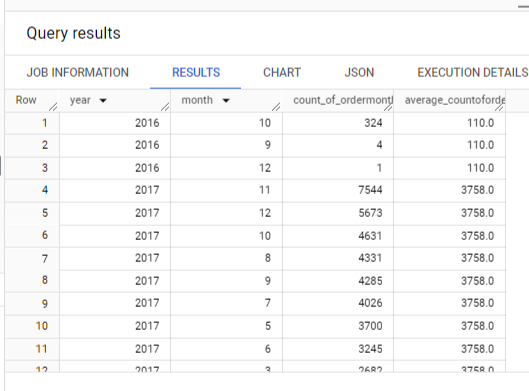
 order by 1,2)

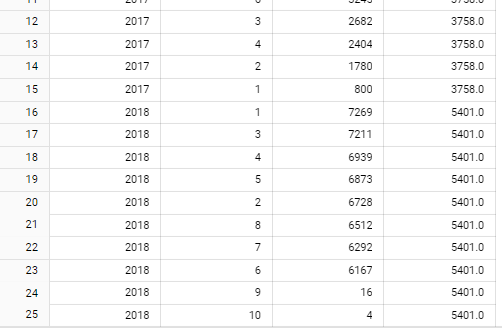
 select year,month,count\_of\_ordermonthwise,

 round(avg(count\_of\_ordermonthwise)over(partition by year)) as average\_countoforder\_yearwise

 from cte

 order by 1,3 desc





**Inference:** If we notice the year wise average count of order from the column (“average\_countoforder”)

We can see that it is increasing year wise, but when we focus year wise monthly count of order , we can notice

These following observations:

1. In **2016-sep** the order count was extremely low –it may be because Target just started its operation in Brazil in 2016.so initially it may be due to lack of proper setup or customer recognition the sales were low.but if we focus on “**2017 –sep**” there are a significant amount of order can be seen. But “2018-sep” again it is low, because Target was closing its operation this year, so it may not taking any extra order.
2. In **October-November-December**—we can observe progressive amount of order, It may due to the famous festivals in Brazil like Black Friday, Christmas, New Year's Eve which drive significant retail activity in Brazil.
3. Rest of the months we can see the orders amount increased year & month wise.

**Q2.** **Can we see some kind of monthly seasonality in terms of the no. of orders being placed?**

Yes ! In **October-November-December**—we can observe progressive amount of order, It may due to the famous

festivals in Brazil like Black Friday, Christmas, New Year's Eve which drive significant retail activity in Brazil.

In “**2017-November”**  orders increased significantly.

**Q3.** **During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)—**

**Query-**

with cte as (

select order\_id,order\_purchase\_timestamp,extract(year from order\_purchase\_timestamp) as year,cast(format\_timestamp("%H",timestamp(order\_purchase\_timestamp)) as int64) as time

 from `TARGET.Orders`),

 cte2 as (

 select \*,case when time>=0 and time <7

          then "Dawn"

          when time >=7 and time <13

          then "Mornings"

          when time>=13 and time <18

          then "Afternoon"

          else "Night"

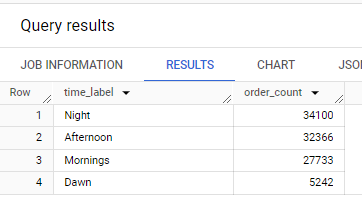
          end as time\_label

          from cte)

select time\_label,count(order\_id) as order\_count from cte2

group by 1

order by 2 desc



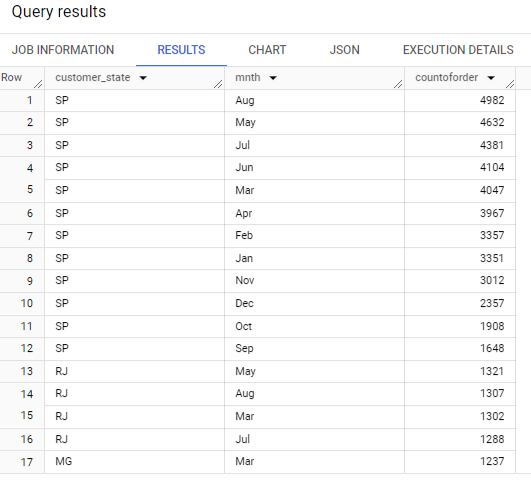
**Inference:**

In the upper table we can see a clear pattern that in every the Brazilian customers mostly place their order at “**Night**” **& “Afternoon” & there is a good amount order placed in morning also”**

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

**Q1.** **Get the month on month no. of orders placed in each state.**

with cte as (

select format\_datetime("%b",a.order\_purchase\_timestamp) as mnth,b.customer\_state,a.order\_id

from `TARGET.Orders` a

left join `TARGET.Customer` b

on a.customer\_id=b.customer\_id)

select customer\_state,mnth,count(order\_id) as countoforder

 from cte

 group by 1,2

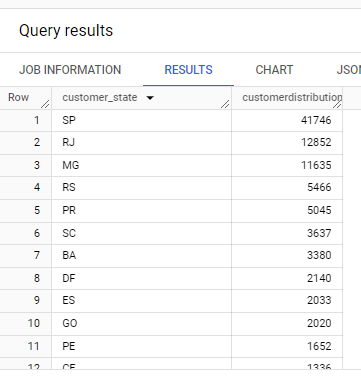
 order by 3 desc

**Inference:**

In the above result , it has been shown the numbers of order placed in each state from highest to lowest for each state.

“SP” is the state where the highest amount of order has been placed.

**Q1.** **How are the customers distributed across all the states:**

select customer\_state,count(customer\_id) as customerdistribution

from `TARGET.Customer`

GROUP BY 1

order by 2 desc

**Inference:** From above result it can be seen how the customers are distributed over states. E.g.-“SP” state has the highest numbers of customers etc.

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

**Q1.** **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 to get the cost of orders table--** with cte as (

select a.year,sum(payment\_value) as ordercost

from (select order\_id ,year,mnth from

(select order\_id,cast(extract(year from order\_purchase\_timestamp) as int64) as year,extract(month from order\_purchase\_timestamp) as m,format\_datetime("%b",order\_purchase\_timestamp) as mnth

from `TARGET.Orders`) v

where  year >=2017 and year<=2018 and m>=1 and m<=8) as a

join `TARGET.Payments`  as b

on a.order\_id=b.order\_id

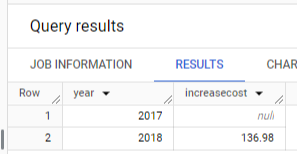
group by 1

order by 1)

select year,round(((ordercost-lag(ordercost,1)over(order by ordercost))

                    /(lag(ordercost,1)over(order by ordercost)))\*100,2) as increasecost

                    from cte

****

**Inference**: From the above result we can easily see that there are 136.98 approx **137**% order cost increased from the year “2017” to “2018”.

**Q2**. **Calculate the Total & Average value of order price for each state**

**Query:**

select c.customer\_state,round(sum(price),2) as total ,round(avg(price),2) as average

from `TARGET.Customer` c

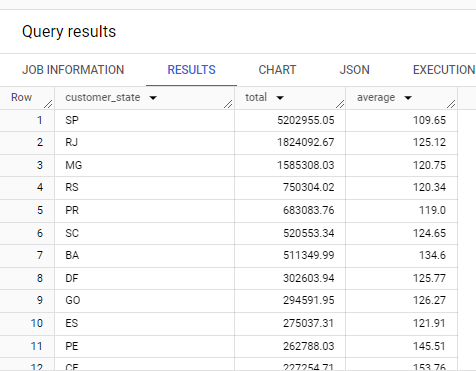
join `TARGET.Orders` o

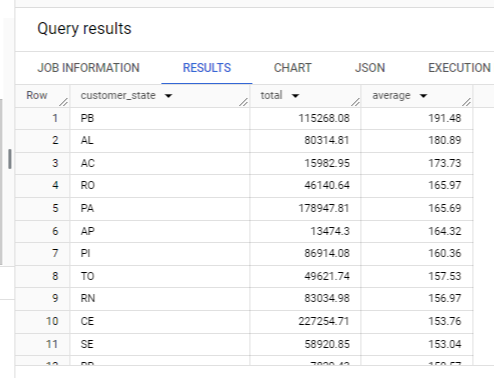
on c.customer\_id=o.customer\_id

join `TARGET.Order\_item` ot

on o.order\_id=ot.order\_id

group by 1





**Inference:** There are two types result we can see the following observations--  
 In the first case where state wise order total price from high to low -- we observe that the state “**SP**” has the

highest Order price but in the 2nd case where state wise order average price from high to low – we observe that

the state “**PB**” has the highest average order price .

**Q3.** **Calculate the Total & Average value of order freight for each state**

select c.customer\_state,round(sum(freight\_value),2) as total ,round(avg(freight\_value),2) as average

from `TARGET.Customer` c

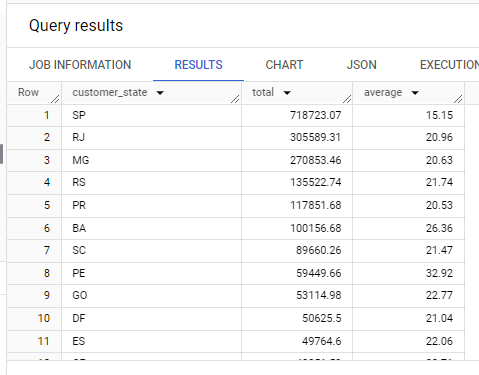
join `TARGET.Orders` o

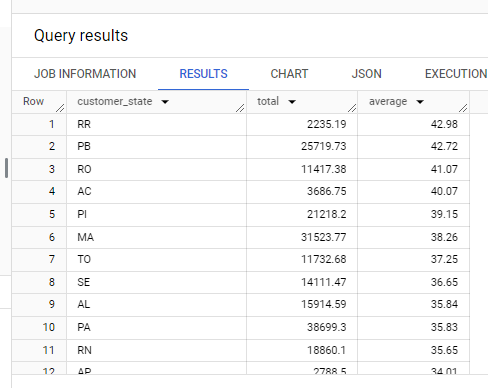
on c.customer\_id=o.customer\_id

join `TARGET.Order\_item` ot

on o.order\_id=ot.order\_id

group by 1

****



**Inference:** There are two types result we can see the following observations--  
 In the first case where state wise total order freight value from high to low -- we observe that the state “**SP**” has

the highest Order price but in the 2nd case where state wise order average order freight value from high to low –

we observe that the state “**RR**” has the highest average order price .

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

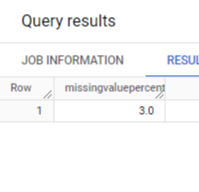
**Q1**. **Find the no. of days taken to deliver each order from the order’s purchase date as delivery time.  
Also, calculate the difference (in days) between the estimated & actual delivery date of an order**

**Query:**

**Checking Null values**-

 select round((count(order\_id)/(select count(order\_id) from `TARGET.Orders` ))\*100) as missingvaluepercentage from `TARGET.Orders`

 where order\_delivered\_customer\_date is null



There are only **3%** missing values present in “**order\_delivered\_customer\_date”** column , so it can be ignored.

**Main query:**

with cte as (select \* from `TARGET.Orders`

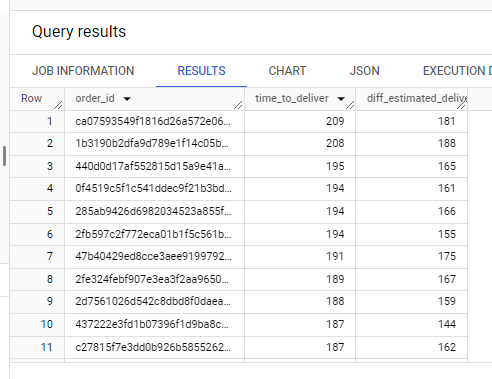
where order\_delivered\_customer\_date is not null)

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

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

 from cte

order by 2 desc



**Inference:** From the upper result we can identify that where delivery time is arranged from high to low we can see that

the highest delivery date of an product is “209” days but it there is a major difference between the delivery days & estimated days .

**Q2.** **Find out the top 5 states with the highest & lowest average freight value**

**Query:**

with cte as (

select c.customer\_state,round(avg(freight\_value),2) as average

from `TARGET.Customer` c

join `TARGET.Orders` o

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

join `TARGET.Order\_item` ot

on o.order\_id=ot.order\_id

group by 1)

select top5state\_highestavgfreightvalue,top5state\_lowestavgfreightvalue

from

(select customer\_state as top5state\_highestavgfreightvalue,

       row\_number()over(order by average desc) as r

       from cte

order by average desc

limit 5) a

join

(select customer\_state as top5state\_lowestavgfreightvalue,

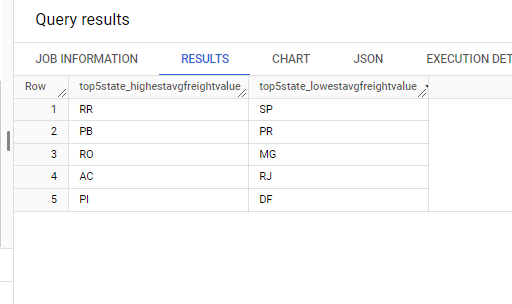
       row\_number()over(order by average) as r

       from cte

order by average

limit 5) b

on a.r=b.r



**Inference:** Here we can see the states with highest average freight values are –[RR,PB,RO,AC,P1]

& the states with lowest average freight values are –[SP,PR,MG,RJ,DF]

**Q4: Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**

**Query:**

with cte as (

select customer\_state, avg(diff\_estimated\_delivery) as average\_deliverydiff

from (select order\_id,

      customer\_id,

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

 from `TARGET.Orders`

 where order\_delivered\_customer\_date is not null

) a

join `TARGET.Customer` b

on a.customer\_id=b.customer\_id

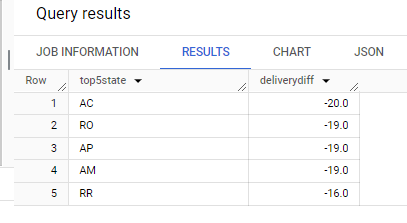
group by 1)

select customer\_state as top5state,round(average\_deliverydiff) as deliverydiff

from cte

order by average\_deliverydiff

limit 5

****

**Inference:** From above query , it can inferred that the top 5 state [AC,RO,AP,AM,RR] where the order delivery is really fast than their estimated delivery . For state AC the actually delivery done 20 days before the estimated delivery.

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

**Q1.** **Find the month on orders placed using different payment types.month no. of**

**Query:**

select mnth,payment\_type,count\_of\_order

from(select format\_timestamp("%b",order\_purchase\_timestamp) as mnth,format\_date("%m",order\_purchase\_timestamp),payment\_type,count(a.order\_id) as count\_of\_order

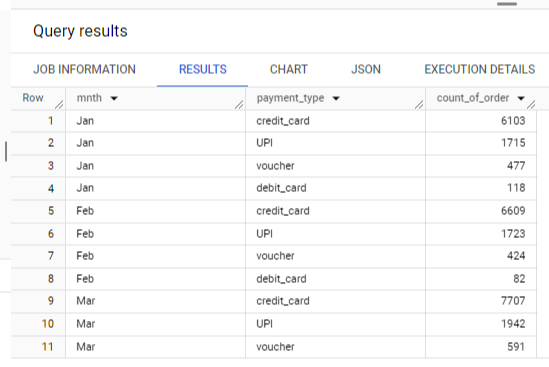
from `TARGET.Orders` a

join `TARGET.Payments` b

on a.order\_id=b.order\_id

group by 1,2,3

order by 2,4 desc)

****

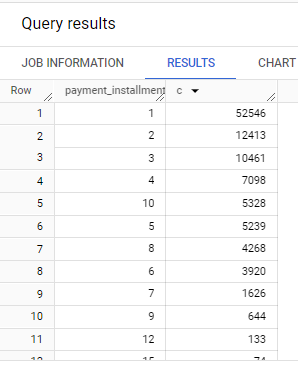
**Inference:** Here we can clearly observe a pattern that on each month most orders have placed using credit card then upi & less orders placed using voucher & debit card.

**Q2. Find the no. of orders placed on the basis of the payment installments that have been paid**

select payment\_installments,count(order\_id) as c from `TARGET.Payments`

group by 1

order by 2 desc

****

**Inference:** Here “c” column denoting the count of order for each payment installment. We can see that most of orders placed for the lowest amount of installments. When this installments increases the count of order decreases.