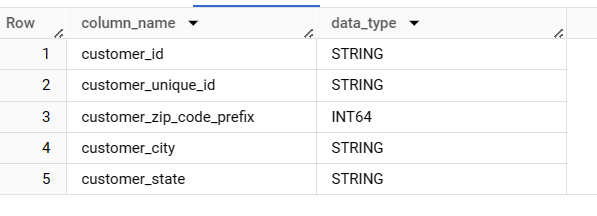
**Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**

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

SELECT column\_name, data\_type FROM

business\_case\_studies.INFORMATION\_SCHEMA.COLUMNS

where table\_name = "customers";



INSIGHTS:

From the customers table, most of the columns are of string type.

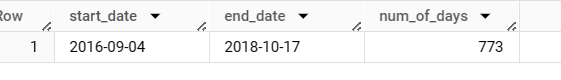
Recommendation: NA

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

SELECT extract(date from min(order\_purchase\_timestamp)) as start\_date , extract(date from max(order\_purchase\_timestamp)) as end\_date,

date\_diff(extract(date from max(order\_purchase\_timestamp)),extract(date from min(order\_purchase\_timestamp)),day) as num\_of\_days

from `business\_case\_studies.orders`



INSIGHTS:

In the given data, the first order is placed on ‘2016-09-04’

And the last order is placed on ‘2018-10-17’.

We have data of 773 days.

Recommendation: NA

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

Here given period is not given in the question, so assuming the period between the start date and end date from 2nd question.

select count(distinct customer\_city) as city\_count, count(distinct customer\_state) as state\_count from `business\_case\_studies.orders` o join

`business\_case\_studies.customers` c on o.customer\_id = c.customer\_id

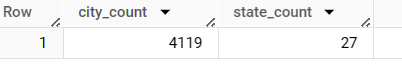
where extract(date from o.order\_purchase\_timestamp) between

(SELECT extract(date from min(order\_purchase\_timestamp)) as start\_date from `business\_case\_studies.orders`)

and

(SELECT extract(date from max(order\_purchase\_timestamp)) as end\_date

from `business\_case\_studies.orders`);



INSIGHTS:

In the given data, customers that are placed orders are from 27 states which has 4119 cities.

Recommendation: NA

**IN DEPTH EXPLORATION**

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

with t1 as(select \*, lag(total\_orders) over(order by year) as lag\_orders from

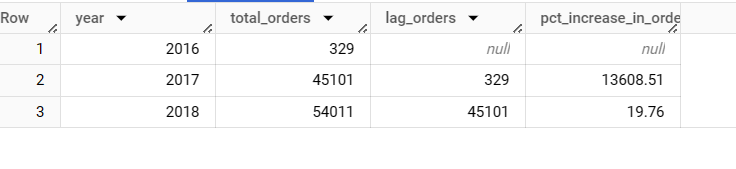
(select extract(year from order\_purchase\_timestamp) as year, count(order\_id) as total\_orders from `business\_case\_studies.orders`

group by year

order by year) order  by year)

select \*, round(((t1.total\_orders-lag\_orders)/lag\_orders)\*100,2) as pct\_increase\_in\_orders from t1

output:



Insights:

We have increasing trend over the years.

from 2016 to 2017, there is a large increase in orders placing by more than 13000%

and from 2017 to 2018, the increase pct in orders is around 20%.

We don’t have the data of orders in most of 2016. While decision making it will be not appropriate.

Recommendation: NA

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

with t1 as

(

  select distinct extract(month from order\_purchase\_timestamp) as month, format\_date("%B",order\_purchase\_timestamp) as month\_name,

count(order\_id) as orders\_for\_2016 from `business\_case\_studies.orders`

where extract(year from order\_purchase\_timestamp)=2016

group by month,month\_name

),

t2 as

(

  select distinct extract(month from order\_purchase\_timestamp) as month, format\_date("%B",order\_purchase\_timestamp) as month\_name,

count(order\_id) as orders\_for\_2017 from `business\_case\_studies.orders`

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

group by month,month\_name

),

t3 as

(

  select distinct extract(month from order\_purchase\_timestamp) as month, format\_date("%B",order\_purchase\_timestamp) as month\_name,

count(order\_id) as orders\_for\_2018 from `business\_case\_studies.orders`

where extract(year from order\_purchase\_timestamp)=2018

group by month,month\_name

),

t4 as(

select distinct extract(month from order\_purchase\_timestamp) as month, format\_date("%B",order\_purchase\_timestamp) as month\_name,

count(order\_id) as orders\_for\_allYears from `business\_case\_studies.orders`

group by month,month\_name

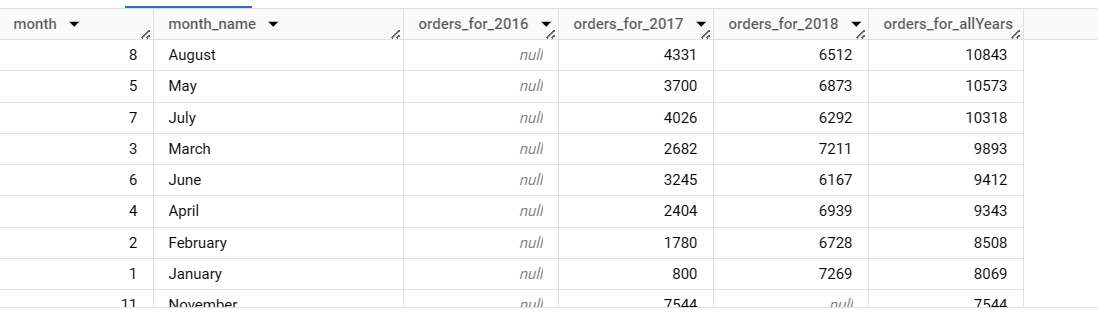
order by month

)

select t4.month, t4.month\_name,orders\_for\_2016, orders\_for\_2017,orders\_for\_2018, orders\_for\_allYears from t4 left join t3 on

t4.month\_name = t3.month\_name left join t2 on t2.month\_name=t4.month\_name left join t1 on t1.month\_name = t4.month\_name

order by orders\_for\_allYears desc;



INSIGHTS:

By seeing the output, we can say that the most of the orders are placed in the month of Aug, May, July.

Less orders are placed in the back end of the year ie. Dec, nov, oct etc

Recommendation:

The company should try to provide offers at the back end of the year

To get the customers attention and make them to engage at that point of time.

1. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)
   1. 0-6 hrs : Dawn
   2. 7-12 hrs : Mornings
   3. 13-18 hrs : Afternoon
   4. 19-23 hrs : Night

with t1 as(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"

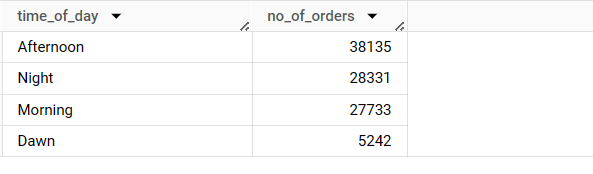
else "Night"

end as time\_of\_day   from `business\_case\_studies.orders`)

select time\_of\_day, count(time\_of\_day) as no\_of\_orders from t1

group by time\_of\_day

order by no\_of\_orders desc;



INSIGHT:

Most of the orders coming in the afternoon followed by night, morning and dawn.

Recommendation:

This data can be useful to the organisation for cost cutting.

If the company wants to decrease the spending, then they can use the resources more at the peak time which is afternoon.

And they can reduce the resources at the dawn as we have very less orders

Comparatively.

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

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

with t1 as(select distinct c.customer\_state, extract(month from order\_purchase\_timestamp) as month, format\_date("%B",order\_purchase\_timestamp) as month\_name,

count(order\_id) over(partition by customer\_state,extract(month from order\_purchase\_timestamp) order by extract(month from order\_purchase\_timestamp)) as no\_of\_orders from `business\_case\_studies.orders` o

join `business\_case\_studies.customers` c using(customer\_id)),

t2 as(

select customer\_state, month, month\_name, no\_of\_orders,

lag(no\_of\_orders) over(partition by customer\_state order by month) as prev\_month\_orders

from t1)

select \*, (no\_of\_orders-prev\_month\_orders) as month\_on\_month\_orders,

sum(no\_of\_orders) over(partition by customer\_state order by month range between unbounded preceding and current row) as cumulative\_orders from t2

order by customer\_state, month



INSIGHTS:

We can see the cumulative data and comparison of data from one month to the previous months. So that we can analyse on the decision making that has taken

Or any impact of certain events in the respective months which is responsible for either increase of orders or decrease of orders.

And in the above case, we are seeing more orders in the months of April and May as

They have semana santa festival.

Our orders increased in the middle of the year mostly.

Recommendation:

At the back end of the year the organisation must try to engage with the customers more as we are seeing less orders comparatively with the middle of the year and starting of the year.

1. How are the customers distributed across all the states?

with t as

(select distinct customer\_state, count(\*) as num\_of\_cust\_from\_state from `business\_case\_studies.customers`

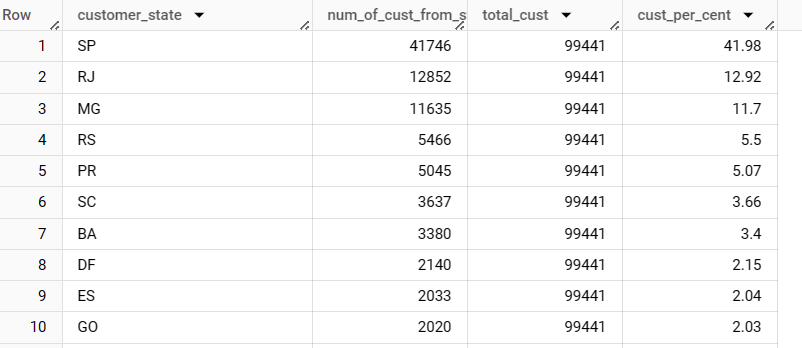
group by customer\_state order by num\_of\_cust\_from\_state desc),

t2 as

(select \*, sum(num\_of\_cust\_from\_state) over() as total\_cust from t)

select \*, round((t2.num\_of\_cust\_from\_state\*100/t2.total\_cust),2) as cust\_per\_cent from t2

order by t2.num\_of\_cust\_from\_state desc;



INSIGHTS:

Most of the customers are from SP which consists of 42% of the total customers

And SP, RJ and MG consists of 65% percent of organisations customers.

Recommendation:

Check what is lacking in the states with lowest customers compared with the states having highest customers and take decisions according.

**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).

select

round(sum(payments\_in\_2017),2) as total\_payment\_2017,

round(sum(payments\_in\_2018),2) as total\_payment\_2018,

round(((sum(payments\_in\_2018)-sum(payments\_in\_2017))/sum(payments\_in\_2017))\*100,4) as increase\_in\_percentage

from

(SELECT

case when extract(year from order\_purchase\_timestamp)=2017 then payment\_value else 0 end as payments\_in\_2017,

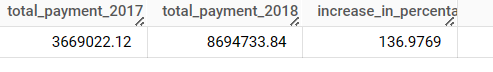
case when extract(year from order\_purchase\_timestamp)=2018 then payment\_value else 0 end as payments\_in\_2018,

FROM `business\_case\_studies.orders` O JOIN `business\_case\_studies.payments` p

using(order\_id)

where extract(month from order\_purchase\_timestamp) between 1 and 8)

)



INSIGHTS:

The payments from 2017 to 2018 increased by 1.36 times. This shows the growth of the orders and the growth of the revenue.

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

select \*,

round((total\_price/total\_orders),3) as avg\_price

from

(select distinct(customer\_state) as state,

count(\*) over(partition by customer\_state) as total\_orders,

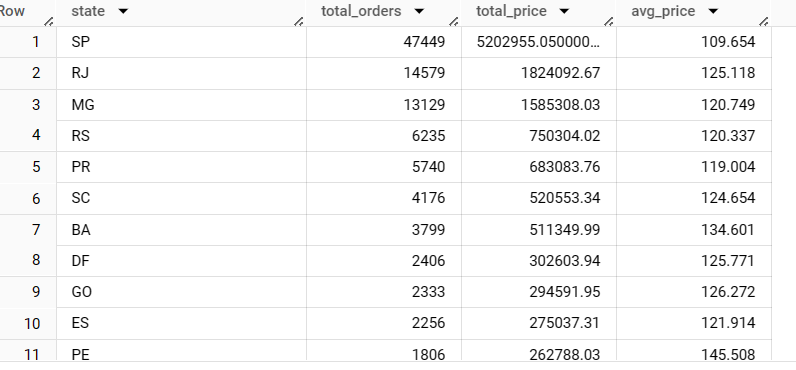
sum(price) over(partition by customer\_state) as total\_price

from

`business\_case\_studies.order\_items` ot join `business\_case\_studies.orders` o

using(order\_id) join `business\_case\_studies.customers` c using(customer\_id)

order by total\_price desc)



INSIGHTS:

SP is the state which has the highest number of orders with total\_price of 5.2million and average price of the order is 110. This is also the state that has highest number of customers and contributing to the company’s revenue by 40%.

Recommendation: NA

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

select \*,

round((total\_freight\_price/total\_orders),3) as avg\_freight\_price

from

(select distinct(customer\_state) as state,

count(\*) over(partition by customer\_state) as total\_orders,

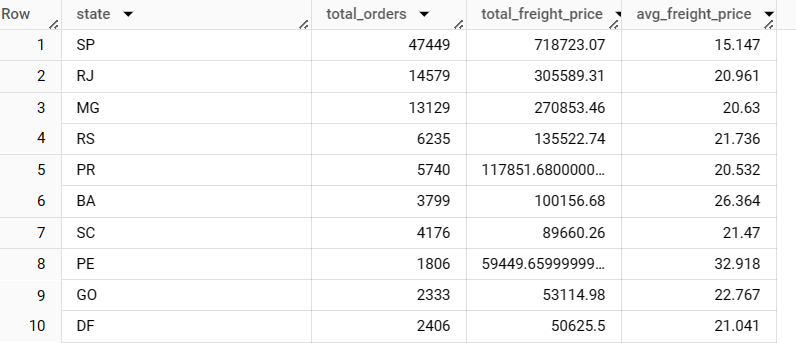
sum(freight\_value) over(partition by customer\_state) as total\_freight\_price

from

`business\_case\_studies.order\_items` ot join `business\_case\_studies.orders` o

using(order\_id) join `business\_case\_studies.customers` c using(customer\_id)

order by total\_freight\_price desc)



INSIGHTS:

SP has the highest freight price(freight price of all orders) with 0.71Million and its average freight price is 15.147.

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

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.

with t1 as

(select order\_id,

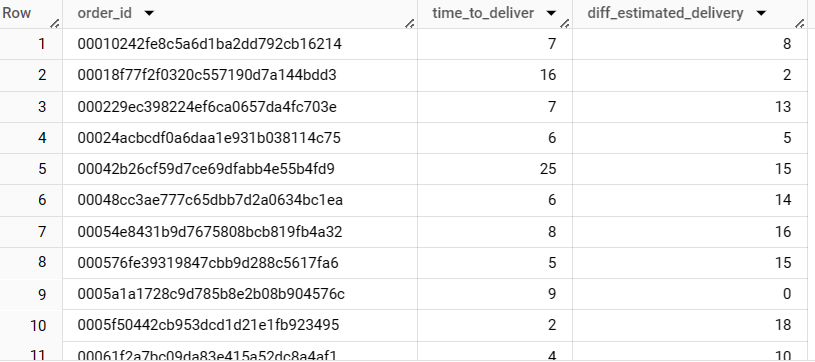
timestamp\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)  as time\_to\_deliver,

timestamp\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,day) as diff\_estimated\_delivery from `business\_case\_studies.orders`

order by order\_id

)

select \* from t1 where t1.time\_to\_deliver is not null;



INSIGHTS:

Diff\_estimated\_delivery gives the time delayed by the logistics to reach the customer

From the organisation.

Recommendation:

If we can get the data by state, then we can improve the performance of logistics where it is needed.

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

with t1 as(select \*,

round((total\_freight\_price/total\_orders),3) as avg\_freight\_price

from

(select distinct(customer\_state) as state,

count(\*) over(partition by customer\_state) as total\_orders,

sum(freight\_value) over(partition by customer\_state) as total\_freight\_price

from

`business\_case\_studies.order\_items` ot join `business\_case\_studies.orders` o

using(order\_id) join `business\_case\_studies.customers` c using(customer\_id)

order by total\_freight\_price desc)

),

t2 as

(select t1.state, avg\_freight\_price, concat("top ", dense\_rank() over(order by t1.avg\_freight\_price desc)) as level1  from t1

order by avg\_freight\_price desc limit 5),

t3 as

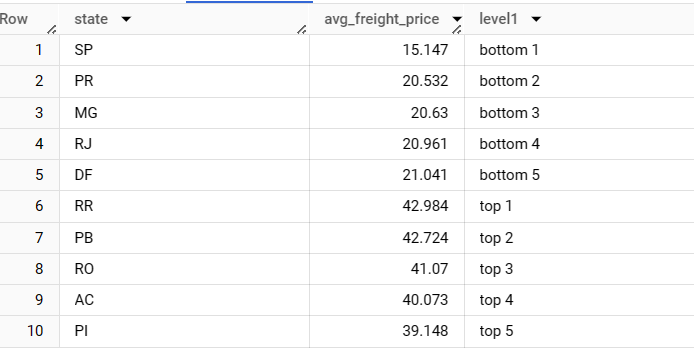
(select t1.state, avg\_freight\_price, concat("bottom ", dense\_rank() over(order by t1.avg\_freight\_price)) as level1  from t1

order by avg\_freight\_price limit 5)

select \* from t2

union all

select \* from t3;



Insights:

We get the avg\_freight\_price for the states and we had filtered it to make the top5 and bottom 5. Customer must pay the freight price apart from the amount he had taken for products. So logistics and tax will play key role. Customer always think to get the low freight price.

Recommendation:

Check the states with high freight price and keep the store in that state which can decrease the freight price and can increase the delivery time which also gets customers appreciation.

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

with t1 as

(select customer\_state,

round((sum(timestamp\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,second)/86400)/count(order\_id)),2)  as avg\_delivery\_time\_for\_state

from `business\_case\_studies.orders`

join `business\_case\_studies.customers` using(customer\_id)

group by customer\_state)

(select \*,

concat("highest ",dense\_rank() over(order by t1.avg\_delivery\_time\_for\_state desc)) as level from t1

order by t1.avg\_delivery\_time\_for\_state

desc limit 5)

union all

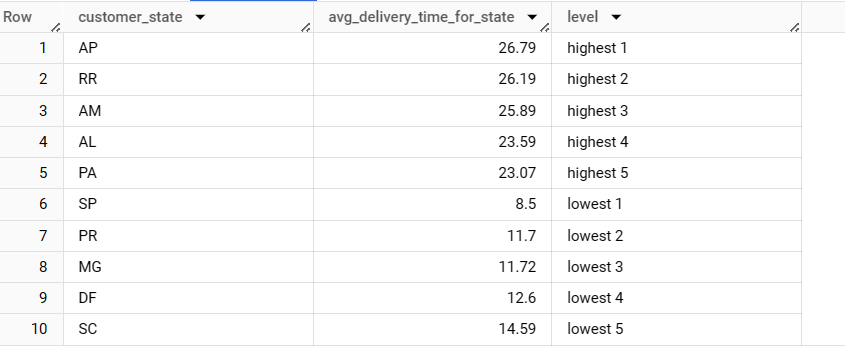
(

select \*,

concat("lowest ",dense\_rank() over(order by t1.avg\_delivery\_time\_for\_state)) as level from t1

order by t1.avg\_delivery\_time\_for\_state

limit 5)



Insights:

As shown in the table, the best performing states in terms of delivery time are SP,PR, etc and with the level highest are not the best performers.

Recommendation:

Place a store in the state if feasible which can make delivery time a lot less, can grab few customers more and reduce the freight prices.

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

with t1 as(select customer\_state,

round((sum(timestamp\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,second)/86400))/(count(order\_id)),2) as avg\_delivery\_time,

round((sum(timestamp\_diff(order\_estimated\_delivery\_date,order\_purchase\_timestamp,second)/86400))/(count(order\_id)),2) as avg\_estimated\_delivery\_time

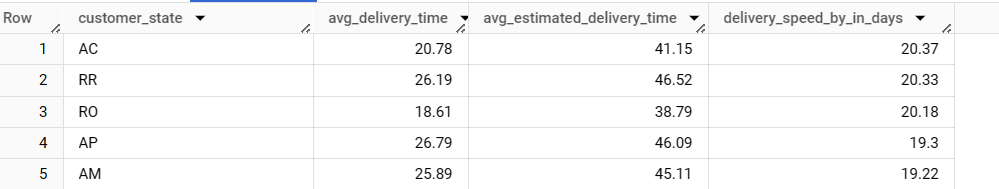
from `business\_case\_studies.orders` join `business\_case\_studies.customers`

using(customer\_id)

group by customer\_state)

select \*, round((t1.avg\_estimated\_delivery\_time - t1.avg\_delivery\_time),2) as delivery\_speed\_by\_in\_days from t1

order by delivery\_speed\_by\_in\_days desc limit 5



INSIGHTS:

We got the data of the states that are performing well in terms of order delivery.

All the 5 states mentioned in the table have speed up the delivery by nearly 20days.

Recommendation: NA

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

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

with t1 as

(select payment\_type,

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

FORMAT\_DATETIME("%B", order\_purchase\_timestamp) as month\_name,

count(order\_id) as number\_of\_orders from `business\_case\_studies.orders` join `business\_case\_studies.payments`

using(order\_id)

group by payment\_type, month, month\_name

order by payment\_type, month),

t2 as(

select \*, lag(number\_of\_orders) over(partition by payment\_type order by month) as prev\_month\_orders

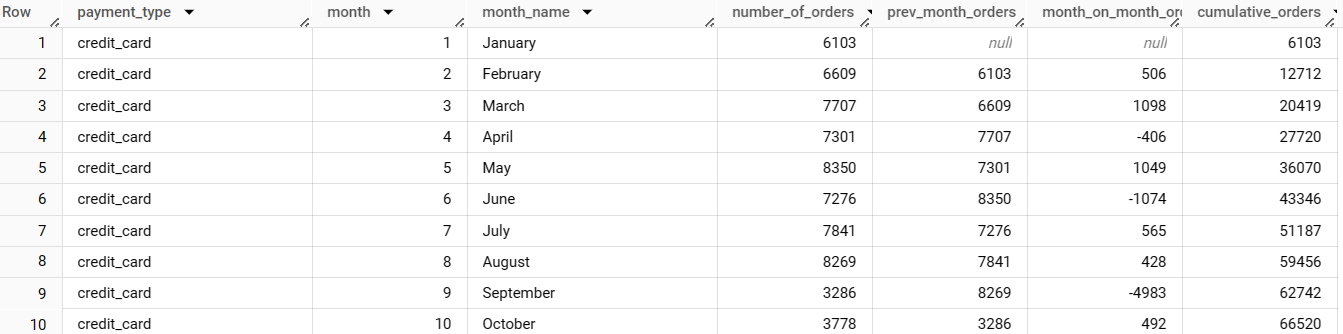
from t1)

select \*,

(number\_of\_orders - prev\_month\_orders) as month\_on\_month\_orders,

sum(number\_of\_orders) over(partition by payment\_type order by month range between unbounded preceding and current row) as cumulative\_orders

 from t2



INSIGHTS:

We have the data of orders for each payment type on monthly basis.

70% percent of the customers used credit card as the mode of the payment.

Done the month to month comparison of number of orders.

Recommendation:

As most of the customers are using credit cards, you might have given any offer.

If the company able to provide any benefits for other mode of payments as well, then there is a chance of increasing in the orders.

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

select payment\_installments, count(\*) as number\_of\_orders from `business\_case\_studies.payments`

group by payment\_installments

order by number\_of\_orders desc;



INSIGHTS:

from the output, we can conclude that majority of the orders are taken with less installments.

According to me, there is a chance that products are either so affordable or just affordable by some group of people(middle/higher class if appropriate). As the lower middle/middle/employee with less salary class have less income, they will choose to pay in more installments. But from the output, the higher the installments, the lower the orders.

Recommendation:

From this we can classify the customers and give them suggestions on buying the products that suits them.