1,1 SELECT

  column\_name,

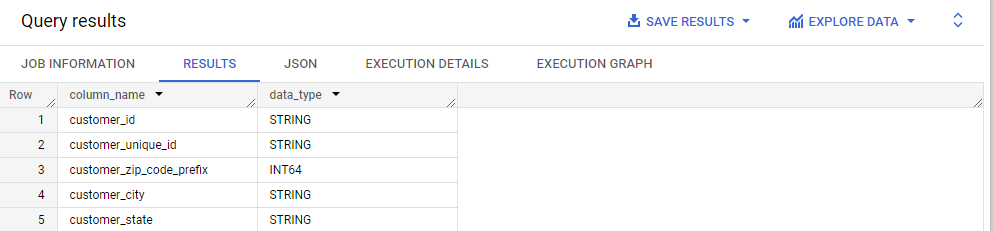
  data\_type

FROM

  sql-databases-demo.case\_study.INFORMATION\_SCHEMA.COLUMNS

WHERE

  table\_name = 'customers';



**Observation**: From this query we can see Data type of all columns in the "customers" table.

1,2 select min(order\_purchase\_timestamp) as from\_date , max(order\_purchase\_timestamp) TO\_date from `case\_study.orders`

A screenshot of a computer

Description automatically generated

**Observation**: From this we can understand that the timestamp ranges from 2016-09-04 to 2019-10-17 .

1,3 select count(distinct customer\_city) city\_count,count(distinct customer\_state) state\_count from `case\_study.customers`

A close-up of a white background

Description automatically generated

**Observation**: From this we can see the count of Cities & States of customers who ordered during the given period.

2,1 select c,

count(\*) from

(select \*,extract(month from order\_purchase\_timestamp) b,extract(year from order\_purchase\_timestamp) c

from case\_study.orders

) a

group by c

order by count(\*) desc

A screen shot of a computer

Description automatically generated

**Observation** : From this we can see the growth of the orders that has been placed over the years

2,2 select b,

count(\*) from

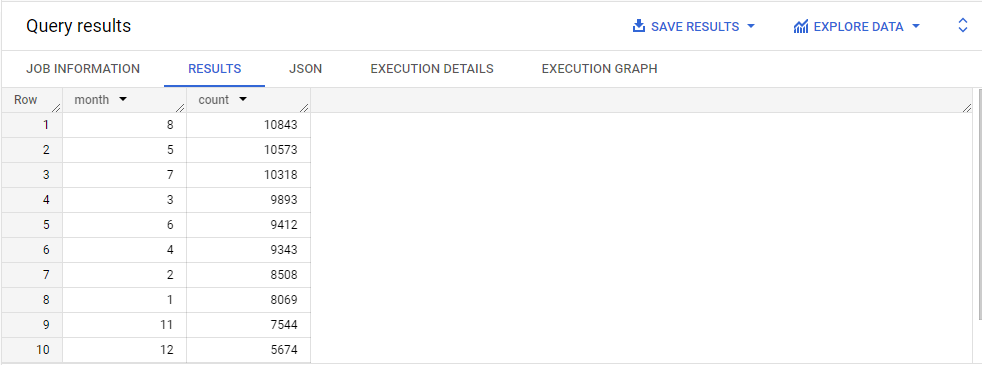
(select \*,extract(month from order\_purchase\_timestamp) b,extract(year from order\_purchase\_timestamp) c

from case\_study.orders

) a

group by b

order by count(\*) desc



additional analysis

select ym,

count(\*) count from

(select \*,concat(extract(month from order\_purchase\_timestamp),'-',extract(year from order\_purchase\_timestamp) ) ym

from case\_study.orders

) a

group by ym

order by count(\*) desc

A screenshot of a computer

Description automatically generated

**Observation:** In the first query you can see data distribution month wise and in the second query you can see the data distribution in year wise monthly record,

So here the insight is every year month august has good number of purchase .

2,3 select result,count(\*) from

(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 'Mornings'

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'

else 'remaining'

end result

from case\_study.orders

) a

group by result

order by count(\*) desc

A white box with black text

Description automatically generated

**Observation:** From this we ca see how order is distributed over the time period.

3,1

select state,date,count(order\_id)orders from

(select o.order\_id,geolocation\_state state,geolocation\_city,concat(extract(month from order\_purchase\_timestamp),'-',extract(year from order\_purchase\_timestamp)) date  from case\_study.orders o join `case\_study.order\_items` oi on o.order\_id=oi.order\_id join `case\_study.sellers` s on

oi.seller\_id=s.seller\_id join `case\_study.geolocation` gl on s.seller\_zip\_code\_prefix=gl.geolocation\_zip\_code\_prefix) a

group by state,date

order by date

A screenshot of a computer

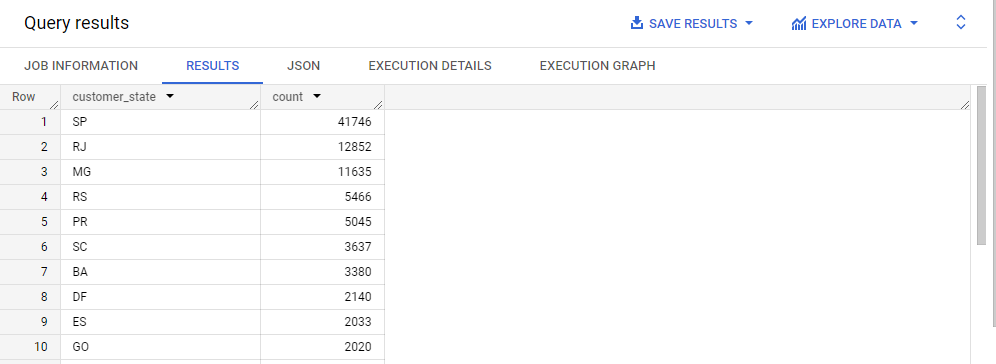
Description automatically generated**observation**: In the above result we can see the month on month number of orders placed in each state.

3,2 customer distribution state and city wise

select customer\_state,count(\*) count from case\_study.customers c

group by customer\_state

order by count(\*) desc



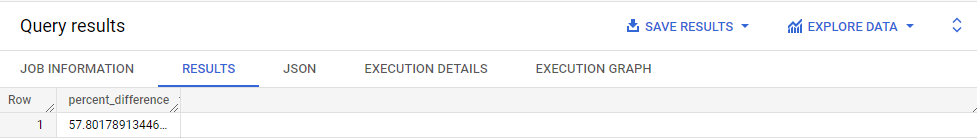
**Observation**: From this we can see the how customers distributed across the states

4,1 select

 ((sum(case when extract(year from o.order\_purchase\_timestamp)=2018 and extract(month from o.order\_purchase\_timestamp) between 1 and 8 then p.payment\_value else 0 end)-sum(case when extract(year from o.order\_purchase\_timestamp)=2017 and extract(month from o.order\_purchase\_timestamp) between 1 and 8 then p.payment\_value else 0 end))/sum(case when extract(year from o.order\_purchase\_timestamp)=2018 and extract(month from o.order\_purchase\_timestamp) between 1 and 8 then p.payment\_value else 0 end))\*100 as percent\_difference

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

where extract(year from o.order\_purchase\_timestamp) between 2017 and 2018 and extract(month from o.order\_purchase\_timestamp)between 1 and 8;



**Observation** : from the above query we can see the percentage increase in the cost of orders from year 2017 to 2018 in respective months

4,2 select state,round(sum(price),2) total,round(sum(price)/count(\*),2) avgr from

(select o.order\_id,geolocation\_state state,geolocation\_city,price from case\_study.orders o join `case\_study.order\_items` oi on o.order\_id=oi.order\_id join `case\_study.sellers` s on

oi.seller\_id=s.seller\_id join `case\_study.geolocation` gl on s.seller\_zip\_code\_prefix=gl.geolocation\_zip\_code\_prefix join `case\_study.payments` p on o.order\_id=p.order\_id ) a

group by state

A screenshot of a computer

Description automatically generated

**Observation**: From the above results we can find the Total & Average value of order price for each state

4,3 select state,round(sum(freight\_value),2) total,round(sum(freight\_value)/count(\*),2) avgr from

(select o.order\_id,geolocation\_state state,geolocation\_city,freight\_value from case\_study.orders o join `case\_study.order\_items` oi on o.order\_id=oi.order\_id join `case\_study.sellers` s on

oi.seller\_id=s.seller\_id join `case\_study.geolocation` gl on s.seller\_zip\_code\_prefix=gl.geolocation\_zip\_code\_prefix join `case\_study.payments` p on o.order\_id=p.order\_id ) a

group by state

A screenshot of a computer

Description automatically generated

**Observation**: from the above results we can see the Total & Average value of order freight for each state.

5,1 select \* from

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

date\_diff(order\_delivered\_customer\_date,order\_estimated\_delivery\_date

,day) diff\_estimated\_delivery

 from case\_study.orders )

A screenshot of a computer

Description automatically generated

**Observation**: From the results we Found the no. of days taken to deliver each order from the order’s purchase date as delivery time.

Also, calculated the difference between the estimated & actual delivery date of an order.

5,2 select \* from

(select geolocation\_state,round(avg(freight\_value),2) avg\_value from case\_study.sellers s join case\_study.order\_items o using(seller\_id) join case\_study.geolocation g on seller\_zip\_code\_prefix=geolocation\_zip\_code\_prefix

group by g.geolocation\_state) a

order by a.avg\_value desc

limit 5

A screenshot of a computer

Description automatically generated

**Observation**: from the above results we can the top 5 states with the highest average freight value

select \* from

(select geolocation\_state,round(avg(freight\_value),2) avg\_value from case\_study.sellers s join case\_study.order\_items o using(seller\_id) join case\_study.geolocation g on seller\_zip\_code\_prefix=geolocation\_zip\_code\_prefix

group by g.geolocation\_state) a

order by a.avg\_value

limit 5

A screenshot of a computer

Description automatically generated

**Observation**: from the above results we can the top 5 states with the lowest average freight value

5,3 select geolocation\_state,avg(time\_to\_deliver) highest\_del\_time from

(select g.geolocation\_state,date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) time\_to\_deliver,

date\_diff(order\_delivered\_customer\_date,order\_estimated\_delivery\_date

,day) diff\_estimated\_delivery

 from case\_study.orders o left join case\_study.customers c using(customer\_id) join case\_study.geolocation g on

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

) a

group by geolocation\_state

order by avg(time\_to\_deliver) desc

limit 5

A screenshot of a computer

Description automatically generated

**Observation:** From the above results we can find out the top 5 states with the highest average delivery time.

select geolocation\_state,avg(time\_to\_deliver) lowest\_del\_time from

(select g.geolocation\_state,date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) time\_to\_deliver,

date\_diff(order\_delivered\_customer\_date,order\_estimated\_delivery\_date

,day) diff\_estimated\_delivery

 from case\_study.orders o left join case\_study.customers c using(customer\_id) join case\_study.geolocation g on

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

) a

group by geolocation\_state

order by avg(time\_to\_deliver)

limit 5

A screenshot of a computer

Description automatically generated

**Observation:** From the above results we can find out the top 5 states with the lowest average delivery time.

5,4 select customer\_state,avg(diff\_estimated\_delivery) date\_diff from

(select \*,date\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date

,day) diff\_estimated\_delivery

 from case\_study.orders join case\_study.customers c using(customer\_id) )

 group by customer\_state

order by avg(diff\_estimated\_delivery) desc

limit 5

A screenshot of a computer

Description automatically generated

**Observation**: From the above results we can Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.

6,1 select month,payment\_type, count(\*) from

(select \*,extract(month from order\_purchase\_timestamp) month

from case\_study.orders join `case\_study.payments` using(order\_id)) a

group by month,payment\_type



**Observation** : From above results we can find the month on month no. of orders placed using different payment types.

6,2 select count(distinct(order\_id)) as no\_of\_orders,

payment\_installments

from `case\_study.payments`

where payment\_installments >=1

group by payment\_installments

order by payment\_installments ;

A screen shot of a computer

Description automatically generated

**Observation**: From above results we can find the no. of orders placed on the basis of the payment installments that have been paid.

**Actionable insights**:

* The dataset time ranges between 2016 sept -2018 oct.
* We have customer from 4119 cities and 27 states.
* Over the years customers count rate got drastically increased.
* In in month of august we can see good number of purchases every year.
* Most of the orders were placed in the afternoon.
* The percentage difference between the year 2017-2018 is 57% approx.
* State PR has highest average cost of purchase.
* State CE has highest average freight value.
* State AP takes more time to deliver the orders.
* State SP take very less time to deliver their orders.
* In the state AL we can see the delivery is faster that the estimated delivery time.

**Recommendation:**

* Need to increase the resource since the order count was drastically increased over the period of year.
* Need to provide better service in the month of august, since it has good order numbers over the years.
* Most of the transactions were cashless so need to provide proper gateway security for money transactions.
* Must improve the delivery time in states such as SP, PR, MG, DF, SC
* Must do marketing to pull more users from states with less users CE, DF, GO.
* Most of the orders were placed during the afternoon time so need to provide better service to avoid interruption