***SQL BUSINESS CASE***

1.**The time period for which the order data is given :**

select x.first\_day,

x.last\_day,

datediff(x.last\_day,x.first\_day) as active\_days

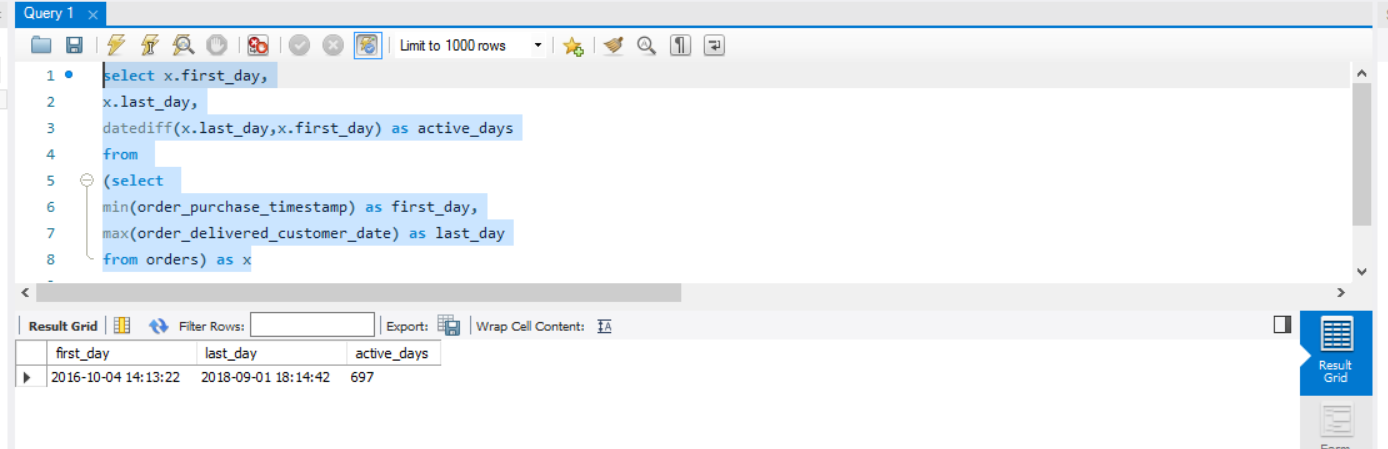
from

(select

min(order\_purchase\_timestamp) as first\_day,

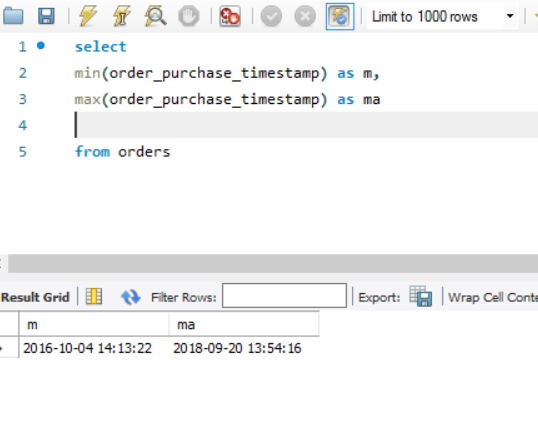
max(order\_delivered\_customer\_date) as last\_day

from orders) as x



2.Cities and States of Customers Ordered During the Given Period

* First I capture the minimum and maximum value of order\_purchase\_timestamp from the orders table to get the given period of customer orders



* Then using the max date of order\_purchase time stamp I put the condition and wrote my final query-

select

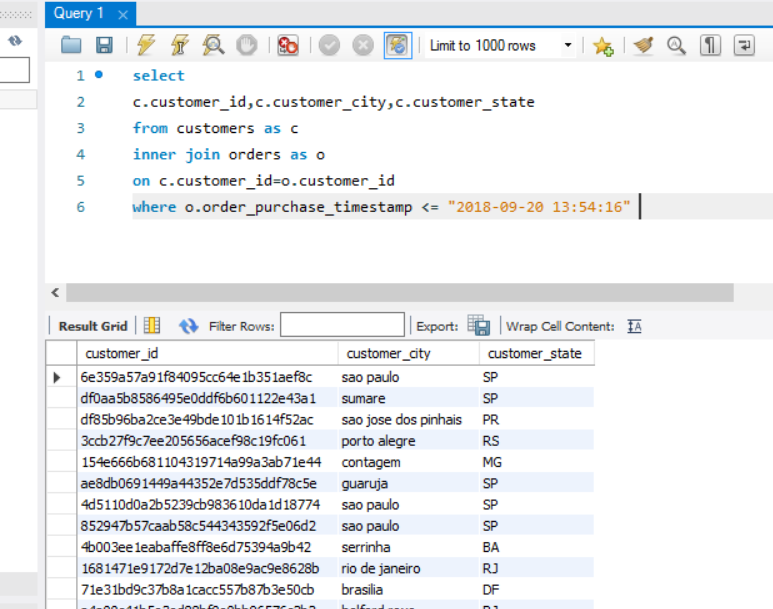
c.customer\_id,c.customer\_city,c.customer\_state

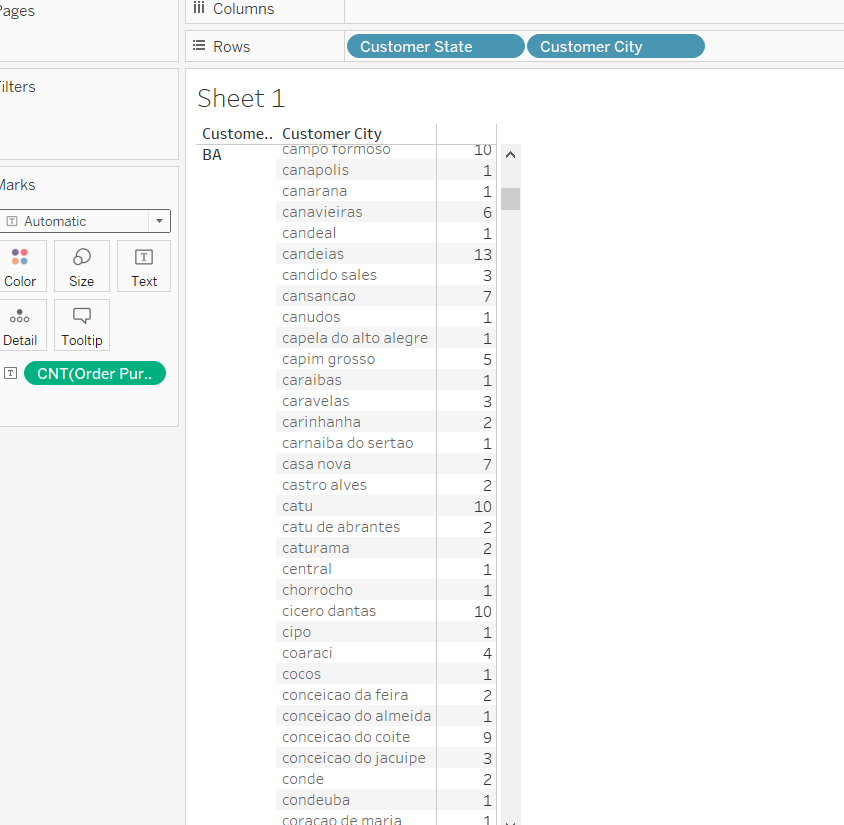
from customers as c

inner join orders as o

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

where o.order\_purchase\_timestamp <= "2018-09-20 13:54:16"





Is there any growing trend for e-commerce in brazil?

select

sum(x.price),x.Month from

(select

shipping\_limit\_date,price,

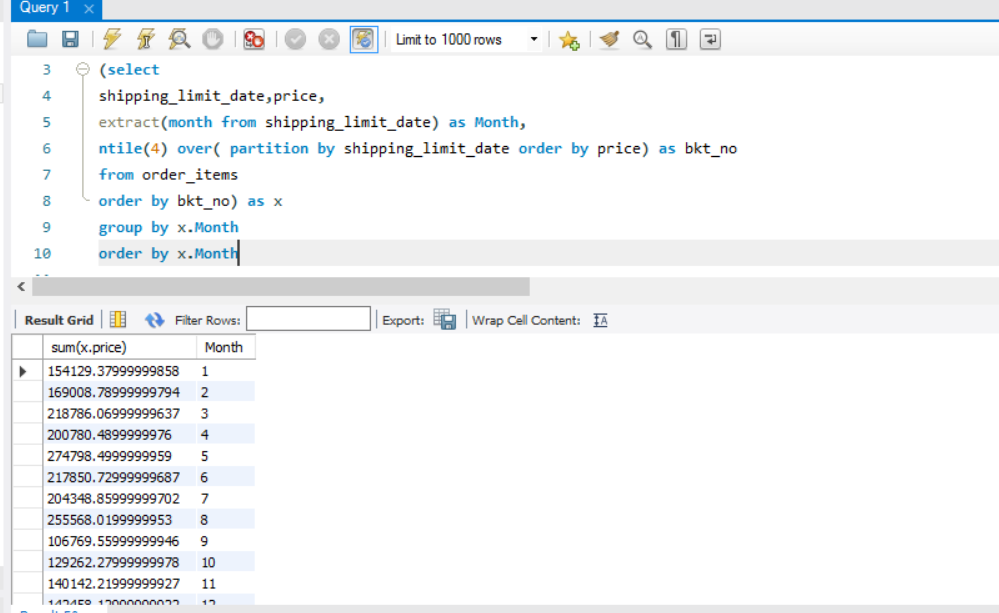
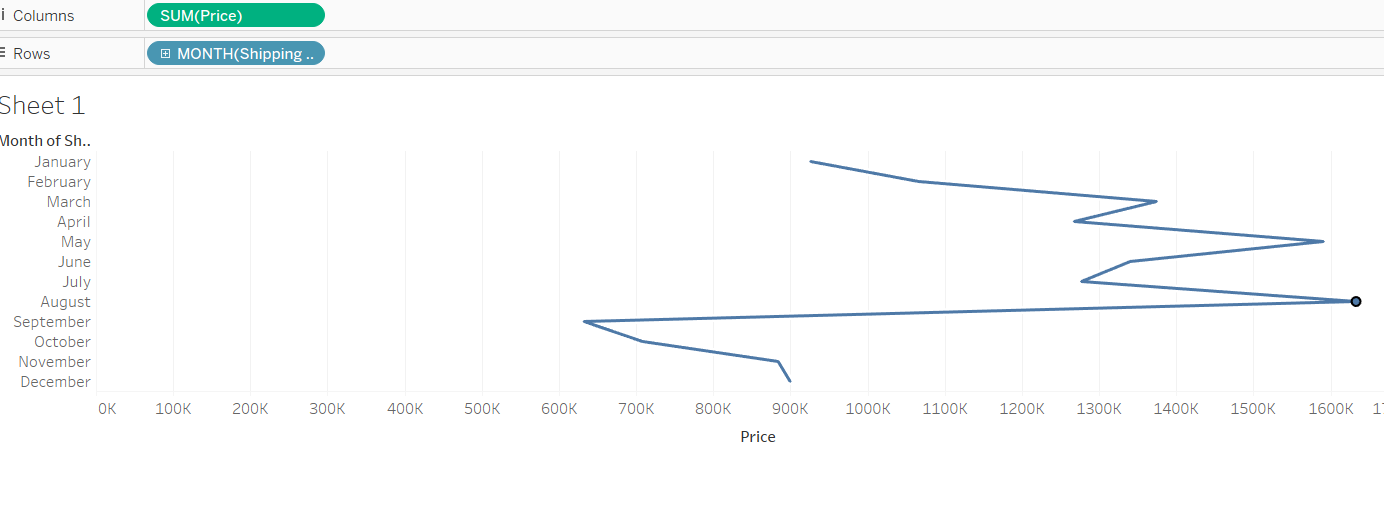
extract(month from shipping\_limit\_date) as Month,

ntile(4) over( partition by shipping\_limit\_date order by price) as bkt\_no

from order\_items

order by bkt\_no) as x

group by x.Month



From this example we can see that a sharp rise in the month of August where the sales are maximum and it is decreasing after the month of August.

Hence,the peak seasonality occurs in the month of August.

**1.Time in which Brazilian Customers Tend to buy-**

select

count(x.order\_purchase\_timestamp),x.timing

from

(select

order\_id,customer\_id,order\_purchase\_timestamp,

extract(hour from order\_purchase\_timestamp) as h,

case

when extract(hour from order\_purchase\_timestamp) <6 and extract(hour from order\_purchase\_timestamp)>5

then "Dawn"

when extract(hour from order\_purchase\_timestamp)>=6 and extract(hour from order\_purchase\_timestamp)<12

then "Morning"

when extract(hour from order\_purchase\_timestamp)>=12 and extract(hour from order\_purchase\_timestamp)<16

then "Afternoon"

when extract(hour from order\_purchase\_timestamp)>=16 and extract(hour from order\_purchase\_timestamp)<=20

then "evening"

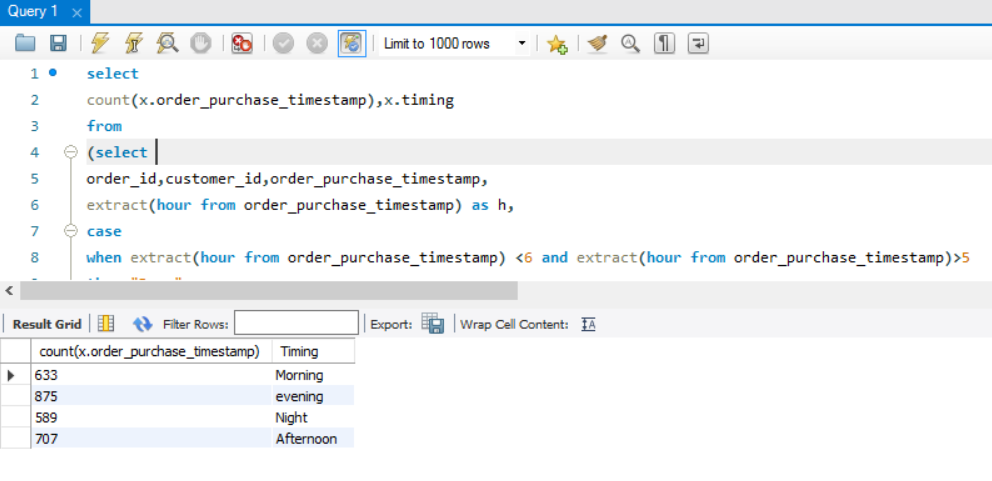
when extract(hour from order\_purchase\_timestamp)>20 or extract(hour from order\_purchase\_timestamp)<=4 or extract(hour from order\_purchase\_timestamp)>4

then "Night"

end as Timing

from orders ) as x

group by x.Timing



So from the above query we can analyse that maximum order purchase happen in the Evening.

**5.Month on Month Orders by State-**

select

x.Month,c.customer\_state,sum( x.order\_id) as Monthly\_orders from

(select

order\_purchase\_timestamp,customer\_id,order\_id,

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

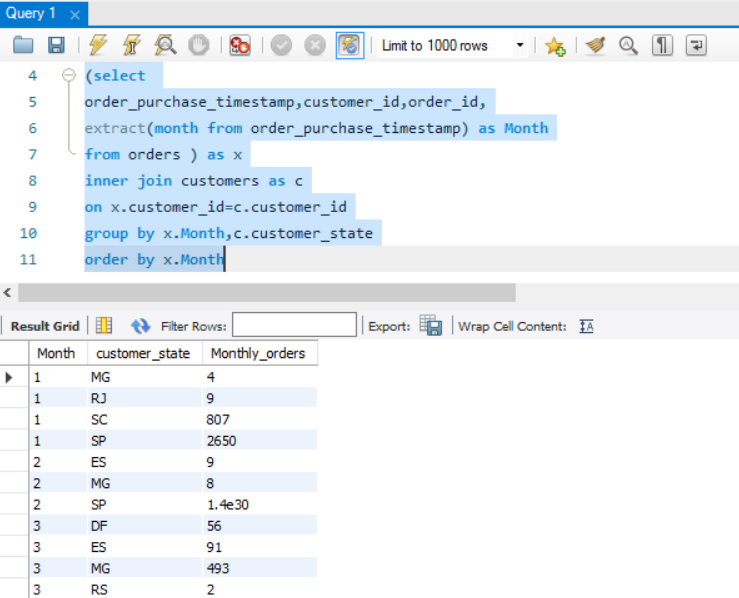
from orders ) as x

inner join customers as c

on x.customer\_id=c.customer\_id

group by x.Month,c.customer\_state

order by x.Month



**6.Distribution of Customers across the states of Brazil-**

select

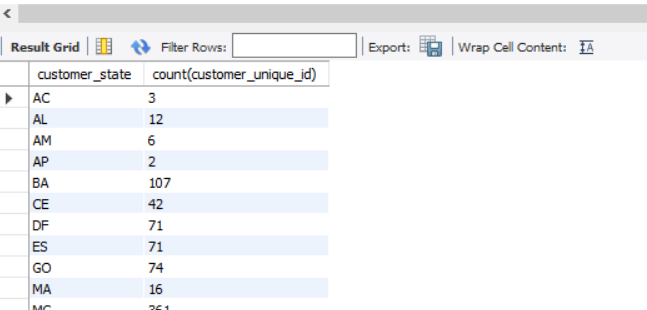
customer\_state,count(customer\_unique\_id)

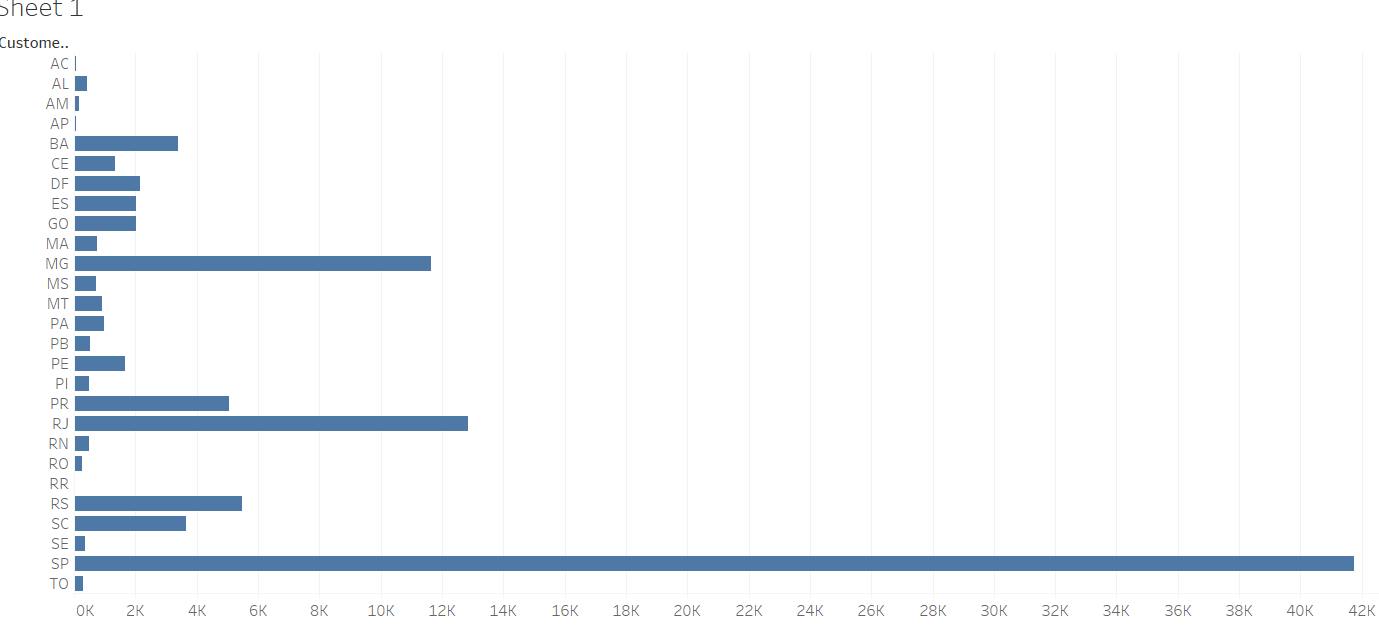
from

customers

group by customer\_state

order by customer\_state





**7.% increase in cost of orders from 2017 to 2018-**

select

extract(month from o.shipping\_limit\_date) as Month,

extract(year from o.shipping\_limit\_date) as Year,

sum(p.payment\_value),

lead(sum(p.payment\_value),1) over(partition by extract(month from o.shipping\_limit\_date) order by extract(year from o.shipping\_limit\_date) ) as Next\_value,

round(( (lead(sum(p.payment\_value),1) over(partition by extract(month from o.shipping\_limit\_date) order by extract(year from o.shipping\_limit\_date) )-sum(p.payment\_value))/ sum(p.payment\_value) ) \* 100,2) as percentage\_increase\_decrease

from

order\_items as o

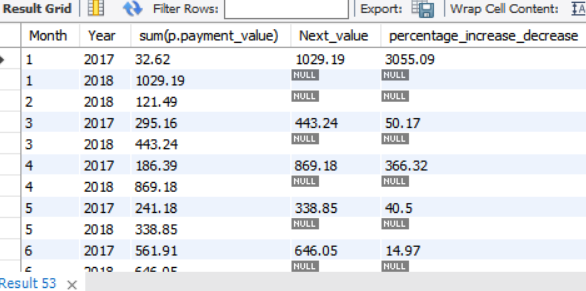
join payments as p

on o.order\_id=p.order\_id

where extract(month from o.shipping\_limit\_date) between "1" and "8"

group by Month,Year

order by Month,Year



**8.Mean and Sum of price and Freight value by Customer state**

select

c.customer\_state,round(avg(o.price),2) as avg\_price,round(sum(o.price),2) as sum\_price,round(avg(o.freight\_value),2) as avg\_freight,round(sum(o.freight\_value),2) as sum\_freight

from

order\_items as o

inner join

orders as o1

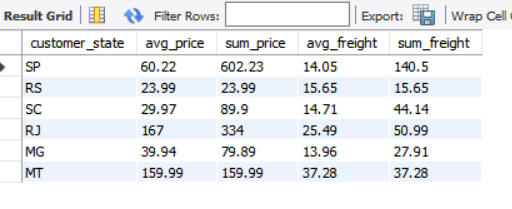
on o.order\_id=o1.order\_id

inner join

customers as c

on c.customer\_id=o1.customer\_id

group by c.customer\_state



**9.No of Days between Purchasing,delivering and estimated delivery:**

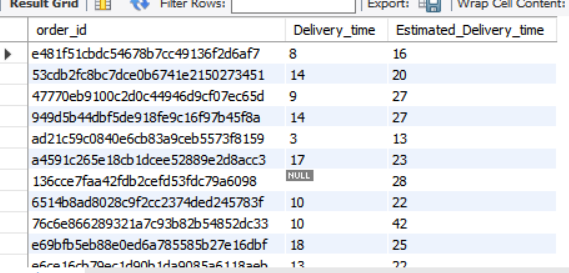
select

order\_id,

datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp) as Delivery\_time,

datediff(order\_estimated\_delivery\_date,order\_purchase\_timestamp) as Estimated\_Delivery\_time

from orders



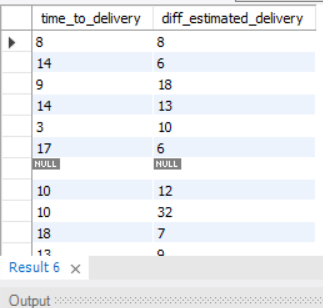
**10.Time to delivery and diff\_estimated delivery:**

select

datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp) as time\_to\_delivery,

datediff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date) as diff\_estimated\_delivery

from orders



**11.Mean of Freight\_value,time to delivery,diff\_estimated\_delivery by state:**

select

c.customer\_state,

round(avg(o1.freight\_value),2) as avg\_freight\_value,

round(avg(datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp)),2) as avg\_Delivery\_time,

round(avg(datediff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date)),2) as avg\_Estimated\_Delivery\_time

from order\_items as o1

join

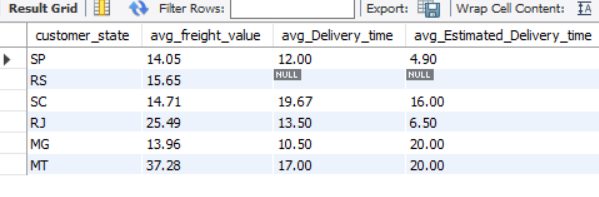
orders as o

on o1.order\_id=o.order\_id

join customers as c

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

group by c.customer\_state



**12.Top 5 states with highest/lowest average freight\_value**

**FOR TOP 5(HIGHEST TO LOWEST):**

select

c.customer\_state,

avg(o1.freight\_value) as Top\_5

from customers as c

inner join orders as o

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

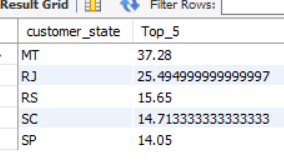
join order\_items as o1

on o.order\_id=o1.order\_id

group by c.customer\_state

order by Top\_5 desc

limit 5



**FOR LOWEST TO HIGHEST(AVERAGE FREIGHT\_VALUE):**

select

c.customer\_state,

avg(o1.freight\_value) as Top\_5

from customers as c

inner join orders as o

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

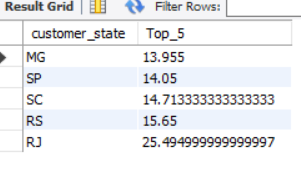
join order\_items as o1

on o.order\_id=o1.order\_id

group by c.customer\_state

order by Top\_5

limit 5



**12.Top 5 states with highest/lowest Average time to delivery:**

**-For Top 5 Highest to lowest:**

select

customer\_state,

avg(datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp)) as Top\_5

from

customers as c

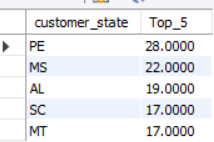
inner join orders as o

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

group by customer\_state

order by Top\_5 desc

limit 5



**-For Top 5 Lowest to Highest:**

select

customer\_state,

avg(datediff(order\_delivered\_customer\_date,order\_purchase\_timestamp)) as Top\_5

from

customers as c

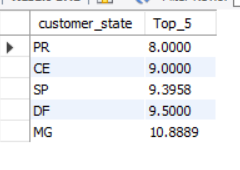
inner join orders as o

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

group by customer\_state

order by Top\_5

limit 5



**13.Top 5 States where delivery is very fast/Not so fast compared to estimated delivery:**

select

c.customer\_state,

round(avg(datediff(o.order\_delivered\_customer\_date,order\_purchase\_timestamp)),2) as delivery\_time,

round(avg(datediff(o.order\_estimated\_delivery\_date,order\_purchase\_timestamp)),2) as estimated\_delivery\_time

from

customers as c

inner join orders as o

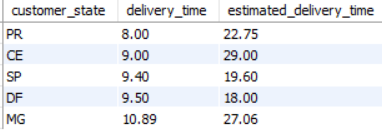
on c.customer\_id=o.customer\_id

group by c.customer\_state

having round(avg(datediff(o.order\_delivered\_customer\_date,order\_purchase\_timestamp)),2)<round(avg(datediff(o.order\_estimated\_delivery\_date,order\_purchase\_timestamp)),2)

order by delivery\_time

limit 5



**14.Count of Orders based on payment Installments:**

select

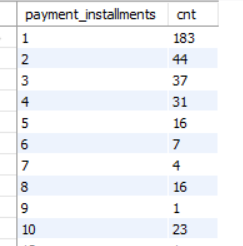
payment\_installments,

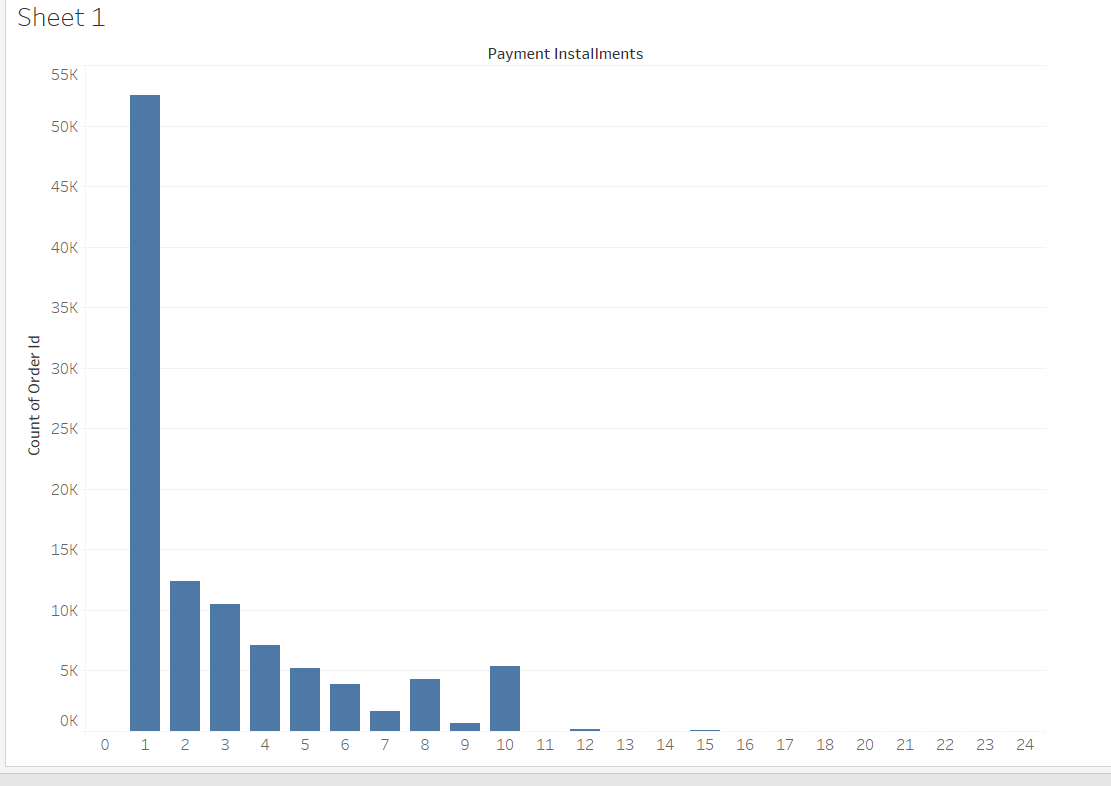
count(\*) as cnt

from payments

group by payment\_installments

order by payment\_installments





**15.Data Type of columns in a table:**

SELECT TABLE\_NAME,COLUMN\_NAME,DATA\_TYPE from INFORMATION\_SCHEMA.COLUMNS where

table\_schema = 'customer\_orders'

