# **Business Case: Target SQL**

This business case has data of 100k orders from 2016 to 2018 made at Target, Brazil. It is Americas leading retailer business chain.

Data is available in 8 tables, which gives information about orders from different dimensions like status of order, payment details, location and time of the order , customer who made the purchase, items in the order, product details, seller information of the products, order reviews etc.

**Analysis**

What does good look like:

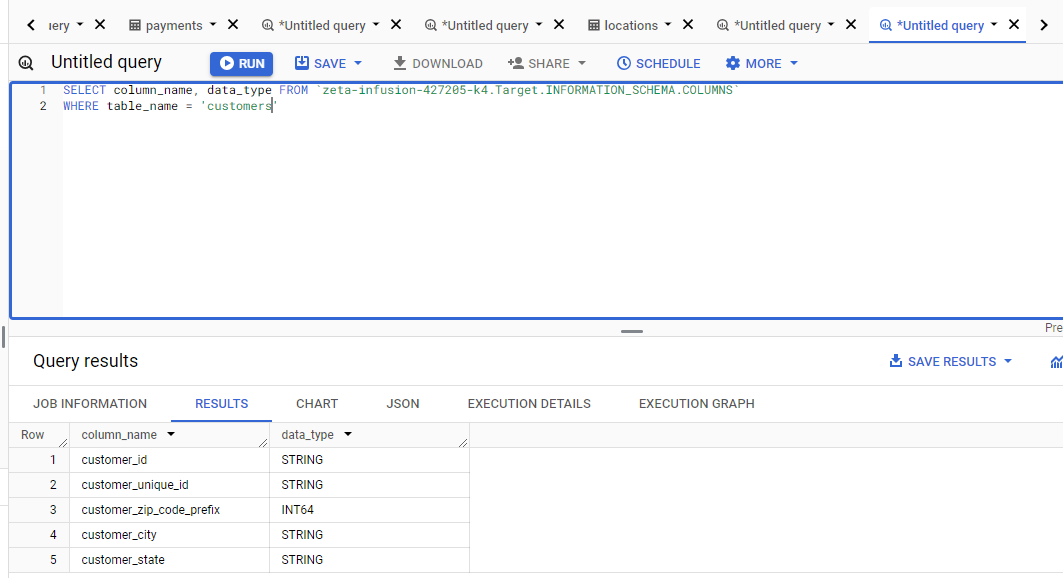
Question-1 Data type of all columns in the "customers" table.

Answer-

 Select column\_name, data\_type

from `zeta-infusion-427205-k4.Target.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name = 'customers'

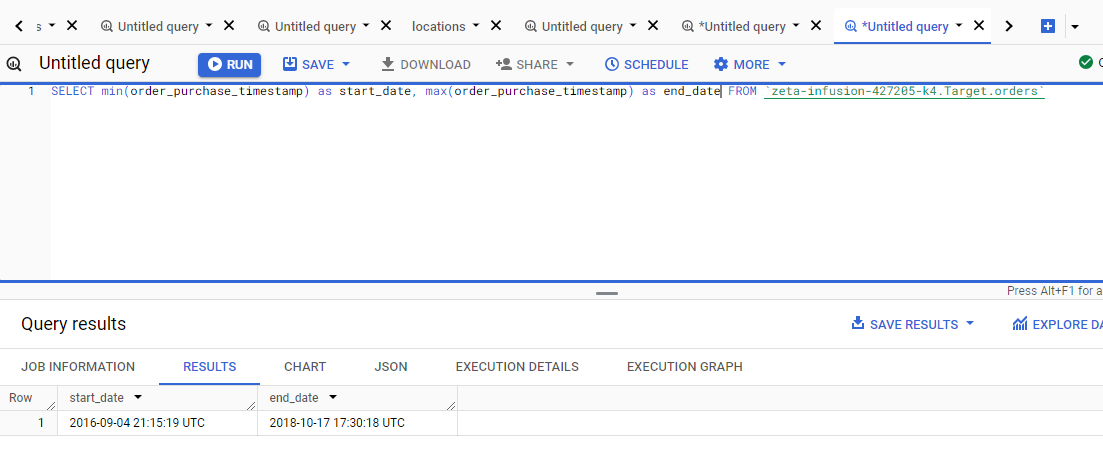


2. Get the time range between which the orders were placed.

Answer-

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

from `zeta-infusion-427205-k4.Target.orders`



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

Answer-

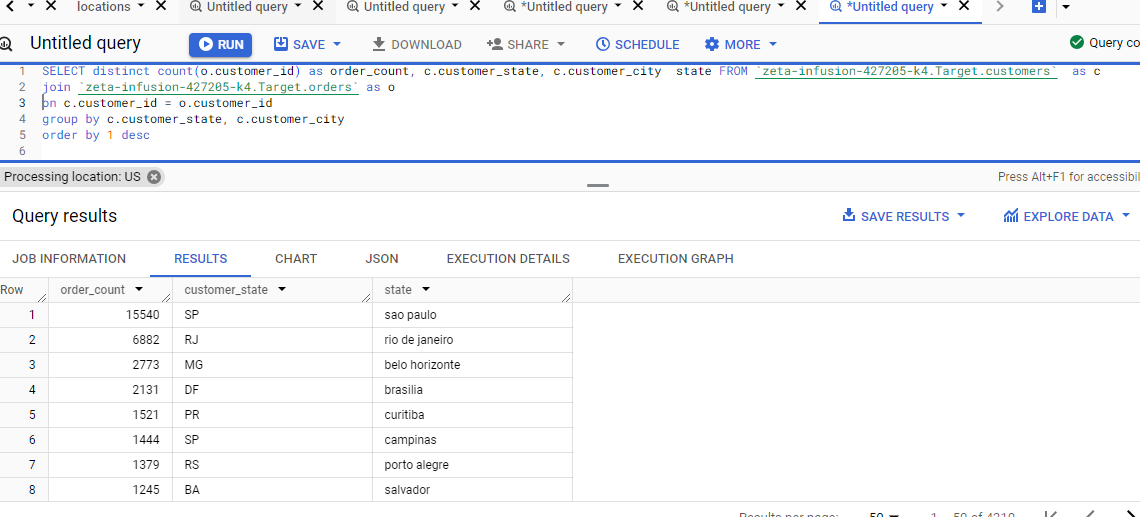
 SELECT distinct count(o.customer\_id) as order\_count, c.customer\_state, c.customer\_city  state FROM `zeta-infusion-427205-k4.Target.customers`  as c

join `zeta-infusion-427205-k4.Target.orders` as o

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

group by c.customer\_state, c.customer\_city

order by 1 desc



2nd               In\_depth Exploration

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

Answer -

 SELECT  extract(year from o.order\_purchase\_timestamp) as year,

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

count(distinct o.order\_id) as order\_count

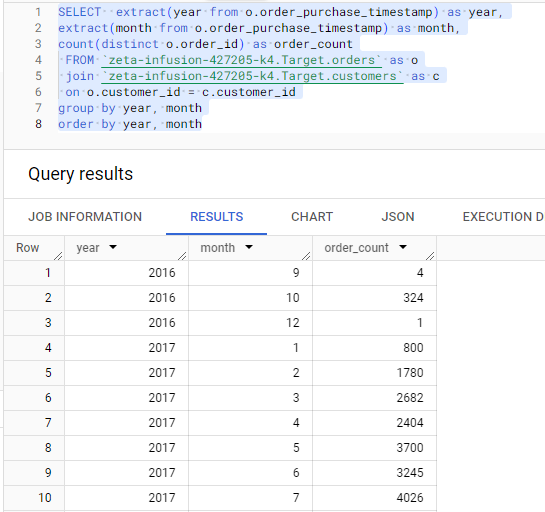
 FROM `zeta-infusion-427205-k4.Target.orders` as o

 join `zeta-infusion-427205-k4.Target.customers` as c

 on o.customer\_id = c.customer\_id

group by year, month

order by year, month



Question 2 Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

Answer-

 SELECT

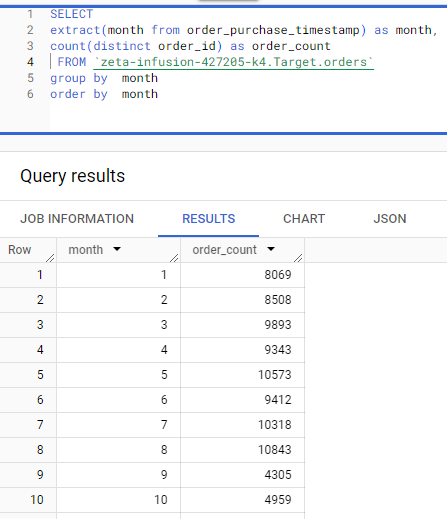
extract(month from order\_purchase\_timestamp) as month,

count(distinct order\_id) as order\_count

 FROM `zeta-infusion-427205-k4.Target.orders`

group by  month

order by  month



Question 3 During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night?

Answer-

SELECT  case when extract(hour from  o.order\_purchase\_timestamp) between 0 and 6 then 'Dawn'

when extract(hour from  o.order\_purchase\_timestamp) between 7 and 12 then 'Mornings'

when extract(hour from  o.order\_purchase\_timestamp) between 13 and 18 then 'Afternoon'

when extract(hour from  o.order\_purchase\_timestamp) between 19 and 23 then 'Night'

end as hour,

count(o.order\_id) as order\_count

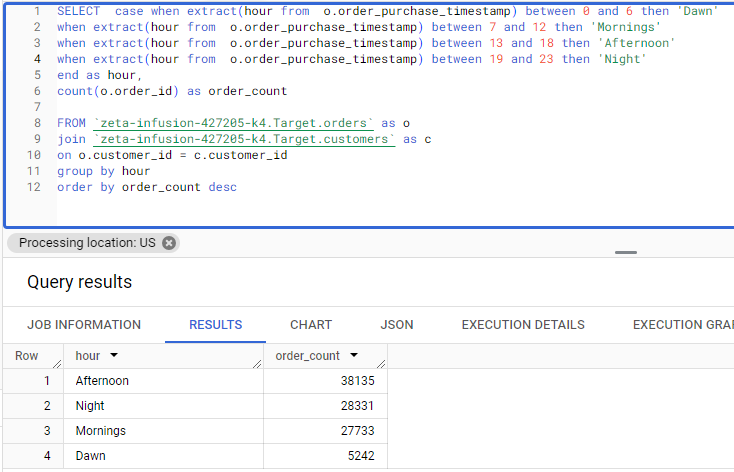
FROM `zeta-infusion-427205-k4.Target.orders` as o

join `zeta-infusion-427205-k4.Target.customers` as c

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

group by hour

order by order\_count desc



Question 3 1 Get the month on month no. of orders placed in each state?

Answer

SELECT  c.customer\_state,

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

count(o.order\_purchase\_timestamp) as order\_count

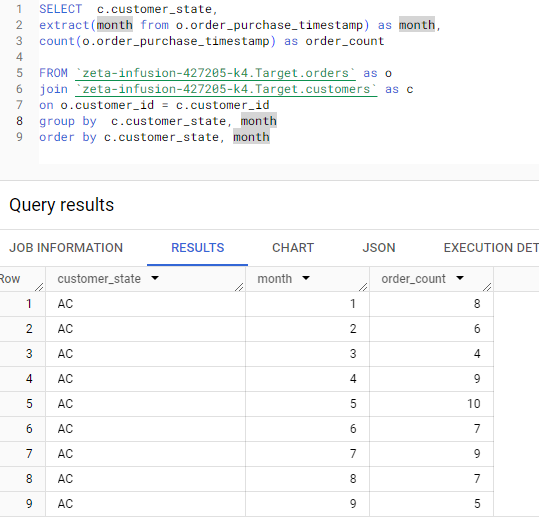
FROM `zeta-infusion-427205-k4.Target.orders` as o

join `zeta-infusion-427205-k4.Target.customers` as c

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

group by  c.customer\_state, month

order by c.customer\_state, month



Question 3 (2) how are the customers distributed across all the states?

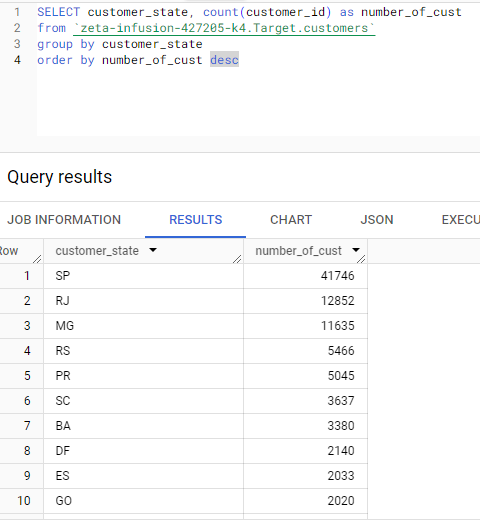
Answer

SELECT customer\_state, count(customer\_id) as number\_of\_cust

from `zeta-infusion-427205-k4.Target.customers`

group by customer\_state

order by number\_of\_cust desc



**Question 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).  
   You can use the "payment\_value" column in the payments table to get the cost of orders.

Answer-

 with order\_costs as

(SELECT   round(sum(p.payment\_value),2) as total\_cost,  extract(year from o.order\_purchase\_timestamp)  as year,

extract(month from o.order\_purchase\_timestamp) as month

FROM `zeta-infusion-427205-k4.Target.payments` as p

join `zeta-infusion-427205-k4.Target.orders` as o

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

where extract(year from o.order\_purchase\_timestamp) in (2017, 2018)

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

group by month, year)

select oc1.month, oc1.total\_cost, oc2.total\_cost as previous\_total\_cost,

round(((oc1.total\_cost-oc2.total\_cost)/oc2.total\_cost)\*100,2) as cost\_increase\_percentage

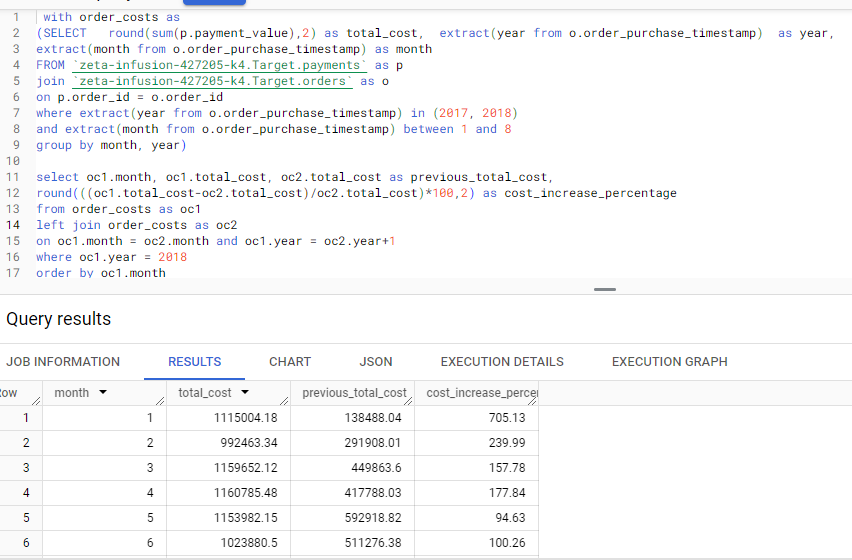
from order\_costs as oc1

left join order\_costs as oc2

on oc1.month = oc2.month and oc1.year = oc2.year+1

where oc1.year = 2018

order by oc1.month



Question 4 (2) Calculate the Total & Average value of order price for each state.

Answer

SELECT c.customer\_state , round(avg(oi.price),2)  as avg\_price, round(sum(oi.price),2) as total\_value

FROM `zeta-infusion-427205-k4.Target.orders` as o

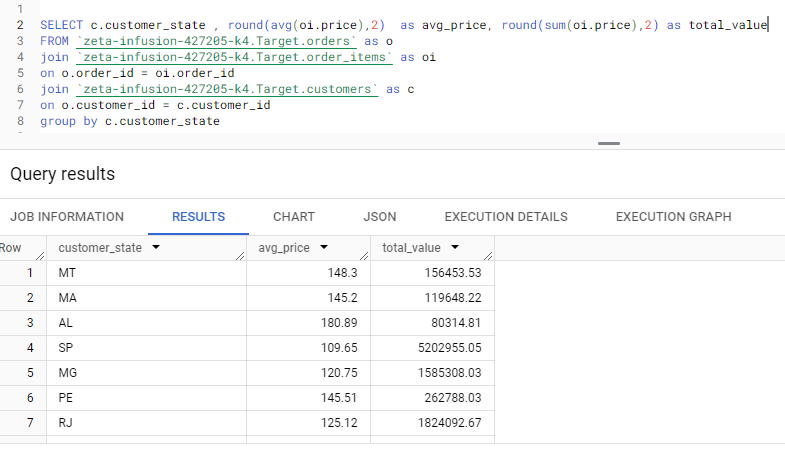
join `zeta-infusion-427205-k4.Target.order\_items` as oi

on o.order\_id = oi.order\_id

join `zeta-infusion-427205-k4.Target.customers` as c

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

group by c.customer\_state



Question 4 (3) Calculate the Total & Average value of order freight for each state.

Answer

SELECT c.customer\_state , round(avg(oi.freight\_value),2)  as avg\_price, round(sum(oi.freight\_value),2) as total\_value

FROM `zeta-infusion-427205-k4.Target.orders` as o

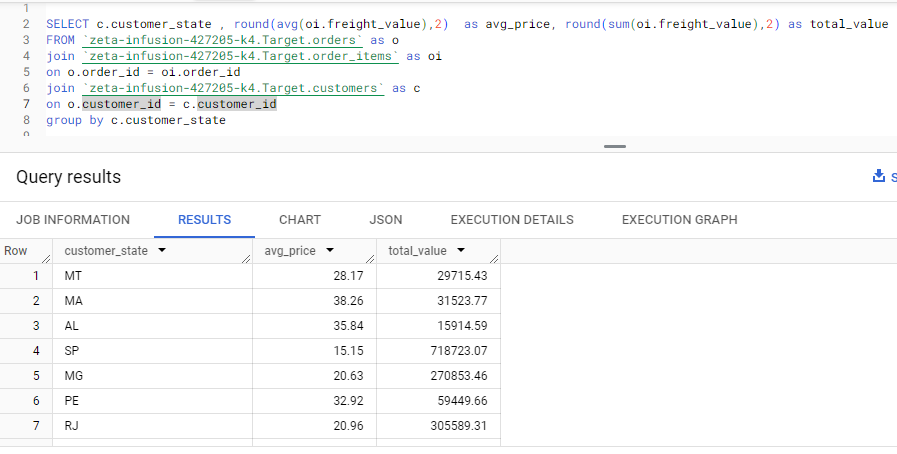
join `zeta-infusion-427205-k4.Target.order\_items` as oi

on o.order\_id = oi.order\_id

join `zeta-infusion-427205-k4.Target.customers` as c

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

group by c.customer\_state



Question 5 (1) Find the no.of days taken to deliver each order from the order's purchase date as delivery time?

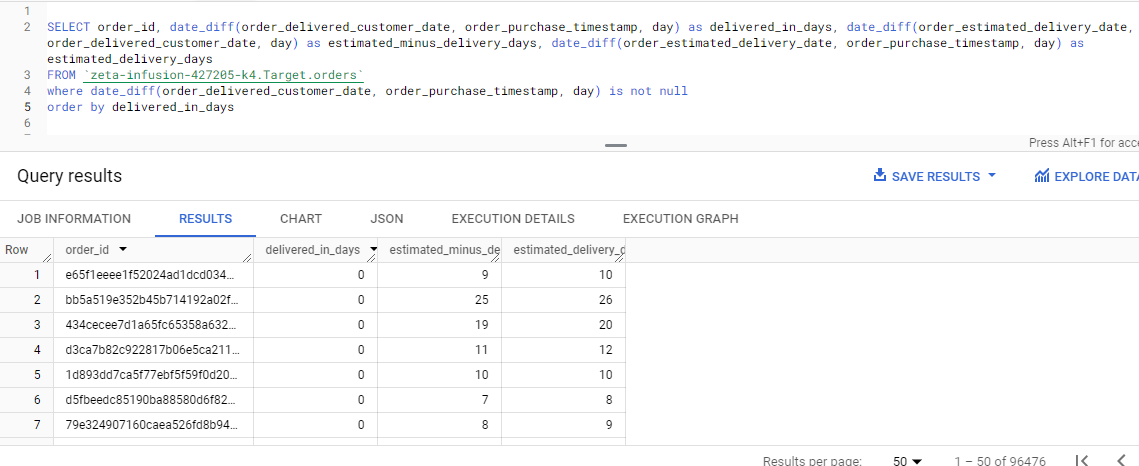
Answer

SELECT order\_id, date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day) as delivered\_in\_days, date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day) as estimated\_minus\_delivery\_days, date\_diff(order\_estimated\_delivery\_date, order\_purchase\_timestamp, day) as estimated\_delivery\_days

FROM `zeta-infusion-427205-k4.Target.orders`

where date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, day) is not null

order by delivered\_in\_days



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

Answer

SELECT  c.customer\_state, round(avg(freight\_value),2) as avg\_freight\_value  FROM `zeta-infusion-427205-k4.Target.order\_items` as oi

join `zeta-infusion-427205-k4.Target.orders` as o

on oi.order\_id = o.order\_id

join `zeta-infusion-427205-k4.Target.customers` as c

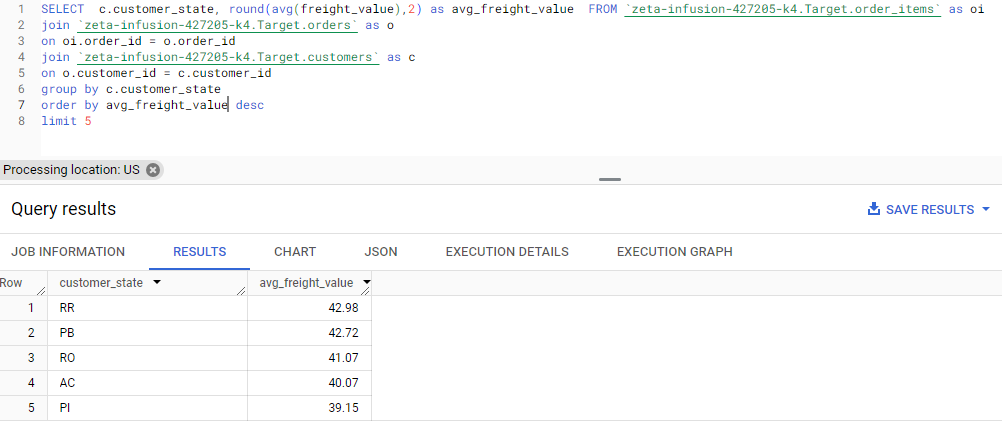
on o.customer\_id = c.customer\_id

group by c.customer\_state

order by avg\_freight\_value desc

limit 5

(Highest Freight Value)



(Lowest Freight Value)

with my\_cte as (

SELECT c.customer\_state, avg(freight\_value) as avg\_freight\_value  FROM `zeta-infusion-427205-k4.Target.order\_items` as oi

join `zeta-infusion-427205-k4.Target.orders` as o

on oi.order\_id = o.order\_id

join `zeta-infusion-427205-k4.Target.customers` as c

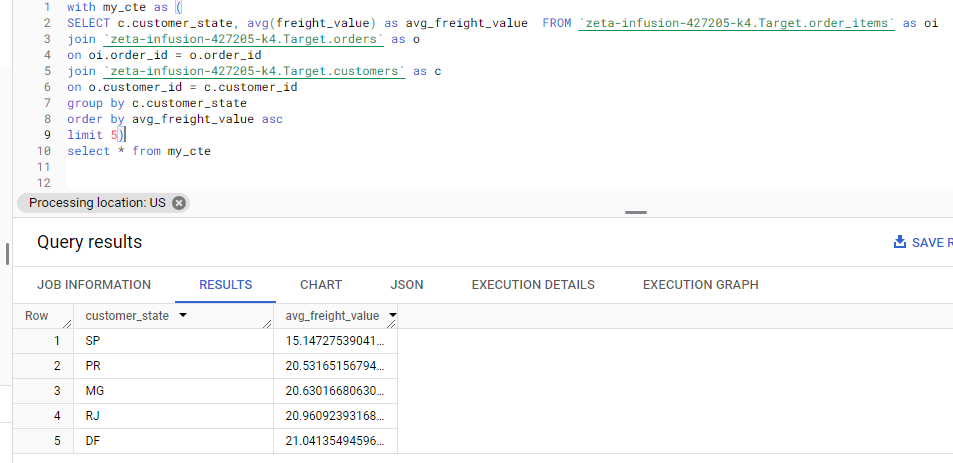
on o.customer\_id = c.customer\_id

group by c.customer\_state

order by avg\_freight\_value asc

limit 5)

select \* from my\_cte



Question 5 (3) Find out the top 5 states with the highest and lowest average delivery time.

Answer

SELECT  c.customer\_state, round(avg(date\_diff(date(order\_delivered\_customer\_date),

date(order\_purchase\_timestamp),day)),2) as avg\_time\_to\_deliver

  FROM `zeta-infusion-427205-k4.Target.orders` as o

join `zeta-infusion-427205-k4.Target.customers` as c

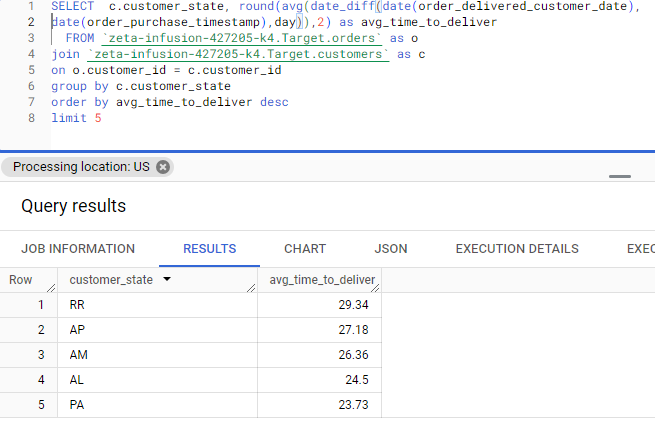
on o.customer\_id = c.customer\_id

group by c.customer\_state

order by avg\_time\_to\_deliver desc

limit 5

(highest average time)



(lowest average delivery time)

SELECT  c.customer\_state, round(avg(date\_diff(date(order\_delivered\_customer\_date),

date(order\_purchase\_timestamp),day)),2) as avg\_time\_to\_deliver

  FROM `zeta-infusion-427205-k4.Target.orders` as o

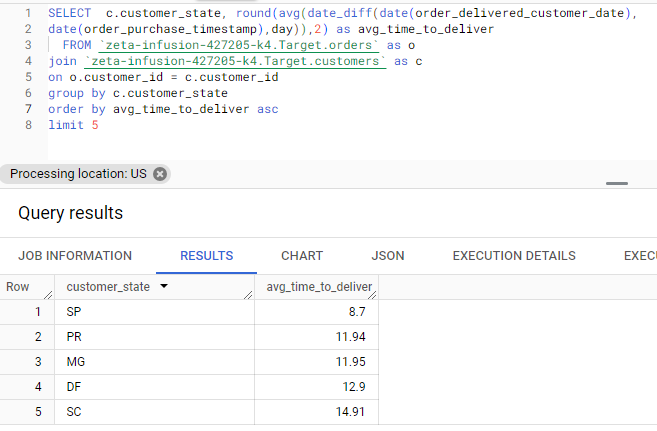
join `zeta-infusion-427205-k4.Target.customers` as c

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

group by c.customer\_state

order by avg\_time\_to\_deliver asc

limit 5



Question 5 (4) Find out the top 5 States where the order delivery is really fast as compared to the estimated date of delivery.

Answer

SELECT  customer\_state, round(avg(date\_diff(date(order\_estimated\_delivery\_date),

 date(order\_delivered\_customer\_date),day)),2) as fast\_delivered

 FROM `zeta-infusion-427205-k4.Target.orders` as o

join `zeta-infusion-427205-k4.Target.customers` as c

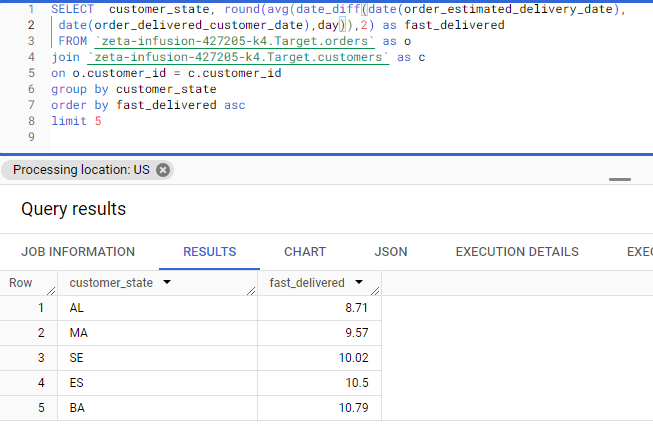
on o.customer\_id = c.customer\_id

group by customer\_state

order by fast\_delivered asc

limit 5

(States where delivery is fast)



(States where delivery is not so fast)

SELECT  customer\_state, round(avg(date\_diff(date(order\_estimated\_delivery\_date),

 date(order\_delivered\_customer\_date),day)),2) as fast\_delivered

 FROM `zeta-infusion-427205-k4.Target.orders` as o

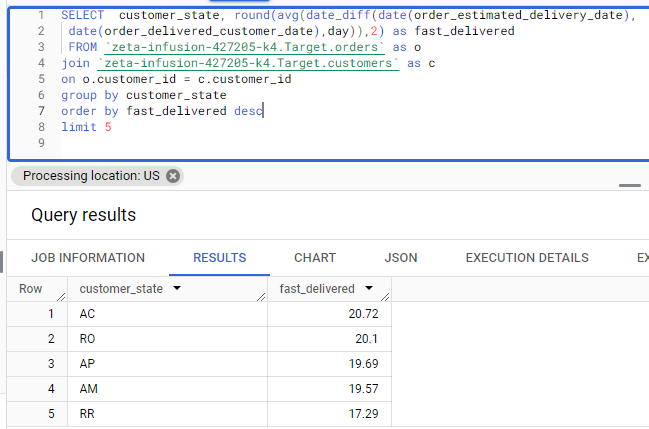
join `zeta-infusion-427205-k4.Target.customers` as c

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

group by customer\_state

order by fast\_delivered desc

limit 5



Question 6 (1) Find the month on month no of orders placed using different payment types

Answer

SELECT  p.payment\_type,

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

count(distinct o.order\_id) as order\_count

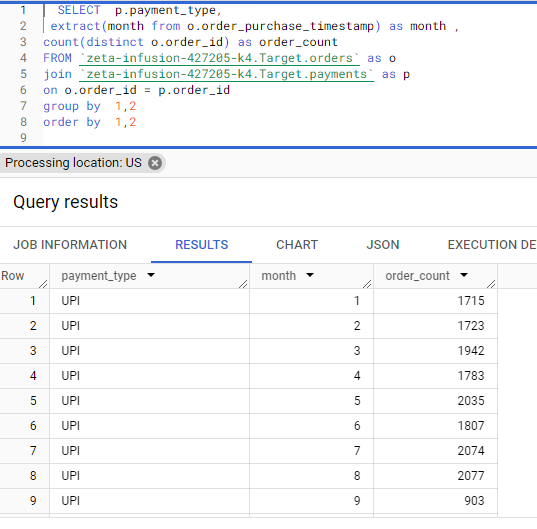
FROM `zeta-infusion-427205-k4.Target.orders` as o

join `zeta-infusion-427205-k4.Target.payments` as p

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

group by  1,2

order by  1,2



Question 6 (2) Find the no. of orders placed on the basis of the payment installments that have been paid.

Answer

SELECT p.payment\_installments , count(distinct o.order\_id) as order\_count

FROM `zeta-infusion-427205-k4.Target.orders` as o

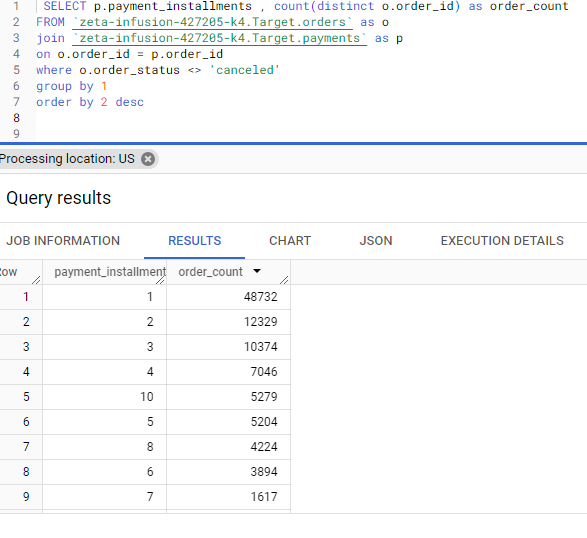
join `zeta-infusion-427205-k4.Target.payments` as p

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

where o.order\_status <> 'canceled'

group by 1

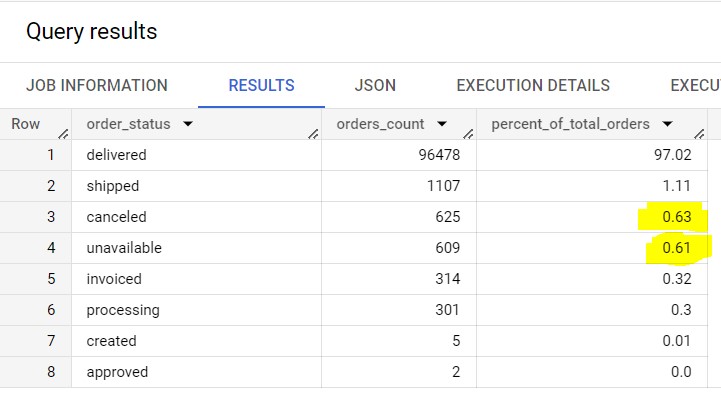
order by 2 desc



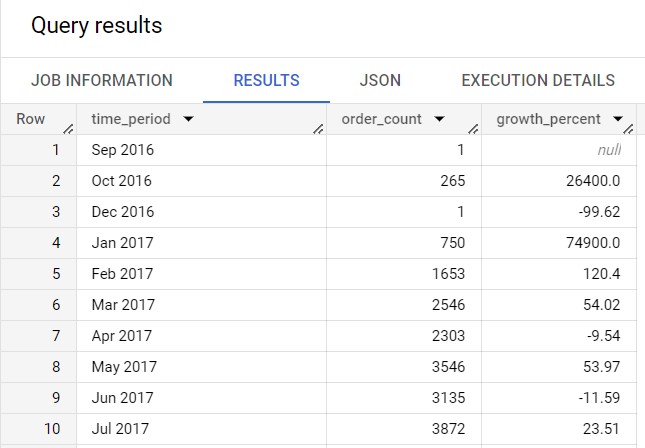
## Actionable Insights

* Total 609 orders were unavailable and 625 orders were cancelled during the given time period, which makes it to be around 1.2 % of total orders.

We can reduce this number by studying the reasons behind order cancellation and items unavailability.



* We can see how the orders trajectory is showing very abrupt increase in orders volume with in very short time. Looking at overall trend, it is seen that business is picking up very fast in brazil so company has to be ready with extra workforce. To avoid high risk, it can consider hiring contractual employees.



* Company received low rating for maximum orders in highlighted states; need to study further about the reasons for customer dissatisfaction to such great extent in these states.

This is the query for counting the number of rating in each state.

SELECT

\*

FROM ( SELECT

c.customer\_state, orv.review\_score

FROM

`target.order\_reviews` orv

JOIN

`target.orders` o

USING

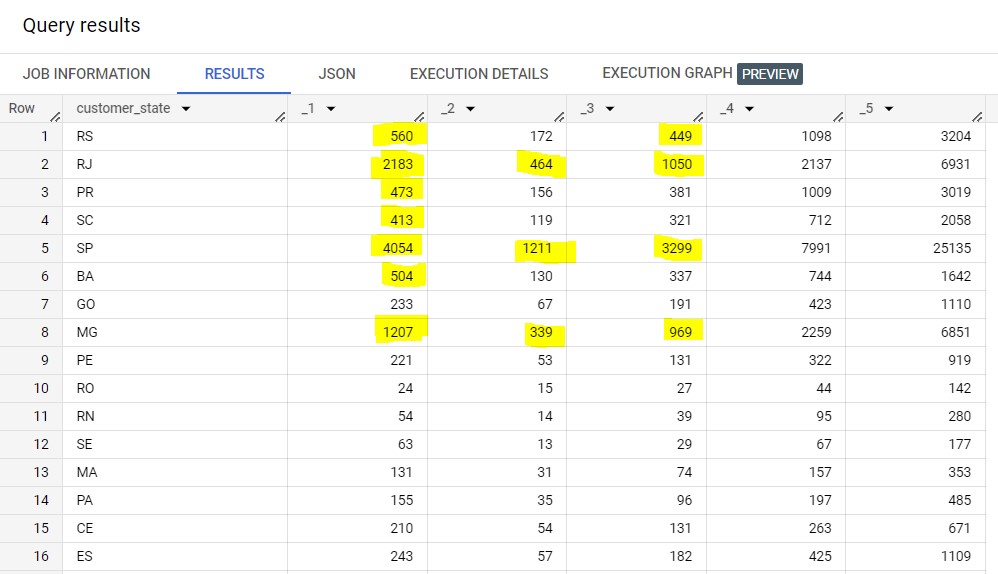
(order\_id)

JOIN

`target.customers` c

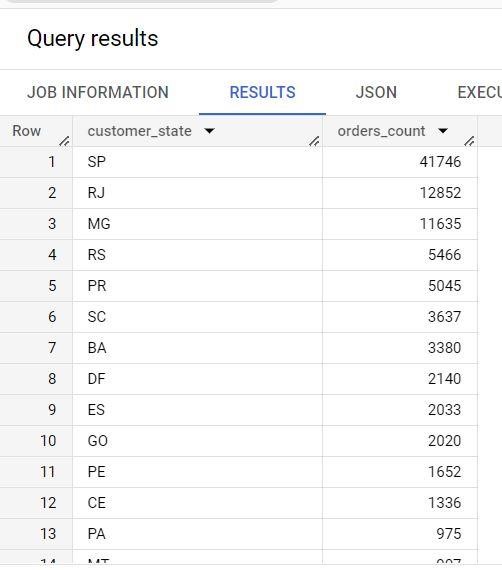
USING

(customer\_id)) PIVOT(COUNT(\*) FOR review\_score IN (1, 2, 3, 4, 5));



## Recommendations

* As Brazilian customers usually tend to buy in afternoon and night, we can increase staff in during this time frame in order to manage the customers’ requests, and services better during this time by reducing workforce of morning and dawn.
* We can see, only 3 state contribute for maximum volume, and rest of the state need to be focused for improving the business.



* Avg delivery time is quite high for most of those states from where company is receiving quite less volume of orders, detailed study is needed further for checking the other reasons behind such low volume of orders from majority of states. Huge delivery time can be the one of the reason and need to work on it.

States with highest average delivery time -

