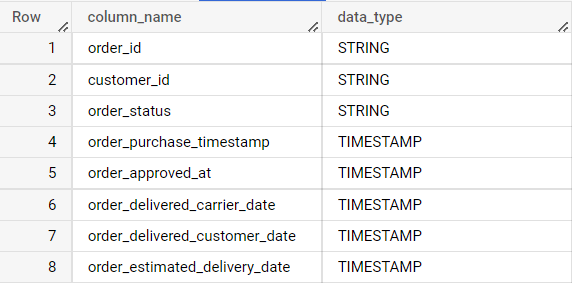
1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset**
   1. **Data type of columns in a table**

select column\_name, data\_type

from `solid-hope-367013.Target\_sql.INFORMATION\_SCHEMA.COLUMNS`

where table\_name = 'orders'



* 1. **Time period for which the data is given**

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

max(order\_purchase\_timestamp) as last\_purchase

from `solid-hope-367013.Target\_sql.orders` limit 10



* 1. **Cities and States covered in the dataset**

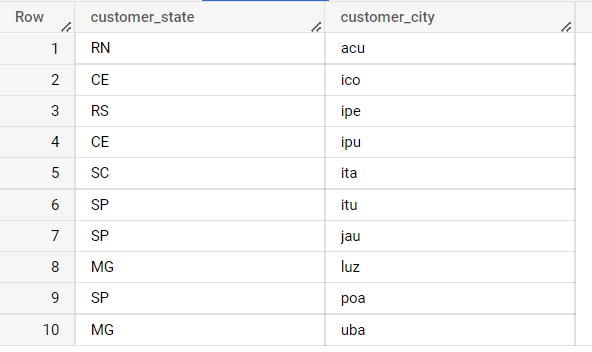
select

  c.customer\_state, c.customer\_city

from `solid-hope-367013.Target\_sql.orders` o  inner join  `solid-hope-367013.Target\_sql.customers` c

on o.customer\_id = c.customer\_id

group by 1,2 limit 10



1. **In-depth Exploration:**
   1. **Is there a growing trend on e-commerce in Brazil? How can we describe a complete scenario? Can we see some seasonality with peaks at specific months?**

select

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

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

  count(distinct (order\_id))

from `solid-hope-367013.Target\_sql.orders`

group by 1,2

order by 1,2



* 1. **What time do Brazilian customers tend to buy (Dawn, Morning, Afternoon or Night)?**

select

  sum(case when hours between 0 and 6 then orders end )as dawn,

  sum(case when hours between 7 and 12 then orders end )as morning,

  sum(case when hours between 13 and 18 then orders end )as eveining,

  sum(case when hours between 19 and 23 then orders end )as night,

  from

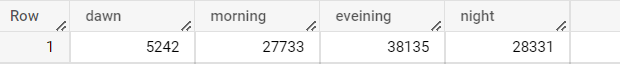
  (select

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

  count(distinct(order\_id)) as orders

  from `solid-hope-367013.Target\_sql.orders`

  group by 1) as a



1. **Evolution of E-commerce orders in the Brazil region:**
   1. **Get month on month orders by region, states**

select

c.customer\_state, c.customer\_city,

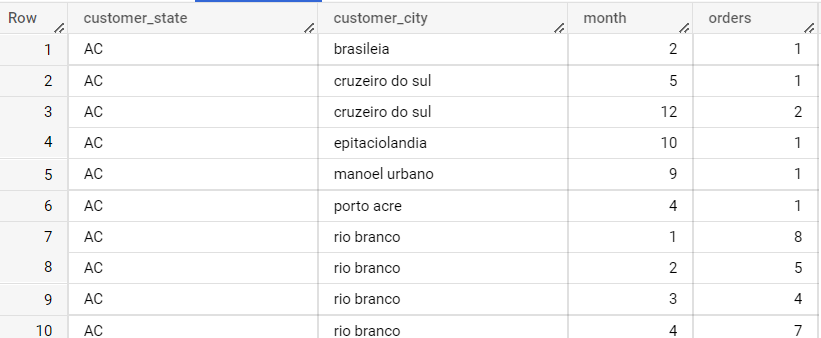
extract(month from order\_purchase\_timestamp) as month,

count(distinct(order\_id)) as orders

from `solid-hope-367013.Target\_sql.orders` o join `solid-hope-367013.Target\_sql.customers` c on o.customer\_id = c.customer\_id

group by 1,2,3

order by 1,2,3



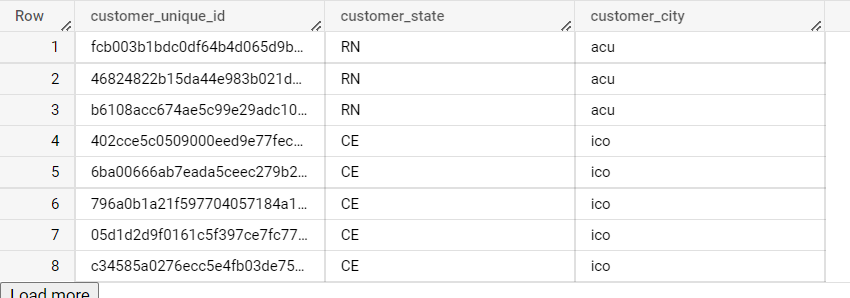
* 1. **How are customers distributed in Brazil**

select

distinct(customer\_unique\_id) ,

customer\_state, customer\_city

from `solid-hope-367013.Target\_sql.customers`



1. **Impact on Economy: Analyze the money movemented by e-commerce by looking at order prices, freight and others.**
   1. **Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)**

with base as (

select \* from `solid-hope-367013.Target\_sql.orders` a

inner join

`solid-hope-367013.Target\_sql.payments` b

on a.order\_id = b.order\_id

where

extract(year from a.order\_purchase\_timestamp) between 2017 and 2018

and

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

),

base\_2 as (

select extract(year from order\_purchase\_timestamp) as year, sum(payment\_value) as cost from base

group by 1

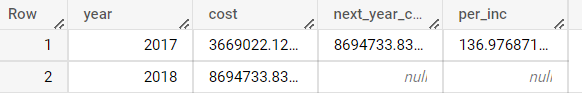
order by 1 asc

),

base\_3 as (

select \*, lead(cost, 1) over (order by year) as next\_year\_cost from base\_2)

select \*, (next\_year\_cost - cost)/ cost \*100 as per\_inc from base\_3



* 1. **Mean & Sum of price and freight value by customer state**

select

c.customer\_state, avg(price) as mean\_of\_price,

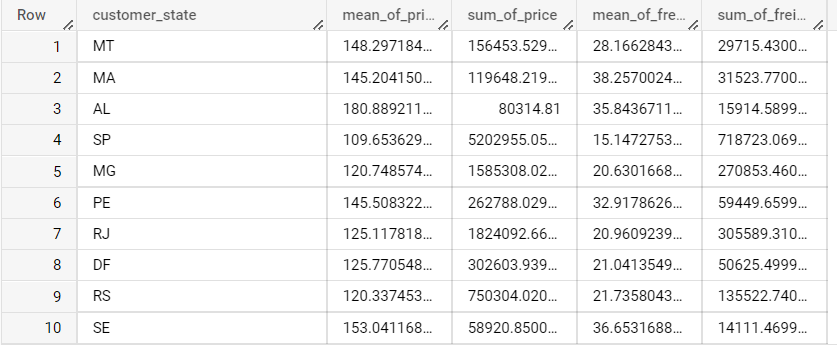
  sum(price) as sum\_of\_price,

  avg(freight\_value) as mean\_of\_freight,

  sum(freight\_value) as sum\_of\_freight

from `Target\_sql.customers` c join `Target\_sql.orders` o on c.customer\_id = o.customer\_id join `Target\_sql.order\_items` oi on o.order\_id = oi.order\_id

group by c.customer\_state

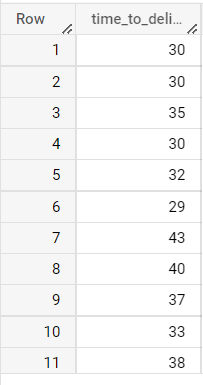


1. Analysis on sales, freight and delivery time
   1. Calculate days between purchasing, delivering and estimated delivery
   2. Create columns:
      * time\_to\_delivery = order\_purchase\_timestamp-order\_delivered\_customer\_date

select

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

 from `Target\_sql.orders`



* + - diff\_estimated\_delivery = order\_estimated\_delivery\_date-order\_delivered\_customer\_date

select

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

 from `Target\_sql.orders`



* 1. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery
  2. **Sort the data to get the following:**
     + **Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5**

select

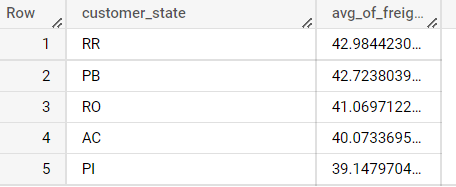
c.customer\_state,

avg(freight\_value) as avg\_of\_freight,

from `Target\_sql.customers` c join `Target\_sql.orders` o on c.customer\_id = o.customer\_id join `Target\_sql.order\_items` oi on o.order\_id = oi.order\_id

group by 1

order by avg\_of\_freight desc limit 5



select

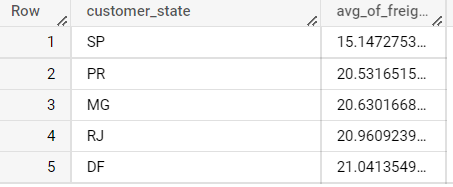
c.customer\_state,

avg(freight\_value) as avg\_of\_freight,

from `Target\_sql.customers` c join `Target\_sql.orders` o on c.customer\_id = o.customer\_id join `Target\_sql.order\_items` oi on o.order\_id = oi.order\_id

group by 1

order by avg\_of\_freight  limit 5



* + - **Top 5 states with highest/lowest average time to delivery**

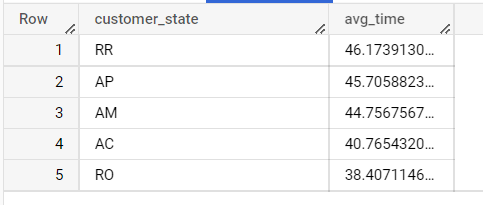
select

c.customer\_state,

  avg(date\_diff(order\_estimated\_delivery\_date,order\_purchase\_timestamp, day)) as avg\_time

 from `Target\_sql.orders` o join `Target\_sql.customers` c on o.customer\_id = c.customer\_id

 group by c.customer\_state

 order by avg\_time desc limit 5   
  


select

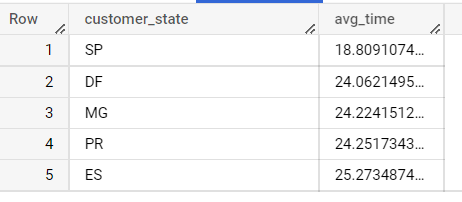
c.customer\_state,

  avg(date\_diff(order\_estimated\_delivery\_date,order\_purchase\_timestamp, day)) as avg\_time

 from `Target\_sql.orders` o join `Target\_sql.customers` c on o.customer\_id = c.customer\_id

 group by c.customer\_state

 order by avg\_time  limit 5



* + - **Top 5 states where delivery is really fast/ not so fast compared to estimated date**  
      **slow:**

select

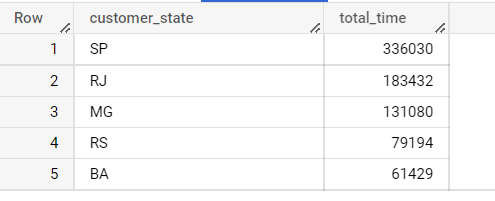
c.customer\_state,

sum(date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp, day)) as total\_time

 from `Target\_sql.orders` o join `Target\_sql.customers` c on o.customer\_id = c.customer\_id

 group by c.customer\_state

 order by total\_time desc limit 5



**Fast:**

select

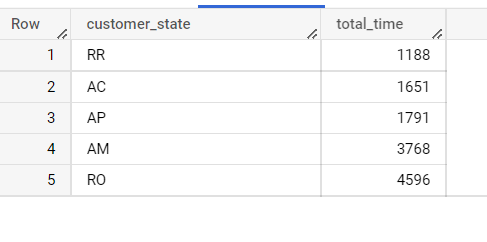
c.customer\_state,

sum(date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp, day)) as total\_time

 from `Target\_sql.orders` o join `Target\_sql.customers` c on o.customer\_id = c.customer\_id

 group by c.customer\_state

 order by total\_time  limit 5



1. **Payment type analysis:**
   1. **Month over Month count of orders for different payment types**

select

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

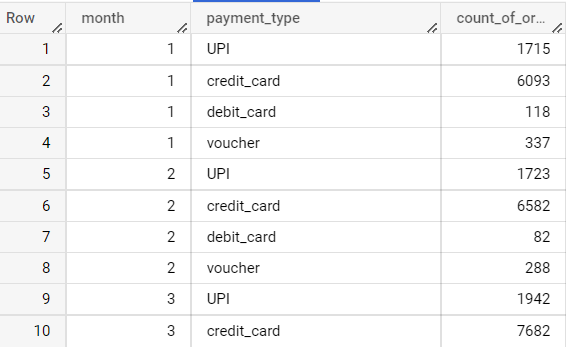
  payment\_type,

  count(distinct (p.order\_id)) as count\_of\_orders

from `solid-hope-367013.Target\_sql.payments` p join `Target\_sql.orders` o on p.order\_id = o.order\_id

group by month, payment\_type

order by month,payment\_type



* 1. **Distribution of payment installments and count of orders**

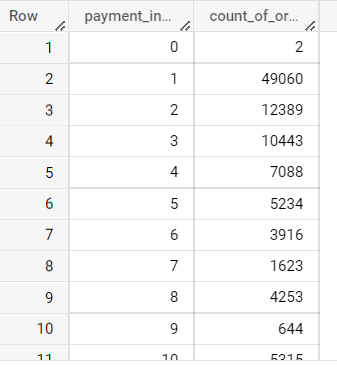
select

payment\_installments,

count(distinct(order\_id)) as count\_of\_orders

from `Target\_sql.payments`

group by payment\_installments



**Actionable Insights:**1. There was an **increasing/ growing trend** on e-commerce in brazil, high sales were recored in the month of **november** 2. Brazilian customers tend to buy more in the **evening**

3.**136.9%** increase in the cost of orders from 2017 to 2018

4. **customers like to use credit cards** for payments  
 **Recommendations:**  
1. as brazil e-commerence is in growing trend, there will be a huge demand in future.  
2.make sure there will be full function and available of products at evenings.  
3. need to make sure that products reach customers before estimated delivery time in some regions.  
4.majority of the customers use credit cards, providing offers on cards can increase sales.  
5.need to deliver the products ASAP