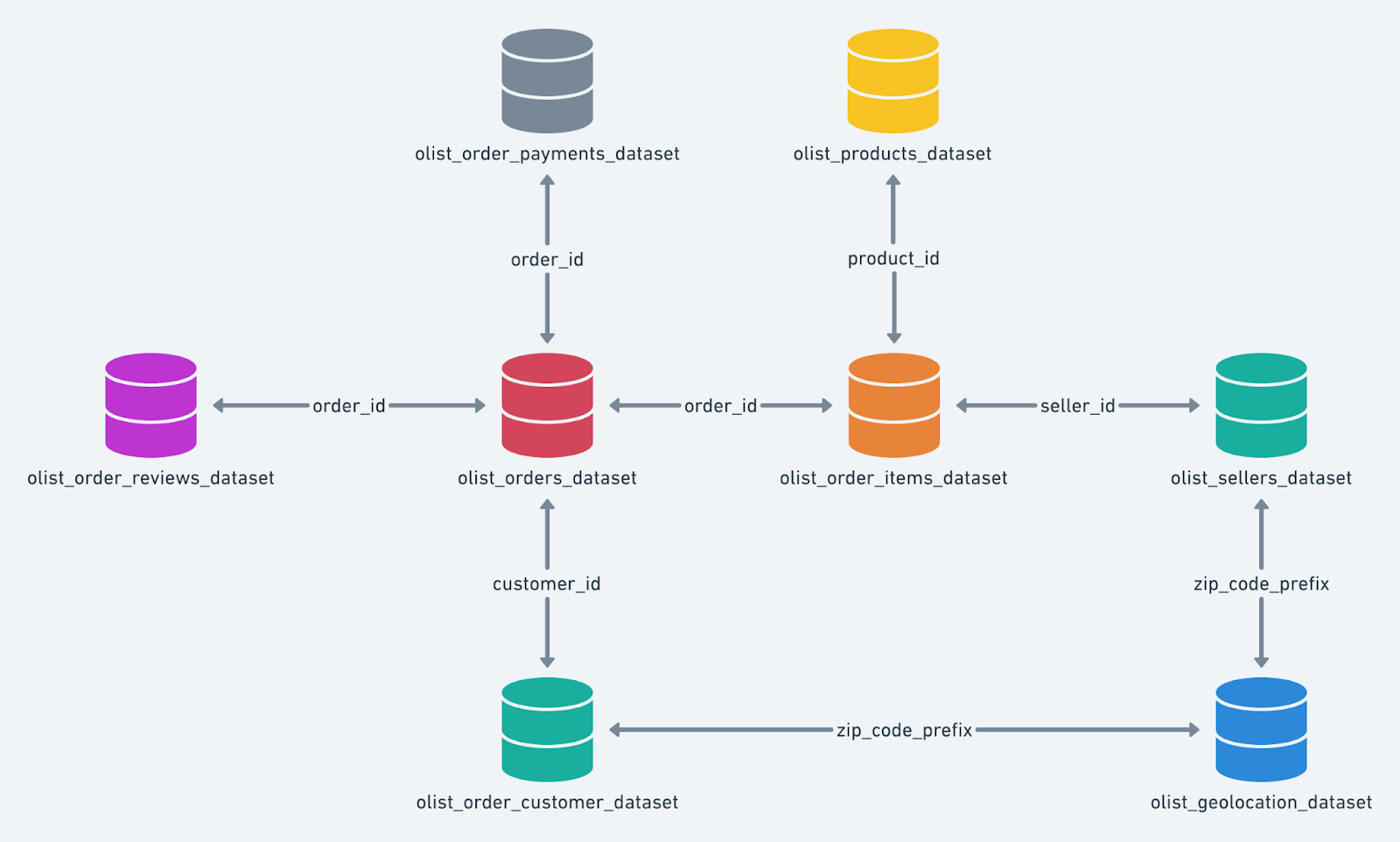
#### **Target – case study using SQL**

**Context**

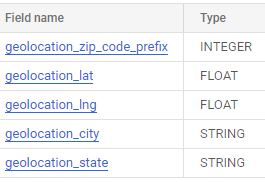
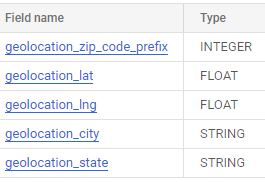
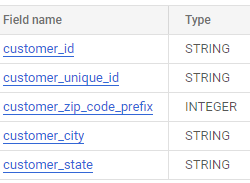
Target is one of the world’s most recognized brands and one of America’s leading retailers. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This business case has information of 100k orders from 2016 to 2018 made at Target in Brazil. Its features allows viewing an order from multiple dimensions: from order status, price, payment and freight performance to customer location, product attributes and finally reviews written by customers.

**High level overview of relationship between datasets:**



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



* 1. Time period for which the data is given

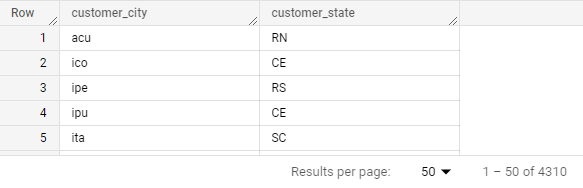
select extract(year from min(order\_purchase\_timestamp)) as min\_year,

extract(year from max(order\_purchase\_timestamp)) as max\_year from `my-project-target-382813.Target.orders` ;



* 1. Cities and States of customers ordered during the given period

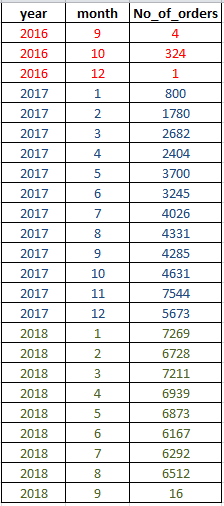
SELECT distinct customer\_city, customer\_state FROM `my-project-target-382813.Target.customers` ;



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(order\_id) as No\_of\_orders  FROM `my-project-target-382813.Target.orders` group by year, month order by  year,month;



**Yes, There’s a Increased growth on e-commerce in brazil. In the year 2016 there was a slow growth and later at the start of 2017 there was good increase in growth when compared to previous year and then in the year 2018 there was a static no. of orders until august.**

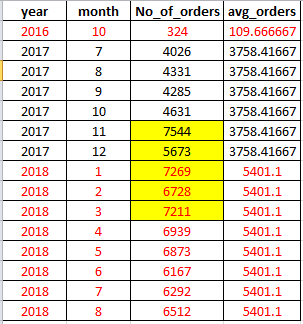
select u.\* from (select  t.\*, avg(No\_of\_orders)over(partition by year) as avg\_orders

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

extract(month from order\_purchase\_timestamp) as month , count(order\_id) as No\_of\_orders,

FROM `my-project-target-382813.Target.orders` group by year, month order by  year,month) t)u

where No\_of\_orders > avg\_orders order by year, month ;

****

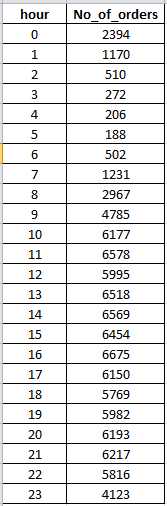
**Yes, At specific months like November and December of 2017, January, February and march of 2018 are peak in no. of orders.**

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

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

count(order\_id) as No\_of\_orders  FROM `my-project-target-382813.Target.orders`

group by hour order by  hour;



select distinct Time\_tend\_to\_buy,sum(No\_of\_orders)over(partition by Time\_tend\_to\_buy)

as sum\_of\_orders

from (with c as (SELECT  extract(hour from order\_purchase\_timestamp) as hour,

count(order\_id) as No\_of\_orders  FROM `my-project-target-382813.Target.orders`

group by hour order by  hour) select \*, case when hour >= 0 and hour < 7 then 'dawn'

                          when hour >= 7 and hour < 11 then 'morning'

                          when hour >= 11 and hour < 18 then 'afternoon'

                          when hour >=18 and hour <= 23 then 'night'

                          end as Time\_tend\_to\_buy from c) u;



**Therefore, Afternoon between 11:00 to 18:00 customers tend to buy more.**

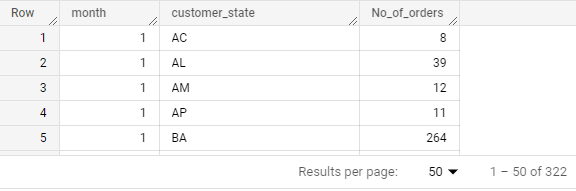
1. Evolution of E-commerce orders in the Brazil region:
   1. Get month on month orders by states

SELECT  extract(month from order\_purchase\_timestamp) as month , B.customer\_state,

count(A.order\_id) as No\_of\_orders  FROM `my-project-target-382813.Target.orders` A

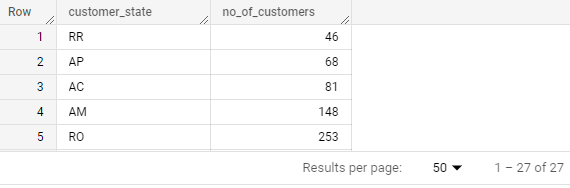
join `my-project-target-382813.Target.customers` B on A.customer\_id = B.customer\_id

group by B.customer\_state, month order by  month, B.customer\_state;



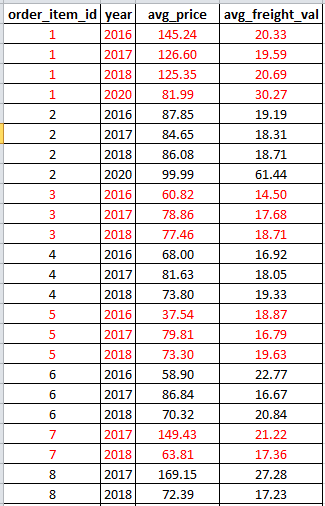
* 1. Distribution of customers across the states in Brazil

select customer\_state, count(customer\_id) as no\_of\_customers from `my-project-target-382813.Target.customers` group by customer\_state order by no\_of\_customers;



1. Impact on Economy: Analyse the money movement by e-commerce by looking at order prices, freight and others.

select order\_item\_id, extract(year from shipping\_limit\_date) as year,  avg(price) as avg\_price, avg(freight\_value) as avg\_freight\_val from `Target.order\_items` group by year, order\_item\_id order by order\_item\_id, year;



* 1. Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only) - You can use “payment\_value” column in payments table

select year, sum(payment\_value) as cost\_of\_orders  from (select extract (year from shipping\_limit\_date) as year,extract (month from shipping\_limit\_date) as month, p.payment\_value from `Target.payments` p join `Target.order\_items` o on o.order\_id = p.order\_id) t where (month >=1 and month<= 8) and year<=2018 group by year order by year;



with c as(select year, sum(payment\_value) as cost\_of\_orders  from (select extract (year from shipping\_limit\_date) as year,extract (month from shipping\_limit\_date) as month, p.payment\_value from `Target.payments` p join `Target.order\_items` o on o.order\_id = p.order\_id) t where (month >=1 and month<= 8) and year<=2018 group by year order by year)

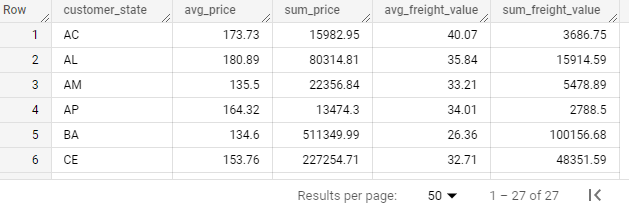
selectround(((cost\_of\_orders -lag(cost\_of\_orders)over(order by year asc ))/lag(cost\_of\_orders)over(order by year asc )) \* 100)

as percentage\_increase from c  ;



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

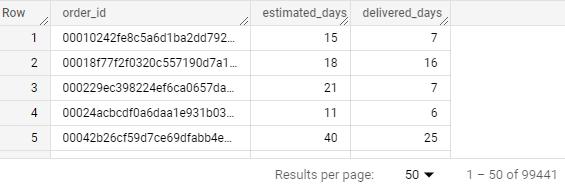
select distinct customer\_state, round(avg(price)over(partition by customer\_state),2) as avg\_price, round(sum(price)over(partition by customer\_state),2) as sum\_price, round(avg(freight\_value)over(partition by customer\_state),2) as avg\_freight\_value, round(sum(freight\_value)over(partition by customer\_state),2) as sum\_freight\_value from Target.customers c join Target.orders o on c.customer\_id = o.customer\_id join Target.order\_items oi on o.order\_id = oi.order\_id order by customer\_state;



5. Analysis on sales, freight and delivery time

1. Calculate days between purchasing, delivering and estimated delivery

select order\_id,date\_diff( order\_estimated\_delivery\_date , order\_purchase\_timestamp, day) as estimated\_days, date\_diff(order\_delivered\_customer\_date , order\_purchase\_timestamp, day) as delivered\_days  from Target.orders;

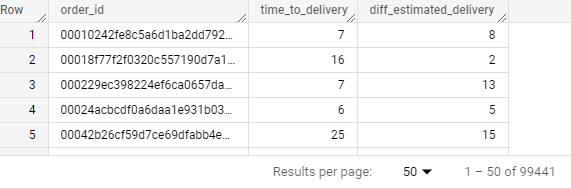


1. Find time\_to\_delivery & diff\_estimated\_delivery.

Formula for the same given below:

* + time\_to\_delivery = order\_purchase\_timestamp - order\_delivered\_customer\_date
  + diff\_estimated\_delivery = order\_estimated\_delivery\_date -order\_delivered\_customer\_date

select order\_id, date\_diff( order\_delivered\_customer\_date , order\_purchase\_timestamp, day) as time\_to\_delivery, date\_diff(order\_estimated\_delivery\_date , order\_delivered\_customer\_date, day) as diff\_estimated\_delivery  from Target.orders order by order\_id;



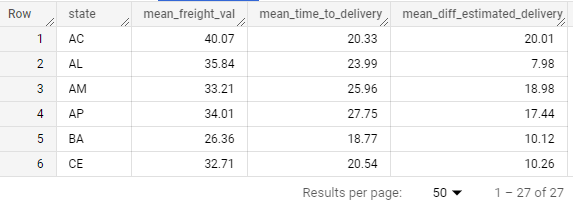
1. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery

select distinct state, mean\_freight\_val, round(avg(time\_to\_delivery)over(partition by state),2) as mean\_time\_to\_delivery, round(avg(diff\_estimated\_delivery)over(partition by state),2) as mean\_diff\_estimated\_delivery from

 (select customer\_state as state, round(avg(freight\_value)over(partition by customer\_state),2) as mean\_freight\_val,

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

from Target.customers c join Target.orders o on c.customer\_id = o.customer\_id join Target.order\_items oi on o.order\_id = oi.order\_id order by customer\_state)t order by state;



1. Sort the data to get the following:

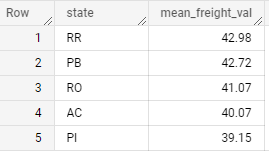
\*\*\*\*\*\*\*\*\*\*\*\* can also be done by considering above table as cte \*\*\*\*\*\*\*\*\*\*\*\*

1. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5

**Highest average freight value :**

select distinct state, mean\_freight\_val  from (select customer\_state as state, round(avg(freight\_value)over(partition by customer\_state),2) as mean\_freight\_valfrom Target.customers c join Target.orders o on c.customer\_id = o.customer\_id join Target.order\_items oi on o.order\_id = oi.order\_id order by customer\_state) t

 order by mean\_freight\_val desc limit 5;



1. Top 5 states with highest/lowest average time to delivery

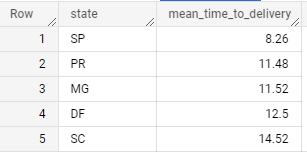
**Lowest average time to delivery :**

select distinct state, round(avg(time\_to\_delivery)over(partition by state),2) as mean\_time\_to\_delivery from

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

  from Target.customers c join Target.orders o on c.customer\_id = o.customer\_id join Target.order\_items oi on o.order\_id = oi.order\_id order by customer\_state)t

order by mean\_time\_to\_delivery asc limit 5;



1. Top 5 states where delivery is really fast/ not so fast compared to estimated date

**Delivery is really fast when then mean\_time\_to\_delivery is less than the mean\_diff\_estimated\_delivery or when difference between them is less than 1.**

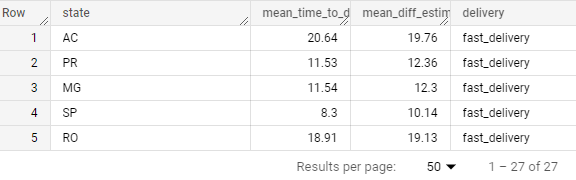
select state , mean\_time\_to\_delivery, mean\_diff\_estimated\_delivery, case when (mean\_time\_to\_delivery < mean\_diff\_estimated\_delivery) or (mean\_time\_to\_delivery - mean\_diff\_estimated\_delivery)<1

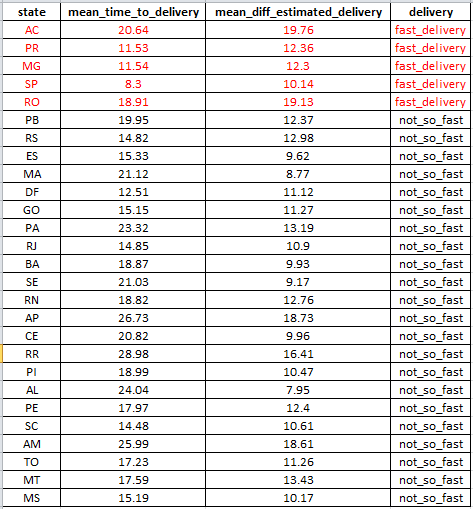
then "fast\_delivery" else "not\_so\_fast" end as delivery

  from (select distinct state, round(avg(time\_to\_delivery)over(partition by state),2) as mean\_time\_to\_delivery, round(avg(diff\_estimated\_delivery)over(partition by state),2) as mean\_diff\_estimated\_delivery from

 (select customer\_state as state, date\_diff( order\_delivered\_customer\_date , order\_purchase\_timestamp, day)  as time\_to\_delivery, date\_diff(order\_estimated\_delivery\_date , order\_delivered\_customer\_date, day) as diff\_estimated\_delivery

  from Target.customers c join Target.orders o on c.customer\_id = o.customer\_id  order by customer\_state)) order by delivery;



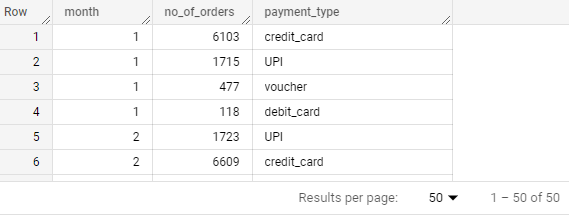


Top 5 states with fast\_delivery marked in RED color out of 27 states.

6. Payment type analysis:

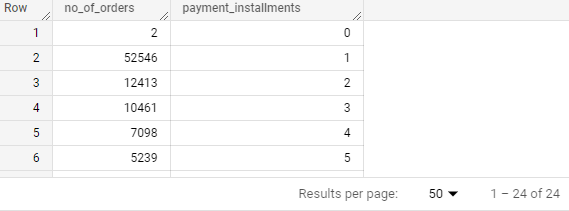
1. Month over Month count of orders for different payment types

select extract(month from(order\_purchase\_timestamp)) as month, count(o.order\_id) as no\_of\_orders,payment\_type from Target.orders o join Target.payments p on o.order\_id = p.order\_id group by month,payment\_type order by month;



1. Count of orders based on the no. of payment instalments.

select  count(o.order\_id) as no\_of\_orders, payment\_installments from Target.orders o join Target.payments p on o.order\_id = p.order\_id group by payment\_installments order by payment\_installments;



**Analysis by,**

**Name : Meghana S**

**Batch : DSML Feb23 Beginner**