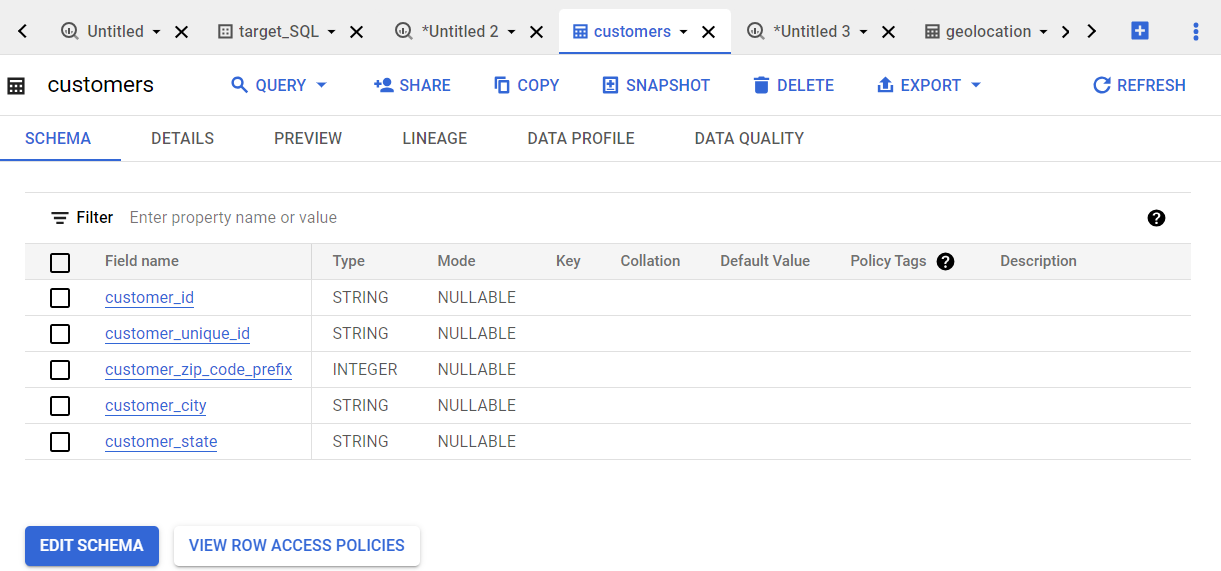
**I. Import the dataset and do usual exploratory analysis steps like checking the**

**structure & characteristics of the dataset.**

1. Data type of all columns in the “customers” table.

**ANS: **

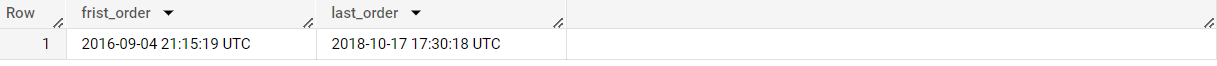
**Insights:** In the data type more string data types

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

**ANS:**

SELECT min(order\_purchase\_timestamp) as frist\_order,

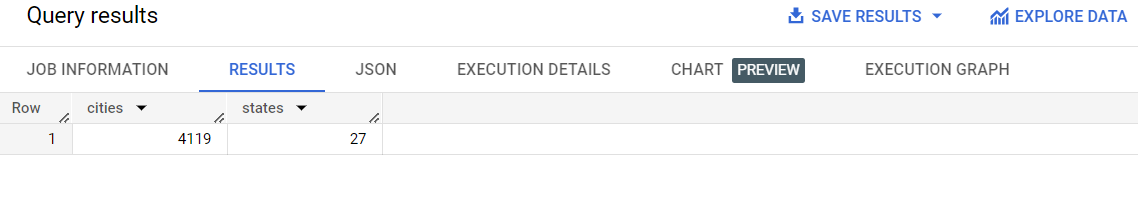
max(order\_purchase\_timestamp) as last\_order FROM `scalar-dsml-sql-raju.target\_SQL.orders`

****

**Insights:** This market is opened for 2 years from 04/09/2016 and last order placed on 17/09/2018.

1. Count the number of Cities and States in our dataset.

**Ans:**  select count(distinct c.customer\_city) cities ,count(distinct c.customer\_state) states from `target\_SQL.customers` c join `target\_SQL.orders` o on c.customer\_id=o.customer\_id



**Insights:** the total number of cities 4119 and 27 states.

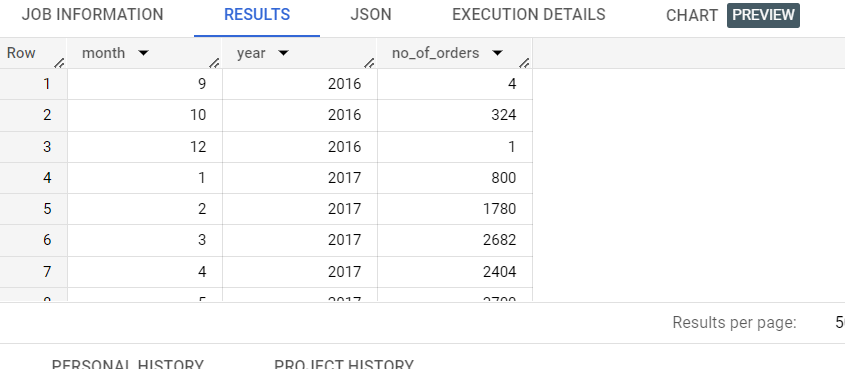
**II. In-depth Exploration:**

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

**ANS:**

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

count(\*) as no\_of\_orders from `target\_SQL.orders` group by month, year order by year asc, month asc



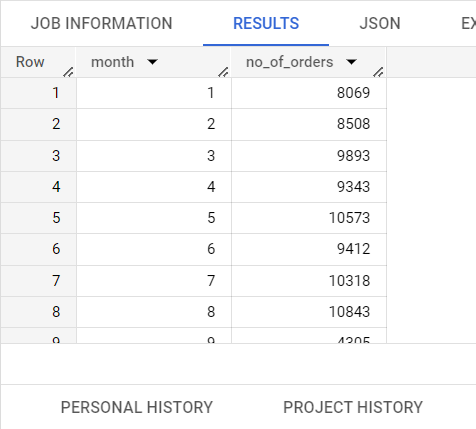
**Insights:** no. of orders in growing trend

**B.** Can we see some kind of monthly seasonality in terms of the no. of orders being

placed?

**Ans:** select extract(month from order\_purchase\_timestamp) as month, count(\*) as no\_of\_orders from `target\_SQL.orders`

group by month order by month



**Insights:**  7,8 months are peaked in no.of orders placed

**C.** During what time of the day, do the Brazilian customers mostly place their

orders? (Dawn, Morning, Afternoon or Night)

● 0-6 hrs : Dawn

● 7-12 hrs : Mornings

● 13-18 hrs : Afternoon

● 19-23 hrs : Night

**ANS:**

select case

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

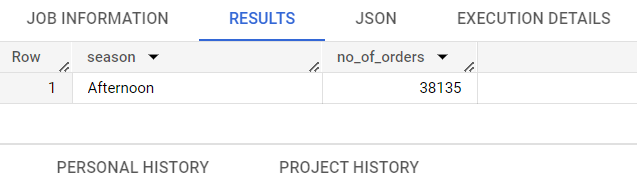
when extract(hour from order\_purchase\_timestamp) between 7 and 12 then 'Morning'

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' end as season,

count(\*) as no\_of\_orders from `target\_SQL.orders` group by season order by no\_of\_orders desc

limit 1



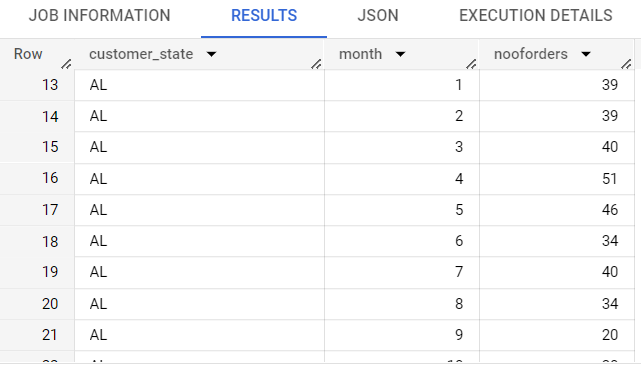
**Insight:**  afternoon is the most placed orders

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**III. Evolution of E-commerce orders in the Brazil region:**

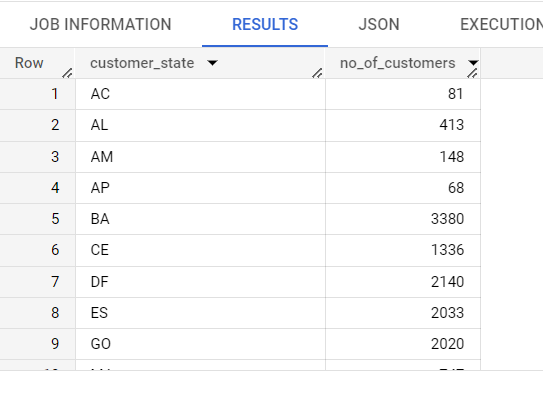
1. Get the month on month no. of orders placed in each state.

**Ans:** select c.customer\_state,extract(month from order\_purchase\_timestamp) as month,count(\*) as no\_of\_orders from `target\_SQL.customers` c join `target\_SQL.orders` o on c.customer\_id=o.customer\_id group by 1,2 order by 1,2

****

1. How are the customers distributed across all the states?

**Ans:** select customer\_state ,count(\*) as no\_of\_customers from `target\_SQL.customers` group by 1 order by 1,2

****

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**IV. Impact on Economy: Analyze the money movement by e-commerce by looking at**

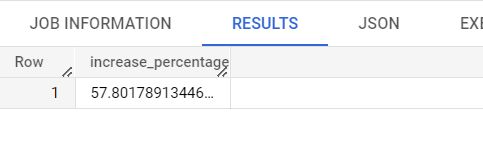
**order prices, freight and others.**

**A.** Get the % increase in the cost of orders from year 2017 to 2018 *(include*

*months between Jan to Aug only).*

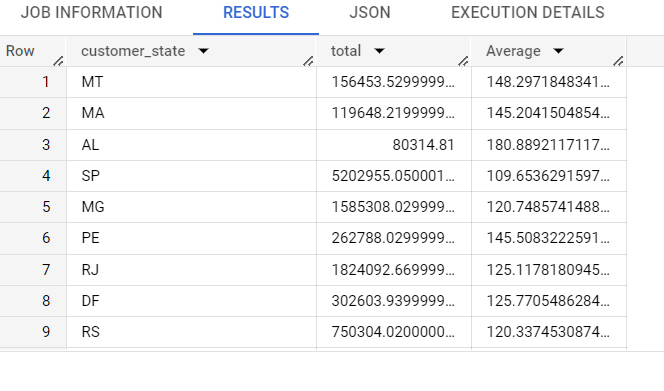
**Ans:** select ((lag\_value-cost\_of\_orders)/lag\_value )\* 100 increase\_percentage from (select year,cost\_of\_orders,lag(cost\_of\_orders) over (order by year desc) lag\_value, from (select extract ( year from order\_purchase\_timestamp) as year,sum(payment\_value) as cost\_of\_orders from `target\_SQL.orders` o join `target\_SQL.payments` p on o.order\_id=p.order\_id where extract ( year from order\_purchase\_timestamp) in (2017,2018) and extract ( month from order\_purchase\_timestamp) in (1,2,3,4,5,6,7,8) group by 1

) sub)anothersub order by 1 desc limit 1

****

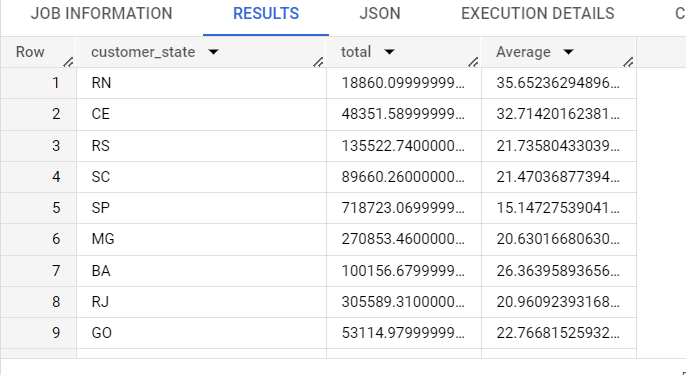
1. Calculate the Total & Average value of order price for each state.

**Ans:** select c.customer\_state,sum(price) as total ,Avg(price) as Average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi on oi.order\_id=o.order\_id group by 1

****

1. Calculate the Total & Average value of order freight for each state.

**Ans:** select c.customer\_state,sum(freight\_value) as total ,Avg(freight\_value) as Average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi on oi.order\_id=o.order\_id group by 1

****

**V. Analysis based on sales, freight and delivery time.**

**A.** Find the no. of days taken to deliver each order from the order’s purchase date

as delivery time.

Also, calculate the difference (in days) between the estimated & actual delivery

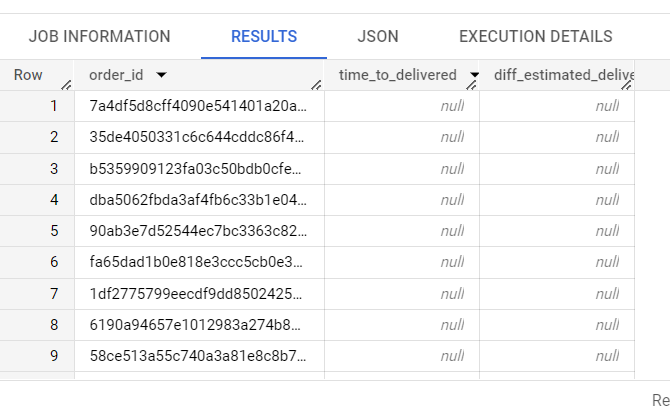
date of an order.

Do this in a single query.

**Ans:** select order\_id, ceil(timestamp\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, hour)/24) as time\_to\_delivered,

ceil(timestamp\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, hour)/24) as diff\_estimated\_delivery from

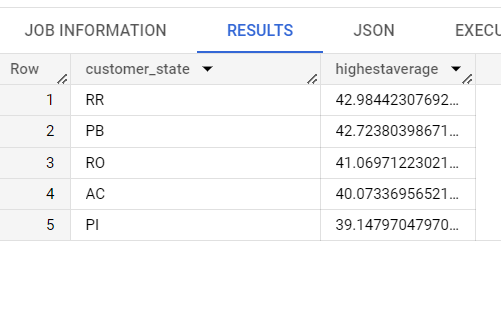
`target\_SQL.orders` order by time\_to\_delivered



**B.** Find out the top 5 states with the highest & lowest average freight value.

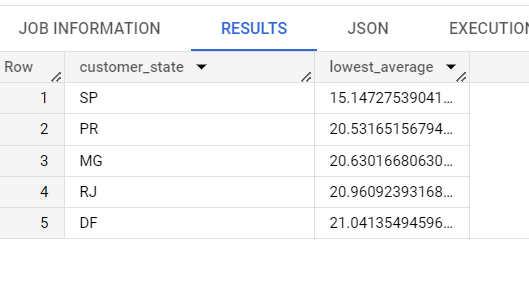
select c.customer\_state,avg(freight\_value) as highest\_average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi

on oi.order\_id=o.order\_id group by 1 order by 2 desc limit 5



select c.customer\_state,avg(freight\_value) as lowest\_average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi

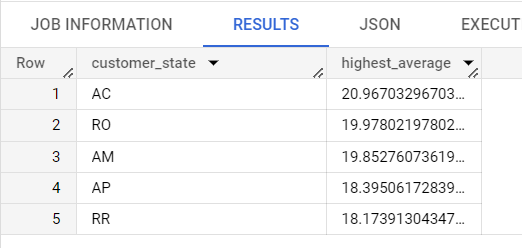
on oi.order\_id=o.order\_id group by 1 order by 2 asc  limit 5

****

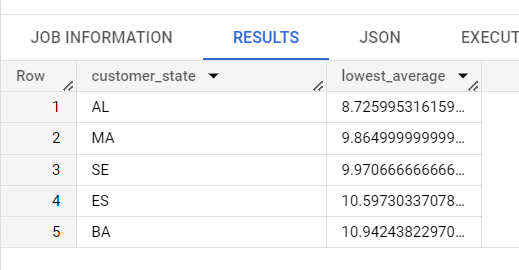
1. Find out the top 5 states with the highest & lowest average delivery time.

**ANS:**

select c.customer\_state,avg(ceil(timestamp\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date ,hour)/24)) as highest\_average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi on oi.order\_id=o.order\_id group by 1 order by 2 desc limit 5

****

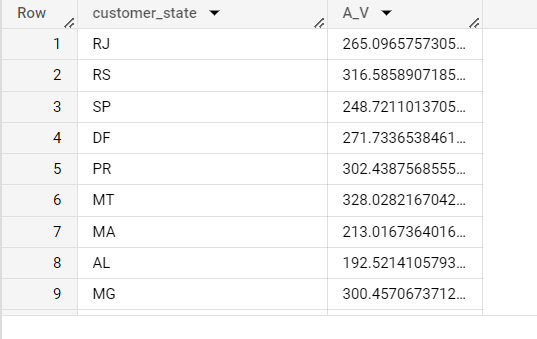
select c.customer\_state,avg(ceil(timestamp\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date ,hour)/24)) as lowest\_average from `target\_SQL.customers` c join `target\_SQL.orders` o on o.customer\_id=c.customer\_id join `target\_SQL.order\_items`as oi on oi.order\_id=o.order\_id group by 1 order by 2 asc limit 5

****

**D.** Find out the top 5 states where the order delivery is really fast as compared to

the estimated date of delivery.

**ANS:** select customer\_state, avg(timestamp\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, hour)) as A\_V from `target\_SQL.orders` o join `target\_SQL.customers` c on c.customer\_id = o.customer\_id group by 1

****

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**VI. Analysis based on the payments:**

1. Find the month on month no. of orders placed using different payment types**.**

**ANS:**

select payment\_type, extract(month from order\_purchase\_timestamp ) as month, count(\*) as no\_of\_orders from `target\_SQL.payments` p join `target\_SQL.orders` o on o.order\_id=p.order\_id group by 1,2 order by 2,1

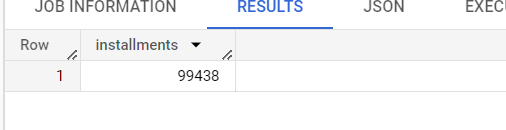
****

**B.** Find the no. of orders placed on the basis of the payment installments that have

been paid.

**ANS:**

select count(distinct(order\_id)) as installments from `target\_SQL.payments` where payment\_installments>0

****