**Business Case: SQL**

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
   2. Time period for which the data is given
   3. Cities and States covered in the dataset

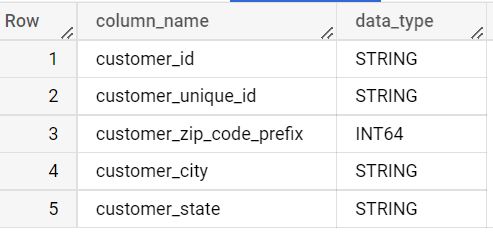
Solution :

* + - * Data type of columns in a table

SELECT column\_name, data\_type

FROM `<CompanySchema>.INFORMATION\_SCHEMA.COLUMNS`

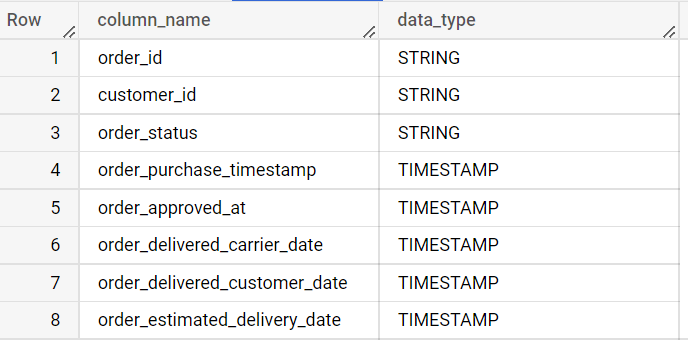
WHERE table\_name="customers";



SELECT column\_name, data\_type

FROM `<CompanySchema>.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name="orders";



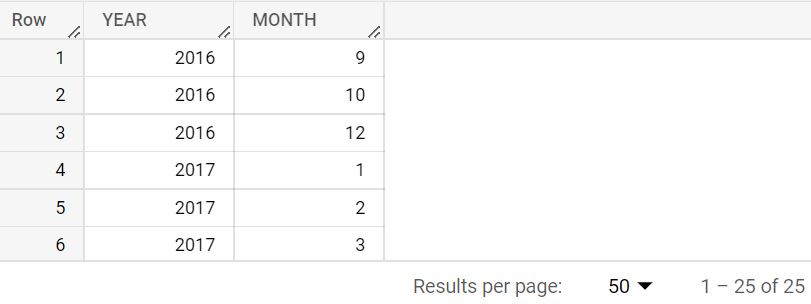
* + - * Time period for which the data is given

select distinct EXTRACT(YEAR from order\_purchase\_timestamp) YEAR,

EXTRACT(MONTH from order\_purchase\_timestamp) MONTH

from <CompanySchema>.orders

order by YEAR,MONTH;

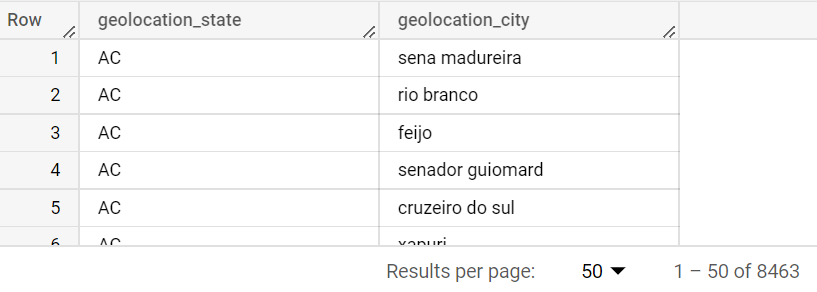


\*\* Data Time Period : Sept,2016 to Oct,2018

* + - * Cities and States covered in the dataset

select distinct geolocation\_state,geolocation\_city from <CompanySchema>.geolocation

order by geolocation\_state;

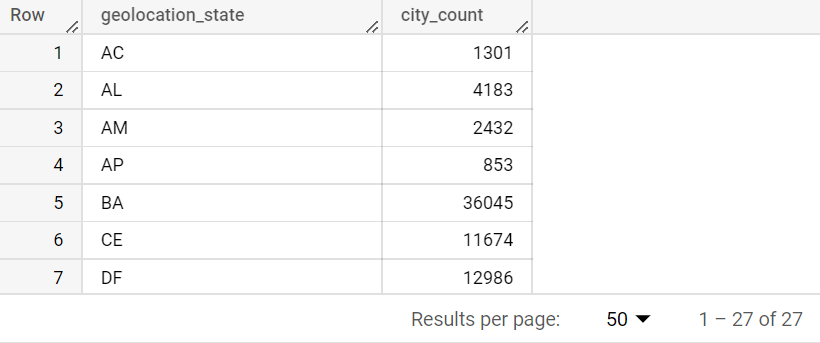


select distinct geolocation\_state,

count(geolocation\_city) as city\_count from <CompanySchema>.geolocation

group by geolocation\_state

order by geolocation\_state;



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

Solution :

* + - * 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 \*,

No\_of\_Order-LAG(No\_of\_Order,1) over(partition by ORDER\_YEAR order by ORDER\_MONTH asc) as growth\_over\_year,

dense\_rank() over(partition by ORDER\_YEAR order by No\_of\_Order desc) as rank\_orderValue,

from

(select ORDER\_YEAR,ORDER\_MONTH,count(order\_id) No\_of\_Order from

(select distinct order\_id,EXTRACT(YEAR  from order\_purchase\_timestamp) as ORDER\_YEAR,

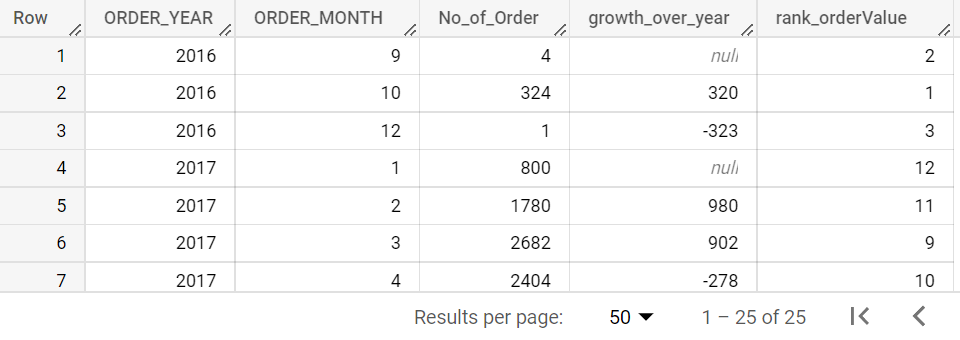
EXTRACT(MONTH  from order\_purchase\_timestamp) AS ORDER\_MONTH

from <CompanySchema>.orders) t

group by ORDER\_YEAR,ORDER\_MONTH

ORDER BY ORDER\_YEAR,ORDER\_MONTH) t2

ORDER BY ORDER\_YEAR,ORDER\_MONTH;



\*\* No. of orders is surely increasing over months with some exceptional dips for few months in between. Positive growth\_over\_year shows No. of Orders getting increased over month.

Checking the rank of No. of orders for each month gives us an idea that order has been on peak around September to November month every year, though for 2018 it’s not the same because it looks like data is not complete for 2018 end of the month in this dataset.

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

select \* from

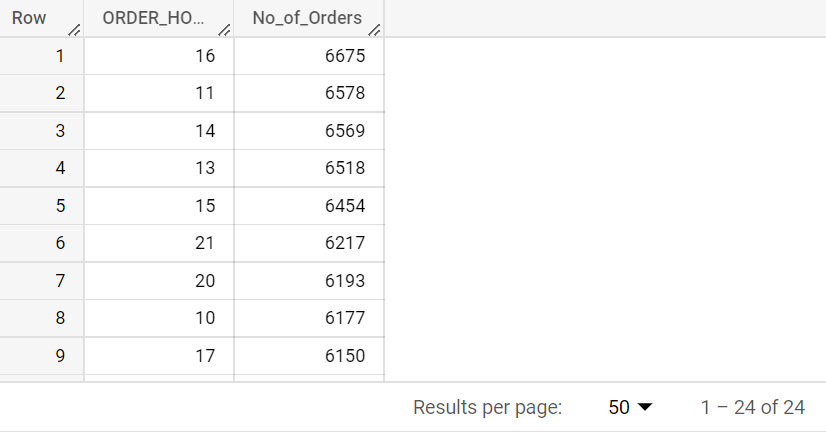
(select EXTRACT(HOUR  from order\_purchase\_timestamp) as ORDER\_HOUR,

count(distinct order\_id) as No\_of\_Orders

from <CompanySchema>.orders

group by EXTRACT(HOUR  from order\_purchase\_timestamp))

order by No\_of\_Orders desc;



\*\* No. of Orders is highest at around 16:00 hour which is Afternoon time, however, No. of orders pick up at around 10 in the morning and till 22:00 – 23:00 hour the volume of order is high.

1. Evolution of E-commerce orders in the Brazil region:
   1. Get month on month orders by region, states
   2. How are customers distributed in Brazil

Solution :

* + - * Get month on month orders by region, states

select distinct customer\_state, customer\_city, ORDER\_YEAR, ORDER\_MONTH,

count(distinct order\_id) over(partition by customer\_state, customer\_city, ORDER\_YEAR,ORDER\_MONTH) No\_of\_Orders FROM

(select c.customer\_state, c.customer\_city, order\_id, EXTRACT(YEAR  from order\_purchase\_timestamp) as ORDER\_YEAR,

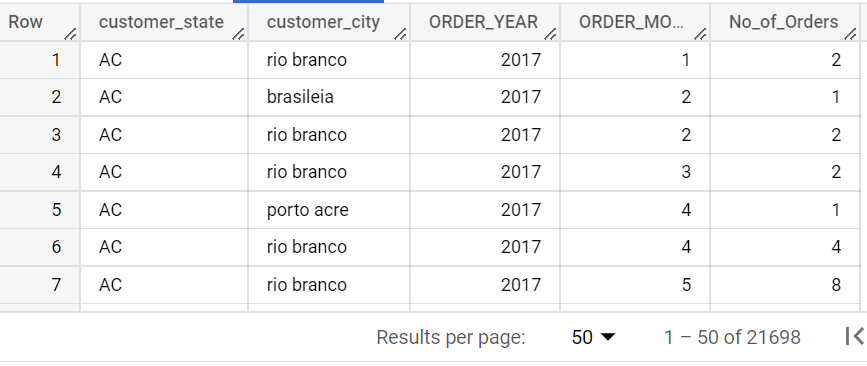
EXTRACT(MONTH  from order\_purchase\_timestamp) AS ORDER\_MONTH

from <CompanySchema>.customers c

left join <CompanySchema>.orders o

on c.customer\_id=o.customer\_id)

order by customer\_state,ORDER\_YEAR,ORDER\_MONTH,customer\_city;



* + - * How are customers distributed in Brazil

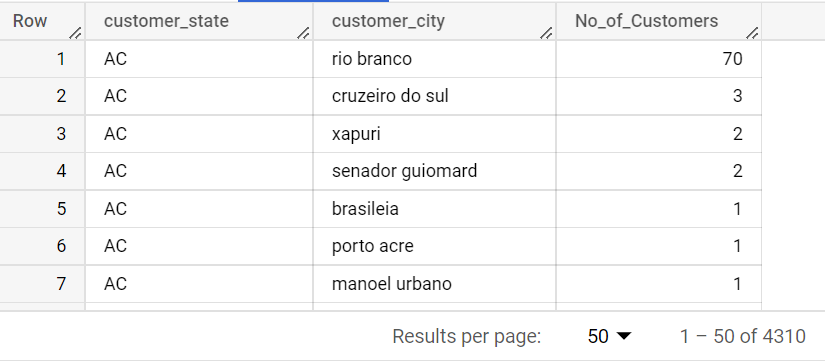
select customer\_state,customer\_city,

count(distinct customer\_id) No\_of\_Customers

from <CompanySchema>.customers

group by customer\_state,customer\_city

order by customer\_state,No\_of\_Customers desc;



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)
   2. Mean & Sum of price and freight value by customer state

Solution :

* + - * Get % increase in cost of orders from 2017 to 2018 (include months between Jan to Aug only)

select ORDER\_YEAR, sum(Order\_cost) OrderCostSum\_overYear,

(sum(Order\_cost)-LAG(sum(Order\_cost)) over(order by sum(Order\_cost)))\*100/LAG(sum(Order\_cost)) over(order by sum(Order\_cost)) PercentValue\_diff

from

(select distinct order\_id, EXTRACT(YEAR from shipping\_limit\_date) ORDER\_YEAR, price as Order\_cost

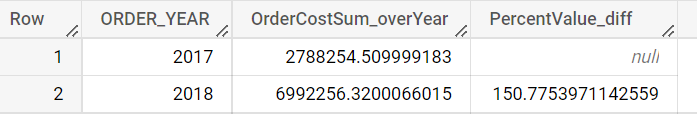
from <CompanySchema>.order\_items

where EXTRACT(MONTH from shipping\_limit\_date) between 1 and 8

and EXTRACT(YEAR from shipping\_limit\_date) in (2017,2018))

group by ORDER\_YEAR

order by ORDER\_YEAR;



\*\* For the months between Jan to July, % increase in sum of cost of orders from 2017 to 2018 is around 150%.

* + - * Mean & Sum of price and freight value by customer state

select customer\_state,sum(price+freight\_value) sum\_CostValue,

avg(price+freight\_value) mean\_CostValue from

(select distinct customer\_state,oi.order\_id,price,freight\_value from

<CompanySchema>.customers c inner join <CompanySchema>.orders o

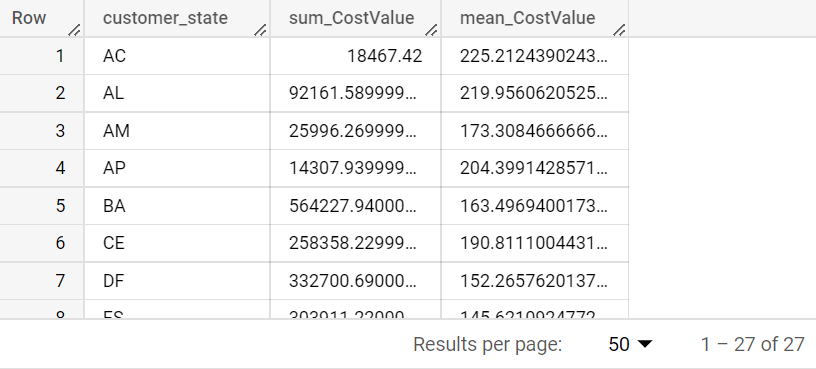
on c.customer\_id=o.customer\_id

inner join <CompanySchema>.order\_items oi

on o.order\_id=oi.order\_id)

group by customer\_state

order by customer\_state;



\*\* For different customer states, sum and mean value for Cost (Price + Freight value) is calculated as above.

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
      * diff\_estimated\_delivery = order\_estimated\_delivery\_date-order\_delivered\_customer\_date
   3. Group data by state, take mean of freight\_value, time\_to\_delivery, diff\_estimated\_delivery
   4. Sort the data to get the following:
      * 1. Top 5 states with highest/lowest average freight value - sort in desc/asc limit 5
        2. Top 5 states with highest/lowest average time to delivery
        3. Top 5 states where delivery is really fast/ not so fast compared to estimated date

Solution :

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

select customer\_state, avg(time\_to\_delivery) mean\_time\_to\_delivery,

avg(diff\_estimated\_delivery) mean\_diff\_estimated\_delivery, avg(freight\_value) mean\_freight\_value from

(select distinct t.\*,freight\_value from

(select distinct customer\_state,o.order\_id,order\_purchase\_timestamp,order\_estimated\_delivery\_date,order\_delivered\_customer\_date,

TIMESTAMP\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) time\_to\_delivery,

TIMESTAMP\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY) diff\_estimated\_delivery

from <CompanySchema>.orders o

join <CompanySchema>.customers c

on o.customer\_id=c.customer\_id) t

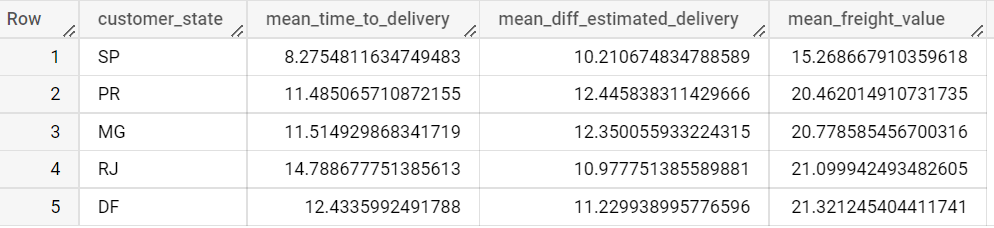
join <CompanySchema>.order\_items oi

on t.order\_id=oi.order\_id)

group by customer\_state

order by mean\_freight\_value

limit 5;



\*\* Mean time to delivery is calculated as average of ‘order delivered to customer’ date minus ‘order purchase date’ over different customer state.

Mean difference of estimated delivery is calculated as average of ‘order estimated delivery date’ minus ‘order delivered customer date’ over different customer state.

Also, Mean Freight value is calculated over each customer state and is ordered by the same in ascending and data limited to count 5.

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

select customer\_state, avg(time\_to\_delivery) mean\_time\_to\_delivery,

avg(diff\_estimated\_delivery) mean\_diff\_estimated\_delivery,avg(freight\_value) mean\_freight\_value from

(select distinct t.\*,freight\_value from

(select distinct customer\_state,o.order\_id,order\_purchase\_timestamp,order\_estimated\_delivery\_date,order\_delivered\_customer\_date,

TIMESTAMP\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) time\_to\_delivery,

TIMESTAMP\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY) diff\_estimated\_delivery

from <CompanySchema>.orders o

join <CompanySchema>.customers c

on o.customer\_id=c.customer\_id) t

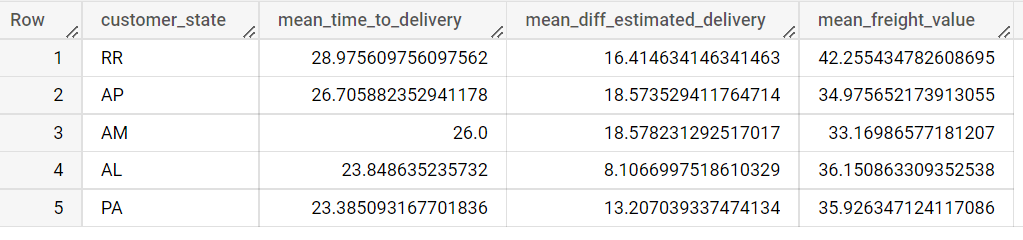
join <CompanySchema>.order\_items oi

on t.order\_id=oi.order\_id)

group by customer\_state

order by mean\_time\_to\_delivery desc

limit 5;



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

select customer\_state, avg(time\_to\_delivery) mean\_time\_to\_delivery,

avg(diff\_estimated\_delivery) mean\_diff\_estimated\_delivery,

avg(freight\_value) mean\_freight\_value from

(select distinct t.\*,freight\_value from

(select distinct customer\_state,o.order\_id,order\_purchase\_timestamp,order\_estimated\_delivery\_date,order\_delivered\_customer\_date,

TIMESTAMP\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp,DAY) time\_to\_delivery,

TIMESTAMP\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,DAY) diff\_estimated\_delivery

from <CompanySchema>.orders o

join <CompanySchema>.customers c

on o.customer\_id=c.customer\_id) t

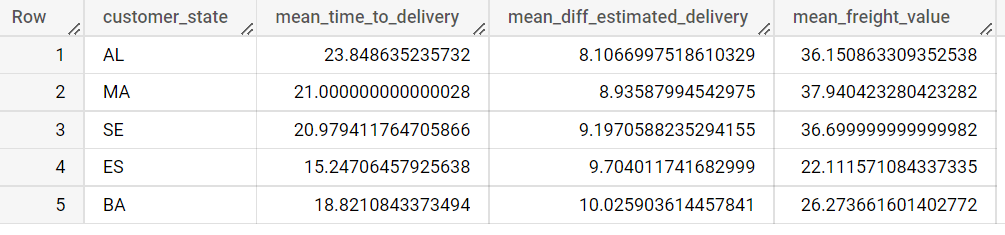
join <CompanySchema>.order\_items oi

on t.order\_id=oi.order\_id)

group by customer\_state

order by mean\_diff\_estimated\_delivery

limit 5;



\*\* AL customer state has least mean difference of delivered order from estimated order.

1. Payment type analysis:
   1. Month over Month count of orders for different payment types
   2. Distribution of payment installments and count of orders

Solution :

* + - * Month over Month count of orders for different payment types

select Order\_Year,Order\_Month,payment\_type,count(order\_id) count\_orders from

(select distinct

EXTRACT(YEAR from order\_purchase\_timestamp) Order\_Year,EXTRACT(MONTH from order\_purchase\_timestamp) Order\_Month,payment\_type,o.order\_id

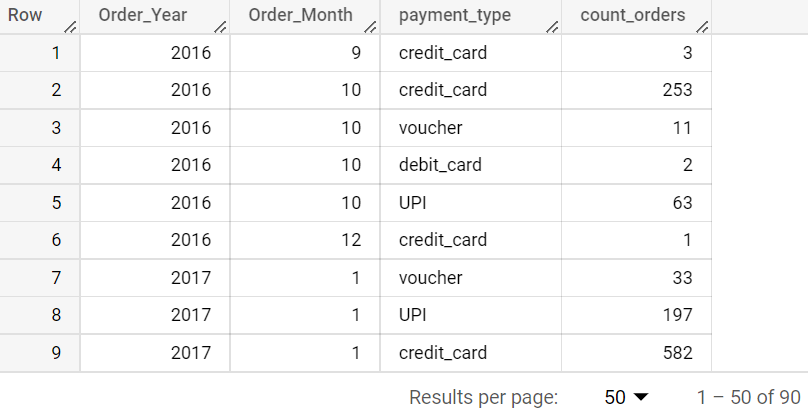
from <CompanySchema>.orders o

inner join <CompanySchema>.payments p

on o.order\_id=p.order\_id)

group by Order\_Year,Order\_Month,payment\_type

order by Order\_Year,Order\_Month;



\*\* From above analysis, we can see that for different payment type over different month of 2016 year and onwards, the count of orders are varying.

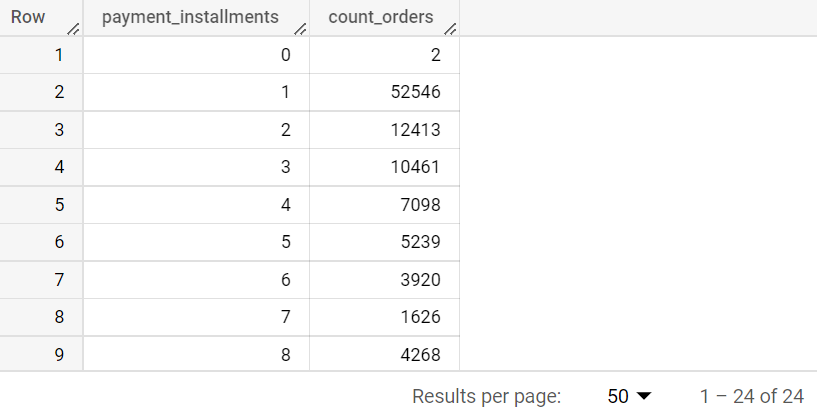
* + - * Distribution of payment installments and count of orders

select payment\_installments, count(order\_id) count\_orders

from <CompanySchema>.payments

group by payment\_installments

order by payment\_installments;



\*\* Count of orders calculation for different payment installments.

1. Actionable Insights
   * + - As we can see that more orders are placed at around 16:00 hour afternoon, so we need to make sure products are available in plenty and should not go out of stock
       - Focus on states where time of delivery is more to bring down the No. of days to be taken to deliver the product
       - Increase delivery frequency/agents to bring down estimated time of delivery for orders
       - For the states, where average freight value is more, we need to further break down the cost to be bear by customers and bring it down
       - Based on the review score, improve the service and quality of products and delivery experience for customers
2. Recommendations
   * + - Around September to November of every year, the customer engagement is more, so we can provide benefits like discounts, vouchers, offers, etc to retain the customers and attract more customers
       - Credit card seems to be used of more often as payment mode, so we can come up with good offers on credit card payments.
       - Also, we can collaborate with more credit card banks for allow payments for customers
       - Looking at number of customers for different cities each state, we can see the engagement and focus on bringing more customers for other cities
       - Looking into count of sellers per city for different states, we can bring in more sellers to better meet demand-supply

\*\*\*\*\* End of Document \*\*\*\*\*