**Business Case: Target SQL**

**Q1. 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:

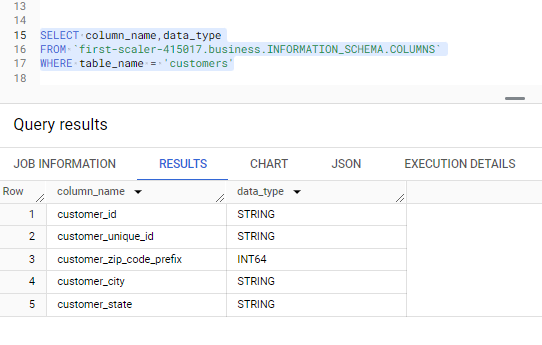
Query:

select column\_name, data\_type

from `first-scaler-415017.business.INFORMATION\_SCHEMA.COLUMNS`

where table\_name = 'customers'

Screenshot: -



Insights: N/A

Recommendation: N/A

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

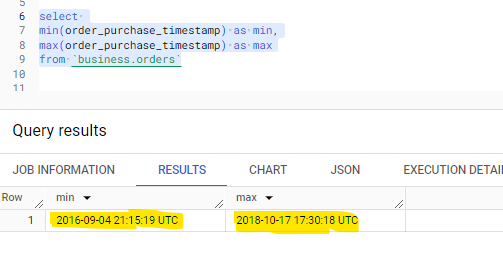
Ans:

Query:

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

from `business.orders`

Screenshot: -



Insights: -

Recommendation: -

1. Count the Cities & States of customers who ordered during the given period.

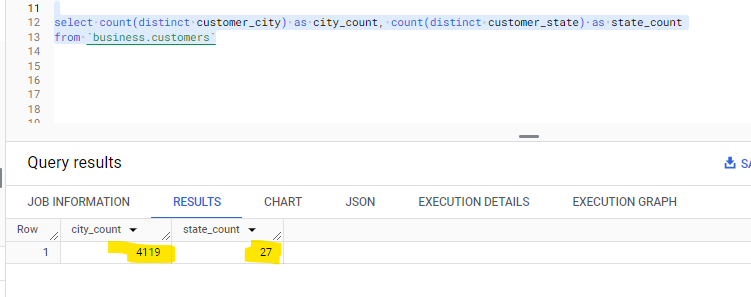
Ans: -

Query: -

select count(distinct customer\_city) as city\_count, count(distinct customer\_state) as state\_count

from `business.customers`

Screenshot: -



Insights: -

Suggestions: -

**Q2. In-depth Exploration:**



Query:

select year, count(\*) as order\_count\_per\_year

from

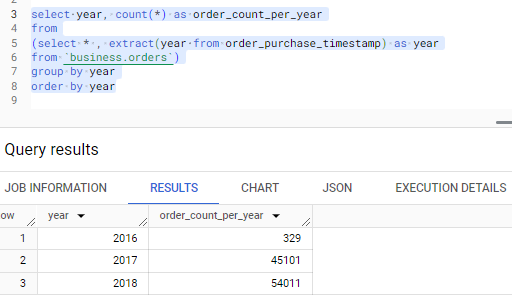
(select \* , extract(year from order\_purchase\_timestamp) as year

from `business.orders`)

group by year

order by year

Screenshot:



Insights:

Yes, the number of orders were increased over the years

In 2016, we have 329 orders,

In 2017, we have 45,101 orders,

in 2018, we have 54,011 orders.

In 2016 we have only 329 orders because we have data from September 2016 to December 2016. i.e., In 2016 first order in the order table was on 2016-09-04 and last order was on 2016-12-23. Also, if we observe carefully the orders in 2016 are very less, because let us assume in 2017 the total orders are 45000 so on an average 3750 orders were placed every month.

Suggestions:-

2.

3.

Query:

select

different\_times, count(different\_times) as count\_different\_times from

(select order\_id,order\_time,order\_purchase\_timestamp,

case

when order\_time between '00:00:00' and '06:59:59'

then  'Dawn'

when order\_time between '07:00:00' and '12:59:59'

then 'Mornings'

when order\_time between '13:00:00' and '18:59:59'

then 'Afternoon'

else 'Night'

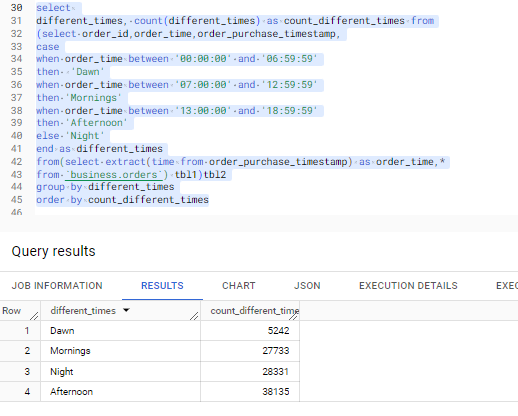
end as different\_times

from(select extract(time from order\_purchase\_timestamp) as order\_time,\*

from `business.orders`) tbl1)tbl2

group by different\_times

Screenshot:



Insights:

Suggestions:

**Q3.** **Evolution of E-commerce orders in the Brazil region:**

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

Query:

select geolocation\_state, order\_month, count(order\_id) as no\_of\_orders from

(select l.geolocation\_state, FORMAT\_DATE('%B', order\_purchase\_timestamp) as order\_month,order\_purchase\_timestamp,o.\*

from `business.orders` o

join `business.customers` c

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

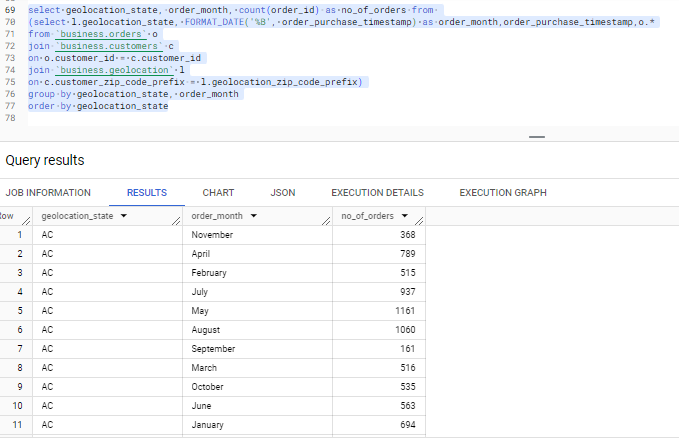
join `business.geolocation` l

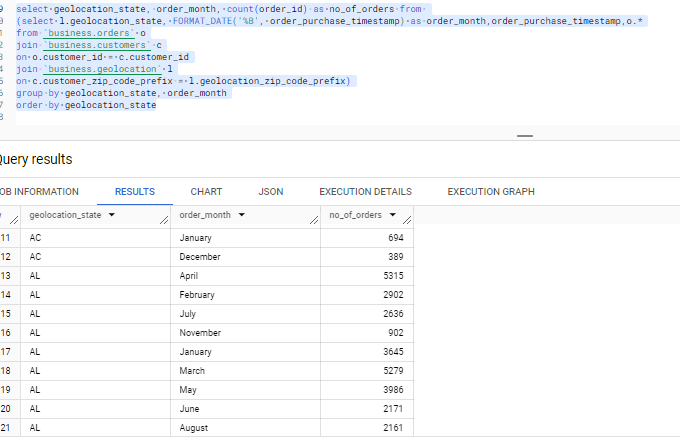
on c.customer\_zip\_code\_prefix = l.geolocation\_zip\_code\_prefix)

group by geolocation\_state, order\_month

order by geolocation\_state

Screenshot:





Insights:

Suggestions: