**Target SQL Business Case**

**1.Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset**

**1.1 Data type of all columns in the "customers" table.**

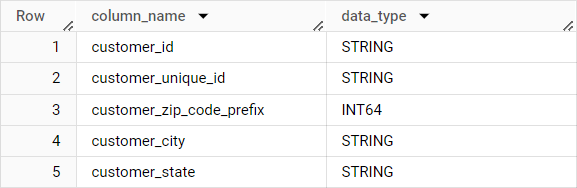
**Query:**

SELECT column\_name, data\_type

FROM sql-project-415807.TARGET. INFORMATION\_SCHEMA.COLUMNS

WHERE table\_name = 'customers';

**Output:**



**Report:**

This report shows the data types of all columns that are present in the ‘customers’ table. Here column\_name represents the name of the columns in the customers table and data\_type represents the datatype of the columns.

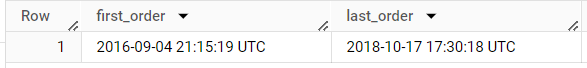
**1.2. Get the time range between which the orders were placed**.

**Query:**

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

max (order\_purchase\_timestamp) as last\_order from TARGET.orders;

**Output:**

****

**Report:**

This report shows the date & time when the first and last order placed by the customers over the period (2016,2017,2018). Here first\_order represents the first order placed by the customer and last\_order represents the last order placed by the customer over the period (2016,2017,2018).

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

**Query:**

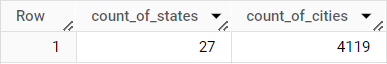
select count (distinct customer\_state) as count\_of\_states,

count (distinct customer\_city) as count\_of\_cities

from TARGET.orders as o join TARGET.customers as c

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

**Output:**



**Report:**

This report shows the number of cities & states of customers who ordered during the given period (2016,2017,2018). Here count\_of\_states represents the count of unique states of the customers who ordered and count\_of\_cities represents the count of unique cities of the customers who ordered.

**2.In-depth Exploration:**

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

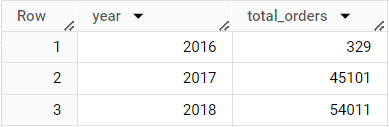
**Answer:** Yes

**Query:**

select extract (year from order\_purchase\_timestamp) as year, count(order\_id) as total\_orders from TARGET.orders

group by year order by year asc;

**Output:**



**Report:**

The above query and sample output represent the number of orders placed in the respective years. Here year represents the year and total\_orders represents the total orders placed in the respective years, by viewing the above data we can see that orders placed in 2017 and 2018 are lot more than the orders placed in 2016 and so we can say that number of orders gets gradually increased compared to their previous years

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

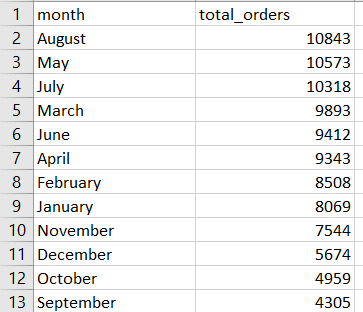
**Answer:** Yes

**Query:**

select format\_date ('%B’, order\_purchase\_timestamp)

as month, count(order\_id) as total\_orders from TARGET.orders group by month order by total\_orders desc;

**Output:**



**Report:**

The above output represents the total number of orders placed in every months of every year (2016,2017,2018). Here month represents the month of the year and total\_orders represents the total number of orders placed in every month. We can see that in August, May, July orders placed by the customers had crossed above 10000, we can say that the total orders placed in the first seven months had an average of 8500 whereas for last five months orders placed had been decreased with an average of 6500

**2.3. 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**

**Answer:** Night

**Query:**

with t1 as

(select FORMAT\_DATE ('%H',order\_purchase\_timestamp) as Hour, count(order\_id) as total\_orders from TARGET.orders

group by Hour order by total\_orders desc)

select case

when Hour in ('1','2','3','4','5','6’) then 'Dawn'

when Hour in ('7','8','9','10','11','12') then 'Mornings'

when Hour in ('13','14','15','16','17','18’) then 'Afternoon'

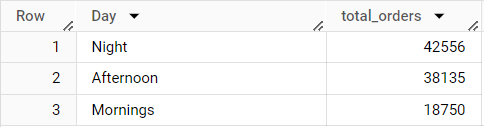
else 'Night'

end as Day, sum(total\_orders) as total\_orders

from t1

group by Day order by total\_orders desc limit 1;

**Output:**



**Report:**

This report shows about total number of orders placed in every day based on different modes (Dawn, Morning, Afternoon, Night). Here Day represents the different modes in the day and total\_orders represents the total number of orders placed in every mode. We can see that the orders placed by the customers are more in Night (19:00-23:59) and so there is no order have been placed in Dawn (00:00-06:00).

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

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

**Query:**

select customer\_state, format\_date ('%B’, order\_purchase\_timestamp) as month, count(order\_id) as Total\_orders

from TARGET.orders as o join TARGET.customers as c on o. customer\_id=c. customer\_id

group by customer\_state, month order by 1 asc,3 desc;

**Sample Output:**



**Actual output in csv:**



**Report:**

This report shows the total number of orders placed by the customers in their respective state on respective months in the given period of time (2016,2017,2018). Here customer\_state represents the state of the customers who ordered, month represents in which customer placed the order and Total\_orders represent the no. of orders placed in each state, in each month by the customers.

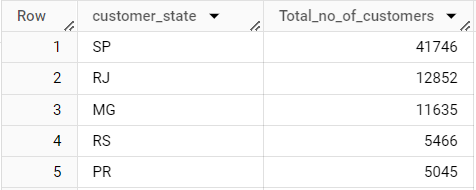
**3.2 How are the customers distributed across all the states?**

**Query:**

select customer\_state, count(customer\_unique\_id) as Total\_no\_of\_customers

from TARGET.customers group by customer\_state order by 2 desc;

**Sample Output:**



**Actual output in csv:**



**Report:**

This report shows the states of all customers who placed the orders over the period (2016,2017,2018). Here customer\_state represents the states of all customers who placed the orders and Total\_no\_of\_customers represents the number of customers who placed the orders in the respective states. We can say state SP had the highest customers which is above 40000 followed by RJ, MG etc.

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

**4.1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).**

**Query:**

with t1 as

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

sum(payment\_value) as total\_payment

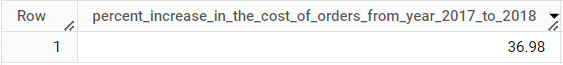
from TARGET.orders as o join TARGET.payments as p on o. order\_id=p. order\_id

where extract (month from order\_purchase\_timestamp) <9 group by year)

select round(((nth\_value(total\_payment,2) over (order by year asc)-nth\_value(total\_payment,1) over (order by year asc))/nth\_value(total\_payment,1) over (order by year asc)) \*100-100,2) as percent\_increase\_in\_the\_cost\_of\_orders\_from\_year\_2017\_to\_2018

from t1 order by percent\_increase\_in\_the\_cost\_of\_orders\_from\_year\_2017\_to\_2018 desc limit 1;

**Output:**



**Report:**

The above query and output display the percentage increase in the cost of orders from year 2017 to 2018 by including months from January to August and we can say that most of the customers had placed their orders in the first six months (January to August) and there is a great increase in the cost of orders comparing from 2017 to2018.

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

**Query:**

select customer\_state, round(sum(payment\_value),2) as Total\_value\_of\_order\_price,

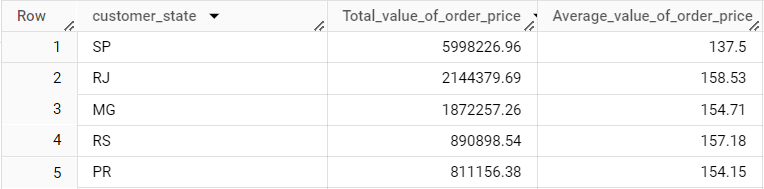
round(avg(payment\_value),2) as Average\_value\_of\_order\_price

from TARGET.orders as o join TARGET.customers as c

 on o. customer\_id=c. customer\_id join TARGET.payments as p on p. order\_id=o. order\_id

group by customer\_state order by 2 desc;

**Sample Output:**



**Actual output in csv:**

****

**Report:**

This report shows the total value of order price and average value of order price of each state over the period (206,2017,2018). Here total value of order price represent the sum of cost of each order which are placed in the each state and the average value of order price represent the average amount spent in the each order in each state . we can see that the state SP has highest total value of order price and the lowest average value of order price.

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

**Query:**

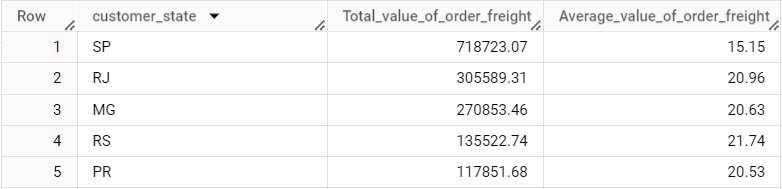
select customer\_state, round(sum(freight\_value),2) as Total\_value\_of\_order\_freight,

round(avg(freight\_value),2) as Average\_value\_of\_order\_freight

from TARGET.orders as o join TARGET.customers as c on o. customer\_id=c. customer\_id

 join TARGET.order\_items as oi on oi. order\_id=o. order\_id group by customer\_state order by 2 desc;

**Sample Output:**



**Actual output in csv:**

****

**Report:**

This report shows the total value of order freight and average value of order freight of each state over the period (206,2017,2018). Here total value of order freight represents the sum of shipping charges of each order which are placed in each state and the average value of order freight represent the average shipping charge for each order in each state. we can see that the state SP has highest total value of order freight value and the lowest average value of order freight value.

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

**5.1 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.**

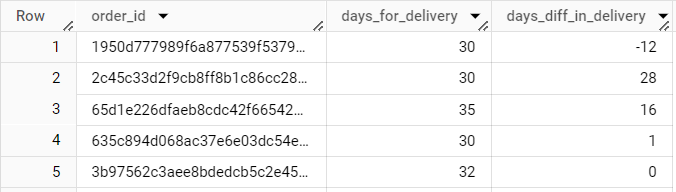
**Query:**

select order\_id, datetime\_diff (order\_delivered\_customer\_date, order\_purchase\_timestamp, day) as days\_for\_delivery,

datetime\_diff (order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day) as days\_diff\_in\_delivery

from TARGET.orders;

**Sample Output:**



**Actual output in csv:**



**Report:**

This report shows the no. of days taken to deliver each order and the difference (in days) between the estimated & actual delivery date of an order. Here in table days\_of\_delivery represents the number of days taken to deliver the product of each order and the days\_diff\_in\_delivery represent the days difference between estimated delivery and actual delivery ,If the value of days\_diff\_in\_delivery is negative it shows that the product is delivered late or if it is positive the product is delivered early and if it is zero the product is delivered on exact time

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

**5.2.1 Top 5 states with the highest average freight value**

**Query:**

select customer\_state, round(avg(freight\_value),2) as highest\_average\_freight\_value

from TARGET. Customers as c join TARGET.orders as o on c. customer\_id=o. customer\_id;

join TARGET.order\_items as oi on oi. order\_id=o. order\_id

group by customer\_state order by 2 desc limit 5;

**Output:**



**Report:**

This report shows the top 5 states with the highest average freight value. Here highest\_average\_freight\_value represents the average freight value in decreasing order and it shows the average freight value of each order in each state and customer\_state represents the top 5 states with the highest average freight value in currency. We see that state RR has the highest average freight value and the state RR and PB has almost same average freight value

**5.2.2 Top 5 states with the lowest average freight value**

**Query:**

select customer\_state, round(avg(freight\_value),2) as lowest\_average\_freight\_value

from TARGET.customers as c join TARGET.orders as o on c. customer\_id=o. customer\_id

join TARGET.order\_items as oi on oi. order\_id=o.order\_id

group by customer\_state order by 2 asc limit 5;

**Output:**



**Report:**

This report shows the top 5 states with the lowest average freight value. Here lowest\_average\_freight\_value represents the average freight value in increasing order and it shows the average freight value of each order in each state and customer\_state represents the top 5 states with the lowest average freight value in currency. We see that state SP has the lowest average freight value.

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

**5.3.1 Top 5 states with the highest average delivery time**

**Query:**

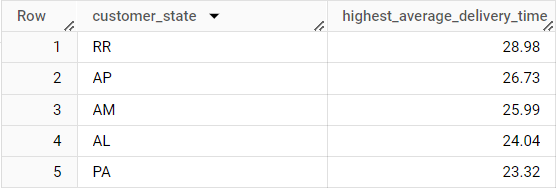
select customer\_state,round(avg(datetime\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)),2)

as highest\_average\_delivery\_time

from TARGET.customers as c join TARGET.orders as o on c.customer\_id=o.customer\_id

group by customer\_state order by 2 desc limit 5;

**Output:**



**Report:**

This report shows the top 5 states with the highest average delivery time Here highest\_average\_delivery\_time represents the average delivery time in decreasing order and it shows the average delivery period of each order in each state and customer\_state represents the top 5 states with the highest average delivery time in days. We see that state RR has the highest average delivery time.

**5.3.2 Top 5 states with the lowest average delivery time**

**Query:**

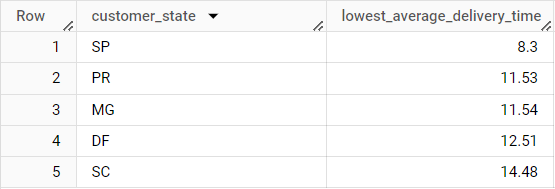
Select customer\_state,round(avg(datetime\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day)),2)

as lowest\_average\_delivery\_time

from TARGET.customers as c join TARGET.orders as o on c.customer\_id=o.customer\_id

group by customer\_state order by 2 asc limit 5;

**Output:**



**Report:**

This report shows the top 5 states with the lowest average delivery time Here lowest\_average\_delivery\_time represents the average delivery time in increasing order and it shows the average delivery period of each order in each state and customer\_state represents the top 5 states with the lowest average delivery time in days . We see that state SP has the lowest average delivery time.

**5.4 Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**

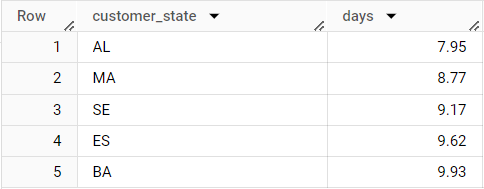
**Query:**

select customer\_state,avg(datetime\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,day))as days

from TARGET.customers as c join TARGET.orders as o on c.customer\_id=o.customer\_id

group by customer\_state order by days asc limit 5;

**Output:**



**Report:**

This report shows the top 5 states where the order delivery is really fast as compared to the estimated date of delivery. Here customer\_state represents the top 5 states with the fastest order delivery time and days represents the average order delivery time of total orders placed from each state in days. We see that state AC has the fastest order delivery time.

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

**Query:**

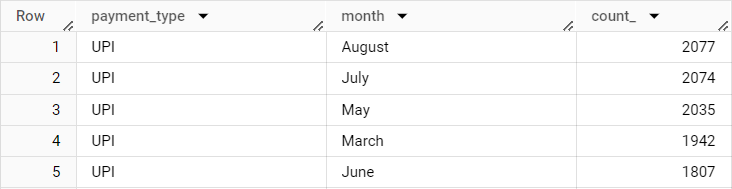
select payment\_type, format\_datetime('%B', order\_purchase\_timestamp) as month,

count(payment\_type) as count\_

from TARGET.orders as o join TARGET.payments as p on o.order\_id=p.order\_id

group by payment\_type, month order by 1 asc ,3 desc;

**Sample Output:**



**Actual output in csv:**

****

**Report:**

This report shows the count the no. of orders placed using different payment methods in each month over the past years (2016,2017,2018). Here payment\_type represents the type of payments done by customers, month represents the month in year and count\_ represents the total count of payment types done in each month.

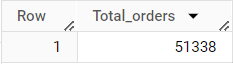
**6.2 Find the no. of orders placed on the basis of the payment installments that have been paid.**

**Query:**

select count(payment\_sequential>=1) as Total\_orders from

TARGET.payments where payment\_installments>0;

**Output:**



**Report:**

This report shows the count the no. of orders placed based on the no. of payment installments where at least one installment has been successfully paid.