BUSINESS CASE STUDY :-SQL MODULE

Context:

Target is a globally renowned brand and a prominent retailer in the United States. 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 particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

**Problem Statement:**

Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

**What does 'good' look like?**

1.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.

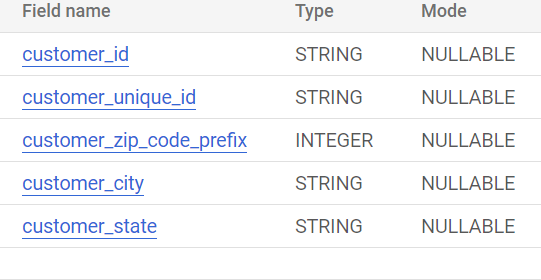
ANSWER:

“customers” table have 5 different attributes or columns and their datatypes:

1.customer\_id- STRING; 2.customer\_unique\_id-STRING;

3. customer\_zip\_code\_prefix-INTEGER;4.customer\_city-STRING;

4.customer\_state-STRING;



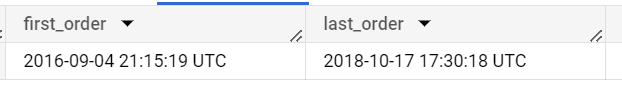
2. Get the time range between which the orders were placed.

ANSWER:

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

      max(order\_purchase\_timestamp) as last\_order

from `projectTarget.orders`



This is the time range of orders from 4th sep, 2016 to 17th oct,2018. First order comes at night 11.15p.m and end order comes in 5.30 a.m.

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

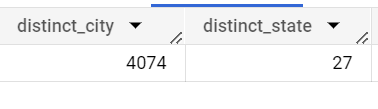
ANSWER:

SELECT count(distinct c.customer\_city) as distinct\_city,

       count(distinct c.customer\_state) as distinct\_state

from `projectTarget.customers` as c inner join `projectTarget.Geolocation` as G

on c.customer\_zip\_code\_prefix=G.geolocation\_zip\_code\_prefix



In this case, customers table has customer\_city and customer\_state columns which can count easily distinct cities and states. But Geolocation table contains both the details of customer and seller cities and states . So in this case we inner join customers table and geolocation table to count the customer details.

SELECT count(distinct customer\_city) as distinct\_city,

count(distinct customer\_state) as distinct\_state

from `projectTarget.customers`er details.

2. **In-depth Exploration:**

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

**ANSWER:**

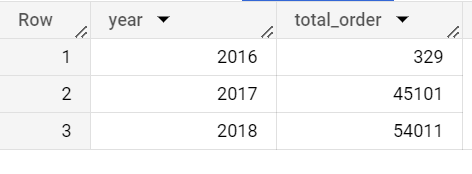
**select extract(year from order\_purchase\_timestamp) as year,**

**count(order\_id) as total\_order**

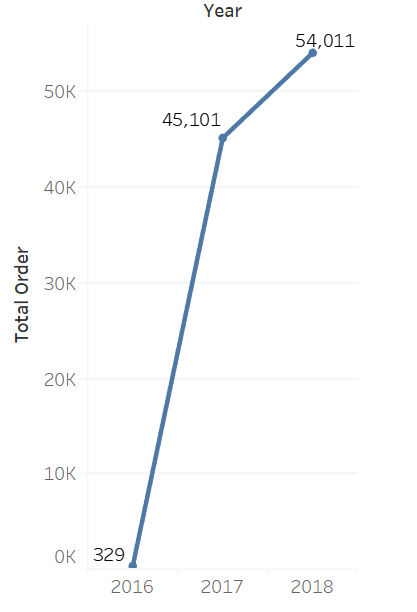
**from `projectTarget.orders`**

**group by year**

**order by year**

****

**In 2016 there is 329 order placed, then 2017 order placed 45101 and in 2018 order placed 54011. With time the order rate is growing. So the statement is TRUE , there is a growing trend in the no. of orders placed over the past years. In tableau , we just visually presents the years and order count details.**

****

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

ANSWER:

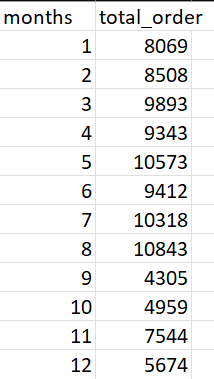
select extract(month from order\_purchase\_timestamp) as months,

count(order\_id) as total\_order

from `projectTarget.orders`

group by months

order by months

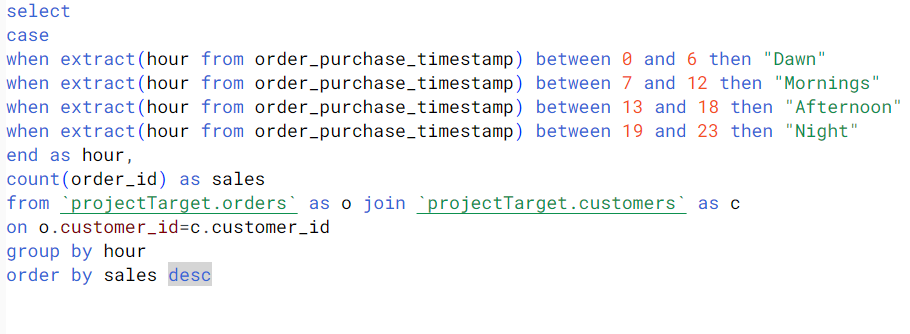
 BUT in this case we can’t surely tell about seasonality properly . data is insufficient to tell about seasonality because it includes after sept 2016 to November 2018. But in brazil sales is higher in July august time but then a certain downfall in sept 2018. Probably there some issue with sales in brazil for target company. But may to august sales is higher , so we can say order is higher in summer times in brazil.

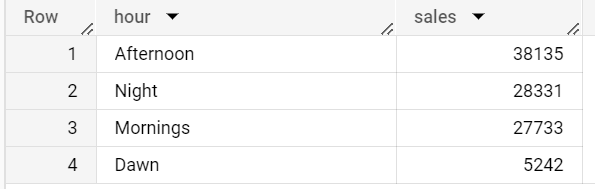


2. 3. During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)

* + 0-6 hrs : Dawn
  + 7-12 hrs : Mornings
  + 13-18 hrs : Afternoon
  + 19-23 hrs : Night

ANSWER:





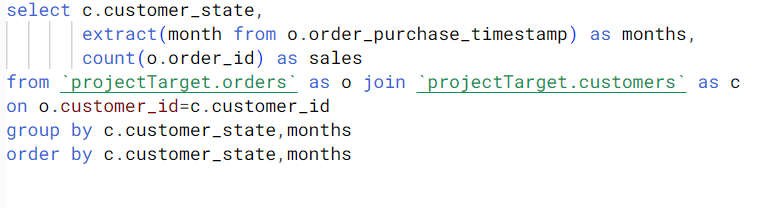
In this problem we can easily assume that the sale in afternoon is highest. So we can say that Brazilian people mainly order the things in daytime. Specifically in afternoon then night then mornings.. after 12 pm they are not very active to order anything.

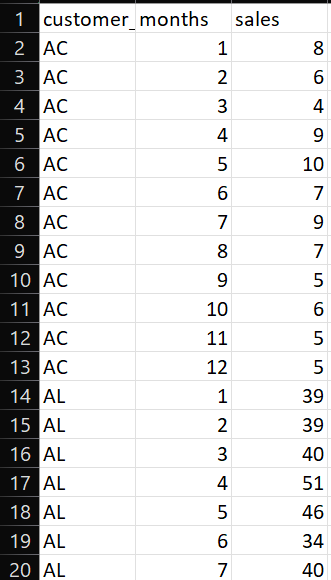
This study gives the TARGET company to provide rightful resources, customer services and availability of products in each stores to provide the customer a fast solutions and fast delivery. TARGET company can save the employment cost and change their marketing strategy to improve their business in brazil.

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

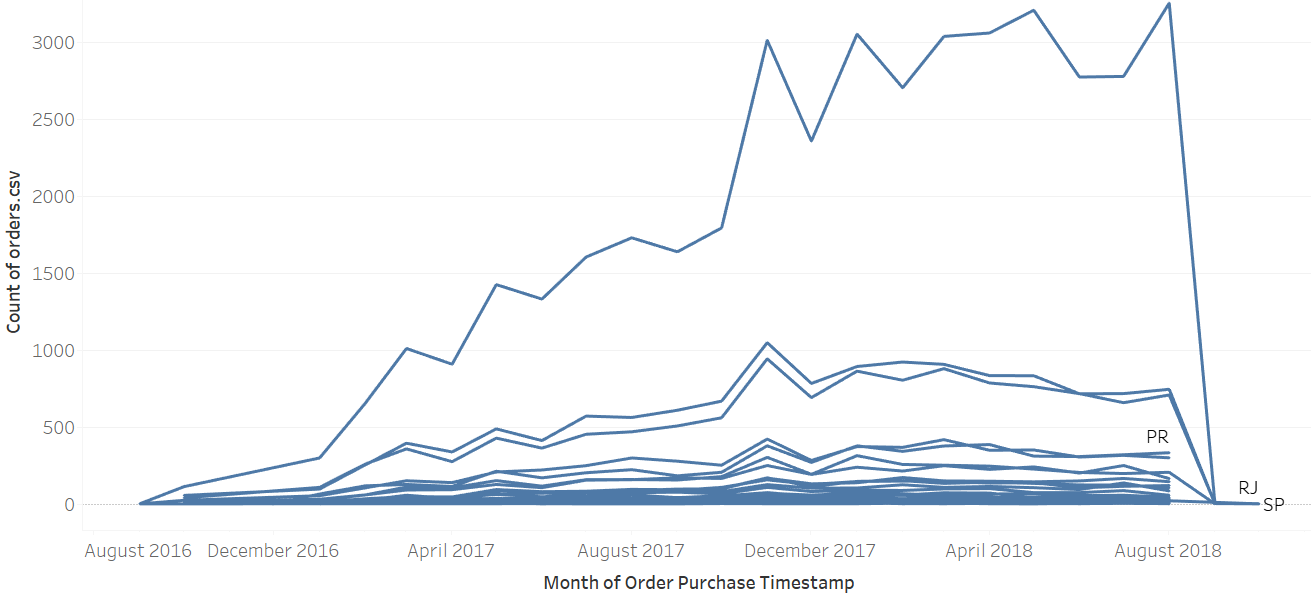
1. Get the month on month no. of orders placed in each state.

ANSWER: in this case we have to join customers and orders table then group by with states and months to count /aggregated the order\_id or sales





In this sql query we can detect month on month orders in Brazil as state wise .



From this analysis we can say that sao paolo state (SP) has the consistently highest order has been given, followed by rio de generio(RJ).

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

ANSWER:

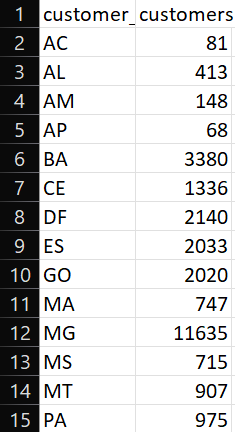
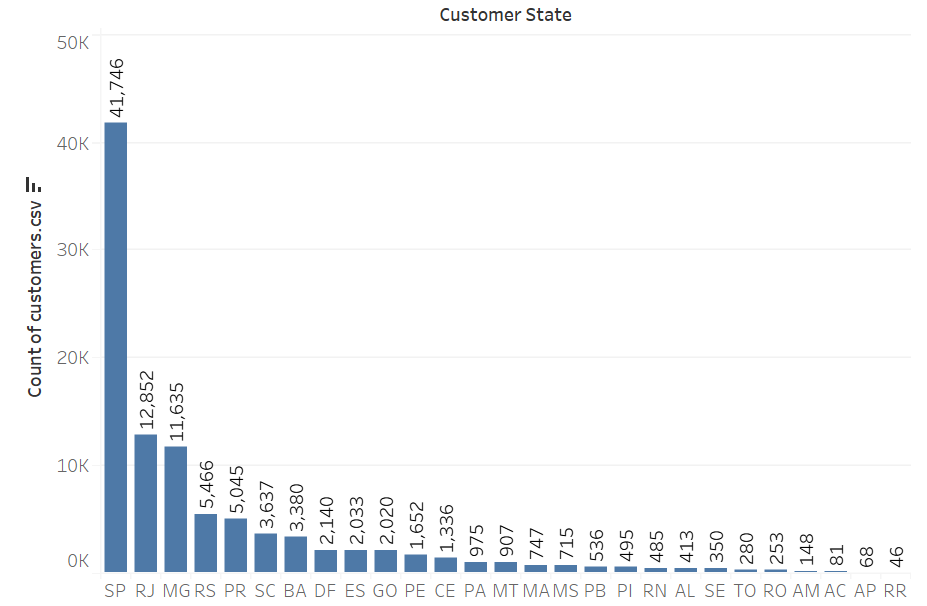
select customer\_state,

       count(customer\_id) as customers

from `projectTarget.customers`

group by customer\_state

order by customer\_state

  
In this we can see state wise customers who can give orders. With tableau we can visualise the state where highest customer is lowest . 

SP state has highest customer and RR state has the lowest customer we can conclude.

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).**

**You can use the "payment\_value" column in the payments table to get the cost of orders.**

**ANSWER :**

SELECT

  EXTRACT(MONTH FROM o.order\_purchase\_timestamp) AS month,

  (

    (

      SUM(CASE WHEN EXTRACT(YEAR FROM o.order\_purchase\_timestamp) = 2018 AND

      EXTRACT(MONTH FROM o.order\_purchase\_timestamp) BETWEEN 1 AND 8 THEN

      p.payment\_value END)

      -

      SUM(CASE WHEN EXTRACT(YEAR FROM o.order\_purchase\_timestamp) = 2017 AND

      EXTRACT(MONTH FROM o.order\_purchase\_timestamp) BETWEEN 1 AND 8 THEN

      p.payment\_value END)

    )

    /

    SUM(CASE WHEN EXTRACT(YEAR FROM o.order\_purchase\_timestamp) = 2017 AND

    EXTRACT(MONTH FROM o.order\_purchase\_timestamp) BETWEEN 1 AND 8 THEN

     p.payment\_value END)

  )\*100 AS percent\_increase

FROM

  `projectTarget.orders` as o

JOIN

  `projectTarget.payments` as p ON o.order\_id = p.order\_id

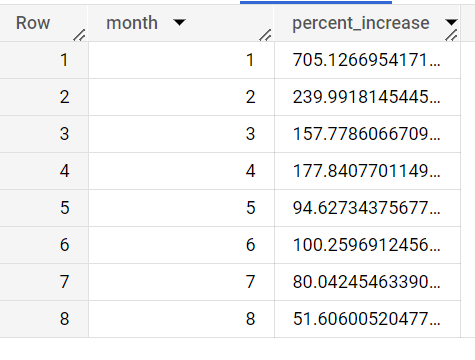
WHERE

  EXTRACT(YEAR FROM o.order\_purchase\_timestamp) IN (2017, 2018) AND

  EXTRACT(MONTH FROM o.order\_purchase\_timestamp) BETWEEN 1 AND 8

GROUP BY month

ORDER BY month;



Percentage increase of cost of order in each months is depicts in here.

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

ANSWER:

select

  c.customer\_state,

  round(avg(i.price), 1) as average\_price,

  round(sum(i.price), 1) as total\_price

from

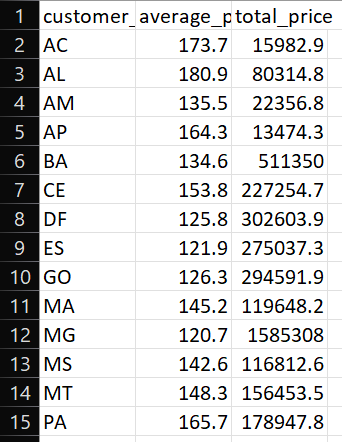
  `projectTarget.orders` as o

join

  `projectTarget.orderitems` as i on o.order\_id = i.order\_id

group by c.customer\_state

order by c.customer\_state;



Here we can conclude average and total value of order price in each state.

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

ANSWER:

select c.customer\_state,

  round(avg(i.freight\_value), 2) as average\_freight\_value,

  round(sum(i.freight\_value), 2) as total\_freight\_value

from `projectTarget.orders` as o

join

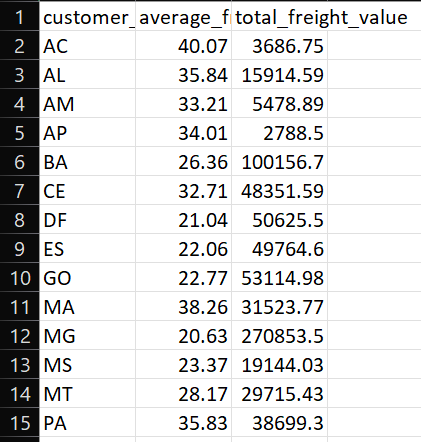
  `projectTarget.orderitems` as i on o.order\_id = i.order\_id

join

  `projectTarget.customers` as c on o.customer\_id = c.customer\_id

group by c.customer\_state

order by c.customer\_state;



**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.  
Do this in a single query.  
  
You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:

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

ANSWER:

Select

order\_id,

  date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, DAY)

  as time\_to\_deliver,

  date\_diff(order\_estimated\_delivery\_date, order\_purchase\_timestamp, DAY)

  as diff\_estimated\_delivery,

  date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, DAY)

  as actual\_delivery\_days

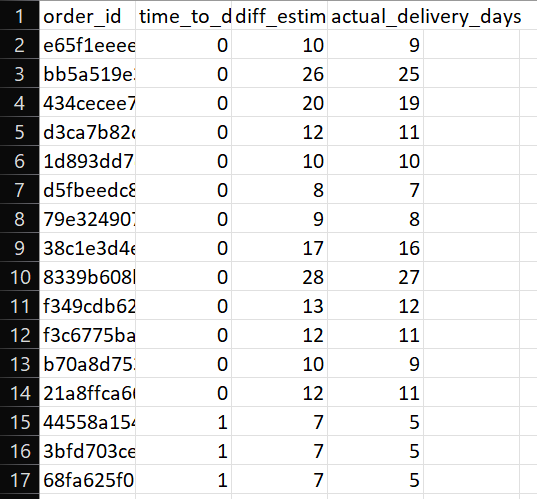
from

  `projectTarget.orders`

where

  date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, DAY) is not null

order by time\_to\_deliver;



To understand order time to deliver and diffence estimated delivery we can easily, we can easily understand actual delivery dates.

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

ANSWER:

Select c.customer\_state,

round(avg(i.freight\_value), 2) as highest\_freight\_value

from

  `projectTarget.orders` as o

join

  `projectTarget.orderitems` as i on o.order\_id = i.order\_id

join

  `projectTarget.customers` as c on o.customer\_id = c.customer\_id

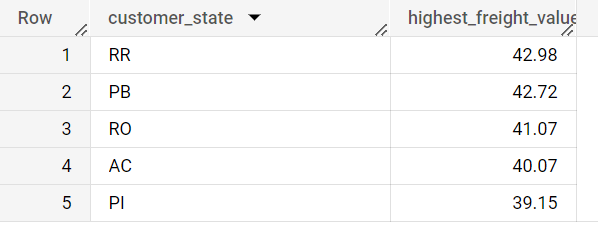
group by

  c.customer\_state

order by

  highest\_freight\_value desc

limit 5



Select

c.customer\_state,

round(avg(i.freight\_value), 2) as lowest\_freight\_value

from

`projectTarget.orders` as o

join

`projectTarget.orderitems` as i on o.order\_id = i.order\_id

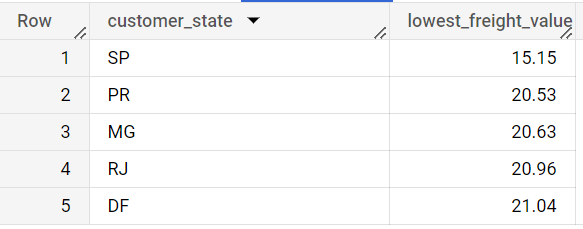
join

`projectTarget.customers` as c on o.customer\_id = c.customer\_id

group by c.customer\_state

order by lowest\_freight\_value asc

limit 5



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

ANSWER:

Select

  c.customer\_state,

  round(avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, DAY)), 2)

  as highest\_avg\_delivery\_time

from

  `projectTarget.orders` as o

join

  `projectTarget.orderitems` as i on o.order\_id = i.order\_id

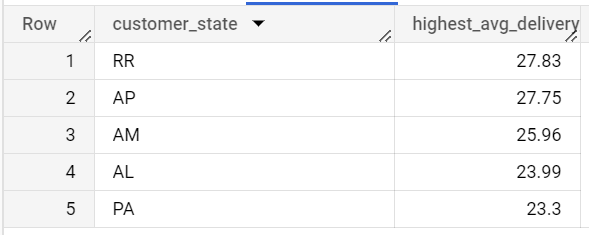
join

  `projectTarget.customers` as c on o.customer\_id = c.customer\_id

group by c.customer\_state

order by highest\_avg\_delivery\_time desc

limit 5;



Select

  c.customer\_state,

  round(avg(date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, DAY)), 2)

  as lowest\_avg\_delivery\_time

from

  `projectTarget.orders` as o

join

  `projectTarget.orderitems` as i on o.order\_id = i.order\_id

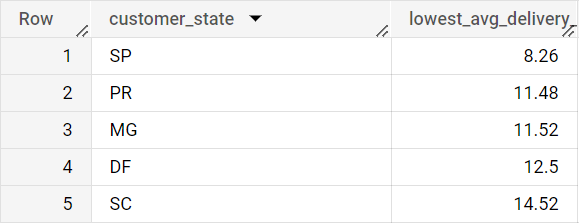
join

  `projectTarget.customers` as c on o.customer\_id = c.customer\_id

group by c.customer\_state

order by lowest\_avg\_delivery\_time asc

limit 5;



In this case we can conclude that SP state has lowest average delivery days and RR states has highest delivery days.

5.4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.  
You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state.

ANSWER:

select

  c.customer\_state,

  round(avg(timestamp\_diff(order\_delivered\_customer\_date,order\_estimated\_delivery\_date, DAY)), 2) as fast\_delivery,

from

  `projectTarget.orders` as o

inner join

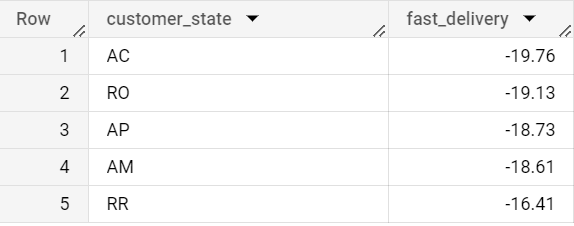
  `projectTarget.customers` as c

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

group by c.customer\_state

ORDER BY fast\_delivery asc

limit 5;



select

c.customer\_state,

round(avg(timestamp\_diff(order\_delivered\_customer\_date,order\_estimated\_delivery\_date, DAY)), 2) as fast\_delivery,

from

`projectTarget.orders` as o

inner join

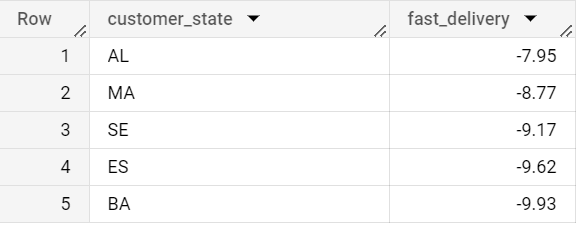
`projectTarget.customers` as c

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

group by c.customer\_state

order by fast\_delivery desc

limit 5;



The top 5 states where the order delivery is really fast as compared to the estimated date of delivery are AC,RO,AP,AM,RR and slow 5 states are AL,MA,SE,ES,BA.

6.Analysis based on the payments:

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

ANSWER:

select

  p.payment\_type,

  extract(MONTH from o.order\_purchase\_timestamp) as month,

  count(distinct o.order\_id) as no\_of\_orders

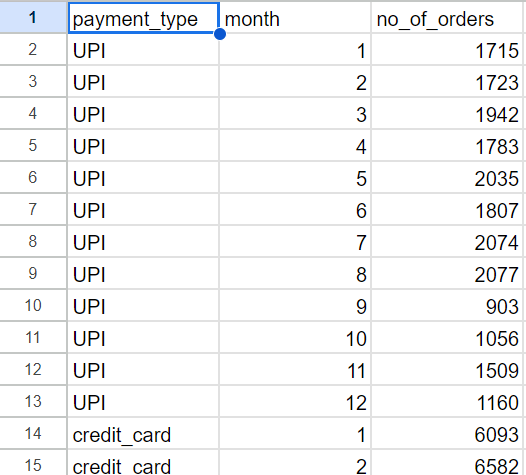
from

  `projectTarget.orders` as o inner join `projectTarget.payments` as p

on o.order\_id = p.order\_id

group by p.payment\_type, month

order by p.payment\_type,month;



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

ANSWER:

select

  p.payment\_installments,

  count(o.order\_id) as no\_of\_order

from

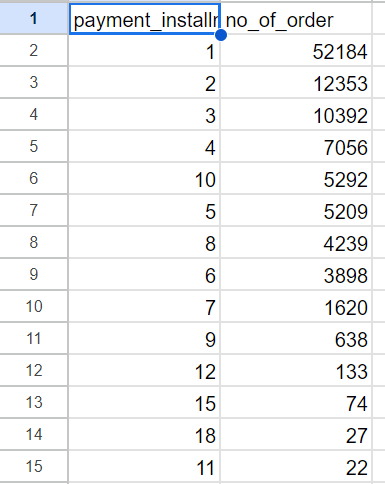
  `projectTarget.orders` as o inner join `projectTarget.payments` as p

on o.order\_id = p.order\_id

where o.order\_status != 'canceled'

group by p.payment\_installments

order by no\_of\_order desc;



These analysis indicates that majority have been choose only one payment installments to order a product. Maily they but it instant full payment of the order by the time of order of the product or delivery time.

So inform the customers about no cost EMI process can increase their order value.

7.Actionable Insights and Recommendations Based on the Analysis.

Actionable Insights:

1. The data indicates that SP state has the highest number of orders ,so there is a possibility to improve in other states through marketing team and promotion.
2. Seasonable sales properly does not detect but there is a smallest growth in orders festive times. But the data time frame is so small to calculate a holistic pattern of sales in festive seasons.

But we can say Brazilian people mainly order the products in afternoon and nights times.so that time we can arrange more employees to gives a better solution for customer.

1. Order counts have been very high in SP(Sao paolo) ,followed by RJ(Rio de generio) state. So improve business is to retain the old customer is a good options and focus on new customer and his choices of products.
2. Faster delivery in rural locations where projected delivery date is very high,

Will create a positive impact in the customer satisfactions and repeating customers.

1. There is a decline in orders during September and October, so that time giving extra discount to the customer will boost the sales .

Recommendations:

1. Improve the speed of delivery by logistic and channel wise shipping process for customer satisfactions and future return of customers.
2. Reduce the cost of delivery and order cost by improving warehouse and season wise stock of products as per customer state and customer city.
3. In some states average freight value is very high , reduce it by providing near warehouse availability.
4. Customer retention can be done through providing extra cashbacks, rewards, loyalty points less delivery cost to improve order count increase. Mostly in those states where order value is very less.
5. In rural areas provide them some information about new technology. Company must improve their technology as per easier user friendly for large section of customers. It will improve customers relation with e-commerce and customer support will be provided with local languages.
6. Choose the area of products which are unpopular in Brazilian markets, make promotion videos with celebrity and provide information to citizen how to use these types of products.
7. Monitoring competitive strategy and marketing with rival companies to improve most customer friendly environment and taking feedback from the customers will be implemented properly .

Conclusions:

in this analysis we figure out the information about the Brazilian market buying patents, production cost, payment preferences, sales trends, audience of customers. By understanding these factors we can take some above actions to improve the business of TARGET company in Brazil.

SP state in Brazil dominating in the e-commerce business, it indicates that other states can improve their audience in e-commerce market in Brazil.

Analyzing customer demographics to improve delivery time and stock of products in each state .

In the time of off season we can give some offers to improve the sales of that time.