Target Business Case

**Problem Statement:**

Target has assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

* **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
* **Data type of all columns in the “Customers” table.**

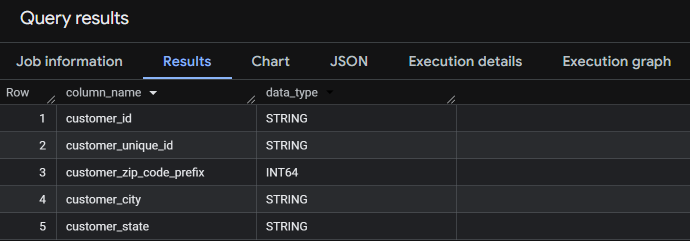
**ANS**:

select column\_name, data\_type from

`TargetP.INFORMATION\_SCHEMA.COLUMNS`

WHERE table\_name = 'customers';

**Screenshot**:



**Insights:**

There are five columns and the data types with their datatypes.

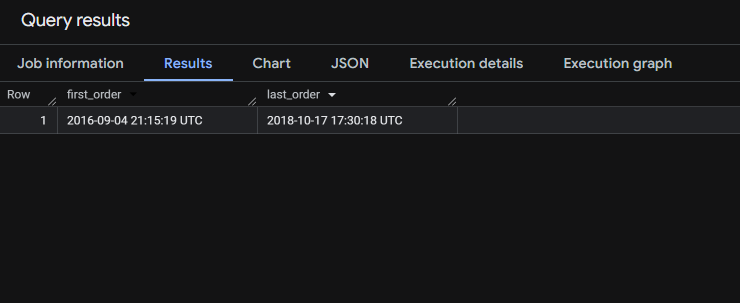
* Get the time range between which the orders were placed.

ANS:

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

`TargetP.orders`;

**Screenshot**:



**Insights:**

**The first order was on ‘2016-09-04' and the last order was on ‘2018-10-17'from the dataset.**

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

ANS:

select count(distinct c.customer\_city) as count\_city, count(distinct c.customer\_state) as count\_state

from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id=o.customer\_id;

**Screenshot:**



**Insights:**

There are 4119 distinct cities and 27 distinct states from where customers have placed orders.

* In-depth Exploration:
* Is there a growing trend in the no. of orders placed over the past years?

**ANS:**

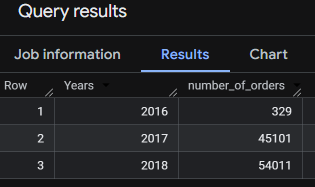
select extract(YEAR from order\_purchase\_timestamp) as Years, count(order\_id) as number\_of\_orders

from

`TargetP.orders` group by extract(YEAR from order\_purchase\_timestamp)

order by extract(YEAR from order\_purchase\_timestamp);

**Screenshot:**



**Insights:**

The trend shows that in the year 2016 the number of orders were at 329, while there was drastic increase in the number of orders for the year 2017. While there is an upward trend in the number of orders in 2018 too, but its not as drastic as when compared to 2016 - 2017.

**Recommendations:**

Figuring out the sales tactics / marketing strategies used in the year 2017 would be useful to keep the number of orders up in the following years to come.

Acquiring new customers, while retaining the old customers would be helpful too with the introduction of discounts, customer purchase bonuses etc.

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

**ANS:**

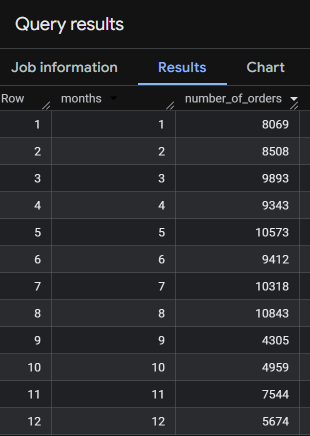
SELECT

extract(MONTH from order\_purchase\_timestamp) AS months,count(order\_id) AS number\_of\_orders

FROM

`TargetP.orders`GROUP BY months ORDER BY months;

**Screenshot:**



**Insights:**

It can be observed that number of orders peaked at months: “May”, “July” and “August” (due to winter holidays in the Brazil). While the least number of orders were at “September” (summer holidays/ sales ends).

The sales peaked on May, July and August due to winter vacations for kids and maybe due to the introduction of winter sales. While

September is “Back to work” season/ end of winters, the major holiday season of December also didn’t see a huge surge in sales. The pre- holiday season of November did see increase in sales.

**Recommendations:**

1.Introduction of more “end of season” sales might help.

2.Introduction of customer loyalty programs.

3. Increase Marketing during major holidays of Christmas and New year.

4. Tailored discounts based on seasons. Eg: An Air conditioners on

discounts during the Summers, while during off-seasons like winters discounts on heaters.

5. Also maintain inventory during the sale period.

* During what time of the day, do the Brazilian customers mostly place their orders?

**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 time\_of\_day,

COUNT(order\_id) AS number\_of\_orders

FROM

`TargetP.orders`

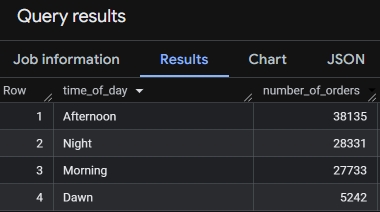
GROUP BY

time\_of\_day

ORDER BY

number\_of\_orders DESC;

**Screenshot:**



**Insights:**

Number of orders peaked in the Afternoon followed by Night and Morning, while the least amount of orders are at Dawn.

**Recommendations:**

1. Introduction of Midnight sales might improve the number of orders placed from 0:00-6:00am (Dawn).

2. Customers are active during the afternoons and night, continue with the pattern of sales to retain customers.

* Evolution of E-commerce orders in the Brazil region:
  + Get the month on month no. Of orders placed in each state.

**ANS:**

select

c.customer\_state,

extract(Month from o.order\_purchase\_timestamp) as months,

extract(Year from o.order\_purchase\_timestamp) as years,

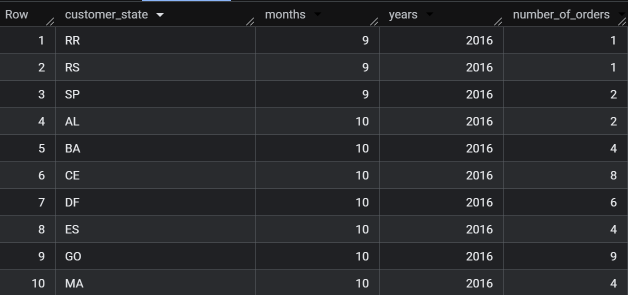
count(\*) as number\_of\_orders

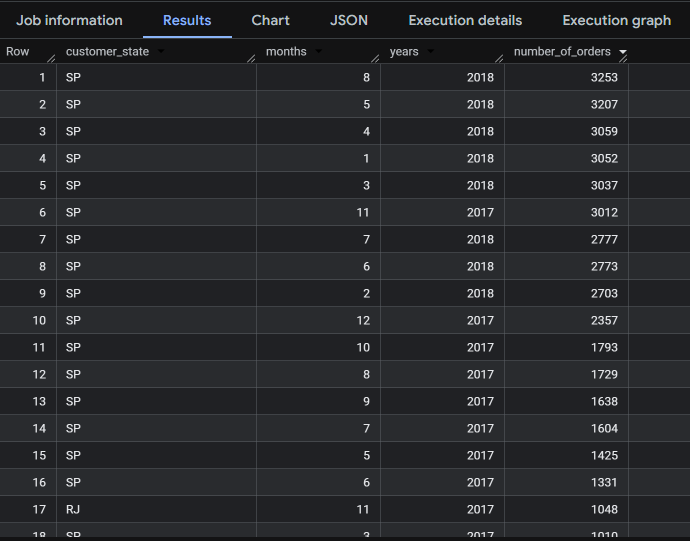
from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id=o.customer\_id

group by c.customer\_state, months,years

order by years, months asc,c.customer\_state;

**Screenshot:**





**Insights:**

1.The month on month number of orders for each states has been recorded.

2.State SP has consistent higher number of orders throughout the year. Also SP has very good sales in JAN 2018 in comparison to other states, with the highest being on August 2018.

3.Order significantly dipped on September motnhs for all states, with a small upward trend in the amount of orders on November

4.Some states have steady order volumes, while others have fluctuating trends.

**Recommendations:**

1. State-Specific Marketing Campaigns

2. Logistics Optimization and inventory planning.

3.Implement/Allocate resources/marketing strategies applied on State SP to other states as well.

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

**ANS:**

select

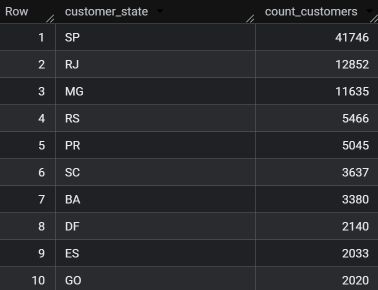
customer\_state , count(customer\_id) as count\_customers

from `TargetP.customers`

group by customer\_state

order by count\_customers desc;

**Screenshot:**



**Insights:**

1. Highest customer base is in the state of SP, followed by RJ and MG.

2.Least amount of customers are in the states of RR,AP,AC.

**Recommendations:**

1. Allocate resources for marketting and advertisements the same which was applied for the state SP.

2. Programs/Sales wise tailored approach to attract more customers in the lower customer base states.

3. Identify spending patterns of customers from the the higher customer bases states so as to incorporate discounts/loyalty programs which would increase the customer base in the lower customer bases states.

* **Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.**
  + **Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).**

**ANS:**

select (total\_values - lag(total\_values) over(order by years))\*100/lag(total\_values) over(order by years) as percentage\_diff

from(

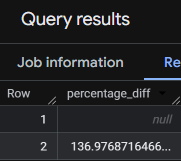
select extract(Year from o.order\_purchase\_timestamp) as years, round(sum(p.payment\_value),3) as total\_values

from `TargetP.payments` p inner join `TargetP.orders` o on p.order\_id=o.order\_id where extract(Year from o.order\_purchase\_timestamp) between 2017 and 2018 and extract(Month from o.order\_purchase\_timestamp)

in (1,2,3,4,5,6,7,8)

group by years) tab1;

**Screenshot:**



**Insights:**

There is a 136.9% increase in the cost of orders from the year 2017 to 2018. indicating strong growth in order revenue. This may reflect an increase in the number of orders or higher-priced products being purchased.

**Recommendations:**

1. A 2.3 times increase in the cost of orders might be good in terms of revenue if it is with respect to the order count,but if it is truly based on just the prices, then may have repercussions.

2. Assess customer feedback for price sensitivity.

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

**ANS:**

select c.customer\_state, round(avg(oi.price),2) as avg\_value,round(sum(oi.price),2) as total\_value

from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id

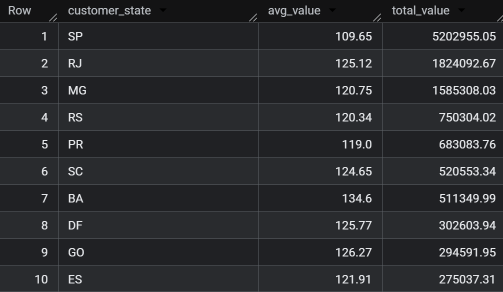
join `TargetP.order\_items` oi on oi.order\_id=o.order\_id

group by c.customer\_state

order by

total\_value DESC;

**Screenshot:**



**Insights:**

1.Highest total value by order price is at Sate SP followed by RJ.

With the least being at RR.

2. While highest average value by order price is at State PB followed by AL. With the least being at SP.

3. SP having the highest total order value has the least average order value indicating that high volume of small value orders are dominant in that region.

**Recommendations:**

1. Segment Marketing by Order Behavior.

2. Promotions based on state’s spending behavior. Also increase customer base and customer retention through loyalty programs.

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

**Ans:**

select c.customer\_state, round(avg(oi.freight\_value),2) as freight\_avg\_value,round(sum(oi.freight\_value),2) as freight\_total\_value

from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id

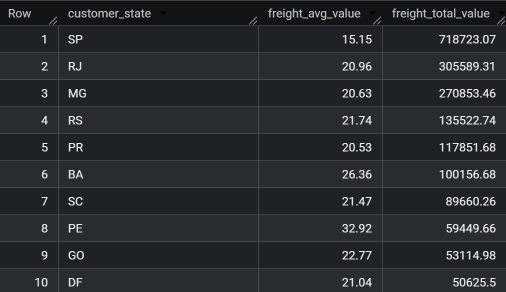
join `TargetP.order\_items` oi on oi.order\_id=o.order\_id

group by c.customer\_state

order by

freight\_total\_value DESC;

**Screenshot:**



**Insights:**

1.Highest total freight value is at SP. Lowest is at RR.

2. Highest average freight value is at RR. Lowest is at SP.

3. Higher average freight value might indicate that the warehouses might be far away from the state in case of RR. While SP has lower average freight value indicateing that hubs/warehouses might be there or nearby.

**Recommendations:**

Increase Customer base where the number of orders and number of customers are less, then accordingly open Warehouses nearby those states.

(Note: Applicable when a threshold of customer base, number of orders have reached) .

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

**ANS:**

select order\_id,

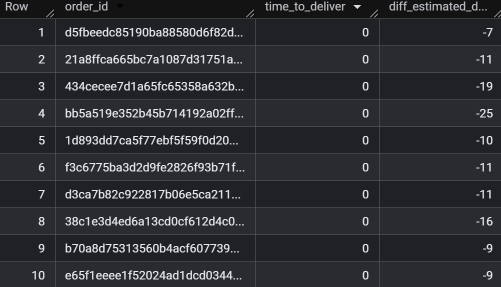
DATETIME\_DIFF(order\_delivered\_customer\_date , order\_purchase\_timestamp,day) as time\_to\_deliver,

DATETIME\_DIFF(order\_delivered\_customer\_date , order\_estimated\_delivery\_date,day) as diff\_estimated\_delivery

from `TargetP.orders` where order\_status ='delivered' and order\_delivered\_customer\_date is not null

order by time\_to\_deliver asc;

**Screenshot:**



**Insights:**

1. There are deliveries delivered either on the same day or within the given time frame.

2.But there are delivers which are not consistent with the delivery speeds. (with many delivered orders not having any info on the table.).

**Recommendations**

1. **Open warehouses or distribution hubs** near high-volume customer regions.

2. Promote **same-day delivery** as a premium service.

3. **Analyze regional delivery performance** to identify areas where delivery is slower.

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

**ANS:**

(select 'Highest' AS freight\_rank,c.customer\_state, round(avg(oi.freight\_value),2) as freight\_avg\_value

from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id

join `TargetP.order\_items` oi on oi.order\_id=o.order\_id

group by c.customer\_state

order by freight\_avg\_value desc limit 5)

union all

(select 'lowest' AS freight\_rank,c.customer\_state, round(avg(oi.freight\_value),2) as freight\_avg\_value

from `TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id

join `TargetP.order\_items` oi on oi.order\_id=o.order\_id

group by c.customer\_state

order by freight\_avg\_value asc limit 5);

**Screenshot:**



**Insights:**

Obtained the top 5 and bottom 5 by average freight value.

Lowest average freight value is at SP, while highest is at

RR.

**Recommendations:**

1.Increase Customer base where the number of orders and number of customers are less, then accordingly open Warehouses/Hubs nearby those states.

2.This not only increases delivery speeds but also reduces cost of freights (transportation).

3.Inventory Management and prior planning of stock Acquistion to meet demands.

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

**ANS:**

(select "Highest" as avg\_delivery\_rank,

c.customer\_state ,round(avg(DATETIME\_DIFF(order\_delivered\_customer\_date , order\_purchase\_timestamp,day)),2) as avg\_delivery\_time

from

`TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id where o.order\_status ='delivered'

group by c.customer\_state

order by avg\_delivery\_time desc limit 5)

union all

(select "Lowest" as avg\_delivery\_rank,

c.customer\_state ,round(avg(DATETIME\_DIFF(order\_delivered\_customer\_date , order\_purchase\_timestamp,day)),2) as avg\_delivery\_time

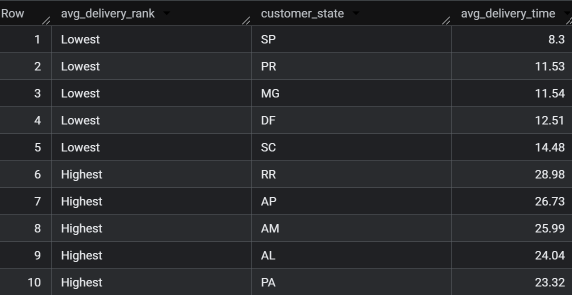
from

`TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id where o.order\_status ='delivered'

group by c.customer\_state

order by avg\_delivery\_time asc limit 5);

**Screenshot:**



**Insights:**

Obtained the top 5 and bottom 5 by average delivery time.

Least average delivery time is at SP, while highest is at RR.

Higher average delivery may indicate that warehouses/Hubs may be far away from the state. While the lesser ones may indicate that warehouses/Hubs are nearby.

**Recommendations:**

1. Openwarehouses **or** distribution **hubs** near high volume customer regions.

2. Promote **same-day delivery** as a premium service.

3. **Analyze regional delivery performance** to identify areas where delivery is slower.

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

**ANS:**

select

c.customer\_state,

round(avg(DATETIME\_DIFF(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,day)),2) as avg\_diff\_dates

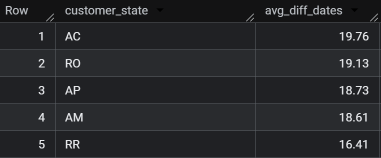
from

`TargetP.customers` c join `TargetP.orders` o on c.customer\_id = o.customer\_id where o.order\_status ='delivered'

group by c.customer\_state

order by avg\_diff\_dates desc limit 5;

**Screenshot:**



**Insights:**

1.Above shows the top 5 states where the deliver is fast as compared to the estimated delivery date.

2.RR state is doing good in terms of delivery time vs estimated time (but this could also mean the number of orders is less, which may not indicate that the delivery system is highly efficient ).

**Recommendations:**

1.Promote same day delivery as a premium option to customers.

2.Priority based shipping.

* **Analysis based on the payments:**
  + **Find the month on month no. of orders placed using different payment types.**

**ANS:**

select

p.payment\_type, extract(month from o.order\_purchase\_timestamp) as months,

extract(year from o.order\_purchase\_timestamp) as year,count(distinct o.order\_id) as number\_of\_orders

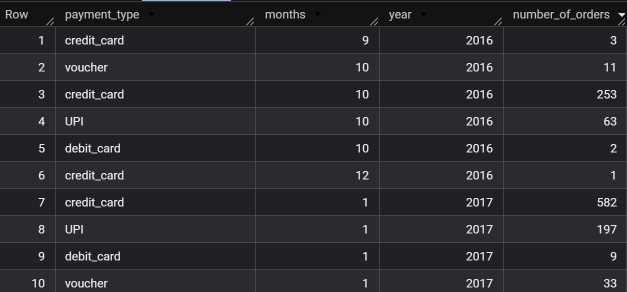
from

`TargetP.orders` o join `TargetP.payments` p on p.order\_id=o.order\_id

group by p.payment\_type,months,year

order by year, months asc;

**Screenshot:**



**Insights:**

1. Debit Cards & Vouchers transactions are less as compared to other payment methods.

2.Credit transactions are dominant.

3. Few records are not defined.

**Recommendations:**

1. Since UPI and credit cards are popular, offer cashback or discounts for debit cards or voucher use to encourage a wider spread of usage.

2. Track Seasonal Payment Trends.

3. Partner up with credit, debit card companies to offer the best value for products and discounts.

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

**ANS: (Without filtering status as delivered).**

SELECT

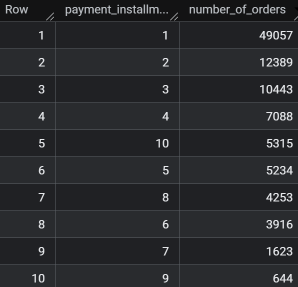
payment\_installments ,count(distinct order\_id) as number\_of\_orders

from `TargetP.payments`

where payment\_value >0

group by payment\_installments order by number\_of\_orders desc;

**Screenshot:**



**ANS: (with filtering status as delivered)**

select

p.payment\_installments ,count(distinct o.order\_id) as number\_of\_orders

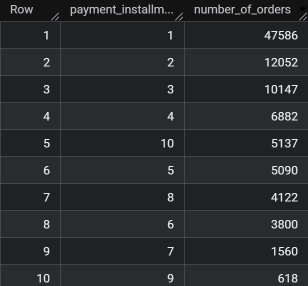
from

`TargetP.orders` o join `TargetP.payments` p on p.order\_id=o.order\_id where o.order\_status ='delivered' and p.payment\_value>0

group by p.payment\_installments

order by number\_of\_orders desc;

**Screenshot:**



**Insights:**

1. Installments Between 2–5 Are Popular.

2. Majority of Orders Are Paid in Full.

3. Very Few Long-Term Installments.

**Recommendations:**

1. Promote Short-Term EMI Plans.

2.Include EMI offers. (cashbacks etc).

3.Users paying in full have higher spending power, leverage premium options/products.