**## 1. In Depth Exploration**

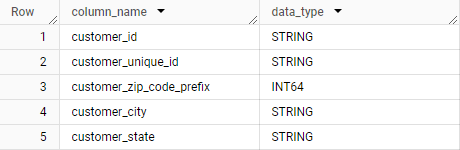
**Q1. Data type of all columns in the “customers” table.**

select

column\_name, data\_type

from `Target`.INFORMATION\_SCHEMA.COLUMNS

where table\_name = 'customers';



**Q2. Get the time range between which the orders were placed.**

select

min(order\_purchase\_timestamp) as First\_order,

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

from `Target.orders`;



**Q3. Count the Cities & States of customers who ordered during the given period.**

select

count(distinct c.customer\_city) as City\_Count,

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

from `Target.customers` c

join `Target.orders` o

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



**## 2. In-depth Exploration**

**Q1. Is there a growing trend in the no. of orders placed over the past years?**

with cte1 as

(select

distinct (year\_month),

count(order\_id) as order\_count

from (

select

\*,

format\_date('%Y-%m',order\_purchase\_timestamp) as year\_month

from `Target.orders`

)

group by 1

order by 1)

select

year\_month,

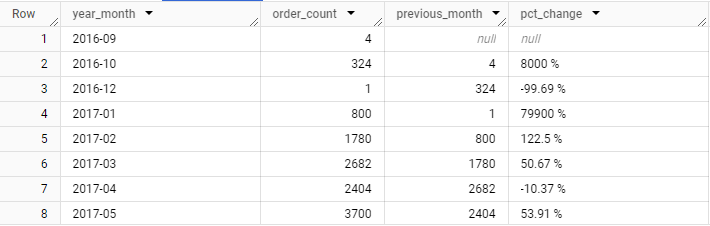
order\_count,

lag(order\_count) over(order by year\_month) as previous\_month,

concat(round(((order\_count - lag(order\_count) over(order by year\_month)) / lag(order\_count) over(order by year\_month))\*100,2)," ","%") as pct\_change

from cte1

order by 1;



**Q2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?**

select

distinct (year\_month),

count(order\_id) as order\_count

from (

select

\*,

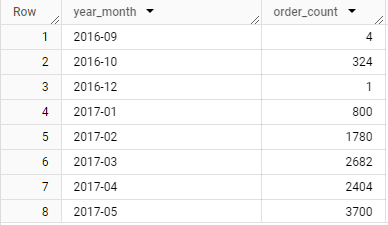
format\_date('%Y-%m',order\_purchase\_timestamp) as year\_month

from `Target.orders`

)

group by 1

order by 1;



**Q3. 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**

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 'Mornings'

when extract(hour from order\_purchase\_timestamp) between 13 and 18 then 'Afternoon'

else 'Night'

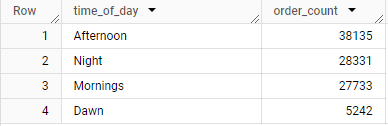
end as time\_of\_day,

count(order\_id) as order\_count

from `Target.orders`

group by 1

order by 2 desc;



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

**Q1. Get the month on month no. of orders placed in each state.**

select

c.customer\_state,

format\_date('%Y-%m',o.order\_purchase\_timestamp) as year\_month,

format\_date('%B', o.order\_purchase\_timestamp) AS month\_name,

count(o.order\_id) as order\_count

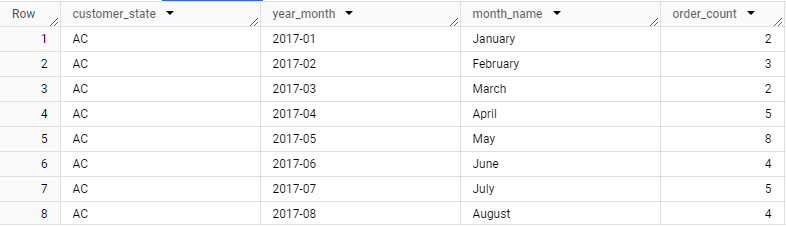
from `Target.customers` c

inner join `Target.orders` o

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

group by 1,2,3

order by 1,2



**Q2. How are the customers distributed across all the states?**

select

c.customer\_state as State\_Name,

count(distinct o.order\_id) as order\_count

from `Target.customers` c

inner join `Target.orders` o

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

group by 1

order by 2 desc;



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

**Q1. Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).**

with cte1 as (

select

format\_date('%Y',o.order\_purchase\_timestamp) as year,

format\_date('%m',o.order\_purchase\_timestamp) as month,

round(sum(p.payment\_value),2) as Total\_Payment

from `Target.orders` o

left join `Target.payments` p

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

group by 1,2

order by 1 asc,2 asc

),

cte2 as (

select

month,

sum(case when year = '2017' then Total\_Payment else 0 end) as Total\_Payment\_2017,

sum(case when year = '2018' then Total\_Payment else 0 end) as Total\_Payment\_2018,

from cte1

where month between '01' and '08'

group by 1

order by 1

)

select

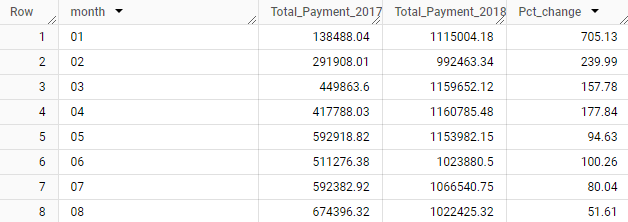
month,

Total\_Payment\_2017,

Total\_Payment\_2018,

round(((Total\_Payment\_2018 - Total\_Payment\_2017)/(Total\_Payment\_2017)) \* 100,2) as Pct\_change

from cte2;



**Q2. Calculate the Total & Average value of order price for each state.**

select

c.customer\_state,

round(sum(p.payment\_value),2) as Total\_Payment,

round(avg(p.payment\_value),2) as Avg\_Payment

from `Target.customers` c

left join `Target.orders` o

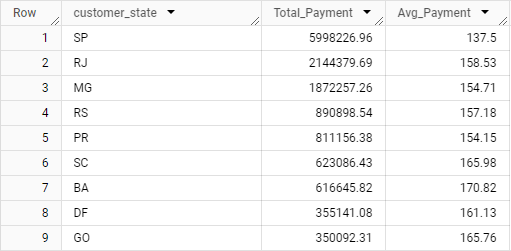
on c.customer\_id = o.customer\_id

left join `Target.payments` p

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

group by 1

order by 1;



**Q3. Calculate the Total & Average value of order freight for each state.**

select

c.customer\_state,

round(sum(o2.freight\_value),2) as Total\_Freight,

round(avg(o2.freight\_value),2) as Avg\_Freight

from `Target.customers` c

left join `Target.orders` o

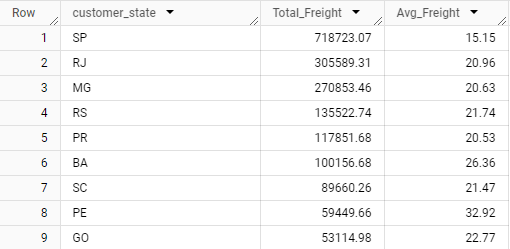
on c.customer\_id = o.customer\_id

left join `Target.order\_items` o2

on o.order\_id = o2.order\_id

group by 1

order by 1;



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

**Q1. 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.**

with cte as (

select

order\_purchase\_timestamp as Order\_date,

order\_delivered\_customer\_date as Delivery\_date,

order\_estimated\_delivery\_date as Estimated\_delivery\_date,

date\_diff(order\_delivered\_customer\_date, order\_purchase\_timestamp, Day) as time\_to\_deliver,

date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, Day) as diff\_estimated\_delivery

from `Target.orders` o

order by 4 asc, 5 asc

)

select

Order\_date,

Delivery\_date,

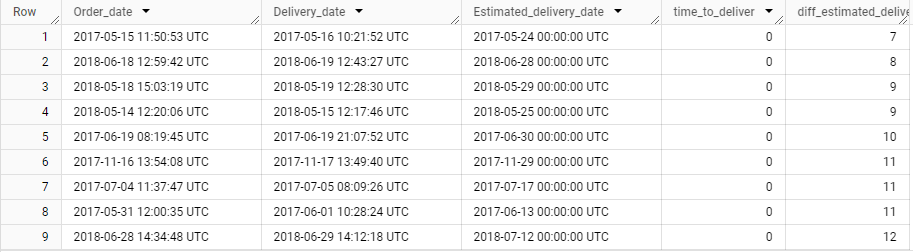
Estimated\_delivery\_date,

time\_to\_deliver,

diff\_estimated\_delivery

from cte

where time\_to\_deliver is not null



**Q2. Find out the top 5 states with the highest & lowest average freight value.**

with cte1 as (

select

c.customer\_state,

round(avg(o2.freight\_value),2) as Avg\_Freight,

dense\_rank() over(order by round(avg(o2.freight\_value),2) desc) as rnk

from `Target.customers` c

left join `Target.orders` o

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

left join `Target.order\_items` o2

on o.order\_id = o2.order\_id

group by 1

),

top5 as (

select

customer\_state,

Avg\_Freight,

rnk

from cte1

where rnk <= 5

),

bottom5 as (

select

customer\_state,

Avg\_Freight,

rnk

from cte1

order by rnk desc

limit 5

)

select

'Top 5' AS Category,

customer\_state,

Avg\_Freight

from top5

union all

select

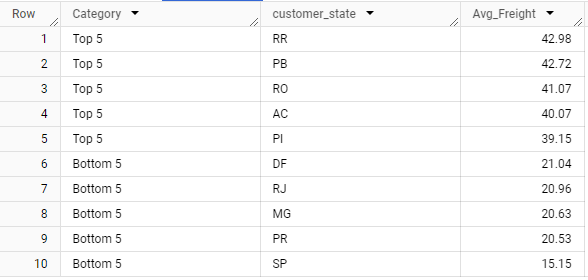
'Bottom 5' AS Category,

customer\_state,

Avg\_Freight

from bottom5

order by Category desc, Avg\_Freight desc;



**Q3. Find out the top 5 states with the highest & lowest average delivery time.**

with cte as (

select

c.customer\_state,

avg(date\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, Day)) as time\_to\_delivery

from `Target.orders` o

right join `Target.customers` c

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

where o.order\_delivered\_customer\_date is not null

group by 1

),

ranked\_data as (

select

customer\_state as state,

time\_to\_delivery as avg\_delivery\_time,

dense\_rank() over(order by time\_to\_delivery asc) as rnk\_asc,

dense\_rank() over(order by time\_to\_delivery desc) as rnk\_desc

from cte

)

select

'Top 5' as category,

state,

avg\_delivery\_time

from ranked\_data

where rnk\_asc <= 5

union all

select

'Bottom 5' as category,

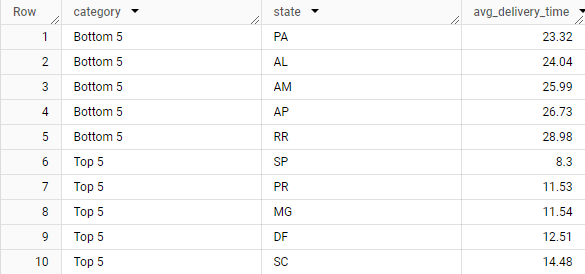
state,

avg\_delivery\_time

from ranked\_data

where rnk\_desc <= 5

order by category, avg\_delivery\_time;



**Q4. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery.**

select

c.customer\_state,

round(avg(datetime\_diff(o.order\_estimated\_delivery\_date, o.order\_delivered\_customer\_date, Day)),2) as diff\_estimated\_delivery

from `Target.customers` c

left join `Target.orders` o

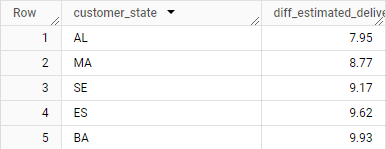
on c.customer\_id = o.customer\_id

where o.order\_delivered\_customer\_date is not null and order\_estimated\_delivery\_date is not null

group by 1

order by 2 asc

limit 5



**## 6. Analysis based on the payments:**

**Q1. Find the month on month no. of orders placed using different payment types.**

select

format\_date('%B-%Y',o.order\_purchase\_timestamp) as month\_year,

p.payment\_type,

count(o.order\_id) as order\_count

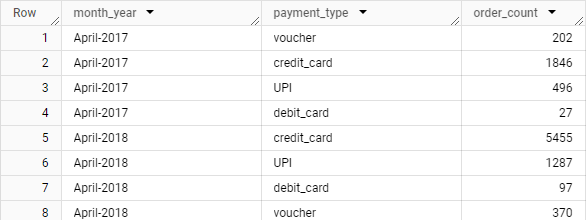
from `Target.orders` o

left join `Target.payments` p

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

group by 1,2

order by 1 asc



**Q2. Find the no. of orders placed on the basis of the payment installments that have been paid.**

select

distinct(p.payment\_installments),

count(o.order\_id) as order\_count

from `Target.orders` o

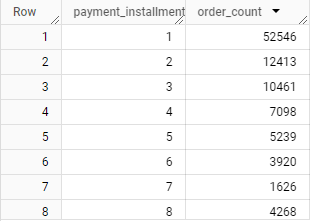
left join `Target.payments` p

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

where p.payment\_installments is not null and p.payment\_installments != 0

group by 1

order by 1 asc



**Insights:**

1. **Seasonal Payment Trends:**
   * Certain months show a higher usage of specific payment methods, potentially linked to promotions or customer behavior.
2. **Installment Usage:**
   * Installments are preferred during large purchases, suggesting customer sensitivity to upfront costs.
3. **Time Range of Orders:**
   * The company's operations span a wide time frame, providing a substantial dataset for analyzing seasonal and yearly trends.
4. **Geographical Spread:**
   * The orders come from diverse cities and states, indicating a good reach across the country.
5. **Delivery Efficiency:**
   * Certain states have significantly faster delivery times compared to the estimated delivery date.
6. **Challenging States:**
   * Some states have consistently high delivery times, possibly due to logistical challenges or remote locations.
7. **Top States for Fast Delivery:**
   * Some states have consistently high delivery times, possibly due to logistical challenges or remote locations.
8. **Freight Costs:**
   * Freight costs vary significantly across states, with some states showing much higher average freight charges.

**Recommendations:**

1. **Payment Campaigns:** 
   * Launch targeted campaigns during peak months, offering incentives for specific payment types.
2. **Upsell Opportunities:** 
   * Promote high-value products with attractive installment options to increase average order value.
3. **Expand Marketing Campaigns:** 
   * Focus marketing efforts during peak ordering times (e.g., afternoon and evening) to boost sales.
4. **Target Underrepresented States:** 
   * Identify and target states with low customer activity to improve market penetration.
5. **Improve Logistics:**
   * Focus on states with slower delivery times by optimizing routes, increasing warehouse coverage, or collaborating with local courier services.
6. **Reward Fast Deliveries:**
   * Promote fast delivery states as examples of company reliability in marketing campaigns.
7. **Leverage Delivery Success:** 
   * Highlight states with exceptional delivery performance as part of customer testimonials and case studies.
8. **Optimize Freight Costs:** 
   * Negotiate better shipping rates with logistics partners for states with high freight costs.