**Business Case: Target**

**Ques 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.**

**Ans.**

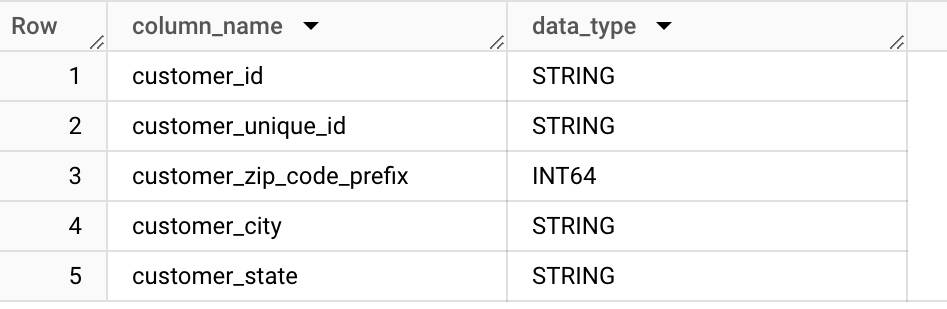
SELECT

column\_name,

data\_type

FROM target-401004.target.INFORMATION\_SCHEMA.COLUMNS

WHERE table\_name = 'customers'



* By looking at the output of the query, it is clear that customer\_id , customer\_unique\_id, customer\_city, customer\_state is of string data type ,where in customer\_zip\_code\_prefix is integer data type.

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

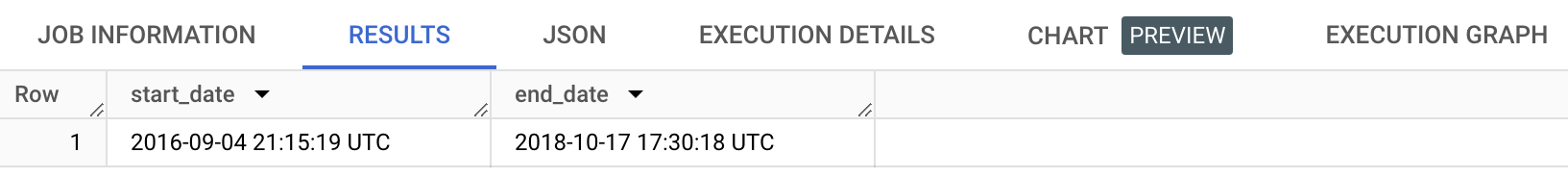
**Ans.**

SELECT

MIN (order\_purchase\_timestamp) AS start\_date,

MAX (order\_purchase\_timestamp) AS end\_date

FROM `target.orders`



* By looking at the output of the query, it is clear that orders placing was started from “September 4 ,2016” and ended on “August 17,2018”.

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

**Ans**.

select

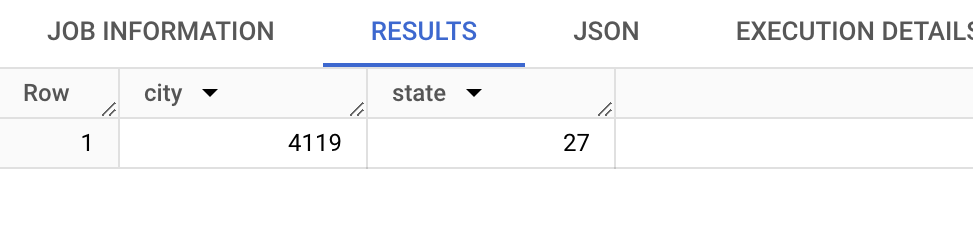
count(distinct customer\_city) as city,

count(distinct customer\_state) as state

from `target.orders` o

inner join `target.customers` c

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



* The output of the query provides us the different cities and different states of different customers who have ordered the products.

**Ques.2 In-depth Exploration:**

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

Ans.

select

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

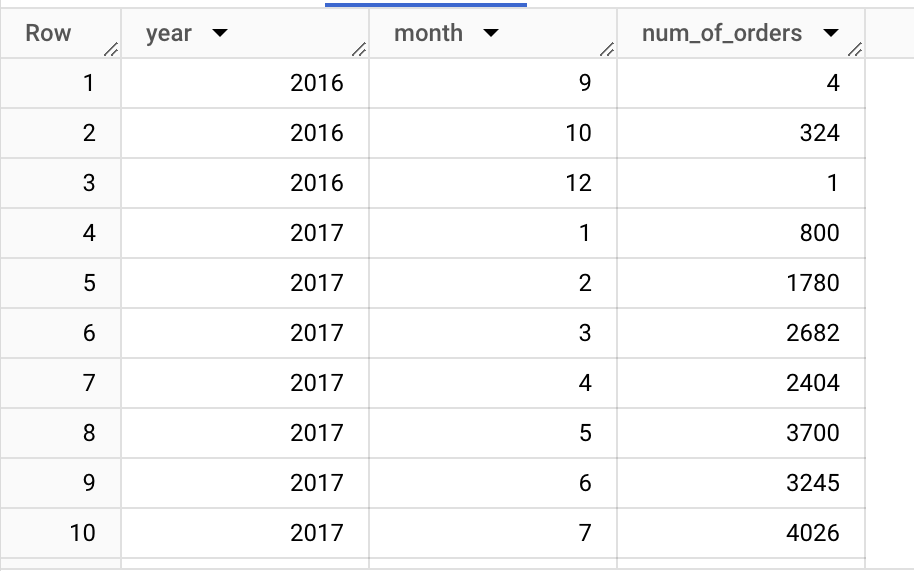
extract(month from order\_purchase\_timestamp ) as month,

count(order\_id) as num\_of\_orders

from `target.orders`

group by 1,2

order by 1,2



* The output shows that orders per month and per year are growing gradually.So yes, we can say that there a growing trend in the no. of orders placed over the past years .

**2.** **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 month,

count(order\_id) as num\_of\_orders

from `target.orders`

group by 1

order by 1



* There is indeed some type of monthly seasonality in the number of orders being placed.
* There seems to be increase in the number of orders during certain months, followed by a decrease in other. Understanding these patterns can help businesses
* Months 5 (May) and 8 (August) have the highest number of orders (10573 and 10843, respectively)

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

**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 Mornings'

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

when extract(hour from order\_purchase\_timestamp ) between 19 and 24 then 'Night'

else 'unknown'

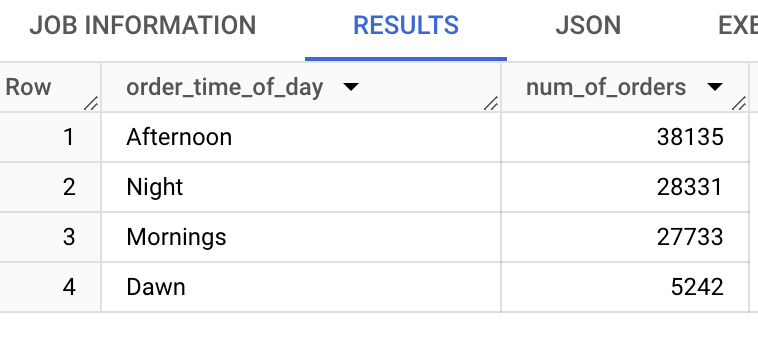
end as order\_time\_of\_day,

count(order\_id) as num\_of\_orders

from `target.orders`

group by 1

order by 2 desc



* During ‘Afternoon’, the Brazilian customers mostly place their orders. Businesses can use this information to time marketing campaigns or promotions during peak order times.

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

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

**Ans.**

select

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

c.customer\_state,

count(order\_id) as num\_of\_orders

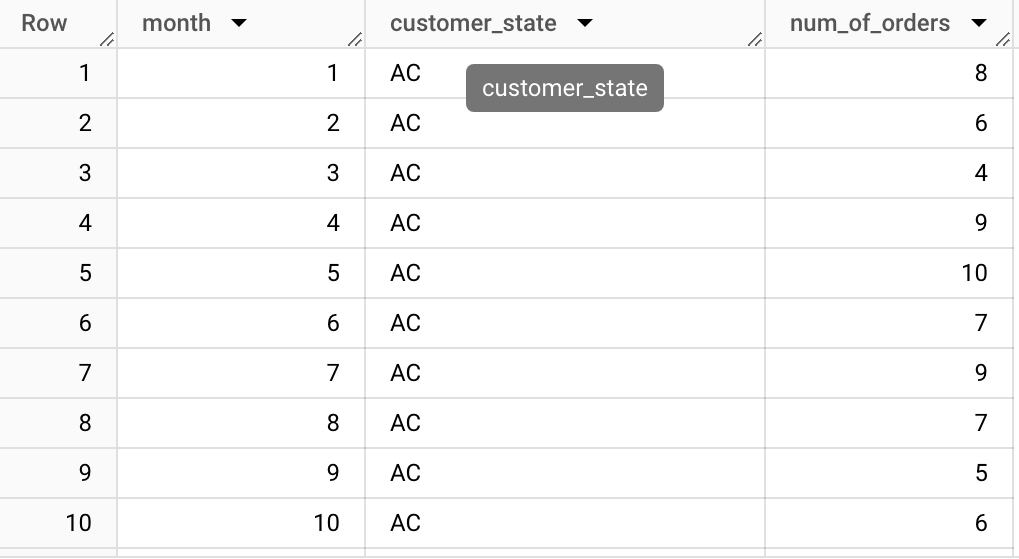
from `target.orders` o

inner join `target.customers` c

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

group by 1,2

order by 2,1



* The output shows that SP, RJ, MG have the highest numbers of orders month on month, and AP, RR, MM have the lowest numbers of orders month on month.
* The high numbers of orders month on month indicate strong and stable demand.

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

**ANS.** select

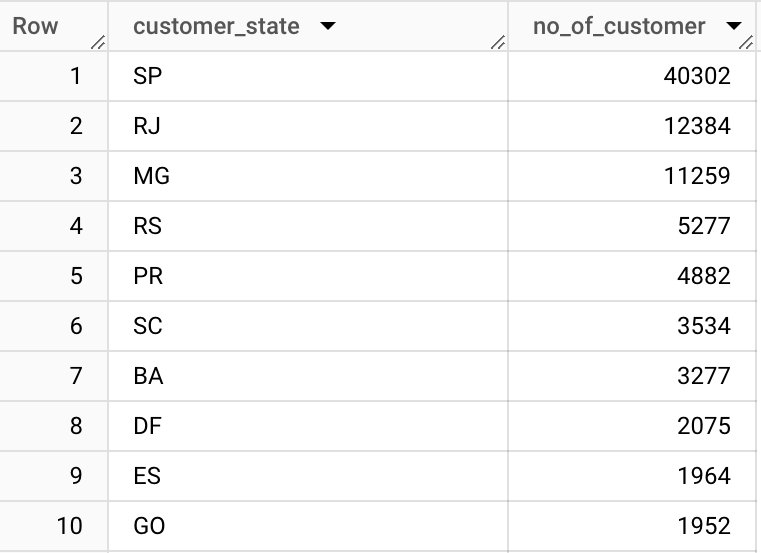
customer\_state,

count(distinct customer\_unique\_id) as no\_of\_customer

from `target.customers`

group by 1

order by 2 desc



* The output provides that states SP, RJ, MG have the most number of customers. This can be a positive for a business.
* States RR, AP, AC have the lowest number of customers, a negative from the company point of view. These states require a more targeted or region-specific marketing and customer service approach.

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

**1. Get the % increase in the cost of orders from 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.**

**ANS;**

with final as(

select format\_date('%Y',

order\_purchase\_timestamp)as date,

sum(p.payment\_value) as cost\_of\_orders

from `target.payments` p

inner join `target.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)between 1 and 8

group by 1

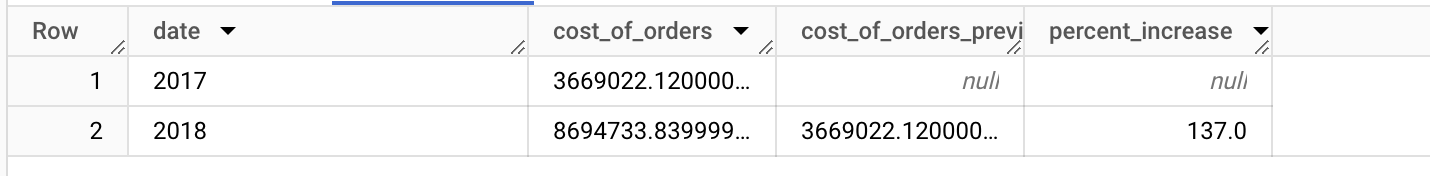
order by 1)

select \* ,lag(cost\_of\_orders) over(order by date) as cost\_of\_orders\_previous\_month,

round(100\*(cost\_of\_orders - lag(cost\_of\_orders)over(order by date))/lag(cost\_of\_orders)over(order by date),1) as percent\_increase

from final

order by date



* The cost of orders in 2018 increased by approximately 137.0% compared to 2017. This indicates a substantial growth in orders between the two years, indicate positive business growth or increased customer demand

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

**ANS:**

select

c.customer\_state ,

round(sum(p.payment\_value))as total\_price,

round(avg(p.payment\_value))as average\_price

from `target.payments` p

inner join `target.orders` o

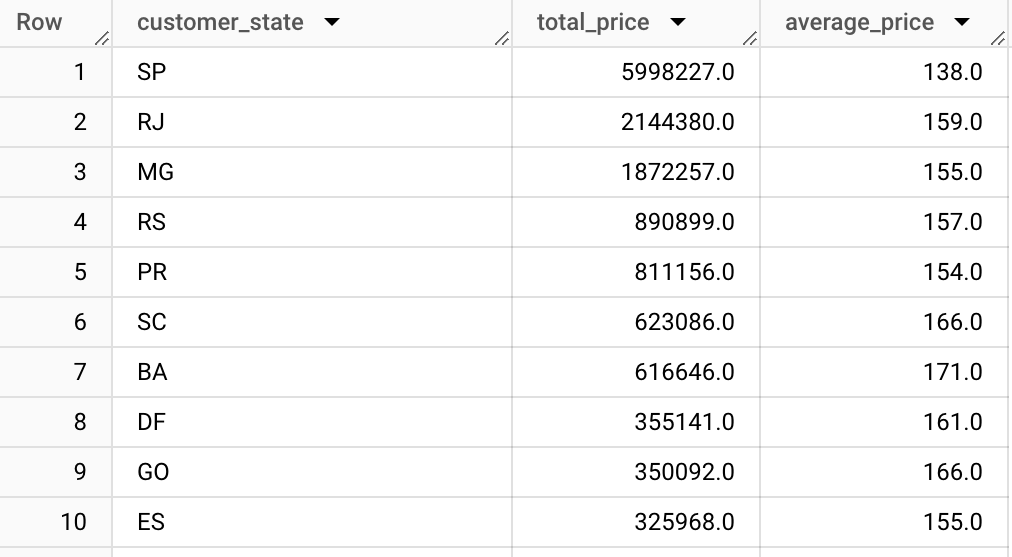
on p.order\_id = o.order\_id

inner join `target.customers` c

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

group by 1

order by 2 desc



* SP, RJ, MG remain the states with the highest total price.
* While SP has the lowest average price, this suggests that SP customers are purchasing lower value.

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

**ANS:** select c.customer\_state as State,

round(sum(i.freight\_value)) as Total\_Price,

round(avg(i.freight\_value)) as Avg\_Price

from `target.customers`c

inner join `target.orders`o

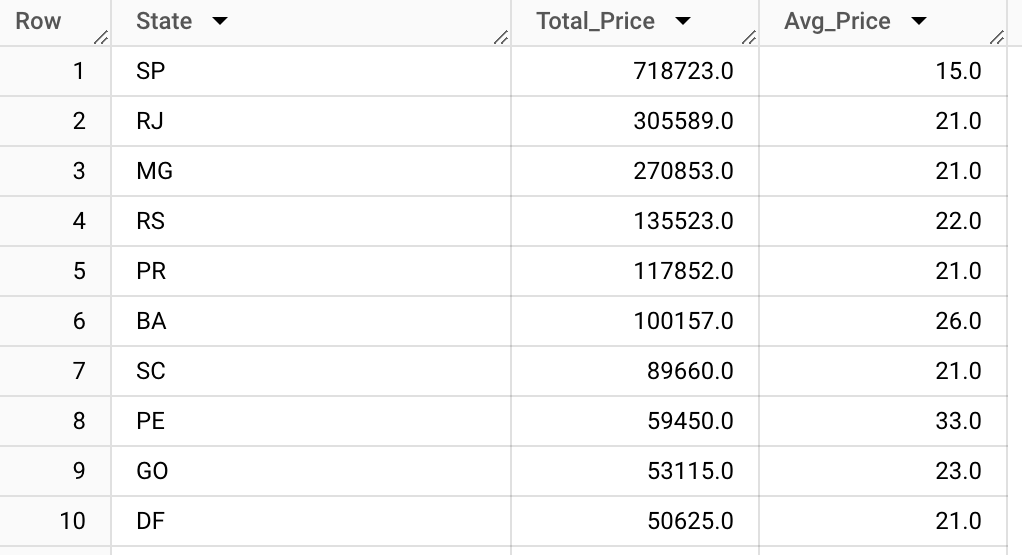
on c.customer\_id = o.customer\_id

inner join `target.order\_items`i

on o.order\_id = i.order\_id

group by c.customer\_state

order by 2 desc



* The variations in freight value can be influenced by factors such as geographical location, distance from the distribution center, transportation infrastructure.
* AC, AP, RR have the lowest total freight values which implies that shipping to and from these states is less expensive.
* SP, RJ, MG have the highest total freight values, which means that shipping to and from these states is more expensive.

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

**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.**

**ANS:**

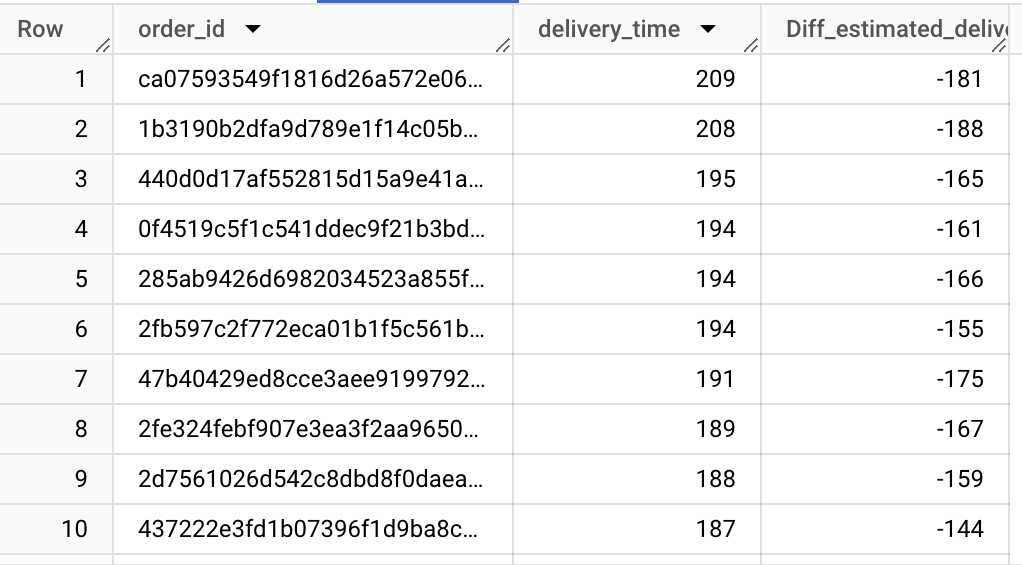
select order\_id,

date\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp, day) as delivery\_time,

date\_diff(order\_estimated\_delivery\_date, order\_delivered\_customer\_date, day) as Diff\_estimated\_delivery

from `target.orders`

order by 2 desc



* The results show that it is taking a long time to deliver the orders to the customer, this could be the concern for the customers' satisfaction and it's essential to investigate the reason behind such extended delivery time.
* Many orders were delivered earlier than the estimated delivery date, as indicated by negative values in the ‘Diff\_estimated\_delivery’ column. Early deliveries can be a positive customer experience.

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

ANS: WITH AvgFreightValues AS (

SELECT

c.customer\_state,

ROUND(AVG(i.freight\_value)) AS Avg\_Freight\_Value,

ROW\_NUMBER() OVER (ORDER BY AVG(i.freight\_value) DESC) AS HighRank,

ROW\_NUMBER() OVER (ORDER BY AVG(i.freight\_value) ASC) AS LowRank

FROM `target.customers` c

INNER JOIN`target.orders` o

ON c.customer\_id = o.customer\_id

INNER JOIN `target.order\_items` i

ON o.order\_id = i.order\_id

GROUP BY c.customer\_state

)

SELECT

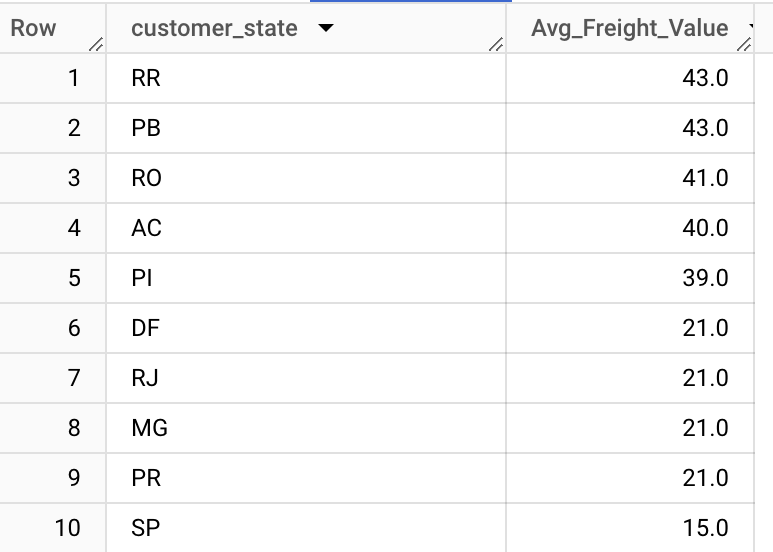
customer\_state,

Avg\_Freight\_Value

FROM AvgFreightValues

WHERE HighRank <= 5 OR LowRank <= 5

ORDER BY HighRank, LowRank;



* States like RR, PB, RO have the heighest average freight value indicating that customers in these states tend to have higher shipping costs per order.
* States like SP, PR, MG have the lowest average freight value indicating that customers in these states tend to have lower shipping costs per order.

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

**ANS:**

WITH AvgDeliveryTime AS (

SELECT

c.customer\_state,

ROUND(AVG(date\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day))) AS Avg\_Delivery\_Time,

ROW\_NUMBER() OVER (ORDER BY AVG(date\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) DESC) AS HighRank,

ROW\_NUMBER() OVER (ORDER BY AVG(date\_diff(o.order\_delivered\_customer\_date, o.order\_purchase\_timestamp, day)) ASC) AS LowRank

FROM `target.customers` c

INNER JOIN`target.orders` o

ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_state)

SELECT

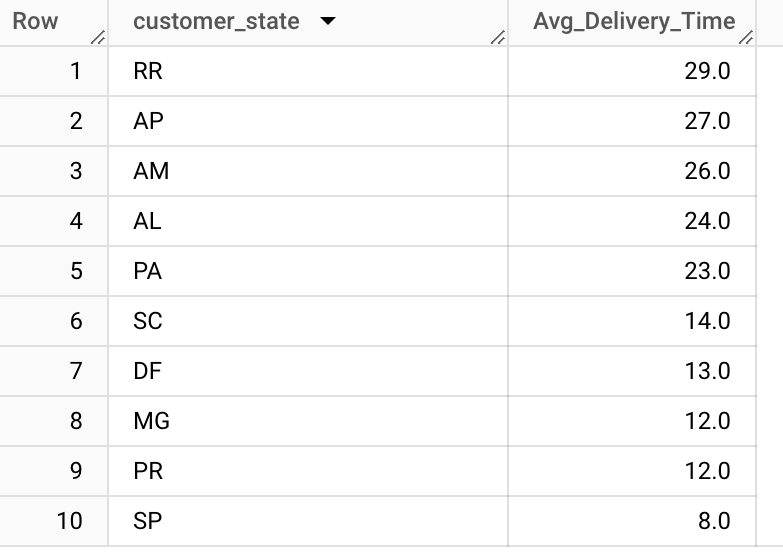
customer\_state,

Avg\_Delivery\_Time

FROM AvgDeliveryTime

WHERE HighRank <= 5 OR LowRank <= 5

ORDER BY HighRank , LowRank ;



* States like SP, PR, MG have the lowest average delivery time, indicating that customers in these states receive their orders quickly.
* States like RR, AP, AM have the highest average delivery time, indicating that customers in these states receive their orders late compared to others.

**4. 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(DATE\_DIFF(o.order\_delivered\_customer\_date, o.order\_estimated\_delivery\_date, day))) AS Delivery\_Time

FROM `target.customers`c

JOIN `target.orders`o

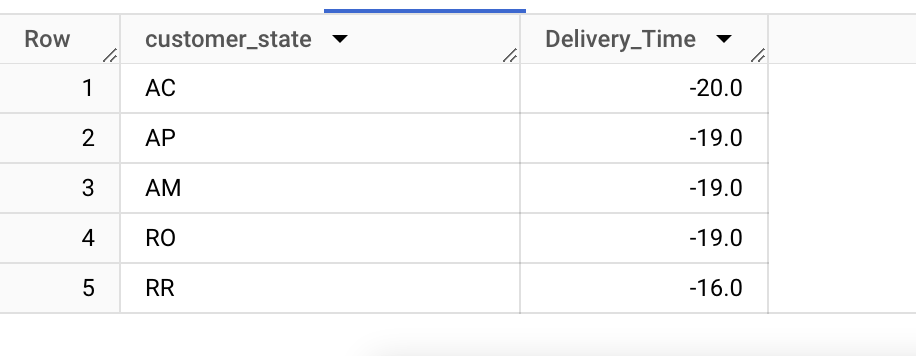
ON c.customer\_id = o.customer\_id

WHERE o.order\_status = 'delivered'

FROUP BY c.customer\_state

ORDER BY 2 ASC

LIMIT 5



* In these states orders were consistently delivered ahead of their estimated delivery time as indicated by negative signs in the “Delivery\_Time” column.
* Data implies that logistics and operations in these states are well organized. This can contribute to a positive customer experience.

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

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

**ANS.**

with final as(

SELECT

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

p.payment\_type,

COUNT(p.order\_id) AS Orders,

FROM `target.orders`o

JOIN `target.payments`p

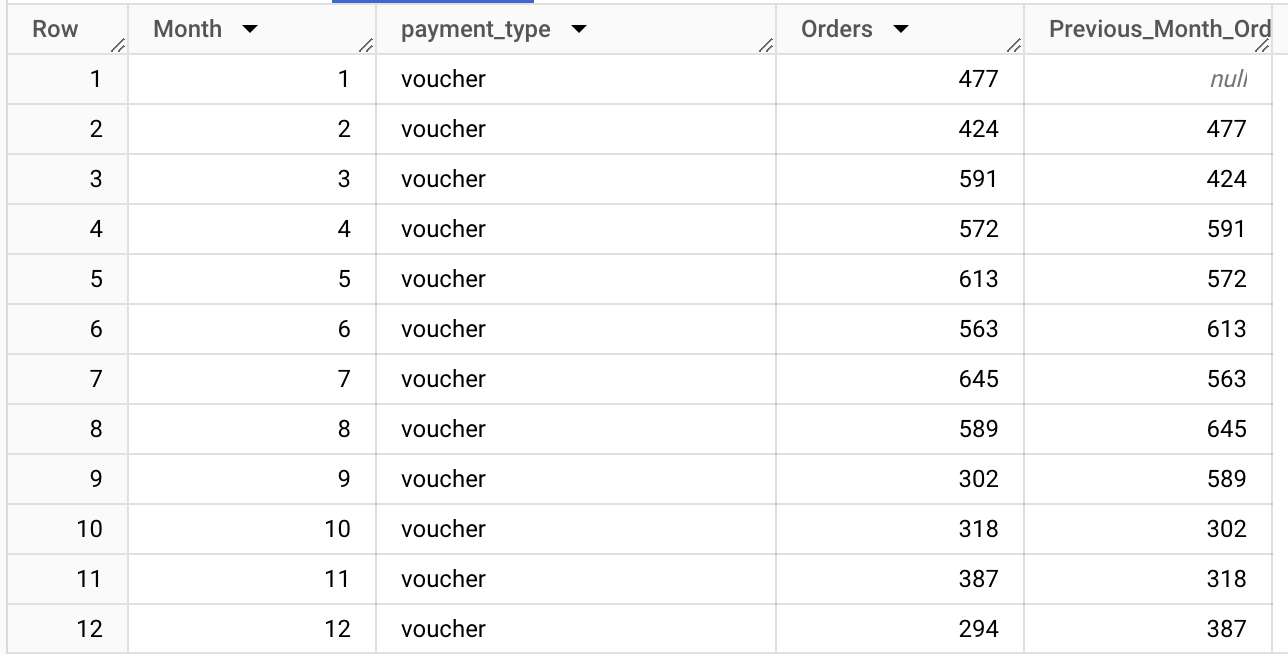
ON o.order\_id = p.order\_id

GROUP BY 1,2

ORDER BY 1 ASC)

select \* ,LAG(Orders) OVER (PARTITION BY payment\_type ORDER BY Month) AS Previous\_Month\_Orders

from final



* The data shows trends in the number of orders for different payment types, including "credit\_card," "UPI," "voucher," and "debit\_card.
* "credit\_card" appears to be a consistently popular payment method, with high order counts in most months.
* "UPI" also exhibits a significant number of orders, particularly in the later months.

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

**Ans:**

SELECT

payment\_installments,

COUNT(order\_id) AS Orders

FROM `target.payments`

WHERE payment\_installments>=1

GROUP BY 1

ORDER BY 1 ASC;



* There are significantly more orders with fewer payment installments, especially with 1, 2, and 3 installments.
* As the number of installments increases, the number of orders decreases. This suggests that many customers prefer to make payments in fewer installments.