**I. Import the dataset and do usual exploratory analysis steps like checking the**

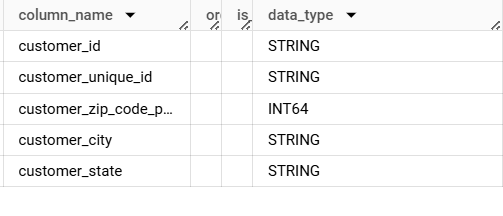
**structure & characteristics of the dataset.**

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

SELECT \*

FROM `aesthetic-fiber-453606-t0.target.INFORMATION\_SCHEMA.COLUMNS

WHERE table\_name = 'customers';



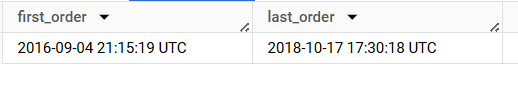
1. Get the time range between which the orders were placed.

SELECT

MIN(order\_purchase\_timestamp) as first\_order,

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

FROM `target.orders`

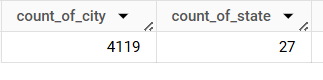
****

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

SELECT

COUNT(DISTINCT(customer\_city)) as count\_of\_city, COUNT(DISTINCT(customer\_state)) as count\_of\_state

FROM `target.customers`



**II. In-depth Exploration:**

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

WITH OrderCounts AS (

    SELECT

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

        count(\*) as No\_of\_orders

    FROM

        `target.orders`

    GROUP BY

        extract(year from order\_purchase\_timestamp)

    ORDER BY

        Years asc

)

SELECT

    Years,

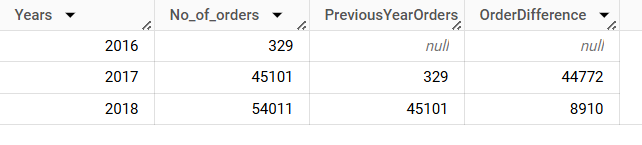
    No\_of\_orders,

    LAG(No\_of\_orders,1) OVER (ORDER BY Years) AS PreviousYearOrders,

    (No\_of\_orders - LAG(No\_of\_orders,1) OVER (ORDER BY Years)) AS OrderDifference

FROM

    OrderCounts;



1. **Increasing Trend in Orders**:

* There is a significant increase in the number of orders from 2016 to 2018.
* **2016** had 329 orders.
* **2017** shows a massive jump to 45,101 orders, indicating a large uptick in demand or business growth during this period.
* **2018** continues the positive trend with a further increase to 54,011 orders.

1. **Annual Order Difference**:

* The increase from 2016 to 2017 is quite substantial, with a difference of 44,772 orders.
* From 2017 to 2018, the increase is smaller compared to the previous year but still significant, with 8,910 more orders.

This trend suggests that there's consistent growth in orders, with a particularly large expansion between 2016 and 2017.

B. Can we see some kind of monthly seasonality in terms of the no. of orders being

Placed?

WITH OrderCounts AS (

    SELECT

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

        count(\*) as No\_of\_orders

    FROM

        `target.orders`

    GROUP BY

        extract(month from order\_purchase\_timestamp)

)

SELECT

    Months,

    No\_of\_orders,

    LAG(No\_of\_orders,1) OVER (ORDER BY Months) AS PreviousMonthOrders,

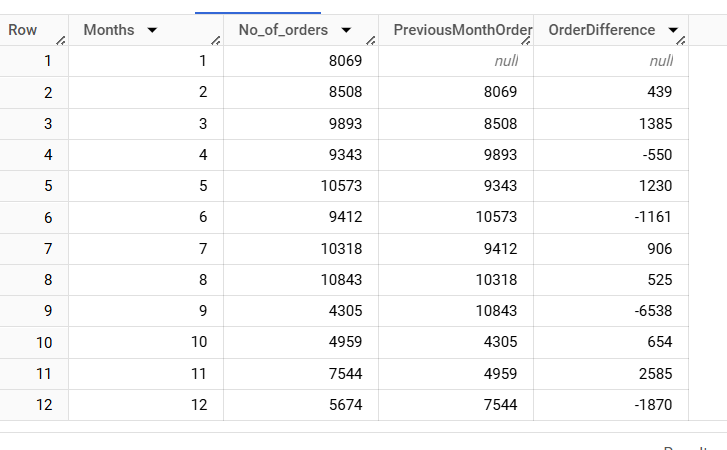
    (No\_of\_orders - LAG(No\_of\_orders,1) OVER (ORDER BY Months)) AS OrderDifference

FROM

    OrderCounts

ORDER BY

    CAST(Months AS INTEGER) ASC;



* **January:** Starting point with 8069 orders.
* **February to August:** Gradual increases in orders observed, peaking in August with 10,843 orders, suggesting a steady rise possibly due to business growth, strategic promotions, or seasonal demands.
* **Significant Trends:**
* **September:** Noticeable dip to 4,305 orders, indicating a significant decrease from August.
* **Fluctuations Toward Year End:**
* **October and November:** Recovery with healthier figures (4,959 in October growing to 7,544 in November).
* **December:** Reduction to 5,674 orders reflecting another drop from prior months, possibly aligning with typical end-of-year slowdowns.

C. 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 'Morning'

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

       then 'Evening'

       else 'Night'

      END AS DayPart,

      count(\*) as No\_of\_orders,

 FROM

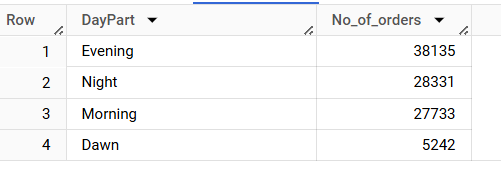
      `target.orders`

 GROUP BY

      DayPart

 ORDER BY

      No\_of\_orders desc



* Brazilian customers mostly place their orders during the Evening.

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

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

SELECT

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

    c.customer\_state,

    count(\*) as No\_of\_orders

  FROM `target.orders` o

  JOIN `target.customers`c on o.customer\_id = c.customer\_id

  GROUP BY

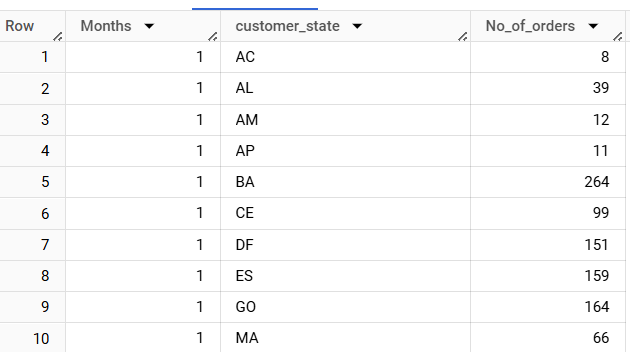
    c.customer\_state,

    Months

  ORDER BY

    Months,

    c.customer\_state;



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

SELECT

  customer\_state,

  count(distinct customer\_unique\_id) AS Customer\_count

FROM `target.customers`

GROUP BY

  customer\_state

ORDER BY

  Customer\_count desc;



* The SP state has more customers followed by RJ and MG.

**IV. Impact on Economy: Analyze the money movement by e-commerce by looking at**

**order prices, freight and others.**

A. Get the % increase in the cost of orders from year 2017 to 2018 (include

months between Jan to Aug only).

WITH YearlyOrderSums AS (

    SELECT

        EXTRACT(YEAR FROM o.order\_purchase\_timestamp) AS OrderYear,

        SUM(p.payment\_value) AS TotalOrderValue

    FROM

        `target.orders` o

    JOIN

        `target.payments` p ON o.order\_id = p.order\_id

    WHERE

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

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

    GROUP BY

        OrderYear

)

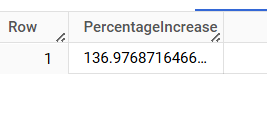
SELECT

    (y2018.TotalOrderValue - y2017.TotalOrderValue) / y2017.TotalOrderValue \* 100 AS PercentageIncrease

FROM

    (SELECT TotalOrderValue FROM YearlyOrderSums WHERE OrderYear = 2018) y2018,

    (SELECT TotalOrderValue FROM YearlyOrderSums WHERE OrderYear = 2017) y2017;



* The cost of orders has increased by around 137% from the year 2017 and 2018 which is huge increase

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

SELECT

  c.customer\_state,

  ROUND(SUM(p.payment\_value),2) As Total,

  ROUND(SUM(p.payment\_value)/COUNT(distinct p.order\_id),2) AS Average

FROM

  `target.payments` p

JOIN

  `target.orders` o on

    o.order\_id = p.order\_id

JOIN

  `target.customers` c on

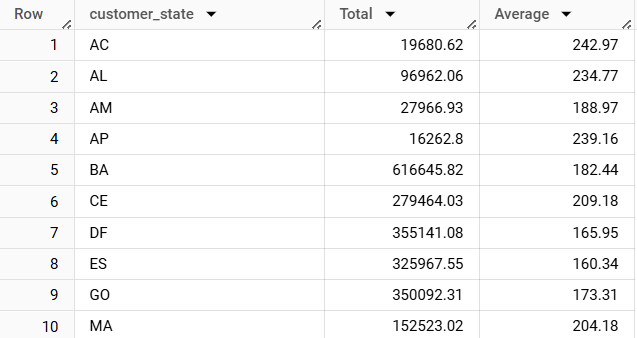
    c.customer\_id = o.customer\_id

GROUP BY

  c.customer\_state

ORDER BY

  c.customer\_state;



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

SELECT

  c.customer\_state,

  ROUND(SUM(f.freight\_value),2) As Total,

  ROUND(SUM(f.freight\_value)/COUNT(distinct f.order\_id),2) AS Average

FROM

  `target.order\_items` f

JOIN

  `target.orders` o on

    o.order\_id = f.order\_id

JOIN

  `target.customers` c on

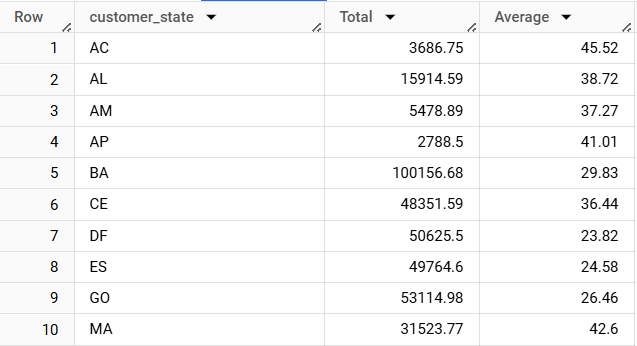
    c.customer\_id = o.customer\_id

GROUP BY

  c.customer\_state

ORDER BY

  c.customer\_state;



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

A. 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.

SELECT

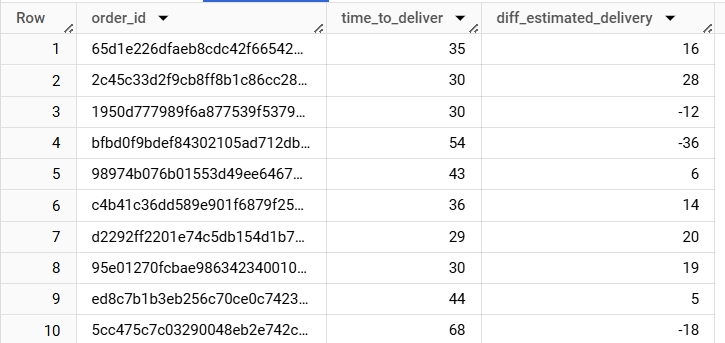
  order\_id,

  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`;



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

With RankedStates AS(

  SELECT

    c.customer\_state,

    AVG(o.freight\_value)AS Avg\_Freight

  FROM

    `target.order\_items` o

  JOIN

    `target.orders` ord ON

      ord.order\_id = o.order\_id

  JOIN

    `target.customers` c ON

      c.customer\_id = ord.customer\_id

  GROUP BY

    c.customer\_state

  ORDER BY

    c.customer\_state

),

 RankedHigh As(

  SELECT

    customer\_state,

    Avg\_Freight,

    ROW\_NUMBER()OVER(ORDER BY Avg\_Freight desc)As Rank\_Desc

  FROM

    RankedStates

  ),

  Ranked\_Low AS(

    SELECT

      customer\_state,

      Avg\_Freight,

    ROW\_NUMBER()OVER(ORDER BY Avg\_Freight ASC)As Rank\_Asc

  FROM

    RankedStates

  )SELECT customer\_state, Avg\_Freight

FROM RankedHigh

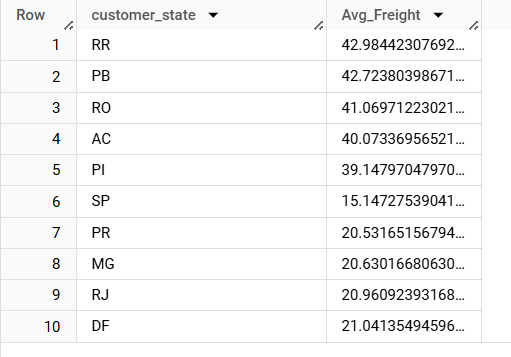
WHERE Rank\_Desc <= 5

UNION ALL

SELECT customer\_state, Avg\_Freight

FROM Ranked\_Low

WHERE Rank\_Asc <= 5;



* The top 5 states with the highest average freight values are RR, PB, RO, AC, and PI
* The top 5 states with the lowest average freight values are SP, PR, MG, RJ, and DF

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

WITH Ranked\_DT AS(

  SELECT c.customer\_state,

    AVG(DATE\_DIFF(order\_delivered\_customer\_date, order\_purchase\_timestamp, DAY)) AS Avg\_DT

  FROM `target.orders` o

  JOIN `target.customers` c ON o.customer\_id = c.customer\_id

  GROUP BY c.customer\_state

  ORDER BY c.customer\_state

),Ranked\_High AS(

  SELECT customer\_state,Avg\_DT,

  ROW\_NUMBER()OVER(ORDER BY Avg\_DT desc) AS rank\_desc

  FROM Ranked\_DT

),Ranked\_Low AS(

  SELECT customer\_state,Avg\_DT,

  ROW\_NUMBER()OVER(ORDER BY Avg\_DT asc) AS rank\_asc

  FROM Ranked\_DT

)SELECT customer\_state, Avg\_DT

FROM Ranked\_High

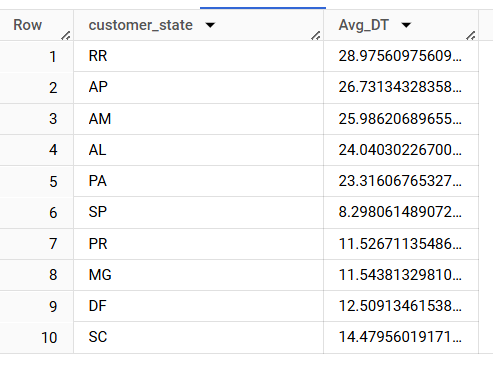
WHERE rank\_desc<=5

UNION ALL

SELECT customer\_state,Avg\_DT

FROM Ranked\_Low

WHERE rank\_asc<=5;



* The top 5 states with the highest average delivery time are RR, AP, AM, AL and PA.
* The top 5 states with the lowest average delivery time are SP, PR, MG, DF, and SC.

D. 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,

AVG(DATE\_DIFF(o.order\_estimated\_delivery\_date,o.order\_delivered\_customer\_date, day)) AS Avg\_Early\_DT

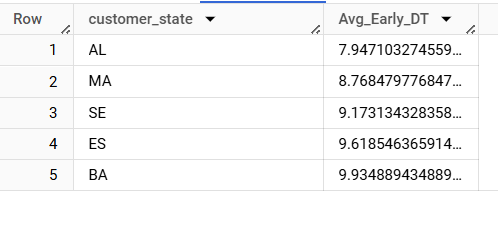
FROM `target.orders` o

JOIN `target.customers`c on o.customer\_id = c.customer\_id

GROUP BY c.customer\_state

ORDER BY Avg\_Early\_DT asc

LIMIT 5;



* The top 5 states where the order delivery is really fast as compared to the estimated date of delivery are AL, MA, SE, ES and BA

VI. Analysis based on the payments:

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

SELECT  EXTRACT(year FROM o.order\_purchase\_timestamp)AS Years,EXTRACT(month FROM o.order\_purchase\_timestamp)AS Months,

p.payment\_type,

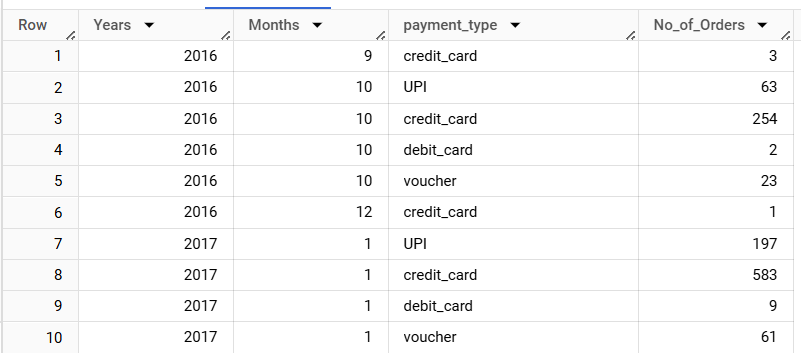
count(\*) AS No\_of\_Orders

FROM `target.orders` o

JOIN `target.payments` p on o.order\_id = p.order\_id

GROUP BY p.payment\_type,Months,Years

ORDER BY Years,Months, p.payment\_type;



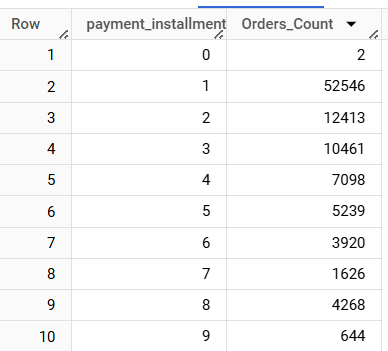
B. Find the no. of orders placed on the basis of the payment installments that have

been paid.

SELECT payment\_installments,count(\*) AS Orders\_Count

FROM `target.payments`

GROUP BY payment\_installments;



**Actionable Insights**

* The increase of orders from 2016 to 2017 is quite substantial, with a difference of 44,772 orders. From 2017 to 2018, the increase is smaller compared to the previous year but still significant, with 8,910 more orders. This trend suggests consistent growth in orders, with a particularly large expansion between 2016 and 2017. So the company has to be ready with extra workforce. To avoid high risk, it can consider hiring contractual employees.
* From September to December, there is a noticeable dip in the orders. So, suggesting for strategic promotions to increase the orders .
* Brazilian customers mostly place their orders during the Evening.
* The SP state has more customers followed by RJ and MG.
* The cost of orders has increased by around 137% from the year 2017 and 2018, which is a huge increase.
* The top 5 states with the highest average freight values are RR, PB, RO, AC, and PI.
* The top 5 states with the lowest average freight values are SP, PR, MG, RJ, and DF.
* The top 5 states with the highest average delivery time are RR, AP, AM, AL and PA.
* The top 5 states with the lowest average delivery time are SP, PR, MG, DF, and SC.
* The top 5 states where the order delivery is really fast as compared to the estimated date of delivery are AL, MA, SE, ES, and BA

**Recommendations:**

* To maintain the growth trajectory without overextending resources, consider hiring contractual employees during peak seasons for flexibility.
* Ensure robust supply chain management to handle increased demand without compromising on delivery quality.
* Implement seasonal promotions and marketing campaigns to boost order volumes.
* Explore partnerships for jointly marketed holiday packages or offers.
* Optimize marketing efforts and customer service responsiveness during these peak evening hours to maximize engagement.
* For states with high customer bases like SP, RJ, and MG, ensure that logistics and customer service operations are finely tuned to meet high demand.
* Consider addressing high freight cost areas such as RR, PB, RO, AC, and PI with optimizations like renegotiated shipping contracts or routing improvements.
* Analyze cost drivers to maintain profit margins. Look into areas like supply chain efficiencies, cost of goods sold, and operational productivity.
* Learn from states like AL, MA, SE, ES, and BA for efficiency to streamline delivery processes in other regions.