**TARGET SQL Case Study**

Software Used: GCP Big Query

|  |  |  |
| --- | --- | --- |
|  | CONTENTS |  |
| 1 | Problem Statements & Answers | 03 |
| 2 | Insights | 15 |
| 3 | Recommendations | 16 |

* 1. **Datatype of all columns in “customers” table.**

SELECT

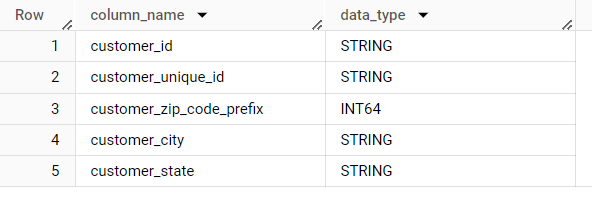
column\_name,

data\_type

FROM Target\_CS.INFORMATION\_SCHEMA.COLUMNS

WHERE table\_name ='customers';

**Output:**



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

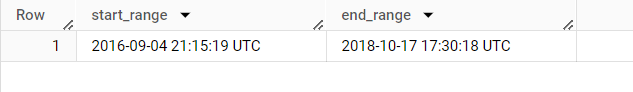
SELECT

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

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

FROM Target\_CS.orders;

**Output:**



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

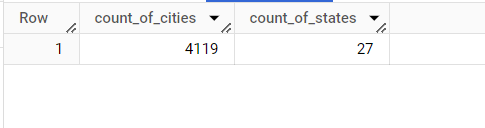
SELECT

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

  COUNT(DISTINCT(customer\_state)) as count\_of\_states

FROM Target\_CS.customers;

**Output:**



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

SELECT

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

COUNT(DISTINCT o.order\_id) AS order\_count

FROM Target\_CS.orders o

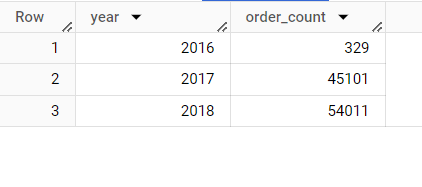
JOIN Target\_CS.customers c

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

GROUP BY 1

ORDER BY 1;

**Output:**



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

SELECT

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

COUNT(DISTINCT order\_id) AS order\_count

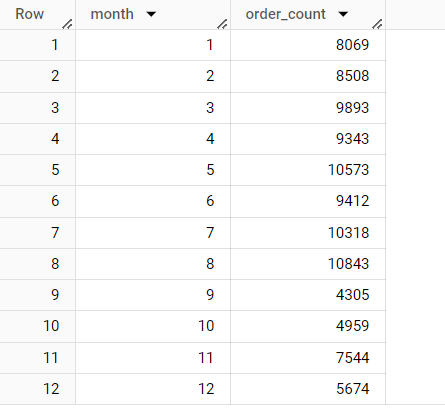
FROM

Target\_CS.orders

GROUP BY month

ORDER BY month;

**Output:**



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

SELECT

CASE

WHEN EXTRACT(HOUR FROM o.order\_purchase\_timestamp) BETWEEN 0 AND 6 THEN

'Dawn'

WHEN EXTRACT(HOUR FROM o.order\_purchase\_timestamp) BETWEEN 7 AND 12 THEN 'Morning'

WHEN EXTRACT(HOUR FROM o.order\_purchase\_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'

WHEN EXTRACT(HOUR FROM o.order\_purchase\_timestamp) BETWEEN 19 AND 23 THEN 'Night'

END AS hour,

COUNT(o.order\_id) AS order\_count

FROM

Target\_CS.orders o

JOIN

Target\_CS.customers c

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

GROUP BY

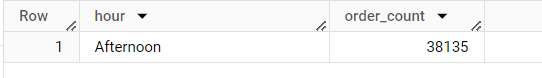
hour

ORDER BY

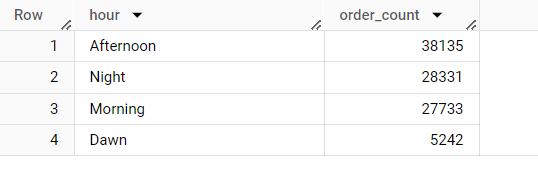
order\_count DESC

LIMIT 1;

**Output:**



**\***Count of orders at certain timings.



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

SELECT

c.customer\_state,

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

COUNT(o.order\_purchase\_timestamp) AS order\_count

FROM

Target\_CS.orders o

JOIN

Target\_CS.customers c

ON

o.customer\_id = c.customer\_id

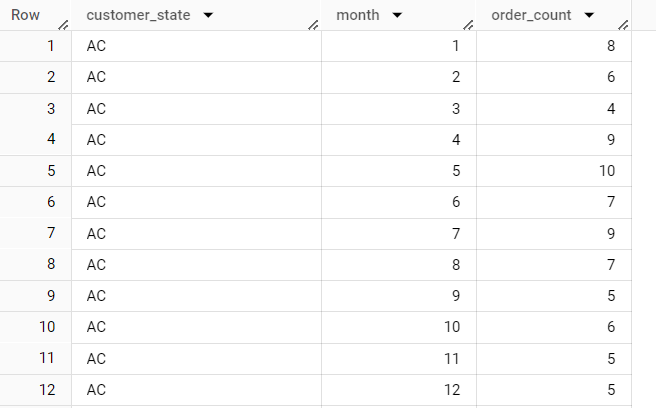
GROUP BY

c.customer\_state, month

ORDER BY

c.customer\_state, month;

**Output:**



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

SELECT

customer\_state,

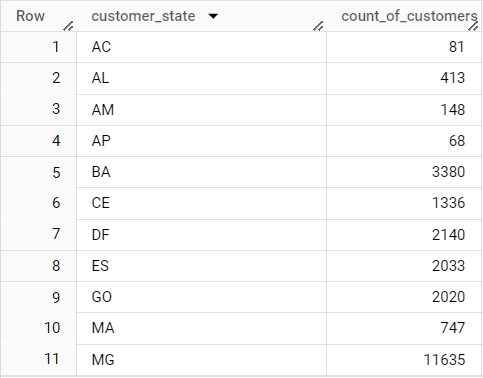
COUNT(customer\_id) AS count\_of\_customers

FROM Target\_CS.customers

GROUP BY customer\_state

ORDER BY customer\_state;

**Output:**

****

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

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 percentage\_increase

FROM Target\_CS.orders o

JOIN Target\_CS.payments p ON o.order\_id = p.order\_id

WHERE

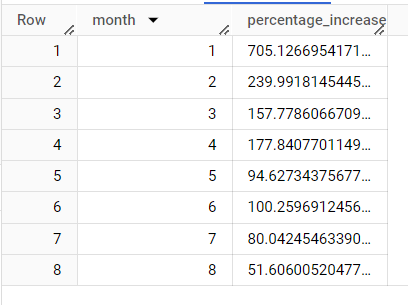
  EXTRACT(YEAR FROM o.order\_purchase\_timestamp) IN (2017, 2018) AND

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

GROUP BY month

ORDER BY month;

**Output:**



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

SELECT

c.customer\_state,

ROUND(SUM(i.price), 2) AS Total\_price,

ROUND(AVG(i.price), 2) AS Average\_price

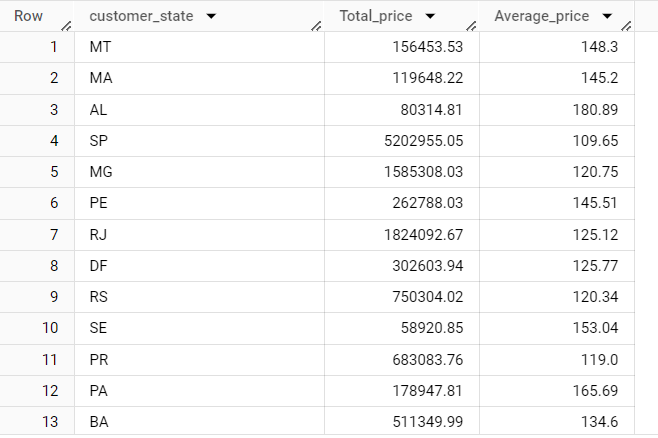
FROM Target\_CS.orders o

JOIN Target\_CS.order\_items i ON o.order\_id = i.order\_id

JOIN Target\_CS.customers c ON o.customer\_id = c.customer\_id

GROUP BY c.customer\_state;

**Output:**



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

SELECT

c.customer\_state,

ROUND(SUM(oi.freight\_value), 2) AS Total\_freight,

ROUND(AVG(oi.freight\_value), 2) AS Average\_freight

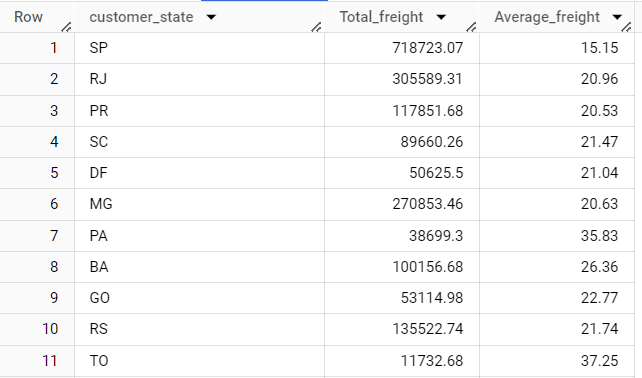
FROM Target\_CS.orders o

JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

JOIN Target\_CS.customers c ON o.customer\_id = c.customer\_id

GROUP BY c.customer\_state;

**Output:**

****

* 1. **Calculate 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.**

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)

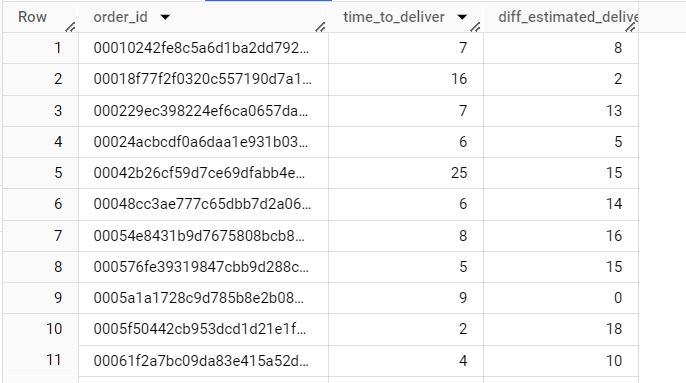
  AS diff\_estimated\_delivery

FROM Target\_CS.orders

WHERE DATE\_DIFF(order\_delivered\_customer\_date, order\_purchase\_timestamp,DAY) IS NOT NULL

ORDER BY order\_id;

**Output:**



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

#Top 5 States with highest average freight value

SELECT

c.customer\_state,

ROUND(AVG(oi.freight\_value),2) as average\_Frieght\_value

FROM Target\_CS.customers c

JOIN Target\_CS.orders o ON c.customer\_id = o.customer\_id

JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

GROUP BY 1

ORDER BY 2 DESC

LIMIT 5;

#Top 5 States with lowest average freight value

SELECT

c.customer\_state,

ROUND(AVG(oi.freight\_value),2) as average\_Frieght\_value

FROM Target\_CS.customers c

JOIN Target\_CS.orders o ON c.customer\_id = o.customer\_id

JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

GROUP BY 1

ORDER BY 2

LIMIT 5;

**Output:**

**Highest freight :**

****

**Lowest freight:**

****

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

#Top 5 states with lowest average time of delivery

SELECT

S.State,

ROUND(AVG(delivery),2) AS average\_Delivery\_Time

FROM(

      SELECT o.order\_id,

      c.customer\_state AS state,

      DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp, DAY)AS delivery

      FROM Target\_CS.orders o

      JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

      JOIN Target\_CS.customers c ON o.customer\_id = c.customer\_id

    ) S

GROUP BY 1

ORDER BY 2

LIMIT 5;

#Top 5 states with highest average time of delivery

SELECT

S.State,

ROUND(AVG(delivery),2) AS average\_Delivery\_Time

FROM(

      SELECT o.order\_id,

      c.customer\_state AS state,

      DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp, DAY)AS delivery

      FROM Target\_CS.orders o

      JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

      JOIN Target\_CS.customers c ON o.customer\_id = c.customer\_id

    ) S

GROUP BY 1

ORDER BY 2 DESC

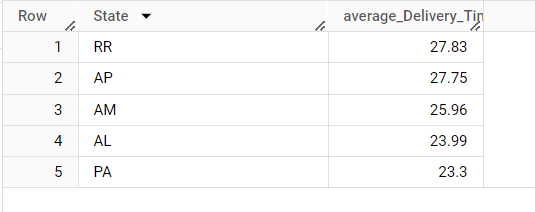
LIMIT 5 ;

**Output:**

**Lowest average delivery time:**



**Highest average delivery time:**



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

SELECT

T.State,

ROUND(estimated\_time - delivery\_time,2) AS Delivery\_days

FROM(

      SELECT

      c.customer\_state AS state,

      DATE\_DIFF(order\_delivered\_customer\_date,order\_purchase\_timestamp, DAY)AS delivery\_time,

      DATE\_DIFF(order\_estimated\_delivery\_date,order\_purchase\_timestamp, DAY)AS estimated\_time

      FROM Target\_CS.orders o

      JOIN Target\_CS.order\_items oi ON o.order\_id = oi.order\_id

      JOIN Target\_CS.customers c ON o.customer\_id = c.customer\_id

    ) T

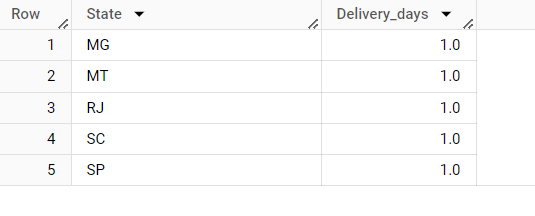
WHERE estimated\_time>delivery\_time

GROUP BY 1, 2

ORDER BY 2

LIMIT 5;

**Output:**

****

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

SELECT

p.payment\_type,

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

COUNT(DISTINCT o.order\_id) AS order\_count

FROM Target\_CS.orders o

JOIN Target\_CS.payments p ON o.order\_id = p.order\_id

GROUP BY 1, 2

ORDER BY 2;

**Output:**



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

SELECT

COUNT(o.order\_id) AS order\_count,

p.payment\_installments

FROM Target\_CS.orders o

JOIN Target\_CS.payments p

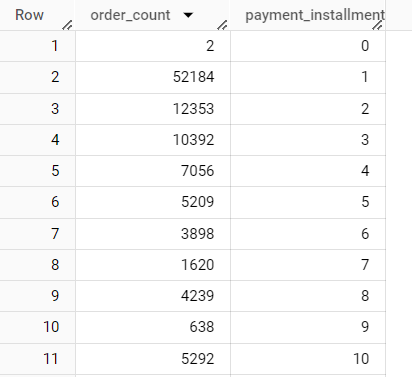
ON o.order\_id = p.order\_id

WHERE o.order\_status <> 'canceled'

GROUP BY 2

ORDER BY 2;

**Output:**



**Insights:**

* The data is recorded from the year 2016 and 2018
* There are in 99441 customers who made transactions over this period.
* There were 4119 cities out of 8000+ cities from all 27 states in Brazil.
* We can observe that there is sudden hike of orders in 2017.
* Usually the highest sales happens during the months of May, July and August with more than 10K+ records.
* Most of the orders come at mid of the day i.e. afternoon
* The data indicates a decline in orders during September and October. Offering discounts or promotions during off-peak seasons can incentivize customers to make purchases during these periods, thus boosting sales.
* States like SP and RJ already have high order counts. To further boost sales and foster brand loyalty, it is recommended to focus on customer retention strategies, such as personalized marketing campaigns, loyalty programs, and exceptional customer service experiences.
* Improving delivery times in areas with longer delivery durations can have a positive impact on customer satisfaction and encourage repeat purchases. Streamlining logistics and implementing efficient shipping processes are key to achieving this.

**Recommendations:**

* Improve logistics and shipping processes to reduce delivery times and enhance customer satisfaction. This includes optimizing warehouse operations, refining shipping routes, and partnering with reliable courier services.
* Implement customer retention strategies to encourage repeat purchases and foster loyalty. This can be achieved through loyalty programs, referral rewards, and personalized offers.
* Evaluate pricing and freight fees to ensure competitiveness in the market while maximizing revenue and profitability. Consider increasing prices or adjusting freight fees as appropriate.
* Invest in technology and infrastructure to enhance the e-commerce experience. This includes implementing chatbots for customer support, improving website performance, and offering personalized product recommendations based on customer behaviour.
* Collaborate with sellers to expand product offerings and improve product quality, catering to diverse customer needs and preferences.