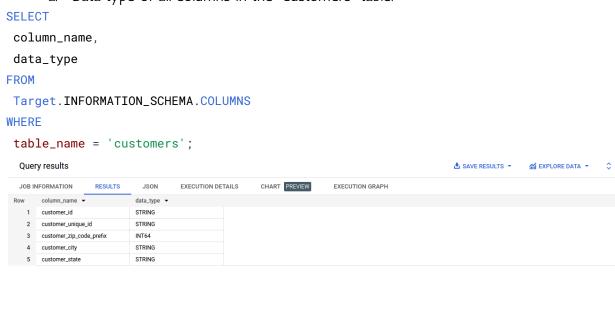
STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

- 1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.
 - a. Data type of all columns in the "customers" table.



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



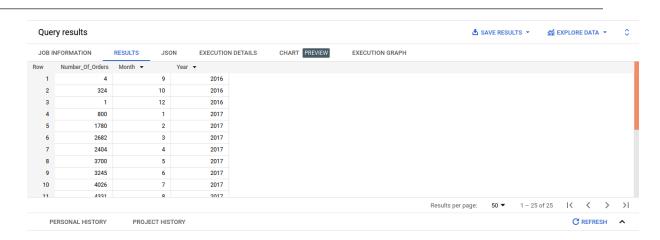
STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

2. In-depth Exploration:

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

```
SELECT
COUNT(order_id) Number_Of_Orders,
EXTRACT(Month FROM order_purchase_timestamp) AS Month,
EXTRACT(YEAR FROM order_purchase_timestamp) AS Year
FROM
`Target.orders`
GROUP BY
Year,
Month
ORDER BY
Year,
Month;
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023



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

```
SELECT
COUNT(order_id) Number_Of_Orders,
EXTRACT(Month FROM order_purchase_timestamp) AS Month,
EXTRACT(YEAR FROM order_purchase_timestamp) AS Year
FROM
`Target.orders`
GROUP BY
Year,
Month
ORDER BY
Year,
Month;
```

Insight - With the given data we can derive from the query results that there is no such seasonality every year. However for year 2017 it was revealed that orders from month 11 is at Peak

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

```
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
 COUNT(order_id) Number_Of_Orders,
 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 23 THEN 'Night'
END
 AS Time_Of_the_Day
FROM
 `Target.orders`
GROUP BY
Time_Of_the_Day
ORDER BY
 Number_Of_Orders;
  Query results

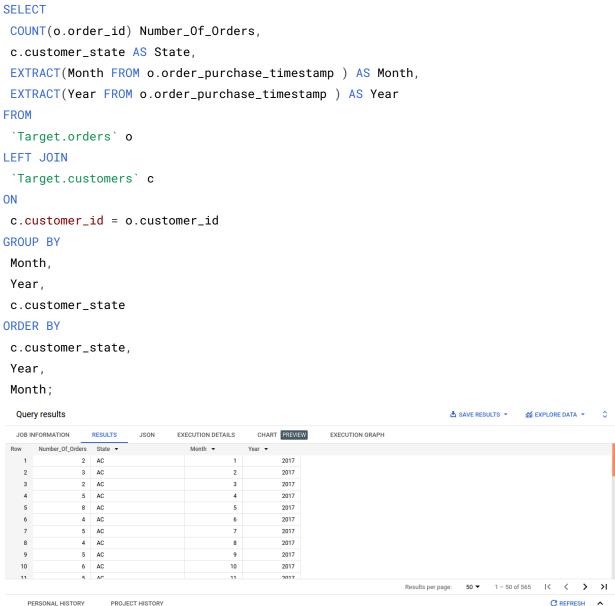
▲ SAVE RESULTS ▼

                                                                                  JOB INFORMATION
             RESULTS JSON EXECUTION DETAILS
                                       CHART PREVIEW
                                                   EXECUTION GRAPH
 Row Number_Of_Orders Time_Of_the_Day ▼
          5242 Dawn
          27733 Mornings
          28331 Night
          38135 Afternoon
```

Insight - The highest orders are placed in the afternoon.

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

- 3. Evolution of E-commerce orders in the Brazil region:
 - a. Get the month on month no. of orders placed in each state. by our customers.



Insight - There is a considerable and significant growth in the no.of placed orders from 2017 onwards. This may be due to major marketing activities and campaigns however we need more data to conclude.

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

b. How are the customers distributed across all the states? **SELECT** COUNT (DISTINCT(customer_unique_id)) AS Number_of_Customers, customer_state AS States FROM `Target.customers` **GROUP BY** Customer_state; Query results JOB INFORMATION RESULTS JSON EXECUTION DETAILS CHART PREVIEW EXECUTION GRAPH Row Number_of_Custome States ▼ 45 RR 67 AP 77 AC 143 AM 240 273 TO 342 SE 474 RN 10 482 PI 510 PR PERSONAL HISTORY PROJECT HISTORY C REFRESH ^

Insight - The state RJ has the most no.of Customers in the list and state RR has the least no.of Customers

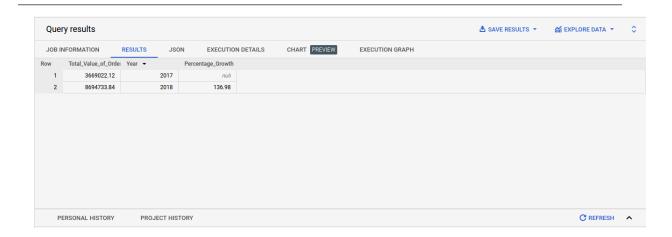
STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

Year;

- 4. 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).

```
SELECT
Total_Value_of_Orders,
Year,
 ROUND((((Total_Value_of_Orders - LAG(Total_Value_of_Orders) OVER(ORDER BY
Year))/LAG(Total_Value_of_Orders) OVER(ORDER BY Year)) *100),2) AS
Percentage_Growth
FROM (
SELECT
   ROUND(SUM(payment_value),2) AS Total_Value_of_Orders,
   EXTRACT (Year
   FROM
     o.order_purchase_timestamp) AS Year,
 FROM
   `Target.orders` o
 LEFT JOIN
   `Target.payments` p
 ON
   p.order_id = o.order_id
   EXTRACT(MONTH FROM o.order_purchase_timestamp ) BETWEEN 1 AND 8
 GROUP BY
   Year
 ORDER BY
  Year )
ORDER BY
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

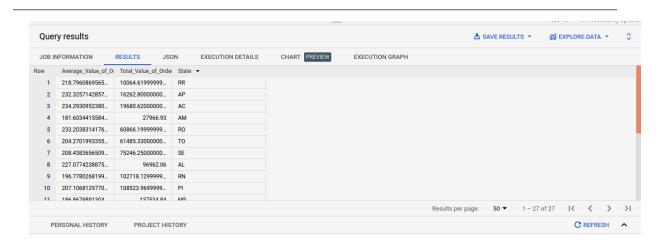


Insight - There is a considerable and significant growth in the no.of placed orders from 2017 onwards. This maybe due to major marketing activities and campaigns however we need more data to conclude

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

```
SELECT
 AVG(payment_value) AS Average_Value_of_Orders,
 SUM(payment_value) AS Total_Value_of_Orders,
 c.customer_state AS State
FROM
 `Target.payments` p
LEFT JOIN
 `Target.orders` o
ON
 o.order_id = p.order_id
LEFT JOIN
 `Target.customers` c
ON
 c.customer_id = o.customer_id
GROUP BY
 State
ORDER BY
Total_Value_of_Orders,
 Average_Value_of_Orders;
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023



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

```
SELECT

AVG(freight_value) AS Avg_Freight_Value_of_Orders,
SUM(freight_value) AS Total_Freight_Value_of_Orders,
c.customer_state AS State

FROM

`Target.order_items` oi

LEFT JOIN

`Target.orders` o

ON

o.order_id = oi.order_id

LEFT JOIN

`Target.customers` c

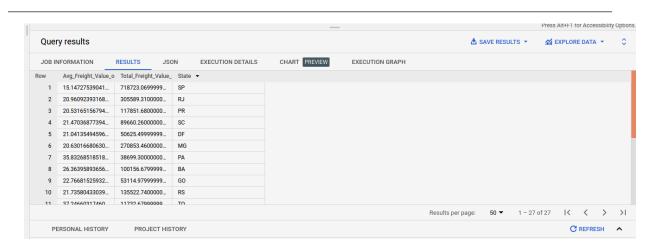
ON

c.customer_id = o.customer_id

GROUP BY

State;
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023



- 5. 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.
 - •time_to_deliver = order_delivered_customer_date order_purchase_timestamp
 - diff_estimated_delivery = order_estimated_delivery_date order_delivered_customer_date

SELECT

```
order_id,
```

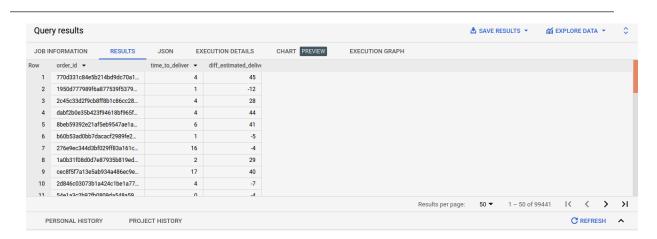
DATE_DIFF(order_delivered_carrier_date, order_purchase_timestamp, day) AS
time_to_deliver,

DATE_DIFF(order_estimated_delivery_date, order_delivered_customer_date, day) AS
diff_estimated_delivery

FROM

`Target.orders`;

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023



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

```
WITH T1 AS(
SELECT
 ROW_NUMBER() OVER (ORDER BY AVG(freight_value)) as RN,
 AVG(freight_value) AS Average_Value_of_Orders,
 c.customer_state AS State
FROM
 `Target.order_items` oi
LEFT JOIN
 `Target.orders` o
 o.order_id = oi.order_id
LEFT JOIN
 `Target.customers` c
 c.customer_id = o.customer_id
GROUP BY
 State
ORDER BY
 Average_value_of_Orders)
SELECT T.*,
CASE
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

```
WHEN RN BETWEEN 1 AND 5 THEN 'BOTTOM'
    ELSE 'TOP'
 END AS TopOrBottom
FROM T1 AS T
WHERE RN IN (1,2,3,4,5) OR RN>(SELECT COUNT(*) - 5 FROM T1);
   Query results

♣ SAVE RESULTS ▼

                                                                                                              JOB INFORMATION RESULTS
                            JSON EXECUTION DETAILS
                                                   CHART PREVIEW
                                                                    EXECUTION GRAPH
                   Average_Value_of_O⊢ State ▼
          1 15.14727539041... SP
                                                   воттом
             2 20.53165156794... PR
3 20.63016680630... MG
                                                   воттом
                                                   воттом
         4 20.96092393168... RJ
        5 21.04135494596... DF
23 39.14797047970... PI
                                                   воттом
                                                   TOP
             25 41.06971223021... RO
26 42.72380398671... PB
                                                    TOP
                                                    TOP
         27 42.98442307692... RR
                      PROJECT HISTORY
                                                                                                                   C REFRESH ^
     PERSONAL HISTORY
```

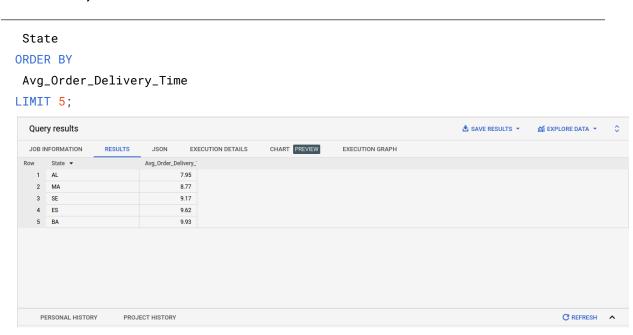
Find out the top 5 states with the highest & lowest average delivery time WITH T2 AS (**SELECT** ROW_NUMBER() OVER (ORDER BY ROUND (AVG(DATE_DIFF(o.order_delivered_carrier_date, o.order_purchase_timestamp, day)), 2)) as RN, c.customer_state AS State, ROUND (AVG(DATE_DIFF(o.order_delivered_carrier_date, o.order_purchase_timestamp, day)), 2) AS Avg_Delivery_Time **FROM** `Target.orders` o LEFT JOIN `Target.customers` c ON c.customer_id = o.customer_id **GROUP BY** State ORDER BY Avg_Delivery_Time

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

```
SELECT T.*,
CASE
   WHEN RN BETWEEN 1 AND 5 THEN 'LOW'
   ELSE 'HIGH'
 END AS TopOrBottom
FROM T2 AS T
WHERE RN IN (1,2,3,4,5) OR RN>(SELECT COUNT(*) - 5 FROM T2);
  Query results
                                                                            JSON EXECUTION DETAILS
  JOB INFORMATION RESULTS
                                          CHART PREVIEW
                                                      EXECUTION GRAPH
                      Avg_Delivery_Time TopOrBottom ▼
                                     2.31 LOW
           2 RO
                                     2.33 LOW
                                      2.68 LOW
                                     2.68 LOW
                                    2.68 LOW
             5 GO
            23 AP
                                      3.04 HIGH
                                     3.11 HIGH
            25 MA
                                     3.13 HIGH
                                      3.15 HIGH
                                     4.02 HIGH
    PERSONAL HISTORY
                  PROJECT HISTORY
                                                                                             C REFRESH ^
```

d. Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery. You can use the difference between the averages of actual & estimated delivery date to figure out how fast the delivery was for each state

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

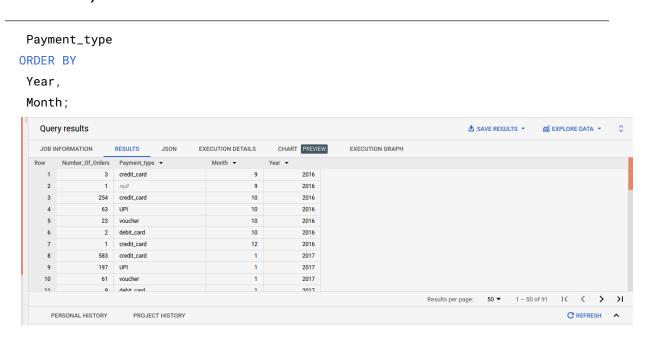


6. Analysis based on the payments:

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

```
SELECT
 COUNT(o.order_id) Number_Of_Orders,
 p.payment_type AS Payment_type,
 EXTRACT(Month
 FROM
   o.order_purchase_timestamp ) AS Month,
 EXTRACT(Year
 FROM
   o.order_purchase_timestamp ) AS Year
FROM
 `Target.orders` o
LEFT JOIN
 `Target.payments` p
ON
 p.order_id = o.order_id
GROUP BY
 Month,
 Year,
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023



Insight - Most of the orders are placed via Credit cards

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

```
SELECT
 COUNT(o.order_id) Number_Of_Orders,
 p.payment_installments AS Payment_installments,
 EXTRACT (Month
 FROM
   o.order_purchase_timestamp ) AS Month,
 EXTRACT(Year
 FROM
   o.order_purchase_timestamp ) AS Year
FROM
 `Target.orders` o
LEFT JOIN
 `Target.payments` p
ON
 p.order_id = o.order_id
WHERE
 p.payment_installments <> 0
GROUP BY
```

STUDENT NAME MANOJ BHAT BUSINESS CASE - TARGET SQL BATCH - DSML JULY 2023

Month,
Year,
Payment_installments
ORDER BY
Year,
Month;

