1.

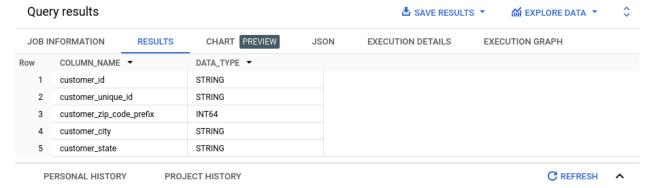
Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.

A)Data type of all columns in the "customers" table.

QUERY:-

```
SELECT
COLUMN_NAME, DATA_TYPE
FROM
    'targetsql-403414.TargetSQL.INFORMATION_SCHEMA.COLUMNS'
WHERE
TABLE_NAME = 'customers';
```

RESULT:-



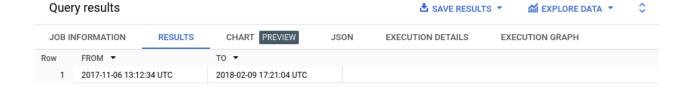
INSIGHTS:-

Most of the columns of CUSTOMERS table are of string type.

B)Get the time range between which the orders were placed.

QUERY:-

```
1 SELECT MIN(order_purchase_timestamp) AS `FROM`, MAX(order_purchase_timestamp) AS `TO` FROM (
2 | SELECT * FROM `targetsql-403414.TargetSQL.orders`
3 | WHERE order_status = 'created'
4 | ORDER BY order_purchase_timestamp
5 );
```

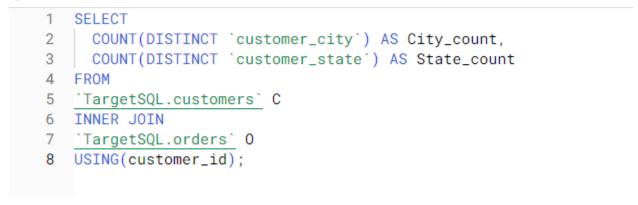


INSIGHT:-

Orders started from 06th Nov 2017 till 9th feb 2018. That is winter period.

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

QUERY:-



RESULT:-



INSIGHT:-

There are total 4119 cities and 27 states where orders were placed.

2.In-depth Exploration

A)Is there a growing trend in the no. of orders placed over the past years? QUERY:-

```
1 SELECT Year, Count(Year) As `Count` FROM
2 (SELECT EXTRACT(YEAR FROM order_purchase_timestamp) AS Year
3 FROM <u>`TargetSQL.orders`</u>)
4 GROUP BY Year
5 ORDER BY 1;
```

Result:-



INSIGHT:-

Yes we can clearly see growth around 20% in the year 2017 and 2018 orders. B)Can we see some kind of monthly seasonality in terms of the no. of orders being placed?

Query:-

```
1 SELECT month, COUNT(*) AS `count` FROM (
2 SELECT EXTRACT(MONTH FROM order_purchase_timestamp) AS month
3 FROM <u>`TargetSQL.orders`</u>
4 )
5 GROUP BY month
6 ORDER BY 2 DESC;
```

RESULT:-

JOB INFORMATION			RESULTS	CHA	JSON	
Row	month ▼		count ▼			
1		8		10843		
2		5		10573		
3		7		10318		
4		3		9893		

INSIGHT:-

In The month August Sales was at its PEAK whereas sales in the months May and July was also High.

E

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

QUERY:-

```
1  SELECT
2  CASE
3     WHEN Hour >= 0 AND Hour <= 6 THEN 'Dawn'
4     WHEN Hour >= 7 AND Hour <= 12 THEN 'Mornings'
5     WHEN Hour >= 13 AND Hour <= 18 THEN 'Afternoon'
6     WHEN Hour >= 19 AND Hour <= 23 THEN 'Night'
7     END
8     AS Daytime,
9     Count(*) AS `number_of_orders`
10     FROM (SELECT EXTRACT(HOUR FROM order_purchase_timestamp) AS Hour
11     FROM `TargetSQL.orders`)
12     GROUP BY 1
13     ORDER BY 2 DESC
14 ;</pre>
```

RESULT:-

Query results ₫ save resu

JOB II	NFORMATION	RESULTS	CHART PREVIEW	JSON	EXECUTION DETAILS
Row	Daytime ▼		number_of_orders •		
1	Afternoon		38135		
2	Night		28331		
3	Mornings		27733		
4	Dawn		5242		

INSIGHT:-

Most of the customers in brazil place their order during Afternoon. Whereas very few people place orders during Dawn time.

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

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

QUERY:-

```
1 SELECT
2 | C.customer_state,
3 | EXTRACT(MONTH FROM 0.order_purchase_timestamp) AS month,
4 | COUNT(C.customer_state) AS `Number Of Orders`
5 | FROM `TargetSQL.orders` O
6 | INNER JOIN `TargetSQL.customers` C
7 | USING(`customer_id`)
8 | GROUP BY 1,2
9 | ORDER BY 1,2;
```

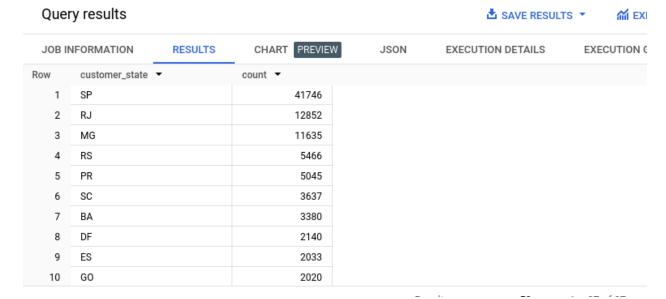
RESULT:-

Query results **≛** SAVE RESULTS ▼ JOB INFORMATION RESULTS CHART PREVIEW JSON EXECUTION DETAILS EXECUTION GRAPH customer_state ▼ month -Number Of Orders ' 1 AC 1 AC 2 6 2 3 4 3 AC AC 4 4 9 5 AC 5 10 AC 6 7 6 7 AC 7 7 8 AC 9 5 AC 10 6 10

INSIGHT:-sales is Highest SP,RJ and MG region.

B)How are the customers distributed across all the states? QUERY:-

```
1 SELECT
2 | `customer_state`,
3 | Count(DISTINCT customer_id) AS count FROM <u>`TargetSQL.customers`</u>
4 GROUP BY 1
5 ORDER BY 2
6 DESC;
```



INSIGHT:-

No. of customers in SP state is 3 times higher than the other state which is ranked at 2nd position which is also the highest in all states.

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).

QUERY:-

```
SELECT YEAR, ROUND(SUM(payment_value),2) AS `COST` FROM (
1
2
3
        EXTRACT(YEAR FROM O.order_purchase_timestamp) AS YEAR,
4
        EXTRACT(MONTH FROM O.order_purchase_timestamp) AS MONTH,
5
        P.payment_value
6
      FROM 'TargetSQL.orders' O INNER JOIN 'TargetSQL.payments' P
     USING(order_id)
7
8
9
   WHERE YEAR IN(2017, 2018) AND MONTH >=8
10 GROUP BY YEAR;
```

RESULT:-



INSIGHTS:-

Order cost in year 2018 is almost 4 times lower than 2017.

B)Calculate the Total & Average value of order price for each state. QUERY:-



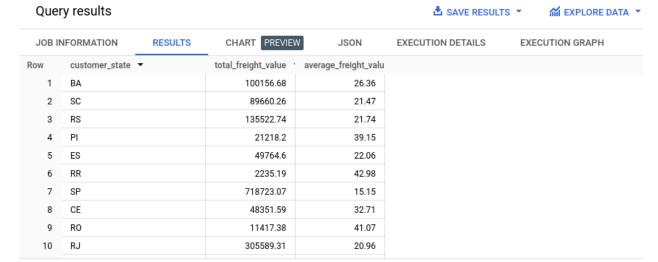
RESULT:-

Que	ry results				₫ SAVE RESULTS	S ¥		\$
JOB I	NFORMATION	RESULTS	CHART PREVIE	W JSON	EXECUTION DETAILS	EXE	CUTION GRAPH	
Row	customer_state	•	Total_Price ▼	Average_Price ▼				
1	MA		152523.02	198.86				
2	RO		60866.2	233.2				
3	AP		16262.8	232.33				
4	AC		19680.62	234.29				
5	SE		75246.25	208.44				
6	то		61485.33	204.27				
7	PI		108523.97	207.11				
8	AM		27966.93	181.6				
9	DF		355141.08	161.13				
10	MG		1872257.26	154.71				

INSIGHT:- SP region has the highest Total price to average price ratio.

C)Calculate the Total & Average value of order freight for each state. QUERY:-

```
1 SELECT
2 DISTINCT C.customer_state,
3 ROUND(SUM(OI.freight_value) OVER(partition by C.customer_state),2) AS total_freight_value,
4 ROUND(SUM(OI.freight_value) OVER (PARTITION BY C.customer_state) / COUNT(*) OVER (PARTITION BY C.customer_state),2)
AS average_freight_value
5 FROM `TargetSQL.order_items` OI
6 INNER JOIN `TargetSQL.orders' O
7 USING(order_id)
8 INNER JOIN `TargetSQL.customers' C
9 USING(customer_id);
```



INSIGHTS:-

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

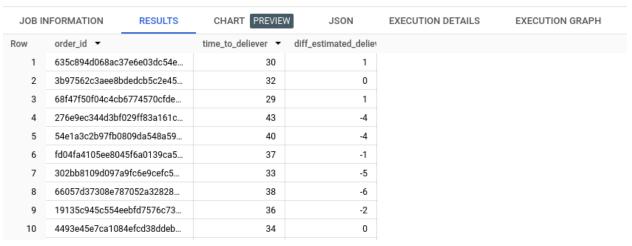
A)Find the no. of days taken to deliver each order from the order's purchase date as delivery time.

Also, calculate the difference (in days) between the estimated & actual delivery date of an order.

QUERY:-

```
1 SELECT
2     order_id,
3     TIMESTAMP_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) AS time_to_deliever,
4     TIMESTAMP_DIFF(order_estimated_delivery_date, order_delivered_customer_date, Day) AS diff_estimated_delievery
5     FROM `TargetSQL.orders`
6     WHERE order_status = 'delivered';
```

RESULT:-



INSIGHTS:-

Most of the order take around 32 days to deliver and are usually late.

B)Find out the top 5 states with the highest & lowest average freight value. QUERY:-

```
1 SELECT * FROM (
      SELECT
     DISTINCT customer_state,
 4 ROUND(SUM(OI.freight_value) OVER (PARTITION BY C.customer_state) / COUNT(*) OVER (PARTITION BY C.customer_state), 2)
    AS average_freight_value
5 FROM `TargetSQL.order_items` OI
6 INNER JOIN `TargetSQL.orders` O
     USING(order_id)
     INNER JOIN 'TargetSQL.customers' C
     USING(customer_id)
10
     ORDER BY 2
     LIMIT 5
11
12 ) A
13 UNION ALL
14 SELECT * FROM (
15
    SELECT
16 DISTINCT customer_state,
17 ROUND(SUM(OI.freight_value) OVER (PARTITION BY C.customer_state) / COUNT(*) OVER (PARTITION BY C.customer_state), 2)
    AS average_freight_value
    FROM `TargetSQL.order_items` OI
INNER JOIN `TargetSQL.orders` O
18
19
20 USING(order_id)
21
      INNER JOIN 'TargetSQL.customers' C
     USING(customer_id)
22
     ORDER BY 2 DESC
    LIMIT 5
24
25 )
```

RESULT:-

JOB II	NFORMATION	RESULTS	CHART PREVIEW	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	customer_state	-	average_freight_valu			
1	SP		15.15			
2	PR		20.53			
3	MG		20.63			
4	RJ		20.96			
5	DF		21.04			
6	PI		39.15			
7	AC		40.07			
8	RO		41.07			
9	PB		42.72			
10	RR		42.98			

INSIGHT:-The lowest average freight value is around 15 for SP whereas highest is 42.98 for RR.

C)Find out the top 5 states with the highest & lowest average delivery time. QUERY:-

RESULT:-

Que	ry results		▲ SAVE RESULT	S ▼		
JOB I	NFORMATION	RESULTS	CHART PREVIEW	JSON	EXECUTION DETAILS	EXECUTION GRAPH
Row	customer_state	•	average_delievery_tii			
1	SP		8.05			
2	PR		11.25			
3	MG		11.27			
4	DF		12.16			
5	SC		14.12			
6	PA		22.62			
7	AL		23.11			
8	AM		25.46			
9	RR		25.83			
10	AP		26.34			

INSIGHT:-

The average delivery time is lowest in 8.05 in SP whereas It is 26.34 in AP.

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

QUERY:-

```
1 SELECT customer_state FROM (
  2 SELECT
  3 DISTINCT customer_state,
  4 FL00R(
  5
      SUM(
  6 TIMESTAMP_DIFF(
       order_delivered_customer_date, order_purchase_timestamp, DAY
  8
  9 ) OVER(PARTITION BY customer_state)/COUNT(*) OVER(PARTITION BY customer_state)
 10
  11 AS actual_delievery_days,
  12 FL00R(
 13 SUM(
 14 TIMESTAMP_DIFF(
       order_estimated_delivery_date, order_purchase_timestamp, DAY
 16
 17 ) OVER(PARTITION BY customer_state)/COUNT(*) OVER(PARTITION BY customer_state)
 18 )
  19 AS estimated_delievery_days
 20 FROM 'TargetSQL.orders'
21 INNER JOIN 'TargetSQL.customers'
  22 USING(customer_id)
  23 WHERE 1 < 2 AND order_status = 'delivered'
 24 ORDER BY 1
  25
 26 LIMIT 5;
```

RESULT:-

Query results



JOB II	NFORMATION	RESULTS	CHART PREVIEW	JSON	EXECUTION DETAILS
Row	customer_state	•			
1	AC				
2	AL				
3	AM				
4	AP				
5	BA				

INSIGHT:-

These are top five states where delivery is really fast even faster than estimated delivery date.

6. Analysis based on the payments:

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

QUERY:-

```
SELECT

--EXTRACT(MONTH-FROM-0.order_purchase_timestamp)-AS-Month,payment_type,-COUNT(*)-AS-Number_of_orders

FROM--TargetSQL.payments--P-LEFT-JOIN

TargetSQL.orders--O

USING(order_id)

GROUP-BY-EXTRACT(MONTH-FROM-order_purchase_timestamp),payment_type
```



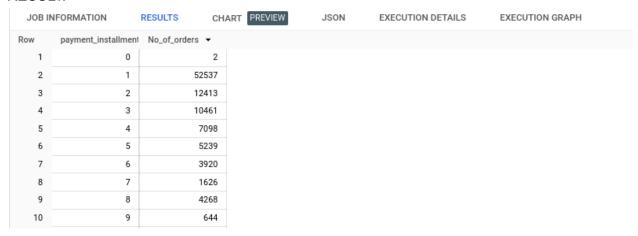
INISGHT:- Most of the orders are coming from credit card.months do not seem to have any effect on no. of orders with respect to payment types.

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

QUERY:-

1 SELECT payment_installments,COUNT(*) AS No_of_orders FROM <u>`TargetSQL.payments`</u>
2 WHERE payment_value > 0
3 GROUP BY payment_installments;

RESULT:-



INSIGHT:-

Very few people place orders on 0 EMIs and option for emi greater than 10 months.