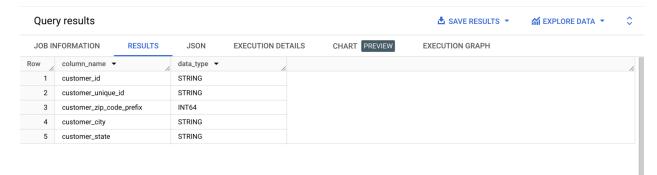
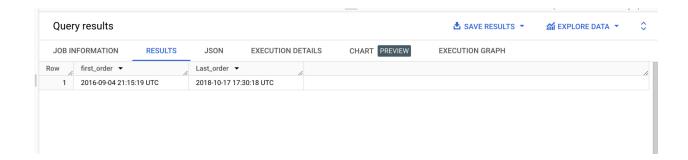
- 1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset.
  - 1.1. Data type of all columns in the "customers" table.

Ans: SELECT column\_name, data\_type
 FROM `target.INFORMATION\_SCHEMA.COLUMNS`
 WHERE table\_name = 'customers';



# INSIGHT - Here we are just understanding the type of data we are given from the source.

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



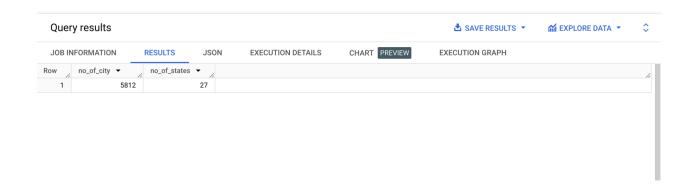
INSIGHT - Here we are understanding what is the range between customer's first order and the last placed order in the given time period.

- 1.2 Customer Support availability Businesses can align their customer support availability with the peak order times. Offering live chat, phone support, or email assistance during the hours when most orders are placed ensures that customers receive prompt assistance when they need it the most, enhancing overall customer satisfaction.
- 1.3 Count the Cities & States of customers who ordered during the given period.

#### INSIGHT -

By counting the unique cities and states, we can gain insights into the geographical reach of your customer base. Understanding where your customers are located geographically helps in tailoring marketing strategies, product offerings, and shipping options to cater to specific regions.

<u>Geographical data can be used to infer certain demographic</u> <u>information. For example, orders from densely populated urban</u> areas might indicate a younger, tech-savvy customer base. On the other hand, orders from rural areas might represent an older demographic. Understanding these demographics can help in creating targeted marketing messages

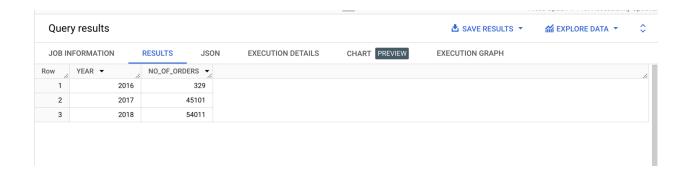


# 2. In-depth Exploration:

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

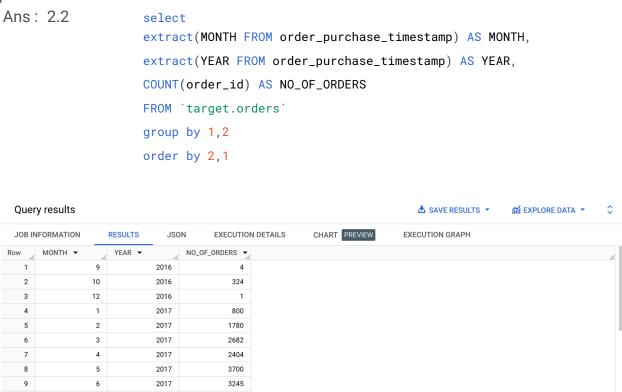
Ans: YES

```
select
extract(YEAR FROM order_purchase_timestamp) AS YEAR,
COUNT(order_id) AS NO_OF_ORDERS
FROM `target.orders`
group by 1
order by 1
```



INSIGHTS - Yes, there is a growing trend in the no. of orders placed over the past years and with growing orders we can say that the company is growing and its customers are satisfied with the offerings.

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



INSIGHTS - <u>Upon analyzing the monthly order data, it is evident that there is no discernible monthly seasonality in the number of orders being placed. The order patterns do not follow a consistent trend or exhibit any noticeable peaks or troughs on a monthly basis.</u>

4026

2017

10

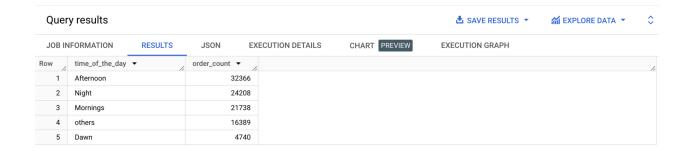
This absence of monthly seasonality indicates that external factors such as holidays, seasons, or special events do not significantly impact the number of orders placed each month.

2.3 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

### ANS: Afternoon

```
SELECT case when T.HOUR >=0 and T.HOUR<6 THEN 'Dawn'
when T.HOUR >=7 and T.HOUR<12 THEN 'Mornings'
when T.HOUR >=13 and T.HOUR<18 THEN 'Afternoon'
when T.HOUR >=19 and T.HOUR<23 THEN 'Night' else 'others'
end as time_of_the_day,
count(order_id) as order_count
FROM
(select order_id, order_purchase_timestamp,
extract(HOUR FROM order_purchase_timestamp) AS HOUR
from `target.orders`) T
group by 1
order by 2 DESC;</pre>
```



INSIGHTS - Upon analyzing the order data from Brazilian customers, it is evident that the majority of orders are placed during the afternoon hours. In the specified time categorization:

## 13:00 to 18:00 (1:00 PM to 6:00 PM): Afternoon

During this period, customers in Brazil show a consistent pattern of placing orders.
This finding suggests that the afternoon is the prime time when Brazilian customers
prefer to make their purchases. Understanding this trend is valuable for businesses, as
it enables targeted marketing efforts and optimized customer service during these
hours to enhance the overall customer experience and potentially increase sales.

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

```
ANS:

SELECT g.geolocation_state,

extract(MONTH FROM order_purchase_timestamp) AS MONTH,

extract(YEAR FROM order_purchase_timestamp) AS YEAR,

count(o.order_id) as no_of_orders

from `target-401008.target.geolocation` g

join `target-401008.target.customers` c

on g.geolocation_zip_code_prefix = c.customer_zip_code_prefix

join `target-401008.target.orders` o

on o.customer_id = c.customer_id

group by 1,2,3

order by 3,2
```

Quer	ry results					<b>\$</b>
JOB IN	NFORMATION RESULTS	JSON EX	EXECUTION DETAILS CHART PREVIEW		EXECUTION GRAPH	
Row	geolocation_state ▼	MONTH ▼	YEAR ▼	no_of_orders ▼		1
1	RR	9	2016	65		
2	RS	9	2016	103		
3	SP	9	2016	492		
4	SE	10	2016	65		
5	AL	10	2016	52		
6	PI	10	2016	56		
7	RR	10	2016	65		
8	ВА	10	2016	292		
9	CE	10	2016	477		
10	DF	10	2016	305		

#### **INSIGHTS** -

# **Regional Order Volume Trends:**

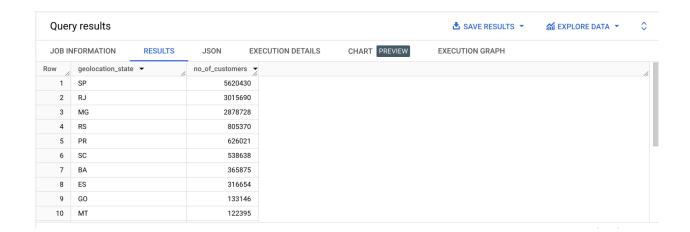
By analyzing the month-on-month orders placed in each state, we have identified significant variations in order volumes across regions. Some states exhibit consistent growth, indicating a robust e-commerce market and potential business opportunities. Conversely, fluctuations or declines in certain states may signify market challenges or changing consumer preferences, demanding targeted strategies for revival.

Studying orders on a month-on-month basis can reveal operational challenges in specific states. Prolonged delivery times or supply chain issues might lead to customer dissatisfaction. Identifying these challenges enables businesses to optimize logistics and streamline operations, ensuring timely deliveries and enhancing customer experience.

3.2 How are the customers distributed across all the states?

```
ANS: Select g.geolocation_state,
```

```
count(c.customer_unique_id) as no_of_customers
from `target-401008.target.geolocation` g
join `target-401008.target.customers` c
on g.geolocation_zip_code_prefix = c.customer_zip_code_prefix
group by g.geolocation_state
order by 2 desc;
```



INSIGHTS - Upon analyzing the customer data across states, a clear pattern emerges: the majority of customers are concentrated in only three states. This concentration indicates that a significant portion of the customer base is localized in specific regions, while other states have comparatively fewer customers.

This finding has significant implications for business strategies and operations.

Focusing marketing efforts, promotional activities, and customer engagement initiatives on these three states can yield higher returns on investment. By tailoring products, services, and marketing campaigns to cater specifically to the preferences and needs of customers in these states, the business can enhance customer satisfaction and loyalty within its primary customer base.

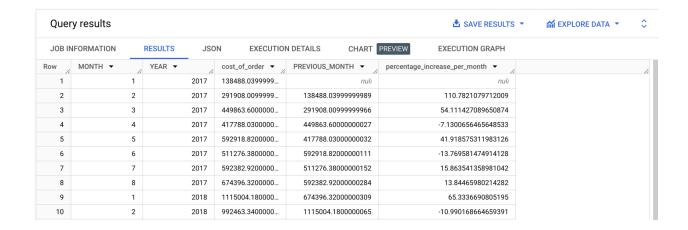
- 4. Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.
  - 4.1 Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only).

You can use the "payment\_value" column in the payments table to get the cost of orders.

```
Ans. with cte as (select extract(MONTH FROM order_purchase_timestamp) AS MONTH,
    extract(YEAR FROM order_purchase_timestamp) AS YEAR, sum(p.payment_value) as

cost_of_order
    from `target.payments` p
    join `target.orders` o
    on o.order_id = p.order_id
    group by 1,2
    order by 2,1)
```

```
select MONTH, YEAR, cost_of_order,
lag(cost_of_order,1) over(ORDER BY YEAR,MONTH) AS PREVIOUS_MONTH,
(cost_of_order- lag(cost_of_order,1) over(ORDER BY YEAR,MONTH))/
lag(cost_of_order,1) over(ORDER BY YEAR,MONTH)*100 as percentage_increase_per_month
    from cte
    where cte.MONTH >=1 and cte.MONTH <=8
    ORDER BY 2,1</pre>
```



# **INSIGHTS** -

#### **Economic Growth and Consumer Confidence:**

An increase in e-commerce spending from 2017 to 2018 indicates a robust economic environment and consumer confidence. When consumers spend more online, it often signifies increased trust in the economy, stable income levels, and positive sentiment about future prospects.

# **E-commerce Market Expansion:**

The significant growth in e-commerce spending suggests the e-commerce market is expanding. More consumers are embracing online shopping, which can be attributed to factors like improved internet infrastructure, widespread smartphone usage, and the convenience offered by online shopping platforms.

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

```
ANS. select g.geolocation_state, SUM(ot.price) As Total_price,

AVG(ot.price) as Average_price

from `target.order_items` ot

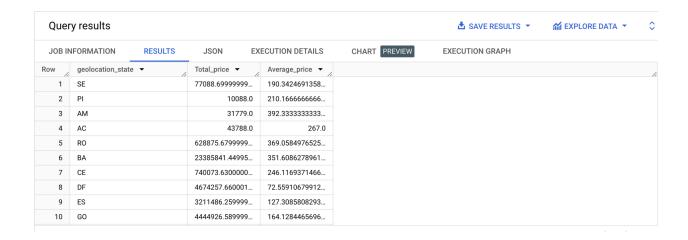
left join `target.sellers` s

on s.seller_id = ot.seller_id

join `target.geolocation` g

on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix

group by 1
```



INSIGHTS - Total Order Price: State SE has the highest total order price, indicating a high volume of sales. States PI and AM also show substantial total order prices, reflecting a significant economic impact in these regions.

Average Order Price: While the State SE has the highest total order price, it's crucial to note the average order price.

States with higher average order prices might signify a more affluent customer base willing to spend more on products, whereas states with lower average order prices may have more price-sensitive customers.

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

```
Ans. select g.geolocation_state, SUM(ot.freight_value) As Total_value,

AVG(ot.freight_value) as Average_value

from `target.order_items` ot

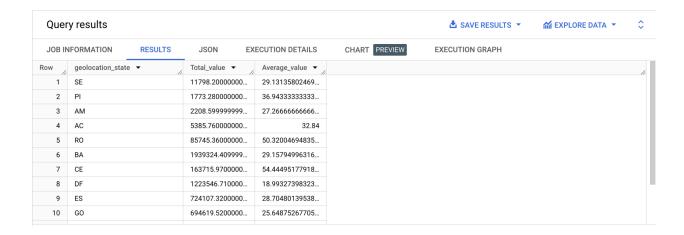
left join `target.sellers` s

on s.seller_id = ot.seller_id

join `target.geolocation` g

on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix

group by 1
```



INSIGHTS - Total Order Freight: States with higher total order freight costs might indicate larger distances between the seller's warehouse and customers' locations.

This information is crucial for logistics and supply chain management, helping businesses optimize shipping routes and reduce transportation costs.

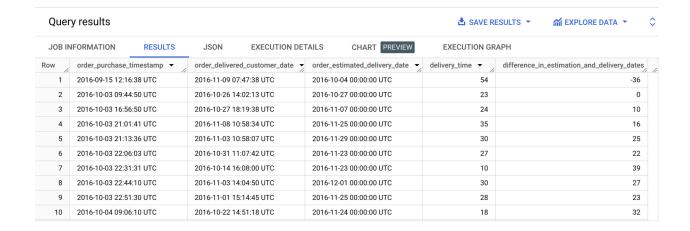
Average Order Freight: Analyzing the average order freight provides insights into the shipping preferences of customers in different states. States with higher average order freight costs might have customers willing to pay more for expedited or specialized shipping services. On the other hand, states with lower average order freight costs might indicate a preference for standard or economical shipping options

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

5.1 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.



## **INSIGHTS - Estimated vs. Actual Delivery Comparison:**

Comparing estimated and actual delivery dates helps in evaluating the accuracy of delivery estimations. A small difference between these dates suggests accurate predictions, while a significant difference might indicate challenges in forecasting delivery times accurately. Understanding these differences can improve customer expectations management and prevent potential dissatisfaction.

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

```
Ans-with cte as

(select g.geolocation_state, avg(freight_value)

as average_freight_value

from `target.order_items` ot

left join `target.sellers` s

on s.seller_id = ot.seller_id

join `target.geolocation` g

on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix

group by 1)

(select geolocation_state, average_freight_value)
```

```
from cte
order by cte.average_freight_value DESC
limit 5)

UNION ALL

(select geolocation_state, average_freight_value from cte
order by cte.average_freight_value ASC
limit 5)
```

sults						
				▲ SAVE RESULTS ▼		<b>\$</b>
MATION RESULTS	JSON EX	ECUTION DETAILS	CHART PREVIEW	EXECUTION GRAPH		
location_state ▼	average_freight_valu					//
	54.44495177918					
	50.32004694835					
	36.94333333333					
	34.69409917355					
	32.84					
	15.93187766357					
	18.43661245059					
	18.93235169996					
	18.99327398323					
	22.10730133627					
	ocation_state ▼ //	54.44495177918 50.32004694835 36.9433333333 34.69409917355 32.84 15.93187766357 18.43661245059 18.93235169996 18.99327398323	54.44495177918 50.32004694835 36.9433333333 34.69409917355 32.84 15.93187766357 18.43661245059 18.93235169996 18.99327398323	54.44495177918 50.32004694835 36.9433333333 34.69409917355 32.84 15.93187766357 18.43661245059 18.93235169996 18.99327398323	54.44495177918 50.32004694835 36.94333333333 34.69409917355 32.84 15.93187766357 18.43661245059 18.93235169996 18.99327398323	54.44495177918 50.32004694835 36.94333333333 34.69409917355 32.84 15.93187766357 18.43661245059 18.93235169996 18.99327398323

INSIGHTS - State CE has the highest average freight value. This might indicate that products delivered to this state are bulkier, heavier, or located far from the distribution centers, leading to higher shipping costs. Businesses operating in this state might consider optimizing their supply chain and exploring local distribution centers to reduce shipping expenses.

<u>State AC rounds up the top 5 list. Exploring partnerships with</u> regional carriers or investing in technology for route

optimization can help reduce freight costs. Businesses might
also consider adjusting pricing strategies to accommodate higher
shipping costs in this state.

State RN has the lowest average freight value. Analyzing the reasons behind low shipping costs in this state can provide valuable insights. Efficient transportation networks, proximity to distribution centers, or strategic partnerships with carriers could be contributing factors. Businesses can leverage these advantages to offer competitive pricing and attract more customers.

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

```
Ans-with dte as

(SELECT g.geolocation_state,

avg(TIMESTAMP_DIFF(order_delivered_customer_date,

order_purchase_timestamp,DAY)) as avg_delivery_time,

from `target.orders` o

join `target.customers` c

on o.customer_id = c.customer_id

join `target.geolocation` g

on c.customer_zip_code_prefix = g.geolocation_zip_code_prefix

where order_delivered_customer_date is not null

group by 1

)

(select geolocation_state, dte.avg_delivery_time

from dte

order by dte.avg_delivery_time desc

limit 5)
```

#### union all

```
(select geolocation_state, dte.avg_delivery_time
from dte
order by dte.avg_delivery_time asc
limit 5)
```

Quer	ry results					<b>™</b> EXPLORE DATA ▼	<b>\$</b>
JOB IN	NFORMATION RESULTS	JSON EX	ECUTION DETAILS	CHART PREVIEW	EXECUTION GRAPH		
Row	geolocation_state ▼	avg_delivery_time					//
1	AP	27.99122623772					
2	AM	24.65119678421					
3	RR	24.52060133630					
4	AL	23.14352789271					
5	PA	22.55023982441					
6	SP	8.470529714190					
7	PR	11.03876404770					
8	MG	11.41862683439					
9	DF	12.49651789233					
10	SC	14.49430832817					

INSIGHTS - State AP has the highest average delivery time. This could be due to various reasons such as logistical challenges, lack of efficient transportation infrastructure, or high order volumes in this state. Businesses should investigate the underlying causes and consider optimizing their supply chain processes and exploring local distribution centers to reduce delivery times.

State SP has the lowest average delivery time. Efficient logistics, strategic placement of distribution centers, and partnerships with reliable carriers might contribute to the quick deliveries in this state. Businesses can capitalize on this advantage by promoting fast and reliable shipping options to attract customers looking for timely deliveries.

5.4 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.

```
Ans-with ata as(
             SELECT g.geolocation_state,
             avg(TIMESTAMP_DIFF(order_delivered_customer_date,
      order_purchase_timestamp, DAY)) as avg_delivery_time,
             avg(TIMESTAMP_DIFF(order_estimated_delivery_date,
      order_delivered_customer_date, DAY)) as estimated_time
             from `target.orders` o
             join `target.customers` c
             on o.customer_id = c.customer_id
             join `target.geolocation` g
             on c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
             where order_delivered_customer_date is not null
             group by 1)
             select geolocation_state, avg_delivery_time,estimated_time,
      (estimated_time -avg_delivery_time ) as prior_est_delivery
             from ata
             order by estimated_time -avg_delivery_time desc
             limit 5
```

, ομιιοι	ess optioner i for Accessibility	г	_					
\$		<b>≛</b> SAVE RESULTS ▼					y results	Quer
		EXECUTION GRAPH	CHART PREVIEW	ECUTION DETAILS	JSON EX	RESULTS	NFORMATION	JOB IN
/			prior_est_delivery y	estimated_time 🕶	avg_delivery_time	• ▼	geolocation_state	Row
			1.760418071575	10.23094778576	8.470529714190		SP	1
			1.531212663053	12.56997671075	11.03876404770		PR	2
			0.929987425172	12.34861425956	11.41862683439		MG	3
			-0.00240101920	18.65209721677	18.65449823598		RO	4
			-1.23642197012	11.26009592221	12.49651789233		DF	5

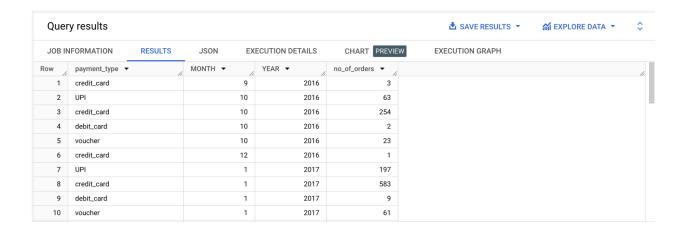
INSIGHTS - State Sp Orders in this state are consistently

delivered much faster than customers anticipate. Understanding
the reasons behind this efficiency, such as optimized logistics
networks or strategic partnerships with local carriers, can
serve as a model for other regions

State Pr follows closely with the second-smallest average difference. Businesses should investigate the practices and processes in this state that contribute to these rapid deliveries. By replicating successful strategies, other states can improve their delivery efficiency and enhance customer satisfaction.

- 6. Analysis based on the payments:
  - 6.1 Find the month on month no. of orders placed using different payment types.

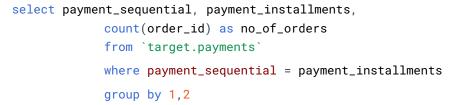
```
ANS select p.payment_type,
    extract(MONTH FROM order_purchase_timestamp) AS MONTH,
    extract(YEAR FROM order_purchase_timestamp) AS YEAR,
    count(o.order_id) as no_of_orders
    from `target.payments` p
    left join `target.orders` o
    on o.order_id = p.order_id
    group by 1,2,3
    order by 3,2,1
```

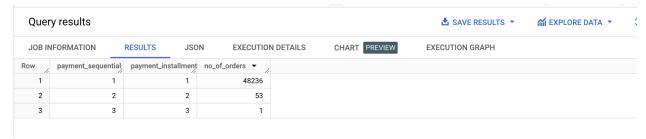


INSIGHTS - Payment Type Preferences: By analyzing the month-on-month trends, you can identify which payment types are consistently popular among customers. For example, credit card payments might show a steady increase over time, indicating a growing trust in online transactions. Understanding these preferences allows businesses to optimize their payment gateway options and user experience

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

#### ANS:





INSIGHTS - Competitive Advantage: Offering flexible payment installment options can serve as a competitive advantage. Businesses can differentiate themselves in the

market by providing hassle-free, transparent, and user-friendly installment plans, attracting customers who prioritize convenient payment methods.