1. **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.

Answer –

```
SELECT column_name,data_type
from
Business_case_submit.INFORMATION_SCHEMA.COLUMNS
where
table_name = 'customers';
```

Screenshot –

Row	column_name	data_type
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

Insight – Data Type customers table will help to analysis and interpretation of data
Recommendation- NA

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

Answer –

```
SELECT min(order_purchase_timestamp) as min,
max(order_purchase_timestamp) as max
FROM `sclaer-dsml-sql-402312.Business_case_submit.orders`
```

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION

Row	min	max
1	2016-09-04 21:15:19 UTC	2018-10-17 17:30:18 UTC

Insights – orders has been placed from 4th sept 2016 to 17th oct 2018
Recommendation - NA

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

Answer –

```
SELECT count(distinct c.customer_city) as count_city,
count(distinct c.customer_state) as count_state FROM Business_case_submit.customers
c
join Business_case_submit.orders o
on c.customer_id = o.customer_id
```

Query results

SAVE RESULT

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS

Row	count_city	count_state
1	4119	27

Insights – 419 city and 27 state orders during a given period.
Recommendation- NA

2. **In-depth Exploration:**

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

Answer-

```
with extracted as
(SELECT extract(MONTH from order_purchase_timestamp) as month,
extract(year from order_purchase_timestamp) as Year,order_id
FROM Business_case_submit.orders)
select month,year,count(distinct order_id) as total_count
from extracted
group by 1,2
order by 1,2
```

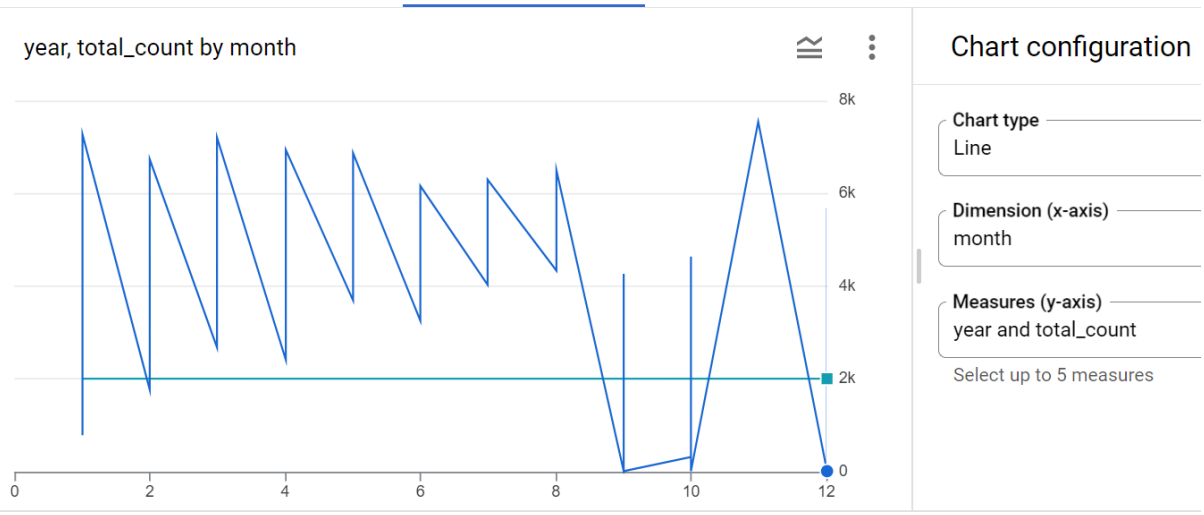
JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS

Row	month	year	total_count
1	1	2017	800
2	1	2018	7269
3	2	2017	1780
4	2	2018	6728
5	3	2017	2682
6	3	2018	7211
7	4	2017	2404
8	4	2018	6939
9	5	2017	3700

Results per page: 50 1 -

Query results SAVE RESULTS EXPLO

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION G



Insights –
1st and 12th month of year having higher order due to festival .also from 2018 has good order compare to 2017
Recommendation -
We can club lower sale item with higher sale item in the above given period so that those product will get ordered.

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

Answer –

```sql
with seasonal as
(SELECT
  EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
  EXTRACT(Year FROM order_purchase_timestamp) AS Year,
  format_datetime('%B',order_purchase_timestamp) as month_name,
  COUNT(DISTINCT order_id) AS total_order
FROM
 Business_case_submit.orders o
GROUP BY 1,2,3)
select year,month,total_order,
LAG(total_order,2) over(order by month) before_month,
LAG(total_order,1) over(order by month) after_month,
Lead(total_order,1) over(order by month) second_month,
Lead(total_order,2) over(order by month) third_month,
from seasonal
order by
  month desc
limit 10
```

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	year	month	total_order	before_month	after_month	second_month	third_month
1	2017	12	5673	4	7544	1	nu
2	2016	12	1	7544	5673	null	nu
3	2017	11	7544	324	4	5673	
4	2017	10	4631	4	16	324	
5	2016	10	324	16	4631	4	754
6	2018	10	4	4631	324	7544	567
7	2017	9	4285	4331	6512	4	1
8	2018	9	16	4285	4	4631	32
9	2016	9	4	6512	4285	16	463
10	2018	8	6512	6292	4331	4285	

Insight - The count of orders generally increases from March to August with fluctuations in between. Notably, there is an increase in orders during February and March

Recommendation – Need some more data or larger data to validate the trends.may be with revenue data

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

Answer –

```sql
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'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 Hour,count(order_id) as total_count
from Business_case_submit.orders o join Business_case_submit.customers c
on o.customer_id = c.customer_id
group by Hour
order by total_count desc
```

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETA

Row	Hour	total_count
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Insight – Seeing the data Brazilian are more active in ecommerce afternoon and night time .Probably they are wrapping up their work and getting leisure time to buy products

Recommendation – Optimize their operation ,identify peak time product buying company can allocate more resource to provide seamless shopping experience

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

Answer –

```sql
select c.customer_state,

extract(month from order_purchase_timestamp) as month,
count(order_id) as total_count
from Business_case_submit.orders o join Business_case_submit.customers c
on o.customer_id = c.customer_id
group by customer_state,month
order by customer_state,month
```

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTI

Row	customer_state	month	total_count
1	AC	1	8
2	AC	2	6
3	AC	3	4
4	AC	4	9
5	AC	5	10
6	AC	6	7
7	AC	7	9
8	AC	8	7
9	AC	9	5
10	AC	10	6

Insight – SP is the highest number of order followed by RJ and MG.It might be people from this region has high paying salary or larger population

Recommendation – we can market more about this ecommerce on lower number of order state ,like billboard with mentioned discount or notification to their phone about any good discount

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

Answer –

```
select customer_state,
count(customer_id) as total_customer
from  Business_case_submit.customers
group by customer_state
order by total_customer desc
```

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	customer_state	total_customer
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020

Insight – SP has the largest ecommerce buying population followed by RJ AND MG

Recommendation – Brand can gain a competitive edge,drive growth for these state so they can deliver quality product and improve customer experience

4.Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

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.

Answer –

```
with cte1 as
(
  SELECT
  extract(month from order_purchase_timestamp) as month,
  round(sum(payment_value),2) as sales_2017 from
  Business_case_submit.orders JOIN Business_case_submit.payments USING(order_id)
  where extract(year from order_purchase_timestamp) = 2017
  and extract(month from order_purchase_timestamp) between 1 and 8
  group by month
  order by month
),
cte2 as
(
  SELECT
  extract(month from order_purchase_timestamp) as month,
  round(sum(payment_value),2) as sales_2018
  FROM Business_case_submit.orders JOIN Business_case_submit.payments
USING(order_id)
  where extract(year from order_purchase_timestamp) = 2018
  and extract(month from order_purchase_timestamp) between 1 and 8
  group by month
  order by month
)

select
month,
sales_2017,
sales_2018,((sales_2018-sales_2017)/sales_2017)*100 as sales_percentage
from cte1 join cte2 using(month)
order by month
```

JOB INFORMATION | RESULTS | CHART | PREVIEW | JSON | EXECUTION DE

Row	month	sales_2017	sales_2018	sales_percentage
1	1	138488.04	1115004.18	705.1266954171...
2	2	291908.01	992463.34	239.9918145445...
3	3	449863.6	1159652.12	157.7786066709...
4	4	417788.03	1160785.48	177.8407701149...
5	5	592918.82	1153982.15	94.62734375677...
6	6	511276.38	1023880.5	100.2596912456...
7	7	592382.92	1066540.75	80.04245463390...
8	8	674396.32	1022425.32	51.60600520477...

Insight – January has the highest no of increase order .May be due to festival or start of the year
Recommendation – Ecommerce can provide more discount or sale on others month to get a good profit

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

Answer –

```
select
round(sum(price),2) Total_price,round(avg(price),2) Avg_price,c.customer_state
from Business_case_submit.orders o join Business_case_submit.order_items oi
on o.order_id = oi.order_id
join Business_case_submit.customers c
on c.customer_id = o.customer_id
group by customer_state
order by total_price desc
```

Query results

SAVE RESULTS

JOB INFORMATION | RESULTS | CHART PREVIEW | JSON | EXECUTION DETAILS | EXEC

Row	Total_price	Avg_price	customer_state
1	5202955.05	109.65	SP
2	1824092.67	125.12	RJ
3	1585308.03	120.75	MG
4	750304.02	120.34	RS
5	683083.76	119.0	PR
6	520553.34	124.65	SC
7	511349.99	134.6	BA
8	302603.94	125.77	DF
9	294591.95	126.27	GO
10	275037.31	121.91	ES

Insight – As per our previous data analysis SP is the highest state total price value but lowest average price product .
Mean while GO,DF,RJ are the highest state average price value

Recommendation – Improvement can be Price strategies, enhance logistics

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

Answer –

```
select
round(sum(freight_value),2) Total_Freight,round(avg(freight_value),2)
Avg_Freight,c.customer_state
from Business_case_submit.orders o join Business_case_submit.order_items oi
on o.order_id = oi.order_id
join Business_case_submit.customers c
on c.customer_id = o.customer_id
group by customer_state
order by total_Freight desc
```

Row	Total_Freight	Avg_Freight	customer_state
1	718723.07	15.15	SP
2	305589.31	20.96	RJ
3	270853.46	20.63	MG
4	135522.74	21.74	RS
5	117851.68	20.53	PR
6	100156.68	26.36	BA
7	89660.26	21.47	SC
8	59449.66	32.92	PE
9	53114.98	22.77	GO
10	50625.5	21.04	DF

Insight –

The analysis reveals interesting findings. While SP has the highest total freight value, it surprisingly has the lowest average freight value among all states and RR,PB,RO has the highest average freight

Recommendation- Improvement can be Price strategies, enhance logistics

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

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.

   You can calculate the delivery time and the difference between the estimated & actual delivery date using the given formula:
   - **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_customer_date,order_purchase_timestamp,day) as
time_to_delivery,
date_diff(order_estimated_delivery_date,order_delivered_customer_date,day) as
diff_estimated_delivery
 from Business_case_submit.orders
where date_diff(order_delivered_customer_date,order_purchase_timestamp,day) is not
null
order by time_to_delivery
```

Row	order_id	time_to_delivery	diff_estimated_deliv
1	e65f1eeee1f52024ad1dcd034...	0	9
2	bb5a519e352b45b714192a02f...	0	25
3	434cecee7d1a65fc65358a632...	0	19
4	d3ca7b82c922817b06e5ca211...	0	11
5	1d893dd7ca5f77ebf5f59f0d20...	0	10
6	d5fbeedc85190ba88580d6f82...	0	7
7	79e324907160caea526fd8b94...	0	8
8	38c1e3d4ed6a13cd0cf612d4c...	0	16
9	8339b608be0d84fca9d8da68b...	0	27
10	f349cdb62f69c3fae5c4d7d3f3...	0	12

Insight – Most delivery is on that day for some particular product,need larger dataset to get to know which state has lesser delivery time

Recommendation - NA

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

Answer –

```
SELECT customer_state,avg_freight_value,
case when t.max <=5 Then 'Max_Freight'
when t.min<=5 then 'Min_Frieght'
end as freight_range from (
select c.customer_state,ROUND(AVG(od.freight_value),2) as avg_freight_value,
dense_rank() over(order by ROUND( AVG(od.freight_value),2) desc) as max,
dense_rank() over(order by ROUND( AVG(od.freight_value),2)) as min,
FROM Business_case_submit.order_items od JOIN Business_case_submit.orders o
ON od.order_id = o.order_id
JOIN Business_case_submit.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state) as t
where t.max <=5 or t.min<=5 order by avg_freight_value desc
```

Row	customer_state	avg_freight_value	freight_range
1	RR	42.98	Max_Freight
2	PB	42.72	Max_Freight
3	RO	41.07	Max_Freight
4	AC	40.07	Max_Freight
5	PI	39.15	Max_Freight
6	DF	21.04	Min_Frieght
7	RJ	20.96	Min_Frieght
8	MG	20.63	Min_Frieght
9	PR	20.53	Min_Frieght
10	SP	15.15	Min_Frieght

Insight –
RR is highest freight value and SP is the lowest
Recommendation –
By leveraging SQL queries and analyzing state-wise patterns, companies can gain valuable insights into sales trends, optimize their logistics, and make data-driven decisions to enhance their overall efficiency and customer experience.

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

Answer-

```
SELECT customer_state,avg_time_to_delivery,
case when t.max <=5 Then 'Max_delivery_time'
when t.min<=5 then 'Min_delivery_time'
end as time_to_delivery_range from (
select c.customer_state,
ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp,
DAY)), 2)
  AS avg_time_to_delivery,
dense_rank() over(order by ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)), 2) desc) as max,
dense_rank() over(order by ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)), 2)asc)  as min,
FROM Business_case_submit.order_items od JOIN Business_case_submit.orders o
ON od.order_id = o.order_id
JOIN Business_case_submit.customers c on o.customer_id = c.customer_id
GROUP BY c.customer_state) as t
where t.max <=5 or t.min<=5 order by avg_time_to_delivery desc
```

Row	customer_state	avg_time_to_delivery	time_to_delivery_range
1	RR	27.83	Max_delivery_time
2	AP	27.75	Max_delivery_time
3	AM	25.96	Max_delivery_time
4	AL	23.99	Max_delivery_time
5	PA	23.3	Max_delivery_time
6	SC	14.52	Min_delivery_time
7	DF	12.5	Min_delivery_time
8	MG	11.52	Min_delivery_time
9	PR	11.48	Min_delivery_time
10	SP	8.26	Min_delivery_time

Insight – RR has the max delivery time and SP has the min delivery time

Recommendation – 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

2. 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.

Answer –

```
WITH delivery_data AS
(SELECT ord.order_id AS order_id,
DATETIME_DIFF(EXTRACT(date FROM ord.order_estimated_delivery_date),
EXTRACT(date FROM ord.order_delivered_customer_date), day)  AS
diff_estimated_delivery
FROM Business_case_submit.orders ord)
SELECT cust.customer_state,ROUND( AVG(diff_estimated_delivery),2)
mean_diff_estimated_delivery
FROM Business_case_submit.order_items ordit
JOIN Business_case_submit.orders ord ON ordit.order_id = ord.order_id
 JOIN Business_case_submit.customers cust ON ord.customer_id = cust.customer_id
JOIN delivery_data dvd ON ord.order_id = dvd.order_id
GROUP BY customer_state
ORDER BY mean_diff_estimated_delivery
LIMIT 5
```

Row	customer_state	mean_diff_estimated
1	AL	8.74
2	MA	9.91
3	SE	10.0
4	ES	10.65
5	BA	10.98

Insight – AL is the fastest delivery state may be it's a big city and demand is high
And BA is the 5th highest

Recommendation – Transportation and supply chain we can more leverage on other state to gain more valuable customer experience with more buying product

6.Analysis based on the payments:

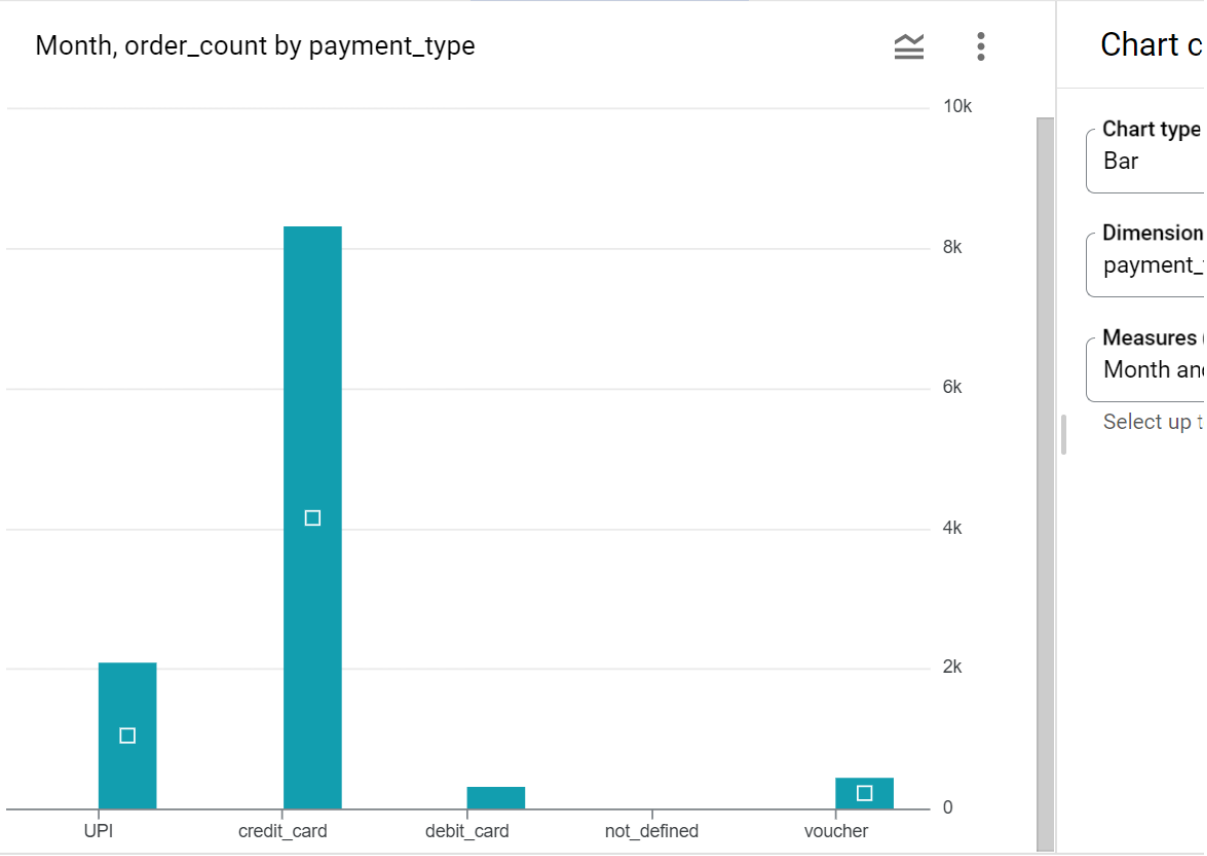
3. Find the month on month no. of orders placed using different payment types.

Answer –

```
extract(MONTH FROM o.order_purchase_timestamp) as Month,
count(distinct o.order_id) as order_count
from Business_case_submit.orders o
join Business_case_submit.payments p
on o.order_id = p.order_id
group by 1,2
ORDER BY 1,2
```

Row	payment_type	Month	order_count
1	UPI	1	1715
2	UPI	2	1723
3	UPI	3	1942
4	UPI	4	1783
5	UPI	5	2035
6	UPI	6	1807
7	UPI	7	2074
8	UPI	8	2077
9	UPI	9	903
10	UPI	10	1056

Results per page: 50 1 – 5



Job history

Insight – Uptrend can see from Jan to august and then september to Novemeber People mostly use Credit card followed by upi and debit card ,may be due credit card point and easy transaction through UPI

Recommendation –

1- **Ecommerce can give additional discount while using those credit card and debit card**
2- **On month sept and oct we can arrange small sales and clubbing with low buying product with high buying product**

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

Answer-

```
select p.payment_installments,

count(o.order_id) as order_count
from Business_case_submit.orders o
join Business_case_submit.payments p
on o.order_id = p.order_id
where o.order_status != 'canceled'
group by 1 ORDER BY 2 desc
```

Row	payment_installment	order_count
1	1	52184
2	2	12353
3	3	10392
4	4	7056
5	10	5292
6	5	5209
7	8	4239
8	6	3898
9	7	1620
10	9	638
11	12	133
12	15	74
13	18	27

Results pe

Job history

Insight – The analysis reveal that majority of order which is max have only 1 payment instalment ,Highest number of instalment is 24 with 18 order

Recommendation – Company can improve payments option, streamline processes and enhance customer experience