Ques1: 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.

SCHEMA | DETAILS | PREVIEW | TABLE EXPLORER PREVIEW

Filter Enter property name or value

Field name	Type	Mode	Key	Collatior
customer_id	STRING	NULLABLE	-	-
customer_unique_id	STRING	NULLABLE	-	-
customer_zip_code_prefix	INTEGER	NULLABLE	-	-
customer_city	STRING	NULLABLE	-	-
customer_state	STRING	NULLABLE	-	-

Ques: /*Get the time range between which the orders were placed */

```
SELECT min(time(order_purchase_timestamp)) as min_time,max(time(order_purchase_timestamp)) as max_time  FROM
`scalar_set.customers` ckp1
inner join `scalar_set.orders` ckp2 on  ckp1.customer_id=ckp2.customer_id;
```

Press Alt+F1 for Accessibilit

ıery results

SAVE RESULTS | OPEN IN

B INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

	min_time	max_time
1	00:00:00	23:59:59

```sql
/*Count the Cities & States of customers who ordered during the given period.*/
WITH min_max_date as

(SELECT min(time(order_purchase_timestamp)) as min_time,max(time(order_purchase_times
`scalar_set.customers` ckp1
inner join `scalar_set.orders` ckp2 on  ckp1.customer_id=ckp2.customer_id),
details as ( SELECT * FROM `scalar_set.customers` ckp1
 inner join `scalar_set.orders` ckp2 on  ckp1.customer_id=ckp2.customer_id)

SELECT count(DISTINCT customer_state) as state_cnt , count(DISTINCT customer_city ) a
from details where time(order_purchase_timestamp)
between (select min_time from min_max_date) and (select max_time from min_max_date);
```

uery results

SAVE RESU

)B INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECU

	state_cnt	city_cnt
1	27	4119

Question2: In-depth Exploration:

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

```sql
SELECT extract(year from order_purchase_timestamp) as year, count(order_id) as order_cnt
from `scalar_set.orders` group by year order by 1;
```

Row	year	order_cnt
1	2016	329
2	2017	45101
3	2018	54011

Yes, the number of orders has steadily increased each year, indicating growing demand. To capitalize on this, consider optimizing inventory, scaling customer support, and targeting peak periods with tailored marketing strategies to further drive growth

Ques: #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(order_id) as order_cnt
from `scalar_set.orders` group by 1 order by 1 ;
```

Row	month	order_cnt
1	8	10843
2	5	10573
3	7	10318
4	3	9893
5	6	9412
6	4	9343

Yes, there is noticeable seasonality, with higher order volumes in May, June, July, and August. To leverage this, consider running targeted promotions, optimizing inventory for peak demand, and planning marketing campaigns ahead of these months for maximum impact.

Ques: 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 order_purchase_timestamp)  >= 1 )and (extract(hour from
order_purchase_timestamp)  <= 6) then "Dawn"
WHEN extract(hour from order_purchase_timestamp)  >= 7 and extract(hour from order_purchase_timestamp) <=12 then
"Morning"
when extract(hour from order_purchase_timestamp)  >= 13 and extract(hour from order_purchase_timestamp)  <= 18
then "Afternoon"
when extract(hour from order_purchase_timestamp) between 19 and  23 then "Night" end as time_day,count(order_id)
as order_cnt
from `scalar_set.orders` group by time_day order by 1;
```

Insight: Brazilian customers primarily place orders in the afternoon, followed by night and morning, with dawn seeing the least activity. To optimize sales, consider scheduling targeted promotions in the afternoon and night, while adjusting staffing and inventory accordingly.

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

```
42  #Get the month on month no. of orders placed in each state.
43  SELECT ckp1.customer_state,extract(month from order_purchase_timestamp) as month_order,count(order_id) as
    order_cnt FROM `scalar_set.customers` ckp1
44  inner join `scalar_set.orders` ckp2 on  ckp1.customer_id=ckp2.customer_id group by 1,2 order by 1,2;
45
```

Press Alt+F1 for Accessi

Query results

SAVE RESULTS | OPEN IN

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

Row	customer_state	month_order	order_cnt
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

#How are the customers distributed across all the states?

```
46  #How are the customers distributed across all the states?
47  SELECT customer_state,count(customer_id) as order_cnt FROM `scalar_set.customers`  group by 1;
```

Press Alt+F

Query results

SAVE RESULTS

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

Row	customer_state	order_cnt
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

Let’s have more insight:

```
#Evolution of E-commerce orders in the Brazil region:
SELECT sum(CASE when t.customer_state IN ("SP","RJ","MG") then order_cnt else 0 end)/sum(order_cnt) from (SELECT
ckp1.customer_state,count(order_id) as order_cnt FROM `scalar_set.customers` ckp1
inner join `scalar_set.orders` ckp2 on  ckp1.customer_id=ckp2.customer_id group by 1 order by order_cnt desc)t;
```

Press Alt+F1 for Accessibility Opt

Query results

SAVE RESULTS | OPEN IN

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

ow	f0_
1	0.666053237598…

Insight: top three states (SP,RJ,MG) contribute to 66.7% of sales in the Brazil region in terms of order placed. We should focus on increasing customer support and inventory supply in these states

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

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

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

SELECT round(((sum(if (substr(order_date,1,4)="2018",cost,0)) - sum(if (substr(order_date,1,4)="2017",cost,0)))/
sum(if (substr(order_date,1,4)="2017",cost,0)))*100,2) as percentage_inc from

(select format_date("%Y-%m", order_purchase_timestamp) as order_date,sum(payment_value) as cost
from `scalar_set.payments` ckp1 inner join `scalar_set.orders` ckp2 on
ckp1.order_id=ckp2.order_id group by 1 having (order_date between "2018-01" and "2018-08") or
(order_date between "2017-01" and "2017-08") order by 1) t;
```

Press Alt+F1 for Accessibility O

Query results

SAVE RESULTS | OPEN IN

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

ow	percentage_inc
1	136.98

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

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

select customer_state,round(sum(payment_value),2) as total,round(avg(payment_value),2) as avg_value
from `scalar_set.payments` ckp1 inner join `scalar_set.orders` ckp2 on
ckp1.order_id=ckp2.order_id inner join `scalar_set.customers` ckp3
on ckp3.customer_id=ckp2.customer_id group by customer_state order by total desc;
```

Press Alt+F1 for Ac

Query results

SAVE RESULTS OPEN

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

	customer_state	total	avg_value
1	SP	5998226.96	137.5
2	RJ	2144379.69	158.53
3	MG	1872257.26	154.71
4	RS	890898.54	157.18
5	PR	811156.38	154.15
6	SC	623086.43	165.98
7	BA	616645.82	170.82

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

```
#Calculate the Total & Average value of order freight for each state.
select customer_state,round(sum(freight_value),2) as total,round(avg(freight_value),2) as avg_value
from `scalar_set.order_items` ckp1 inner join `scalar_set.orders` ckp2 on
ckp1.order_id=ckp2.order_id inner join `scalar_set.customers` ckp3
on  ckp3.customer_id=ckp2.customer_id  group by customer_state order by 2 desc;
```

Press Alt+F1 for Access

Query results

SAVE RESULTS OPEN IN

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

	customer_state	total	avg_value
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	MG	270853.46	20.63
4	RS	135522.74	21.74
5	PR	117851.68	20.53
6	BA	100156.68	26.36

Question 5: Analysis based on sales, freight and delivery time

```sql
#________ Find the no. of days taken to deliver each order from the order's purchase date as delivery time
select order_id,DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,day) as time_to_deliver,
DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date,day) as diff_estimated_delivery
 from `scalar_set.orders`;
```

Press Alt+F1 for Accessibility

Query results

SAVE RESULTS | OPEN IN

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

ow	order_id	time_to_deliver	diff_estimated_delive
1	1950d777989f6a877539f5379...	30	12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	-28
3	65d1e226dfaeb8cdc42f66542...	35	-16
4	635c894d068ac37e6e03dc54e...	30	-1
5	3b97562c3aee8bdedcb5c2e45...	32	0
6	68f47f50f04c4cb6774570cfde...	29	-1
7	276e9ec344d3bf029ff83a161c...	43	4
8	54e1a3c2b97fb0809da548a59	40	4

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

```sql
#________Find out the top 5 states with the highest & lowest average freight value.

With top_5_states_highest as (select seller_state as state,ROUND(avg(freight_value),2) as avg_freight from
`scalar_set.order_items` ckp1
 inner join `scalar_set.sellers`  ckp2
 on ckp1.seller_id=ckp2.seller_id group by seller_state order by 2 DESC limit 5),
 top_5_states_lowest as (select seller_state as state,ROUND(avg(freight_value),2) as avg_freight from `scalar_set.
order_items` ckp1
 inner join `scalar_set.sellers`  ckp2
 on ckp1.seller_id=ckp2.seller_id group by seller_state order by 2 limit 5)

 select concat(state," is among top 5 states with highest average freight value of --", avg_freight) as state
from top_5_states_highest
 union all
 select concat(state," is among top 5 states with lowest average freight value of --", avg_freight) as state from
top_5_states_lowest;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXEC

Row	state
1	RO is among top 5 states with highest average freight value of ___50.91
2	CE is among top 5 states with highest average freight value of ___46.38
3	PB is among top 5 states with highest average freight value of ___39.19
4	PI is among top 5 states with highest average freight value of ___36.94
5	AC is among top 5 states with highest average freight value of ___32.84
6	SP is among top 5 states with lowest average freight value of ___18.45
7	PA is among top 5 states with lowest average freight value of ___19.39
8	RJ is among top 5 states with lowest average freight value of ___19.47
9	DF is among top 5 states with lowest average freight value of ___20.57

```
#________Find out the top 5 states with the highest & lowest average delivery time.
  #______the top 5 states with the highest average delivery time_____

Select customer_state as state,ROUND(AVG(DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,day)),
2) as time_to_deliver from `scalar_set.orders` ckp1 inner join `scalar_set.customers` ckp2
on ckp1.customer_id =ckp2.customer_id group by customer_state order by 2 desc limit 5;
```

Press Alt+F1 for Accessibility Opt

Query results

SAVE RESULTS | OPEN IN

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

tow	state	time_to_deliver
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32

```
#______the top 5 states with the lowest average delivery time_____

Select customer_state as state,ROUND(AVG(DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,day)),
2) as time_to_deliver from `scalar_set.orders` ckp1 inner join `scalar_set.customers` ckp2
on ckp1.customer_id =ckp2.customer_id group by customer_state order by 2 limit 5;
```

Press Alt+F1 for Accessibility Op

Query results

SAVE RESULTS | OPEN IN

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

ow	state	time_to_deliver
1	SP	8.3
2	PR	11.53
3	MG	11.54
4	DF	12.51
5	SC	14.48

```sql
#__________Find out the top 5 states where the order delivery is really fast as compared to the estimated date of delivery._____________

Select customer_state as state, ROUND(AVG(DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date, day)),2) as diff_estimated_delivery from `scalar_set.orders` ckp1 inner join `scalar_set.customers` ckp2 on ckp1.customer_id =ckp2.customer_id group by customer_state order by 2 limit 5;
```

Press Alt+F1 for Accessibility Opti

uery results

SAVE RESULTS | OPEN IN

B INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

	state	diff_estimated_delive
1	AC	-19.76
2	RO	-19.13
3	AP	-18.73
4	AM	-18.61
5	RR	-16.41

Ques6: Analysis based on the payments

```sql
#__________Find the month on month no. of orders placed using different payment types.
SELECT t.month, count(distinct t.payment_type) as count_payment_mode from (SELECT *,extract(month from order_purchase_timestamp) as month from `scalar_set.payments` ckp1 inner join`scalar_set.orders` ckp2 on ckp1.order_id =ckp2.order_id) t
group by 1 order by 1;
```

Press Alt+F1 for Accessibility C

ery results

SAVE RESULTS | OPEN IN

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

	month	count_payment_mod
	1	4
2	2	4
3	3	4
4	4	4
5	5	4

```sql
 #______Find the no. of orders placed on the basis of the payment installments that have been paid._________

#creating a CTE for order_id for which all intsallments are paid where price <= total amount paid
With id_installement_paid as (SELECT ckp1.order_id,cnt,price,total_pymt from `scalar_set.order_items` ckp1 inner
join (
SELECT order_id,count(payment_installments) as cnt,max(payment_value) as total_pymt from `scalar_set.payments`
where payment_installments > 0 group by order_id)ckp2 on ckp1.order_id=ckp2.order_id where price <= total_pymt)
select cnt as installments,count(order_id) as cnt_order from id_installement_paid group by 1 order by 1;
```

Press Alt+F1 for Accessibility Optic

Query results

SAVE RESULTS | OPEN IN

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

Row	installments	cnt_order
1	1	109359
2	2	1416
3	3	92
4	4	30
5	5	4

Insights:

1: Number of orders has steadily increased each year, indicating growing demand. To capitalize on this, consider optimizing inventory, scaling customer support, and targeting peak periods with tailored marketing strategies to further drive growth

2. Brazilian customers primarily place orders in the afternoon, followed by night and morning, with dawn seeing the least activity. To optimize sales, consider scheduling targeted promotions in the afternoon and night, while adjusting staffing and inventory accordingly.

3. Insight: top three states (SP,RJ,MG) contribute to 66.7% of sales in the Brazil region in terms of order placed. We should focus on increasing customer support and inventory supply in these states