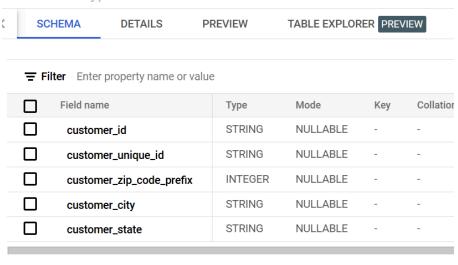
Ques1: Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

1. Data type of all columns in the "customers" table.



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



```
/*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 ) &
  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 RESI
)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?

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 ▼	11	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 <u>`scalar_set.orders`</u> group by 1 order by 1;

Row	month ▼	11	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)

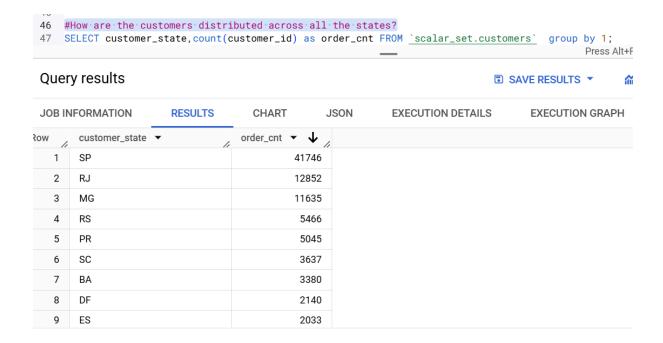
```
SELECT CASE when (extract(hour from order_purchase_timestamp) >= 1 )and (extract(hour from
order_purchase_timestamp) <= 6) then "Dawn"
WHEN extract(hour from order_purchase_timestamp) >= 7 and extract(hour from order_purchase_timestamp) <= 12 then
"Morning"
when extract(hour from order_purchase_timestamp) >= 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 <u>`scalar_set.orders`</u> 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

43	43 SELECT ckp1.customer_state,extract(month from order_purchase_timestamp) as month_order,count(order_id) as order_cnt FROM <u>`scalar_set.customers`</u> ckp1 44 inner join <u>`scalar_set.orders`</u> ckp2 on ckp1.customer_id=ckp2.customer_id group by 1,2 order by 1,2;							
					_		Pres	s Alt+F1 for Accessi
Que	ry results					3	SAVE RESULTS *	M OPEN IN
JOB I	NFORMATION	RESULTS	CHART	JSON	EXECUTION [DETAILS	EXECUTION GRA	APH
Row	customer_state -	. //	month_order ▼	order_cn	t v			
1	AC	,,		1	8			
2	AC			2	6			
3	AC		:	3	4			
4	AC			4	9			
5	AC			5	10			
6	AC			5	7			
7	AC			7	9			
8	AC			R	7			

#How are the customers distributed across all the states?



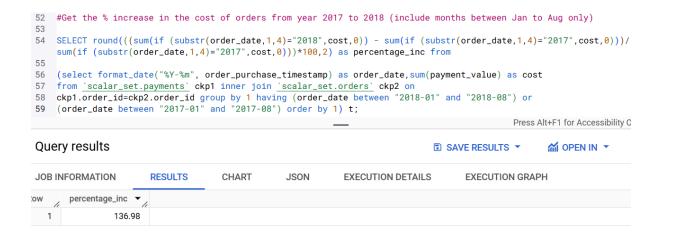
Let's have more insight:



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.

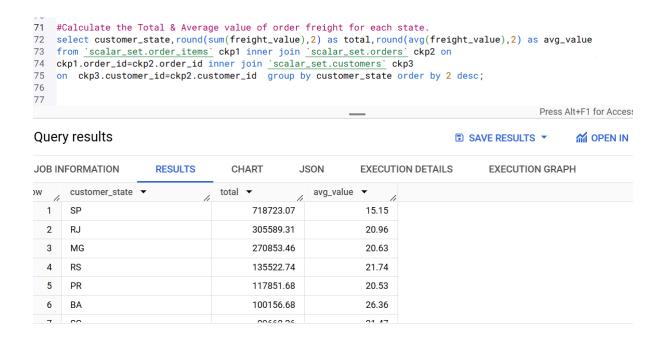
Get the % increase in the cost of orders from year 2017 to 2018 (include months between Jan to Aug only)



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

4 5	#Calculate the Total & Average value of order price for each state.						
<pre>6 select customer_state, round(sum(payment_value),2) as total, round(avg(payment_value),2) as avg_value 7 from 'scalar_set.payments' ckp1 inner join 'scalar_set.orders' ckp2 on ckp1.order_id=ckp2.order_id inner join 'scalar_set.customers' ckp3 9 on ckp3.customer_id=ckp2.customer_id group by customer_state order by total desc;</pre>							
0			_		Pres	s Alt+F1 for Ac	
)ue	ery results			•	SAVE RESULTS ▼		
ОВ	INFORMATION RESULTS	CHART J	ISON EXECUT	ON DETAILS	EXECUTION GRA	APH .	
/	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



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

```
#_____ Find the no. of days taken to deliver each order from the order's purchase date as delivery time
80
       select order_id, DATE_DIFF(order_delivered_customer_date, order_purchase_timestamp,day) as time_to_deliver,
81
       DATE_DIFF(order_delivered_customer_date, order_estimated_delivery_date,day) as diff_estimated_delivery
82
        from <u>`scalar_set.orders`;</u>
83
                                                                                                  Press Alt+F1 for Accessibility
Query results

■ SAVE RESULTS ▼

                                                                                                          ™ OPEN IN ▼
JOB INFORMATION
                        RESULTS
                                       CHART
                                                    JSON
                                                                EXECUTION DETAILS
                                                                                         EXECUTION GRAPH
                                    time_to_deliver ▼
order_id ▼
                                                      diff_estimated_delive
  1
       1950d777989f6a877539f5379...
                                                 30
                                                                   12
  2
       2c45c33d2f9cb8ff8b1c86cc28
                                                 30
                                                                   -28
  3
       65d1e226dfaeb8cdc42f66542...
                                                 35
                                                                   -16
  4
       635c894d068ac37e6e03dc54e...
                                                 30
                                                                    -1
  5
       3b97562c3aee8bdedcb5c2e45...
                                                 32
                                                                    0
       68f47f50f04c4cb6774570cfde...
                                                 29
       276e9ec344d3bf029ff83a161c...
                                                 43
                                                                    4
       54e1a3c2b97fb0809da548a59
                                                 40
```

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

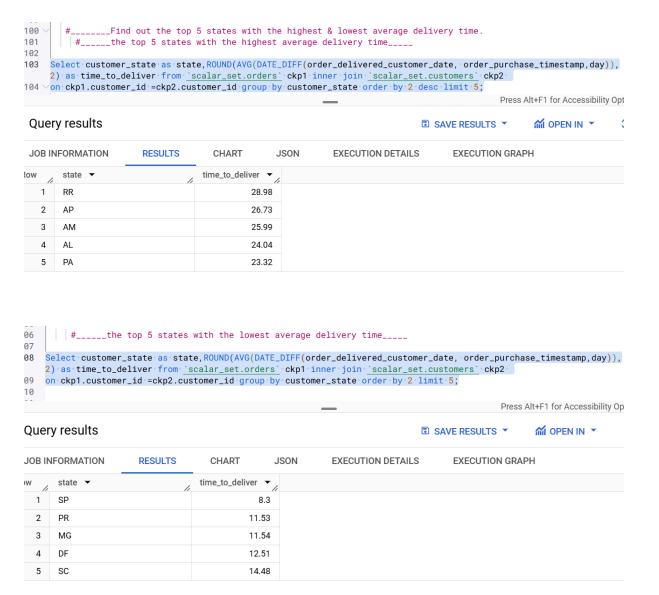
```
#_____Find out the top 5 states with the highest & lowest average freight value.

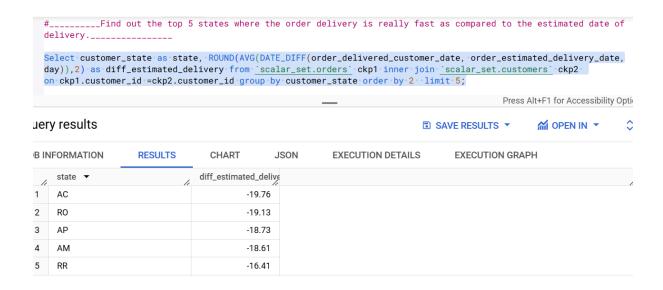
With top_5_states_highest as (select seller_state as state,ROUND(avg(freight_value),2) as avg_freight from 'scalar_set.order_items' ckp1
inner join 'scalar_set.sellers' ckp2
on ckp1.seller_id=ckp2.seller_id group by seller_state order by 2 DESC limit 5),
top_5_states_lowest as (select seller_state as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

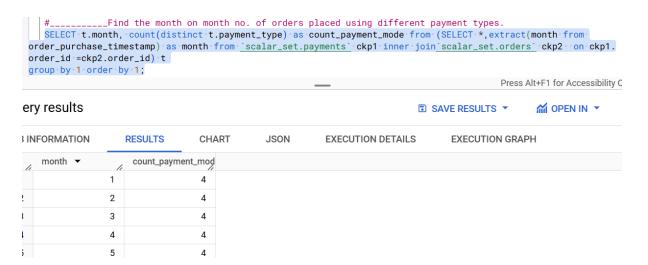
Query results

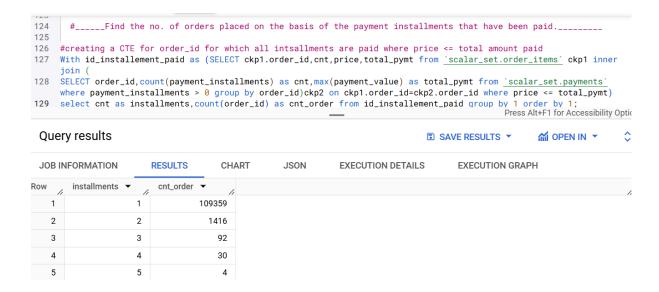
JOB IN	NFORMATION	RESULTS	CHART	JSON	EXEC
Row	state ▼				
1	RO is among top	5 states with highe	st average freigh	t value of50.	91
2	CE is among top	5 states with highe	st average freigh	t value of46.	38
3	PB is among top	5 states with highe	st average freigh	t value of39.	19
4	PI is among top 5	5 states with highes	t average freight	value of36.9	04
5	AC is among top	5 states with highe	st average freigh	t value of32.	84
6	SP is among top	5 states with lowes	t average freight	value of18.4	5
7	PA is among top	5 states with lowes	t average freight	value of19.3	19
8	RJ is among top	5 states with lowes	t average freight	value of19.4	7
9	DF is among top	5 states with lowes	t average freight	value of20.5	57





Ques6: Analysis based on the payments





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