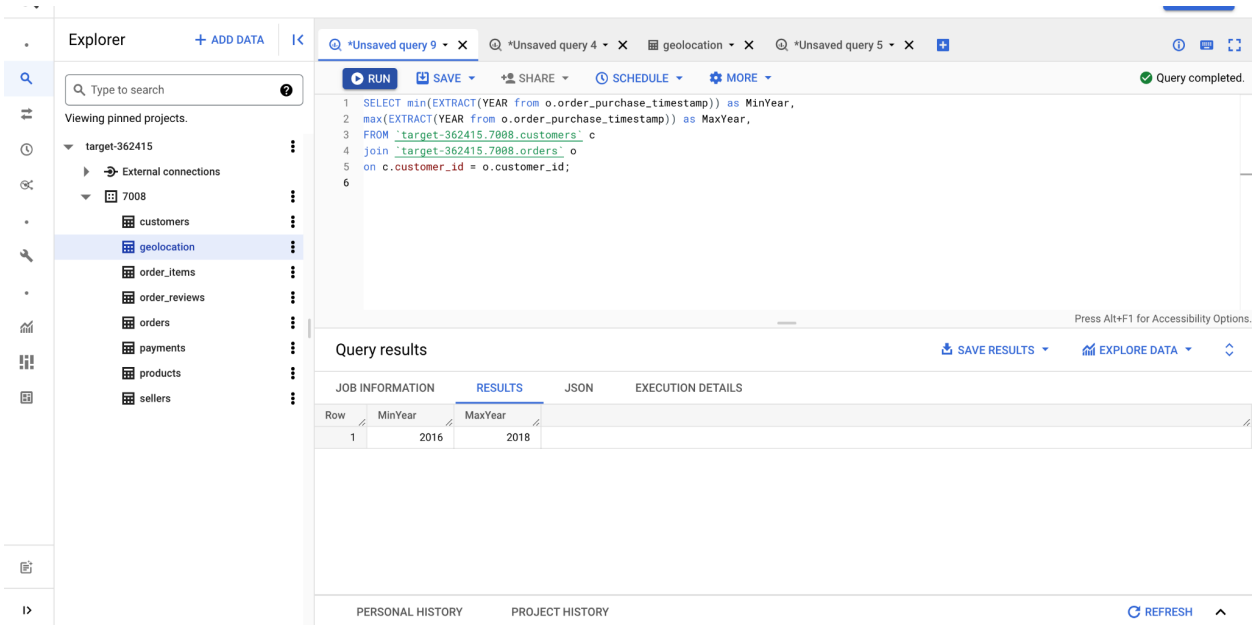
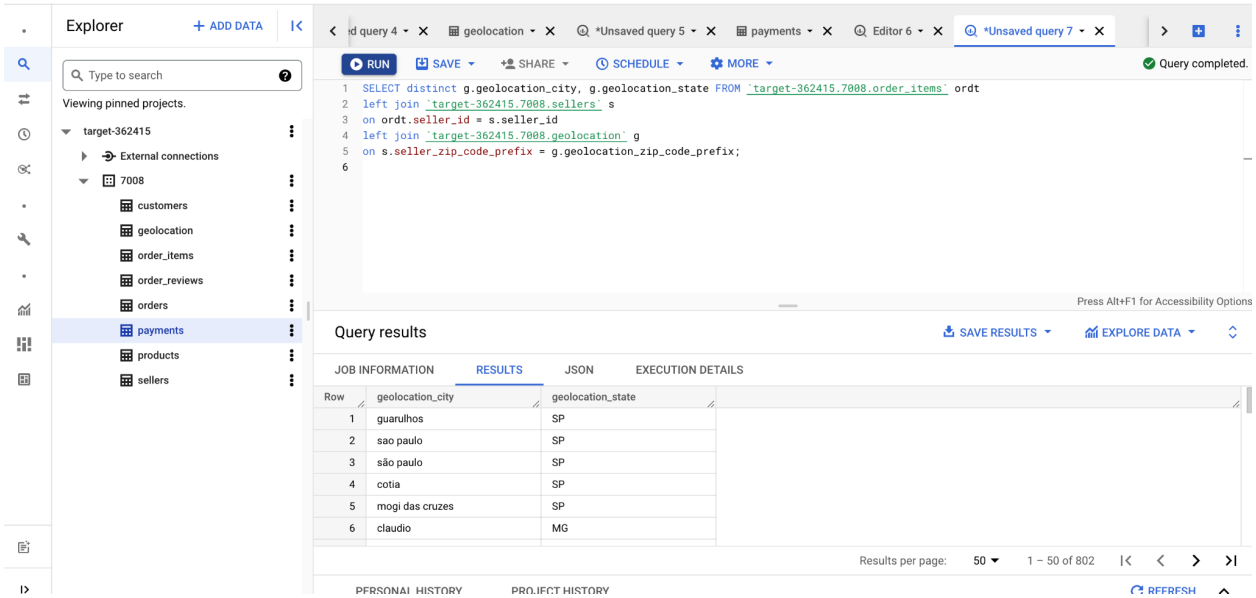
1.

```
SELECT min(EXTRACT(YEAR from o.order_purchase_timestamp)) as MinYear,
max(EXTRACT(YEAR from o.order_purchase_timestamp)) as MaxYear,
FROM `target-362415.7008.customers` c
join `target-362415.7008.orders` o
on c.customer_id = o.customer_id;
```

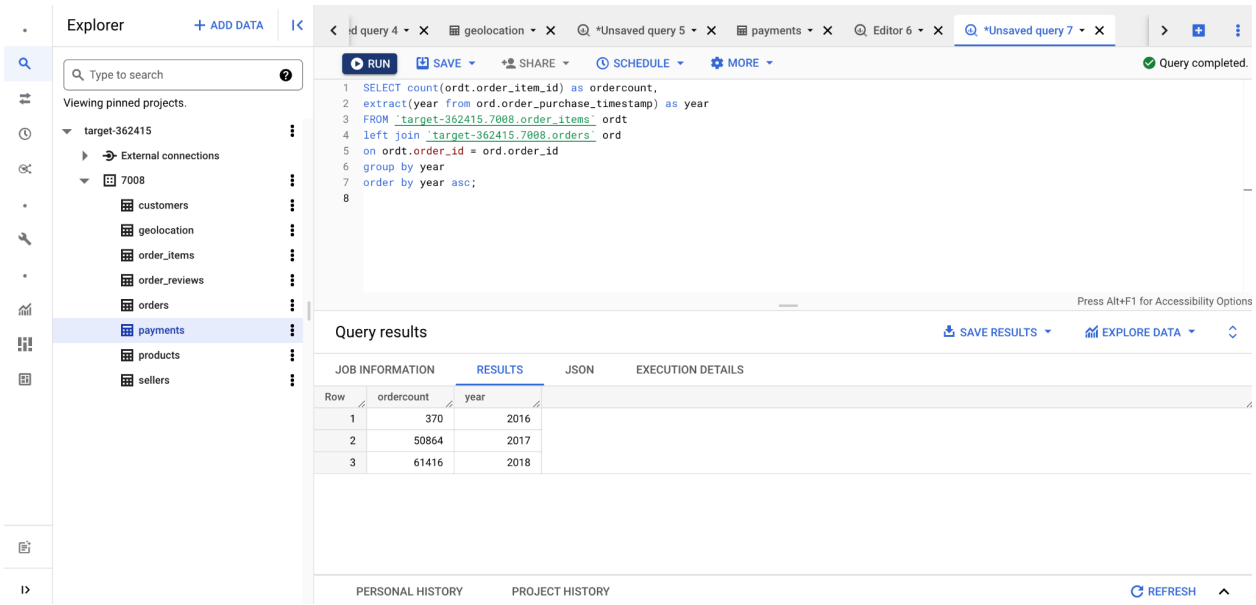


```sql
SELECT distinct g.geolocation_city, g.geolocation_state FROM
`target-362415.7008.order_items` ordt
left join `target-362415.7008.sellers` s
on ordt.seller_id = s.seller_id
left join `target-362415.7008.geolocation` g
on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix;
```

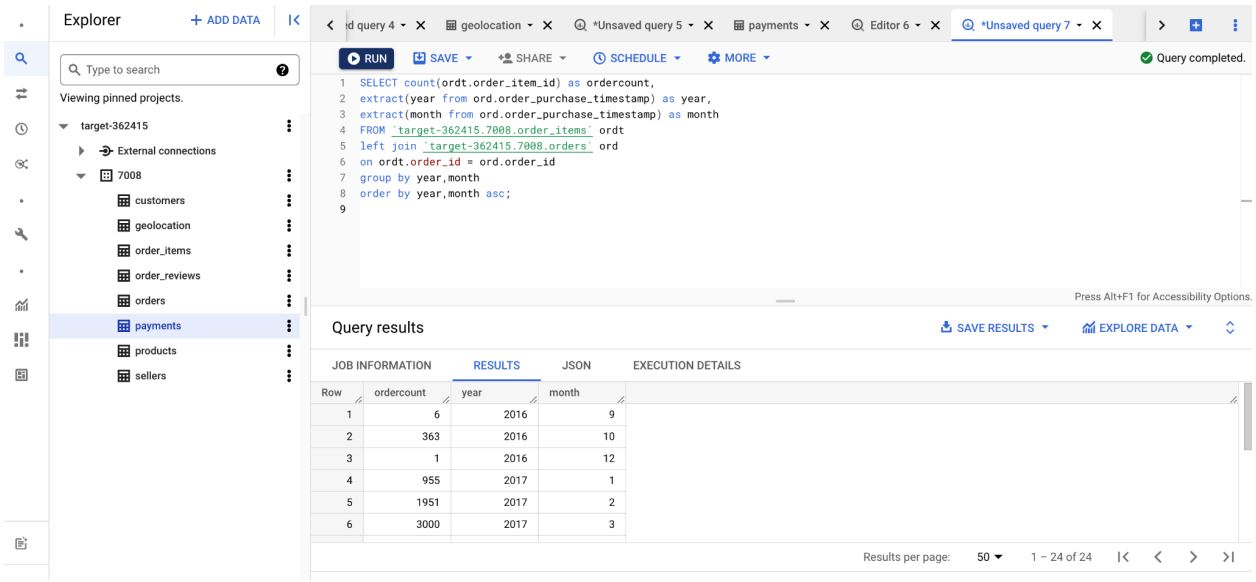


2.

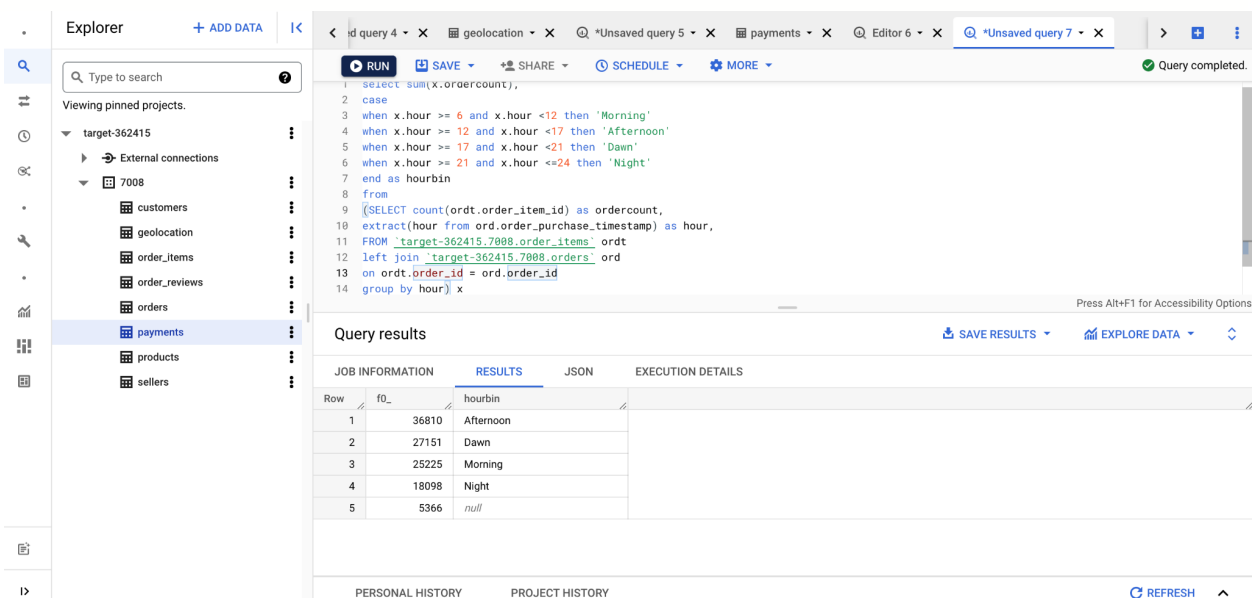
```
SELECT count(ordt.order_item_id) as ordercount,
extract(year from ord.order_purchase_timestamp) as year
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
group by year
order by year asc;
```



```sql
SELECT count(ordt.order_item_id) as ordercount,
extract(year from ord.order_purchase_timestamp) as year,
extract(month from ord.order_purchase_timestamp) as month
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
group by year,month
order by year,month asc;
```

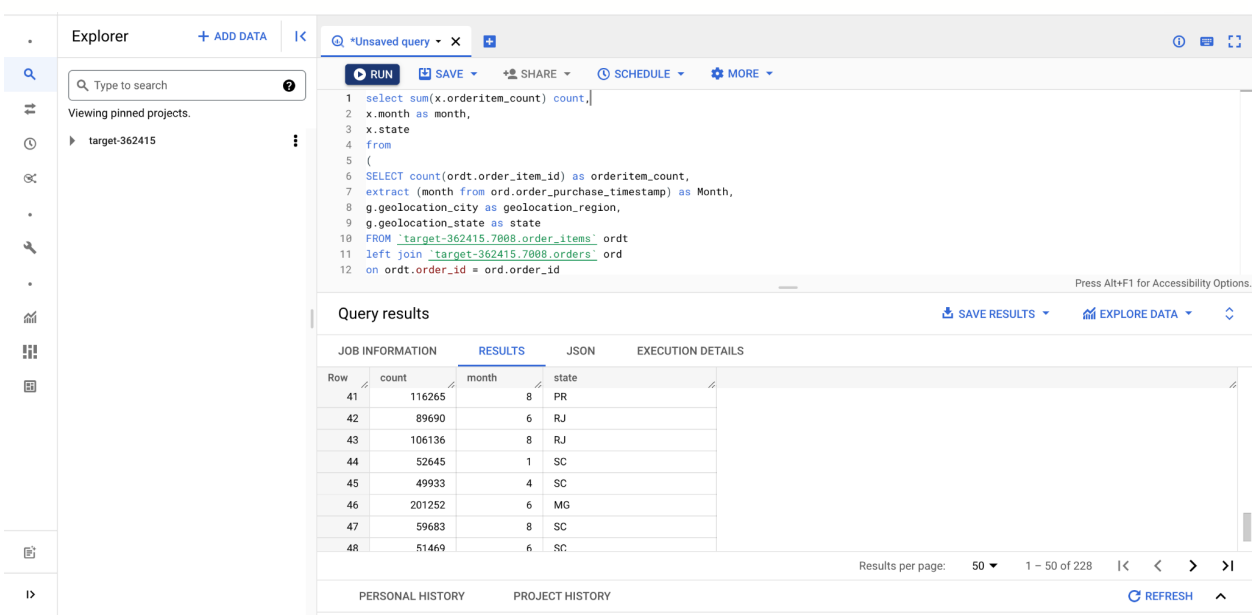


```sql
select sum(x.ordercount),
case
when x.hour >= 6 and x.hour <12 then 'Morning'
when x.hour >= 12 and x.hour <17 then 'Afternoon'
when x.hour >= 17 and x.hour <21 then 'Dawn'
when x.hour >= 21 and x.hour <=24 then 'Night'
end as hourbin
from
(SELECT count(ordt.order_item_id) as ordercount,
extract(hour from ord.order_purchase_timestamp) as hour,
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
group by hour) x
group by hourbin;
```

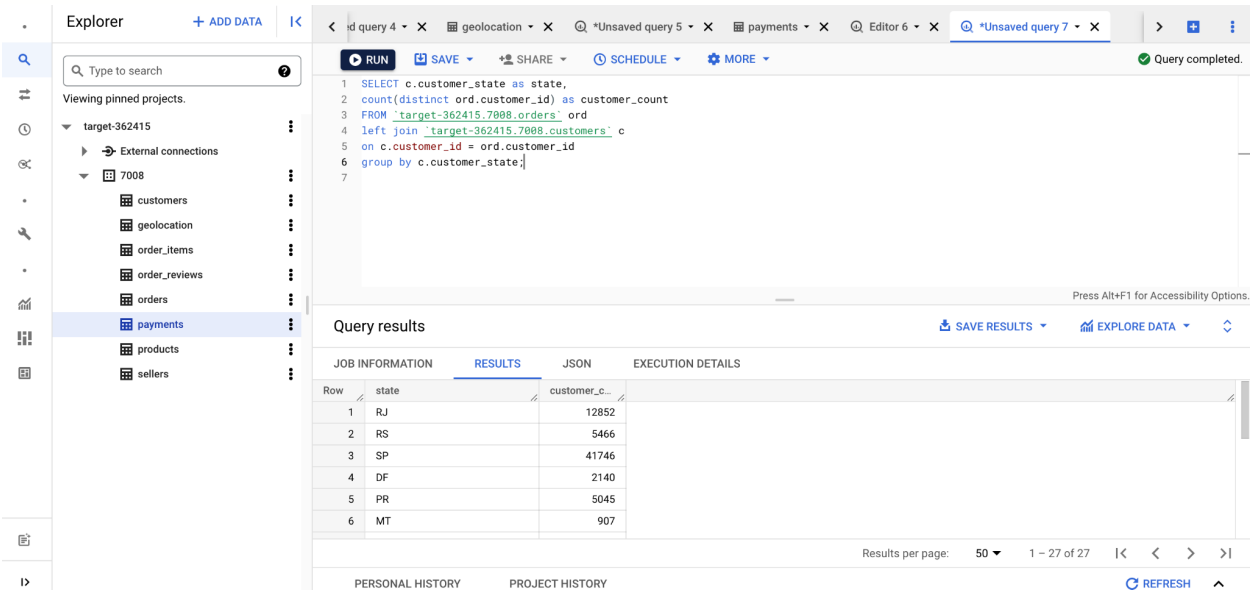


3.

```
select sum(x.orderitem_count) count,
x.month as month,
x.state
from
(
SELECT count(ordt.order_item_id) as orderitem_count,
extract (month from ord.order_purchase_timestamp) as Month,
g.geolocation_city as geolocation_region,
g.geolocation_state as state
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
left join `target-362415.7008.sellers` s
on ordt.seller_id = s.seller_id
left join `target-362415.7008.geolocation` g
on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix
group by state,geolocation_region,Month
order by state,geolocation_region,Month) x
group by x.state,x.month;
```



```sql
SELECT c.customer_state as state,
count(distinct ord.customer_id) as customer_count
FROM `target-362415.7008.orders` ord
left join `target-362415.7008.customers` c
on c.customer_id = ord.customer_id
group by c.customer_state
order by customer_count desc;
```

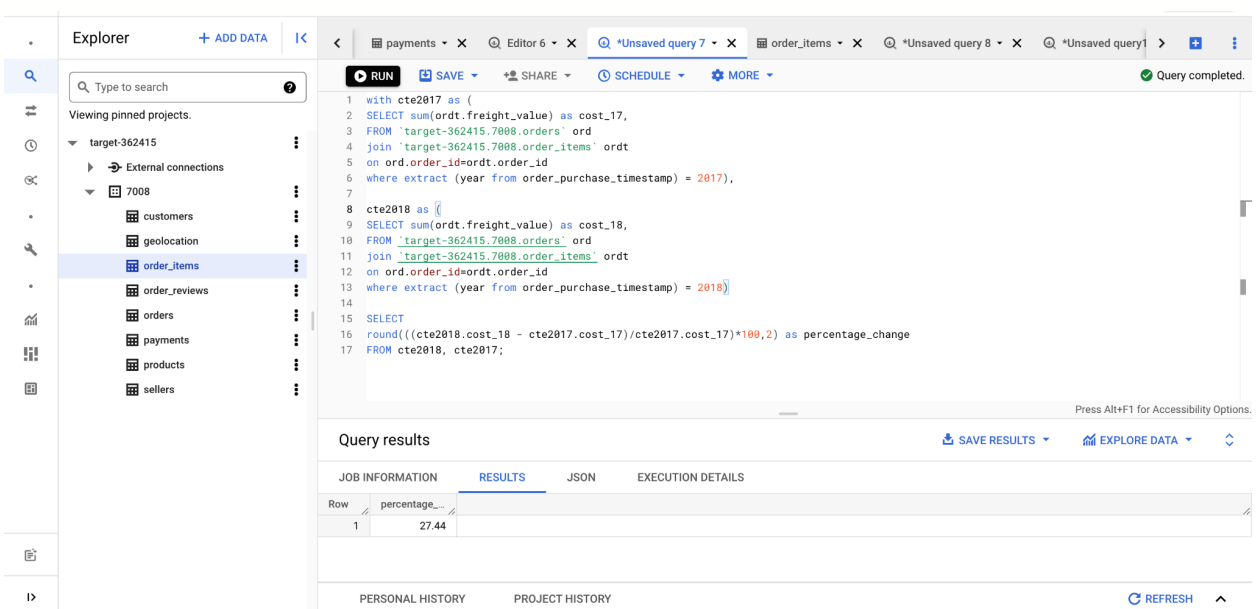


4.

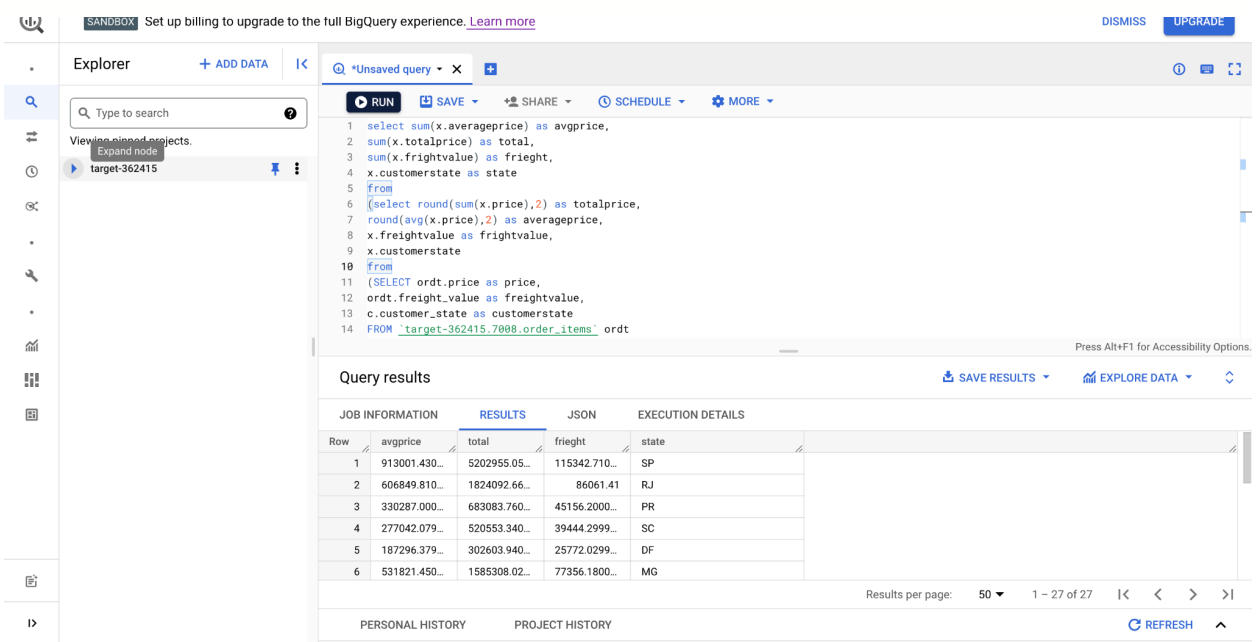
```
with cte2017 as (
SELECT sum(ordt.freight_value) as cost_17,
FROM `target-362415.7008.orders` ord
join `target-362415.7008.order_items` ordt
on ord.order_id=ordt.order_id
where extract (year from order_purchase_timestamp) = 2017),

cte2018 as (
SELECT sum(ordt.freight_value) as cost_18,
FROM `target-362415.7008.orders` ord
join `target-362415.7008.order_items` ordt
on ord.order_id=ordt.order_id
where extract (year from order_purchase_timestamp) = 2018)

SELECT
round(((cte2018.cost_18 - cte2017.cost_17)/cte2017.cost_17)*100,2) as
percentage_change
FROM cte2018, cte2017;
```

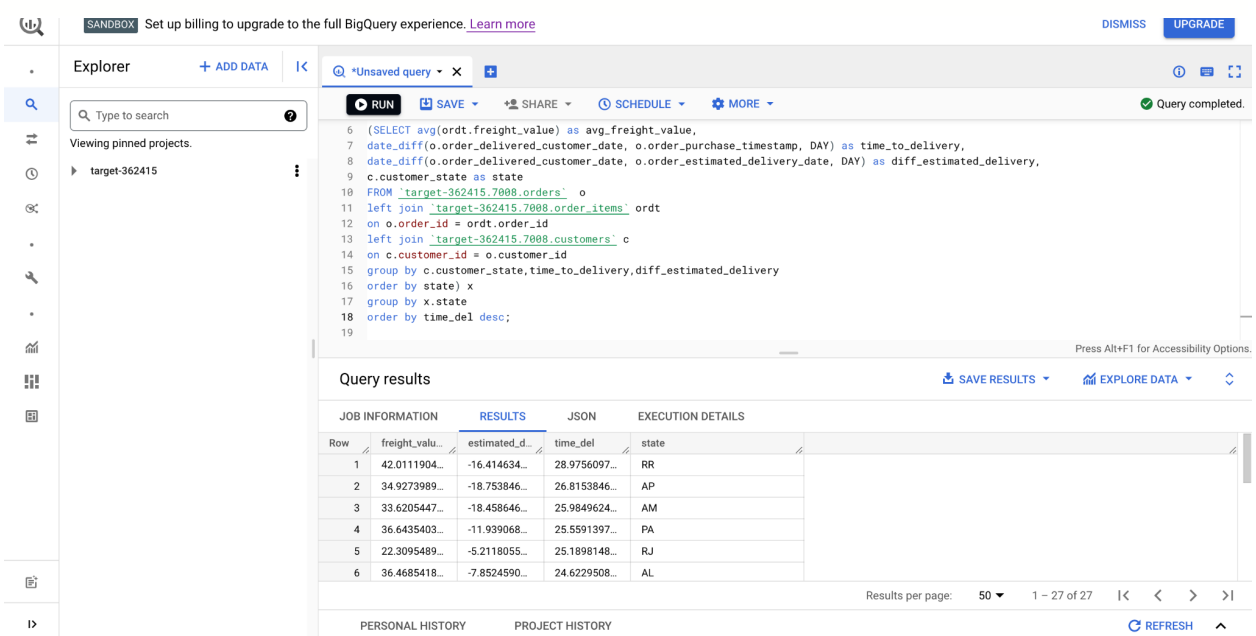


```
select sum(x.averageprice) as avgprice,
sum(x.totalprice) as total,
sum(x.frightvalue) as frieght,
x.customerstate as state
from
(select round(sum(x.price),2) as totalprice,
round(avg(x.price),2) as averageprice,
x.freightvalue as frightvalue,
x.customerstate
from
(SELECT ordt.price as price,
ordt.freight_value as freightvalue,
c.customer_state as customerstate
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
left join `target-362415.7008.customers` c
on c.customer_id = ord.customer_id) x
group by x.freightvalue,x.customerstate) x
group by x.customerstate
order by avgprice desc;
```

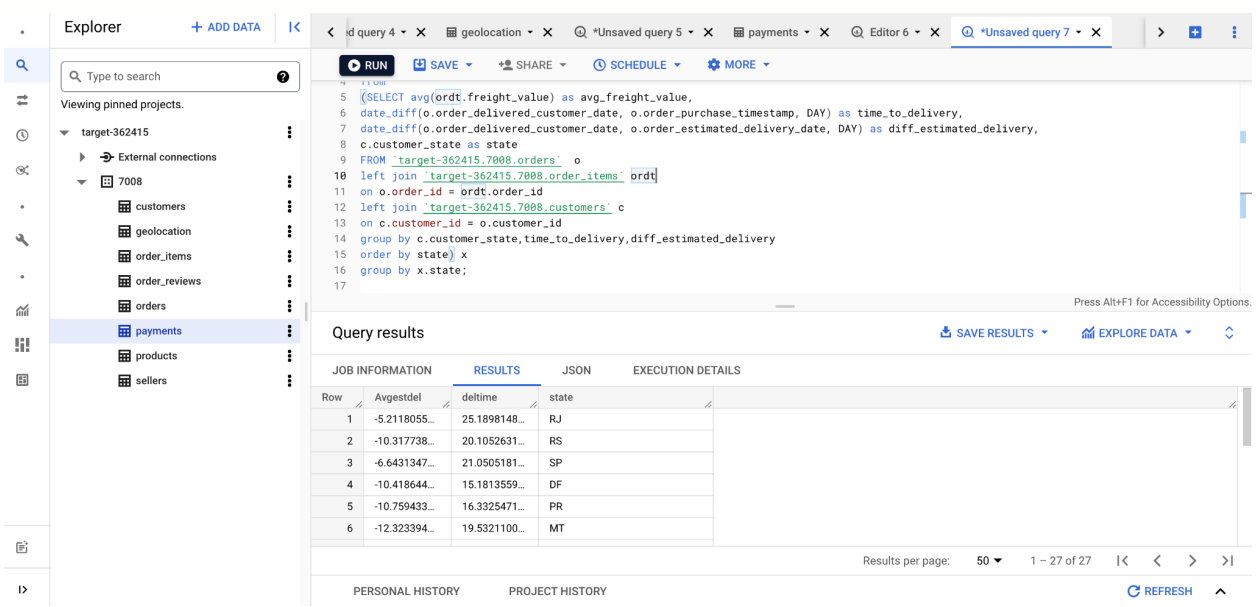


5.

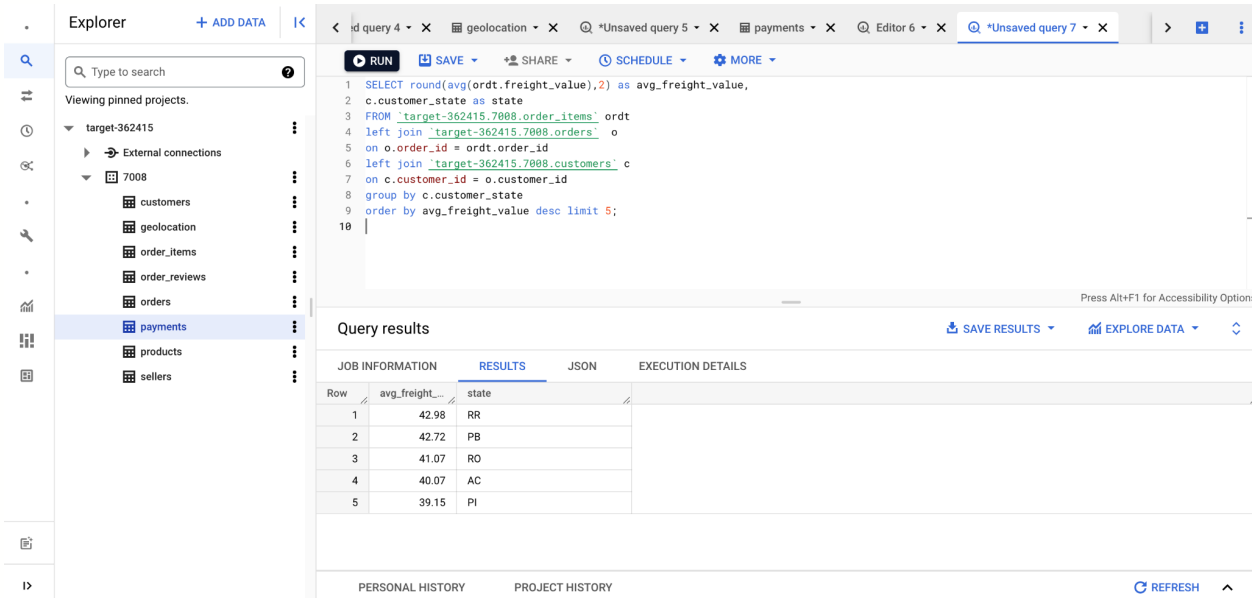
```
select avg(x.avg_freight_value) as freight_values,
avg(x.diff_estimated_delivery) as estimated_del,
avg(x.time_to_delivery) as time_del,
x.state
from
(SELECT avg(ordt.freight_value) as avg_freight_value,
date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY) as
time_to_delivery,
date_diff(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY) as
diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state,time_to_delivery,diff_estimated_delivery
order by state) x
group by x.state
order by time_del desc;
```



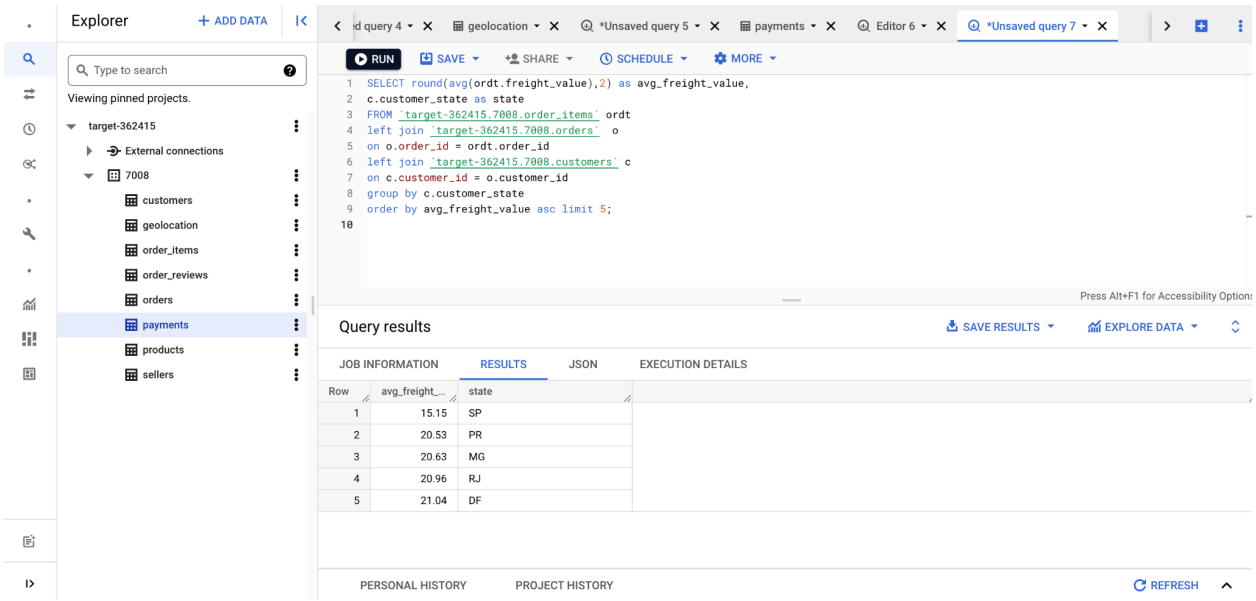
```
select avg(x.diff_estimated_delivery) as Avgestdel,
avg(x.time_to_delivery) as deltime,
x.state
from
(SELECT avg(ordt.freight_value) as avg_freight_value,
date_diff(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY) as
time_to_delivery,
date_diff(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY) as
diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state,time_to_delivery,diff_estimated_delivery
order by state) x
group by x.state;
```



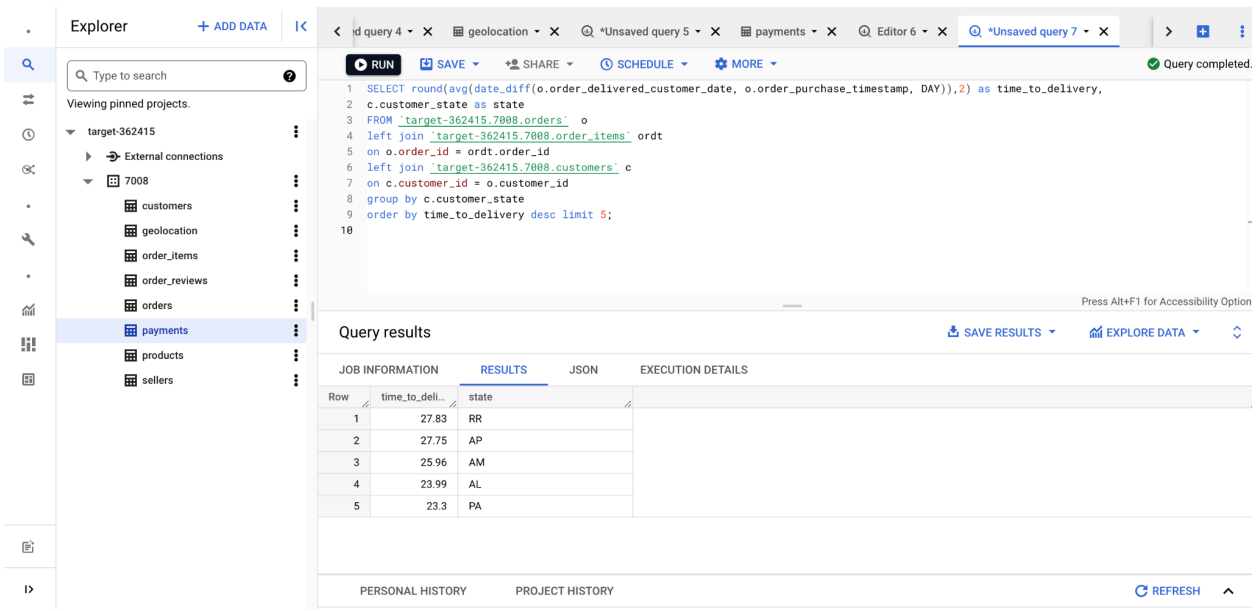
```
SELECT round(avg(ordt.freight_value),2) as avg_freight_value,
c.customer_state as state
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders`  o
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by avg_freight_value desc limit 5;
```



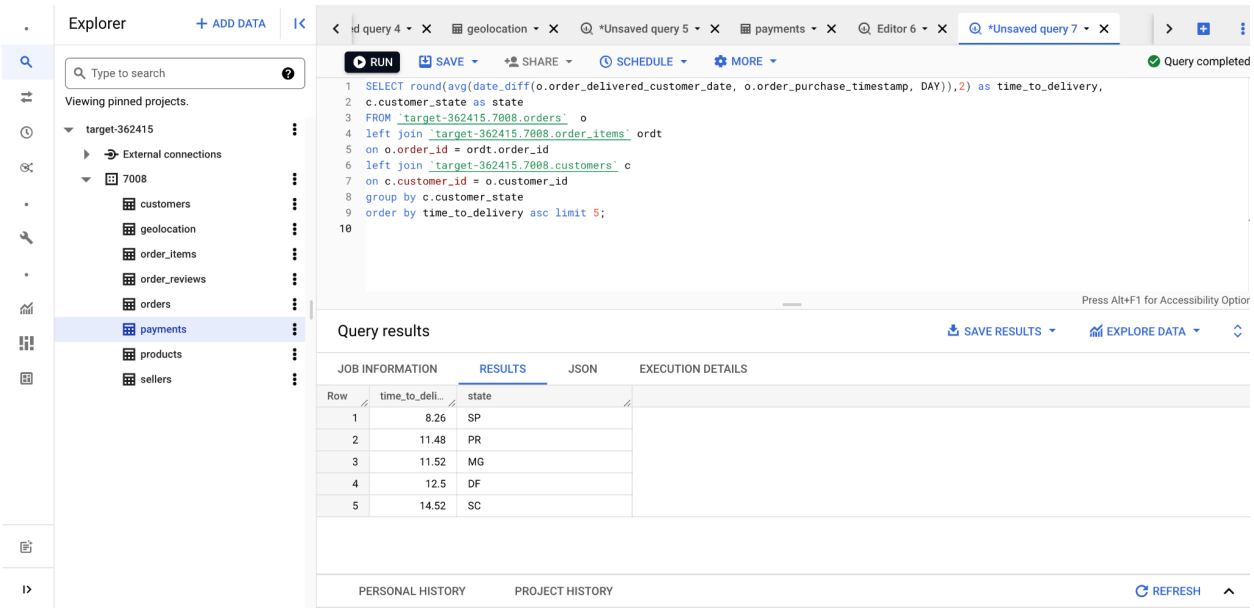
```
SELECT round(avg(ordt.freight_value),2) as avg_freight_value,
c.customer_state as state
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders`  o
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by avg_freight_value asc limit 5;
```



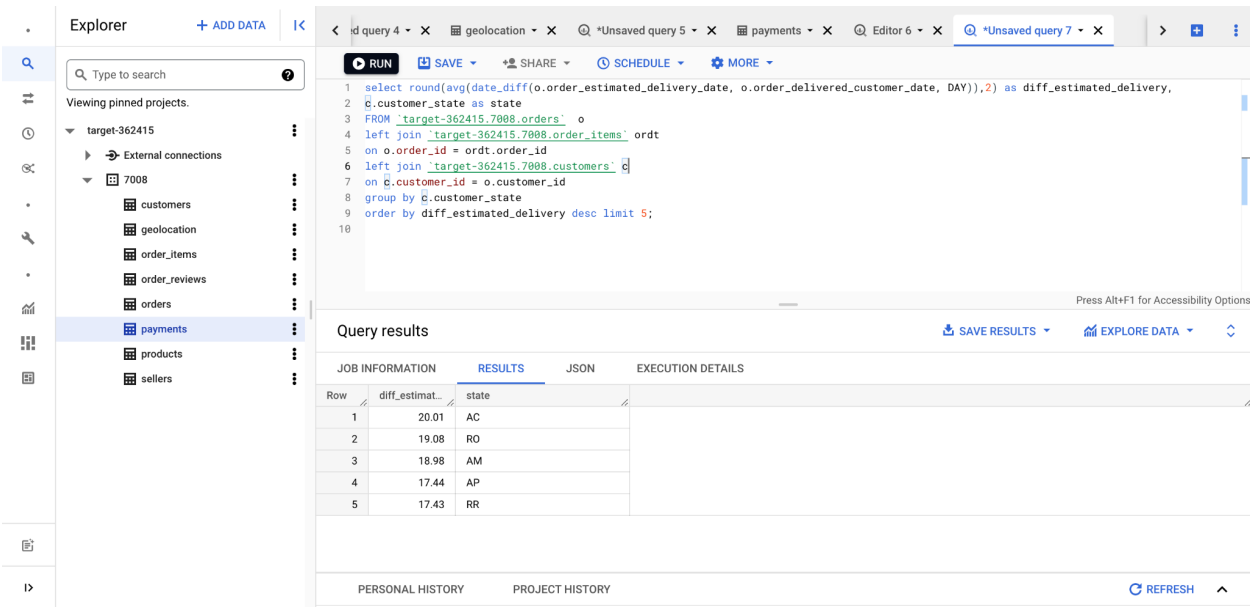
```sql
SELECT round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) as time_to_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by time_to_delivery desc limit 5;
```



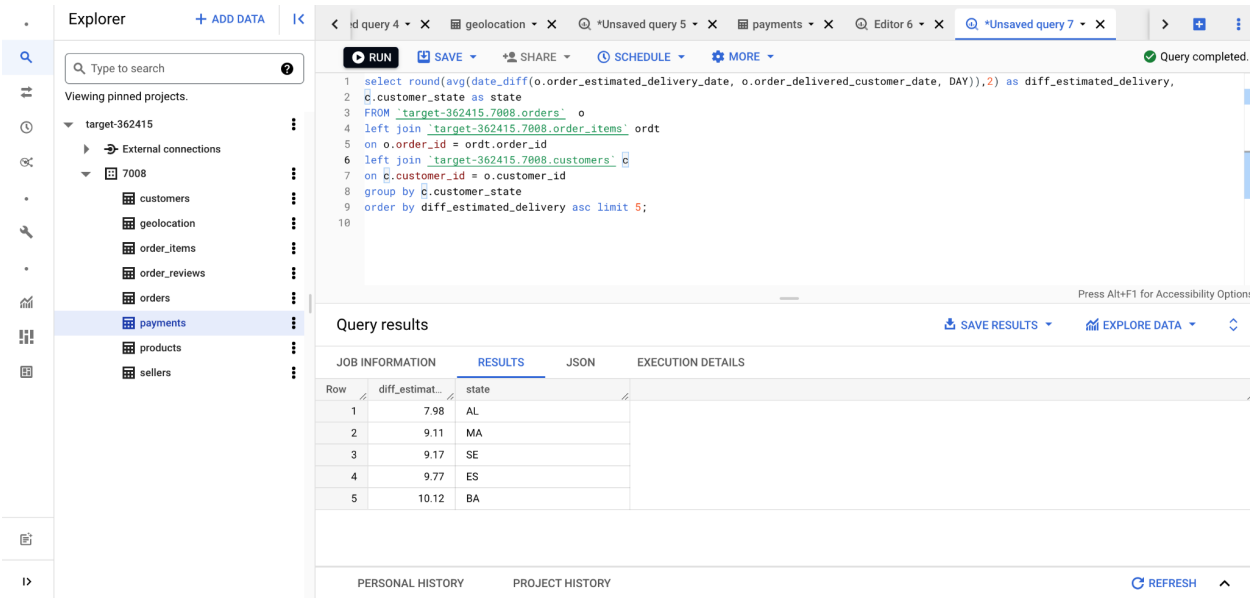
```sql
SELECT round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) as time_to_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by time_to_delivery asc limit 5;
```



```
select round(avg(date_diff(o.order_estimated_delivery_date,
o.order_delivered_customer_date, DAY)),2) as diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by diff_estimated_delivery desc limit 5;
```

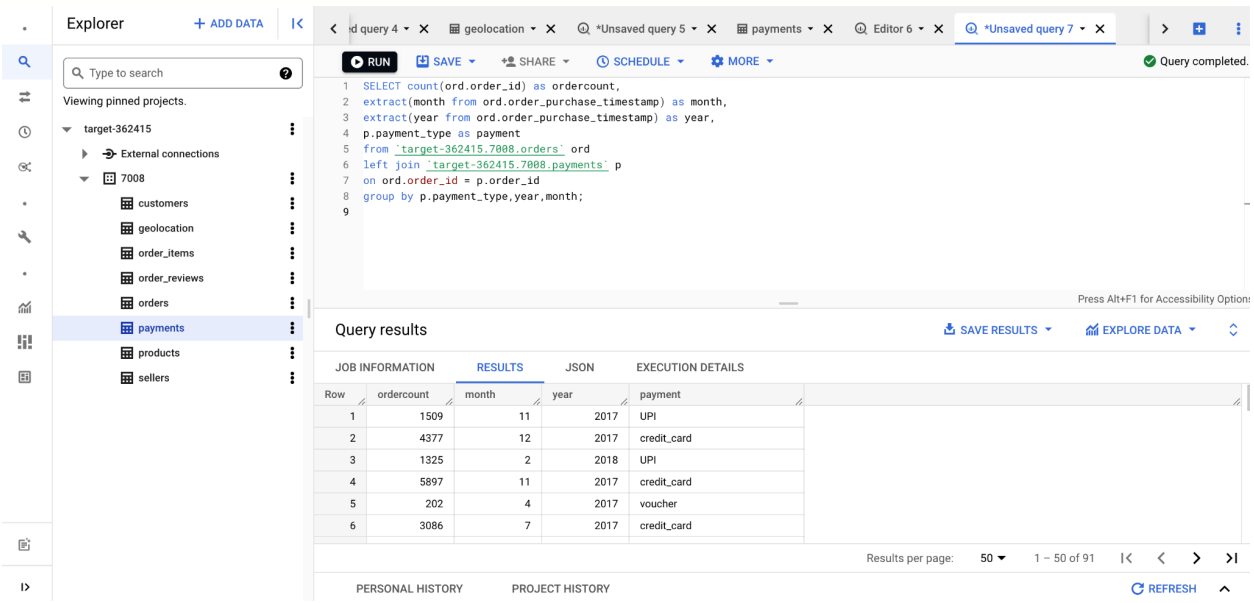


```
select round(avg(date_diff(o.order_estimated_delivery_date,
o.order_delivered_customer_date, DAY)),2) as diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders`  o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers` c
on c.customer_id = o.customer_id
group by c.customer_state
order by diff_estimated_delivery asc limit 5;
```

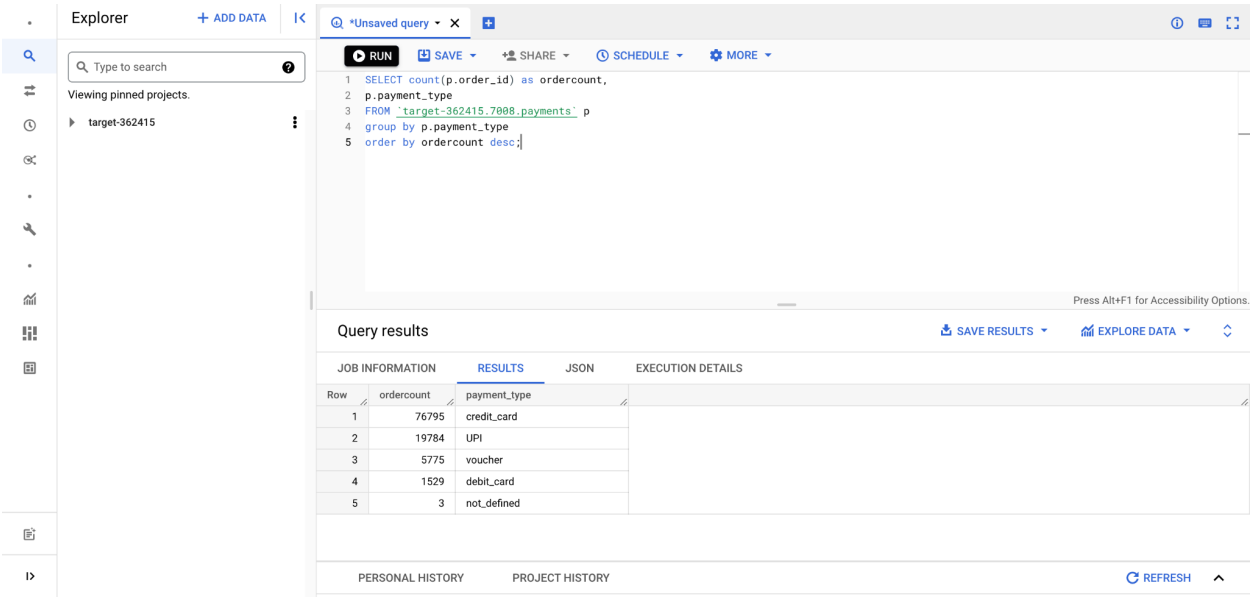


6.

```
SELECT count(ord.order_id) as ordercount,
extract(month from ord.order_purchase_timestamp) as month,
extract(year from ord.order_purchase_timestamp) as year,
p.payment_type as payment
from `target-362415.7008.orders` ord
left join `target-362415.7008.payments` p
on ord.order_id = p.order_id
group by p.payment_type,year,month;
```



```sql
SELECT count(p.order_id) as ordercount,
p.payment_type
FROM `target-362415.7008.payments` p
group by p.payment_type
order by ordercount desc;
```



Initial Analysis/Actionable insights:

1. The order count has significantly increased over the years.
2. The order count peaked in November in 2017 but we are not seeing any seasonality trends.
3. We can observe that the shopping numbers are high during the afternoon.
4. Observed that in SP state the order count decreased to half in the month of October.
5. DF and ES state have the least number of orders.
6. SP and MP have the highest number of orders.
7. SP has the highest number of customers and RR has the least number of customers.
8. Cost percentage increased nearly 27 percent in 2018 then 2017 from the month January to August
9. The average price and freight value is high for state SP and low for state RR.
10. Estimated delivery is negative means the delivery is getting done before the estimated time.
11. The average delivery time is very high in state RR compared to the state DF which has least delhivery time.
12. RR,PB,RO,AC and PI have the highest freight value.
13. SP,PR,MG,RJ and DF have the least freight value.
14. RR,AP,AM,AL and PA have the highest delivery time.
15. SP,PR,MG,DF and SC have the least delivery time.
16. AC,RO,AM,AP and RR have the highest estimated time difference. Means these states are delivering orders very earlier then the estimated delivery time.
17. AL,MA,AC,ES and BA have the least estimated time difference. Means these states are delivering orders when the estimated delivery time is nearer.
18. Customers are more likely to use credit cards for payments.
19. Customers are less likely to use debit cards for payments.

Recommendations:

1. The shopping numbers are higher in the afternoon so we can keep more sales persons in that time period.
2. We can give more offers from DF and ES to increase the order count.
3. We can increase the price in SP and MP as these 2 states have the highest number of orders.
4. We can initiate some offers in state RR for attracting more customers as it has the least number of customers.
5. The cost percentage has increased 27 percent from 2017 to 2018 which is on a higher side. We can offer some discounts to loyal customers.
6. In state RR the average delhivery time is very high which might be the reason for less no of orders from this place. We can reduce the delhivery time for the orders.
7. Similar to state RR other states are there who have high delhivery time which can be reduced.

8. States having the highest estimated time difference means we can mention reduced estimated delivery time which will help our customers a relief by seeing nearer estimated delivery date.
9. To promote other payment methods we can roll out small offers for payment methods which are least used. Like debit card, vouchers.
10. As we can see, customers are using credit cards and UPI more for payments so we should deploy card payment machines and UPI payment methods for easy transactions.
11. As customers are using credit cards there may be more chances of EMIs so we can partner with banks for no cost EMI or less interest rates options.