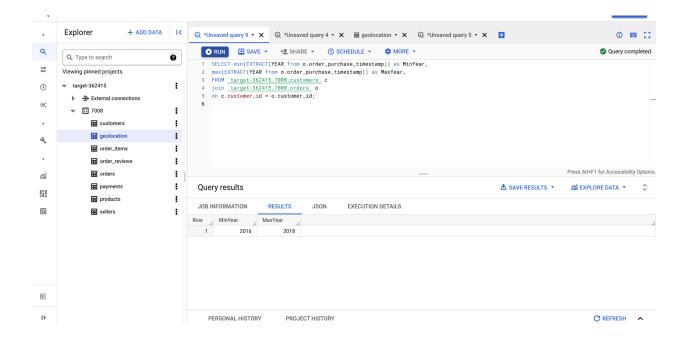
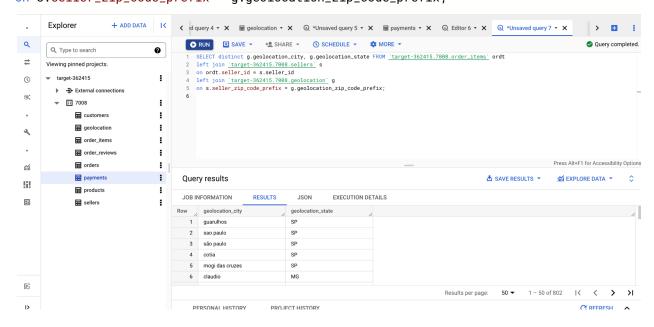
1.

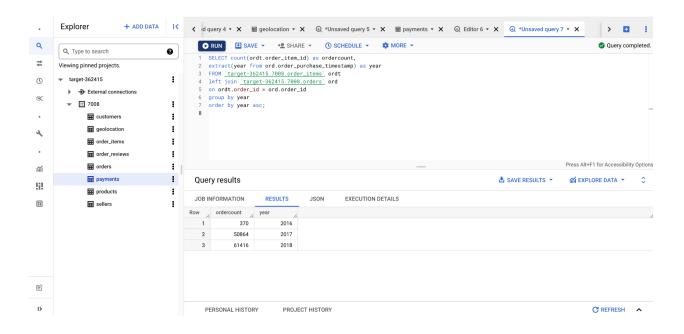
```
SELECT min(EXTRACT(YEAR from o.order_purchase_timestamp)) as MinYear,
max(EXTRACT(YEAR from o.order_purchase_timestamp)) as MaxYear,
FROM `target-362415.7008.customers` c
join `target-362415.7008.orders` o
on c.customer_id = o.customer_id;
```



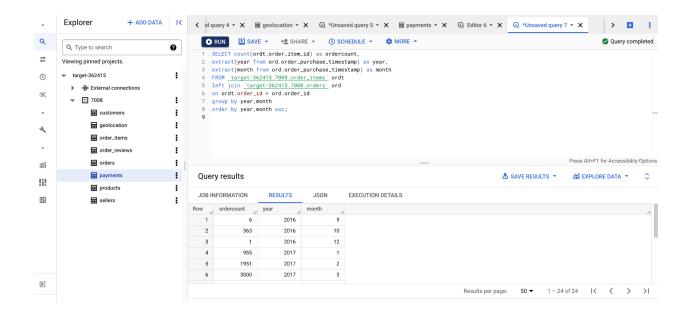
SELECT distinct g.geolocation_city, g.geolocation_state FROM
`target-362415.7008.order_items` ordt
left join `target-362415.7008.sellers` s
on ordt.seller_id = s.seller_id
left join `target-362415.7008.geolocation` g
on s.seller_zip_code_prefix = g.geolocation_zip_code_prefix;



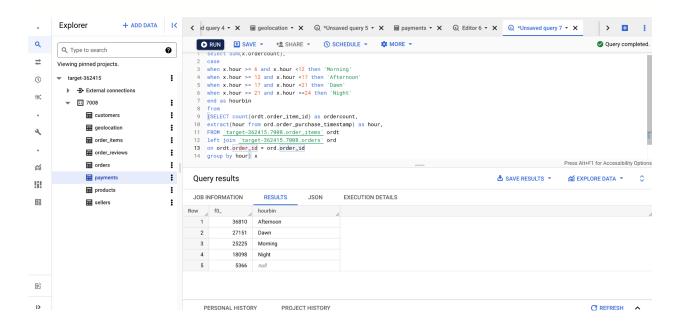
```
SELECT count(ordt.order_item_id) as ordercount,
extract(year from ord.order_purchase_timestamp) as year
FROM `target-362415.7008.order_items` ordt
left join `target-362415.7008.orders` ord
on ordt.order_id = ord.order_id
group by year
order by year asc;
```



```
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;
```

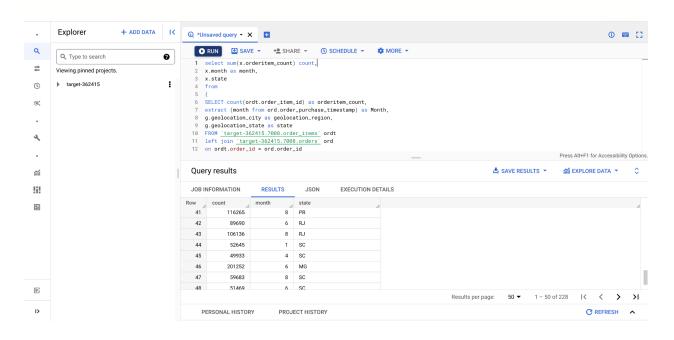


```
select sum(x.ordercount),
case
when x.hour >= 6 and x.hour <12 then 'Morning'
when x.hour >= 12 and x.hour <17 then 'Afternoon'
when x.hour >= 17 and x.hour <21 then 'Dawn'
when x.hour >= 21 and x.hour <=24 then 'Night'
end as hourbin
from
(SELECT count(ordt.order_item_id) as ordercount,
extract(hour 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;</pre>
```

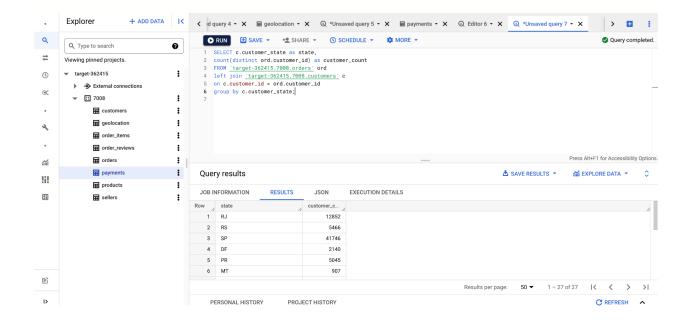


```
3.
```

```
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;
```



```
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;
                       + ADD DATA K G Begyments - X Q Editor 6 - X Q *Unsaved query 7 - X B order_items - X Q *Unsaved query 8 - X Q *Unsaved query > 1
                                        ► RUN SAVE → + SHARE → SCHEDULE → MORE →
 Q
                                                                                                                                    Query completed.
       Q Type to search
                                        1 with cte2017 as (
2 SELECT sum(ordt.freight_value) as cost_17,
3 FROM target-362415.7008.order_items ordt
4 join target-362415.7008.order_items ordt
  =
       ▼ target-362415
  (1)
                                       5 on ord.order_id=ordt.order_id
         ▶ → External connections
                                       6 where extract (year from order_purchase_timestamp) = 2017),
         ₹ 1008
                                 :
                                       8 cte2018 as (
              customers
                                 .
                                      9 SELECT sum(ordt.freight_value) as cost_18,

10 FROM 'target-362415.7008.orders' ord

11 join 'target-362415.7008.order_items' ordt

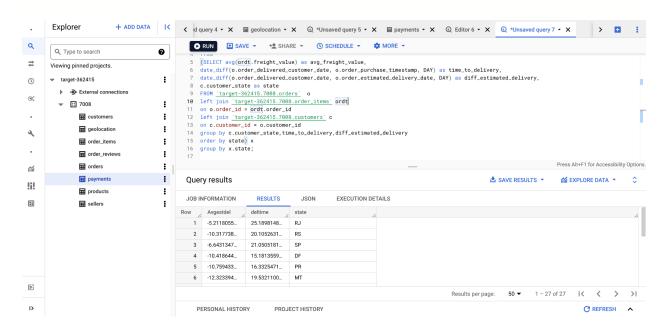
12 on ord.order_id=ordt.order_id
              geolocation
  ٩
            order_items
                                •
              ■ order_reviews
                                       13 where extract (year from order_purchase_timestamp) = 2018)
              orders
                                      14 | S | SELECT | 15 | SELECT | 16 | round(((cte2018.cost_18 - cte2017.cost_17)/cte2017.cost_17)*190,2) as percentage_change | 17 | FROM cte2018, cte2017;
  íííí
                                 1
              payments
                                 :
  99
              products
                                 :
  H
              sellers
                                                                                                                Press Alt+F1 for Accessibility Options
                                                                                                             Query results
                                       JOB INFORMATION
                                                       RESULTS
                                                                          EXECUTION DETAILS
                                                                 JSON
                                               27.44
  Εï
   1>
                                         PERSONAL HISTORY
                                                           PROJECT HISTORY
                                                                                                                                   C REFRESH ^
```

```
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;
         SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
  (II)
                      Explorer
  Q
                                       ► RUN SAVE + + SHARE + (S) SCHEDULE + SMORE +
        Q Type to search
                                         select sum(x.averageprice) as avgprice,
sum(x.totalprice) as total,
sum(x.frightvalue) as frieght,
  \rightleftarrows
  (1)
                                         x.customerstate as state
                                         from
(select round(sum(x.price),2) as totalprice,
  %
                                       7 round(avg(x.price),2) as average x.freightvalue as frightvalue,
                                                         as averageprice,
                                      8 x.freightvalue as frightvalue,
y x.customerstate
10 from
11 (SELECT ordt.price as price,
ordt.freight.value as freightvalue,
c.customer_state as customerstate
  ٩
                                      14 FROM <u>'target-362415.7008.order_items'</u> ordt
  íííí
                                                                                                                 Press Alt+F1 for Accessibility Options.
  99
                                      Ouery results
                                                                                                           Ξi
                                      JOB INFORMATION
                                                      RESULTS
                                                                JSON
                                                                        EXECUTION DETAILS
                                                            frieght
                                     Row avgprice
                                                    total
                                                                      state
                                                   5202955.05...
                                                             115342.710...
                                          913001.430...
                                          606849.810...
                                                    1824092.66...
                                                                86061.41
                                                                        RJ
                                        3 330287.000...
                                                    683083.760...
                                                              45156.2000...
                                                                        PR
                                          277042.079...
                                                    520553.340...
                                                              39444.2999...
                                                                        SC
                                          187296.379... 302603.940...
                                       6 531821.450... 1585308.02... 77356.1800...
  ΕÌ
                                                                                                  Results per page: 50 ▼ 1 – 27 of 27 |⟨ ⟨ > >|
  1>
                                        PERSONAL HISTORY
                                                         PROJECT HISTORY
                                                                                                                               C REFRESH ^
```

```
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;
        SANDBOX Set up billing to upgrade to the full BigQuery experience. Learn more
  (II)
                                                                                                                   DISMISS UPGRADE
                    Explorer
                                                                                                                          ① = []
  Q
                                    RUN SAVE + + SHARE + SCHEDULE + MORE +
                                                                                                                       Query completed.
       Q Type to search
                             0
                                     (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,
       Viewing pinned projects.
  (J)
       ▶ target-362415
                                   Œ.
                                   15 group by c.customer_state,time_to_delivery,diff_estimated_delivery
16 order by state) x
17 group by x.state
18 order by time_del desc;
  a.
                                                                                                       Press Alt+F1 for Accessibility Options.
  ~
                                                                                                  Query results
  989
  Ξi
                                   JOB INFORMATION
                                                                  EXECUTION DETAILS
                                                  RESULTS
                                                          JSON
                                  Row freight_valu... estimated_d... time_del
                                      42.0111904...
                                       34.9273989...
                                                 -18.753846...
                                                          26.8153846...
                                      33.6205447...
                                                -18.458646...
                                                         25.9849624...
                                       36.6435403...
                                                -11.939068...
                                                         25.5591397...
                                                                   PA
                                    5 22.3095489... -5.2118055...
                                                         25.1898148...
                                                                   RJ
                                    6 36.4685418... -7.8524590... 24.6229508...
                                                                                          Results per page: 50 ▼ 1 – 27 of 27 |⟨ ⟨ > >|
  ı>
                                     PERSONAL HISTORY
                                                     PROJECT HISTORY
                                                                                                                      C REFRESH
```

```
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;
        Explorer
                        + ADD DATA K d query 4 * X 📾 geolocation * X Q *Unsaved query 5 * X 📾 payments * X Q Editor 6 * X Q *Unsaved query 7 * X > 💽
 Q
                                          RUN BAVE - SHARE - SCHEDULE - MORE -
        Q Type to search
                                         1 SELECT round(avg(ordt.freight_value),2) as avg_freight_value,
2 c.customer_state as state
3 FROM 'target-362415.7988.order_items' ordt
4 left join 'target-362415.7988.orders' o
5 on o.order_id = ordt.order_id
6 left join 'target-362415.7988.oustomers' c
7 on c.customer_id = o.customer_id
8 orang by c.customer_state.
  =
        Viewing pinned projects.
  (1)
         External connections
  ®
         ▼ ∷ 7008
                                   i
                                         8 group by c.customer_state
9 order by avg_freight_value desc limit 5;
10 |
             customers
                                   .
              geolocation
                                   :
  ٩

    □ order_items

                                   :
              grder_reviews
                                   :
              orders
                                   : |
  ΔÚ
             payments
                                                                                                                     :
                                         Query results
  99
               products
                                   :
                                         JOB INFORMATION
                                                           RESULTS
                                                                     JSON
                                                                              EXECUTION DETAILS
  Ξi
               sellers
                                   i
                                        Row avg_freight_... state
                                                   42.98
                                                         RR
                                                   42.72 PB
                                                   41.07
                                                         RO
                                                   40.07
                                           5
                                                   39.15 PI
```

PERSONAL HISTORY

PROJECT HISTORY

€ REFRESH ^

Ē

```
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;
       Explorer
                   RUN SAVE + + SHARE + SCHEDULE + * MORE +
 Q
      Q Type to search
                                  1 SELECT round(avg(ordt.freight_value),2) as avg_freight_value,
2 c.customer_state as state
3 FROM _target=362415.7088.order_items' ordt
1 left join _target=362415.7088.orders' o
5 on o.order_id = ordt.order_id
      Viewing pinned projects.
  (1)
      :
       ▶ -3- External connections
                                 on c.oustomer_id = o.customer_id
group by c.customer_ste
order by avg_freight_value asc limit 5;
        ▼ 1 7008
                            :
            customers
                            :
            geolocation
                            :
  ٩
           order_items
                            •
            grder_reviews
                                                                                                    Press Alt+F1 for Accessibility Options
            orders
                             : |
  íííí
            payments
                            i
                                                                                               Query results
  99
            products
                             :
                                               RESULTS
                                                       JSON
                                  JOB INFORMATION
                                                                EXECUTION DETAILS
  ##
                                 Row avg_freight_... state
                                          15.15 SP
                                          20.53 PR
                                          20.63 MG
                                          20.96 RJ
                                          21.04 DF
  Ē
```

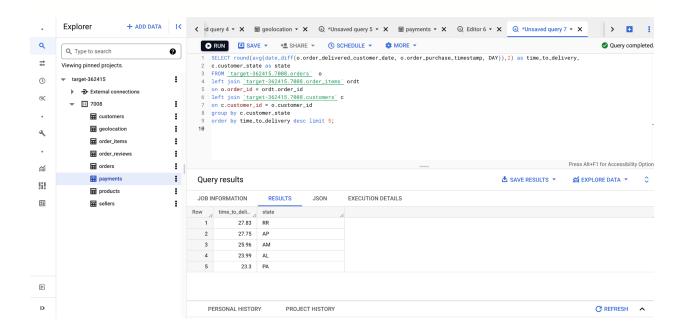
PERSONAL HISTORY

PROJECT HISTORY

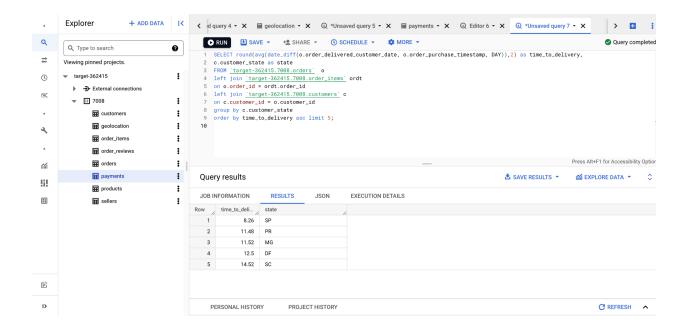
C REFRESH ^

1>

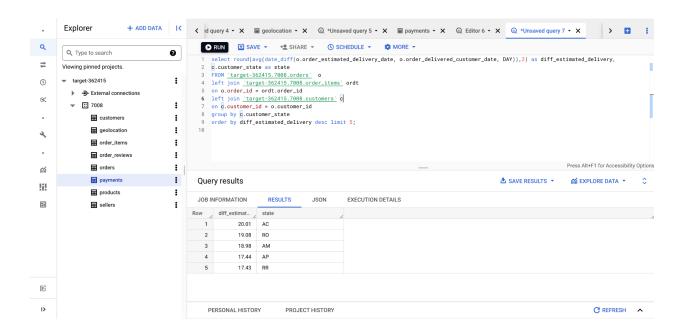
```
SELECT round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) as time_to_delivery,
c.customer_state as state
FROM `target-362415.7008.orders` o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers`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



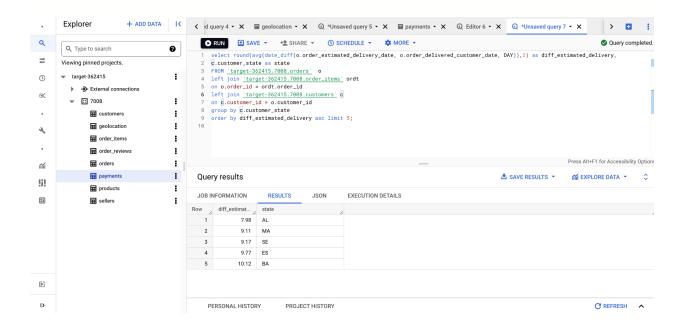
```
SELECT round(avg(date_diff(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) as time_to_delivery,
c.customer_state as state
FROM `target-362415.7008.orders` o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers`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



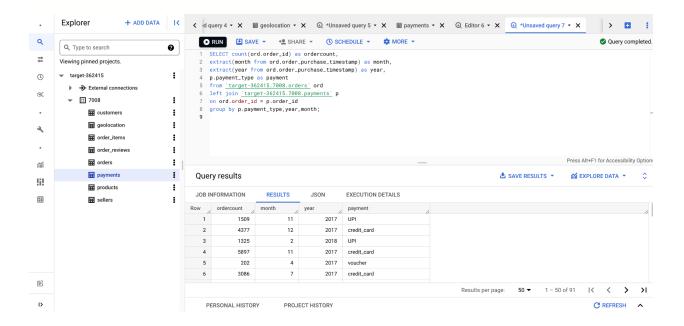
```
select round(avg(date_diff(o.order_estimated_delivery_date,
o.order_delivered_customer_date, DAY)),2) as diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders` o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers`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



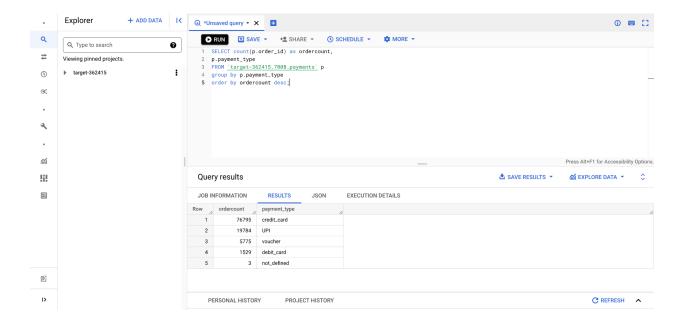
```
select round(avg(date_diff(o.order_estimated_delivery_date,
o.order_delivered_customer_date, DAY)),2) as diff_estimated_delivery,
c.customer_state as state
FROM `target-362415.7008.orders` o
left join `target-362415.7008.order_items` ordt
on o.order_id = ordt.order_id
left join `target-362415.7008.customers`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



```
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;
```



```
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.