**Business Case: Target SQL**

**Context:**

Target is a globally renowned brand and a prominent retailer in the United States. Target makes itself a preferred shopping destination by offering outstanding value, inspiration, innovation and an exceptional guest experience that no other retailer can deliver.

This particular business case focuses on the operations of Target in Brazil and provides insightful information about 100,000 orders placed between 2016 and 2018. The dataset offers a comprehensive view of various dimensions including the order status, price, payment and freight performance, customer location, product attributes, and customer reviews.

By analyzing this extensive dataset, it becomes possible to gain valuable insights into Target's operations in Brazil. The information can shed light on various aspects of the business, such as order processing, pricing strategies, payment and shipping efficiency, customer demographics, product characteristics, and customer satisfaction levels.

**Problem Statement:**

Assuming you are a data analyst/ scientist at Target, you have been assigned the task of analyzing the given dataset to extract valuable insights and provide actionable recommendations.

**What does 'good' look like?**

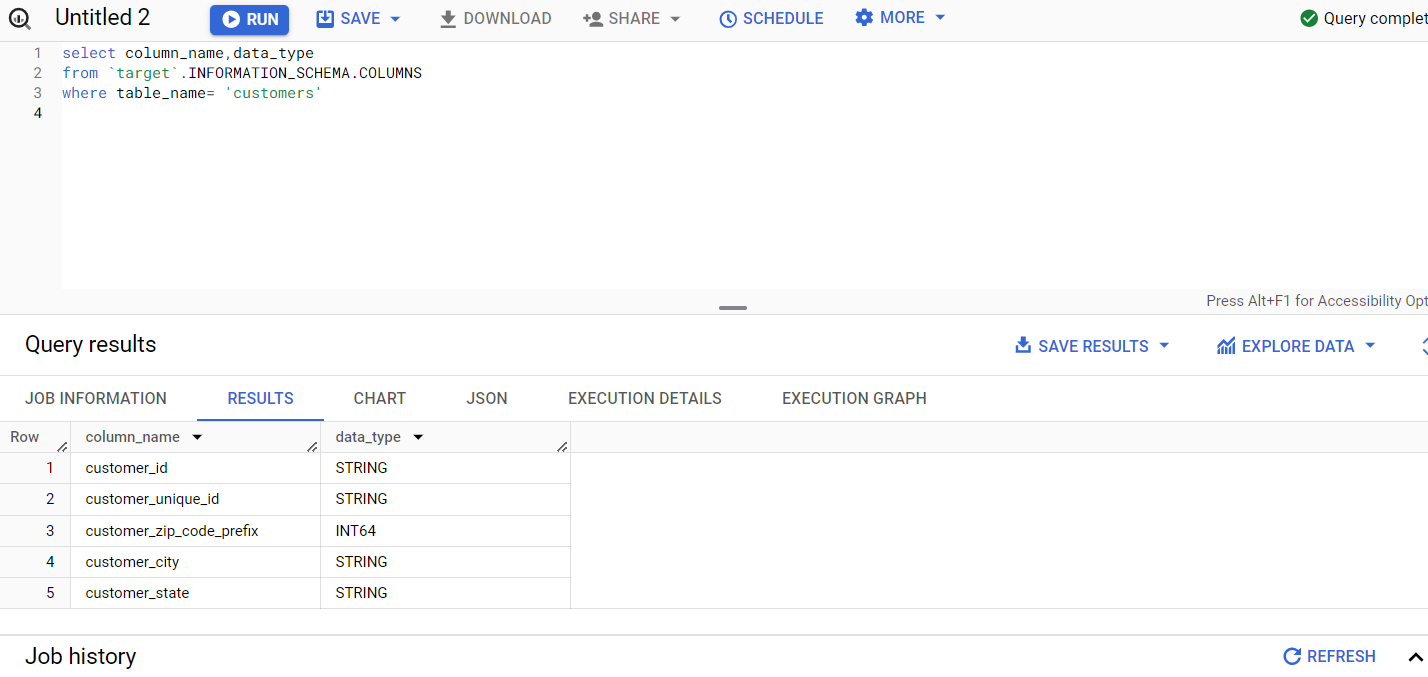
1. **Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:**
   1. Data type of all columns in the "customers" table.
   2. Get the time range between which the orders were placed.
   3. Count the Cities & States of customers who ordered during the given period.

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

select column\_name,data\_type

from `target`.INFORMATION\_SCHEMA.COLUMNS

where table\_name= 'customers'



**Insight**

Here we can see that customer\_id and customer\_unique\_id is STRING as it needs to be unique.

Customer\_city and customer\_state are also string

Customer\_zip\_code\_prefix is INT as it stored the address zip code of customers

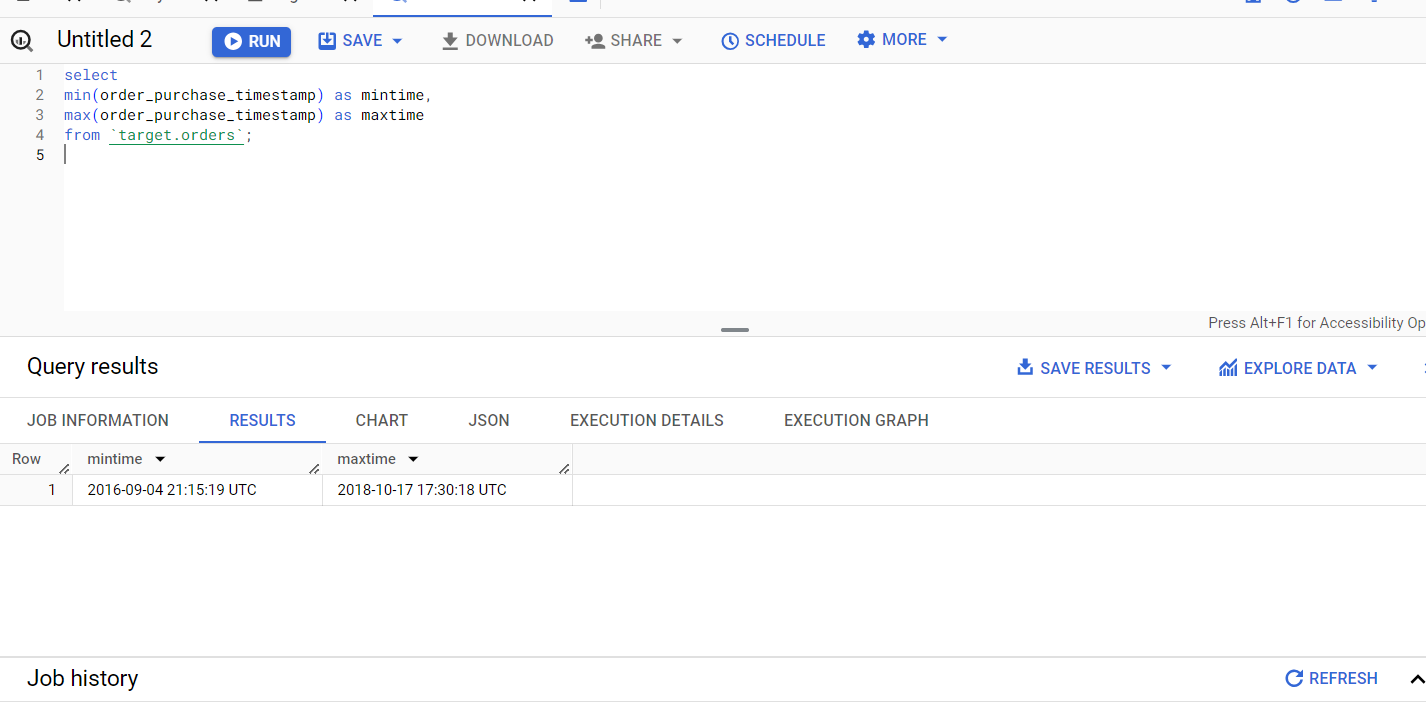
2.Get the time range between which the orders were placed

select

min(order\_purchase\_timestamp) as mintime,

max(order\_purchase\_timestamp) as maxtime

from `target.orders`;



**Insight**

Here using this query to find the date and time of the first and last order placed, which span approximately two years in the given data. The time range is important for overseeing sales performance, identifying trends, peak periods, optimizing seasonal management, marketing strategies, promotions, analyzing customer patterns and behavior, order fulfillment rates, feedback, sales growth, customer reviews, and customer support response times. It can also help identify seasonal trends and preferences, guiding predictions of future demand to improve overall customer satisfaction and achieve operational efficiency.

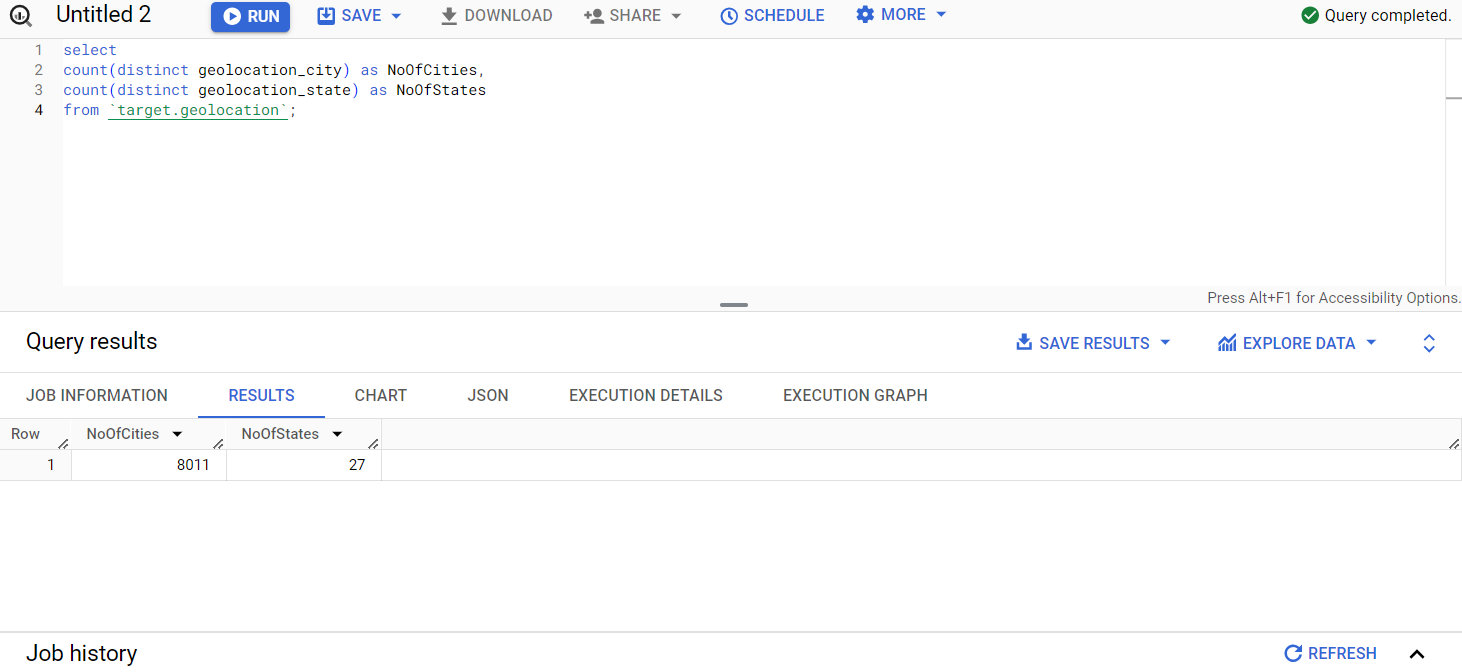
3.Count the number of Cities and States in our dataset

select

count(distinct geolocation\_city) as NoOfCities,

count(distinct geolocation\_state) as NoOfStates

from `target.geolocation`;



**Insight**

Target has a significant presence across various regions in Brazil, indicating a strong market position and a widespread customer base. However, there are around 115 cities missing in the count of unique cities and states in the geolocation dataset. This information is crucial for conducting geographic analysis, identifying opportunities, understanding customer behavior based on location, evaluating regional performance, and making informed decisions regarding resource allocation and strategic planning.

**2.In-depth Exploration:**

* 1. Is there a growing trend in the no. of orders placed over the past years?
  2. Can we see some kind of monthly seasonality in terms of the no. of orders being placed?
  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

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

with cte as

(

select

concat(year,'-',mnth)as period,

sum(no\_of\_orders)as total\_orders,

lag(sum(no\_of\_orders)) over(order by year,mnth)as lagg

from

(

select

extract(year from order\_purchase\_timestamp)as year,

extract(month from order\_purchase\_timestamp)as mnth,

count(\*)as no\_of\_orders

from `target.orders`

where order\_status not in ('canceled','unavailable')

and extract(year from order\_purchase\_timestamp) >= 2017

group by order\_purchase\_timestamp,year,mnth

order by year,mnth

)

group by year,mnth

order by year,mnth

)

select

period,

total\_orders,

Growth\_trend

from

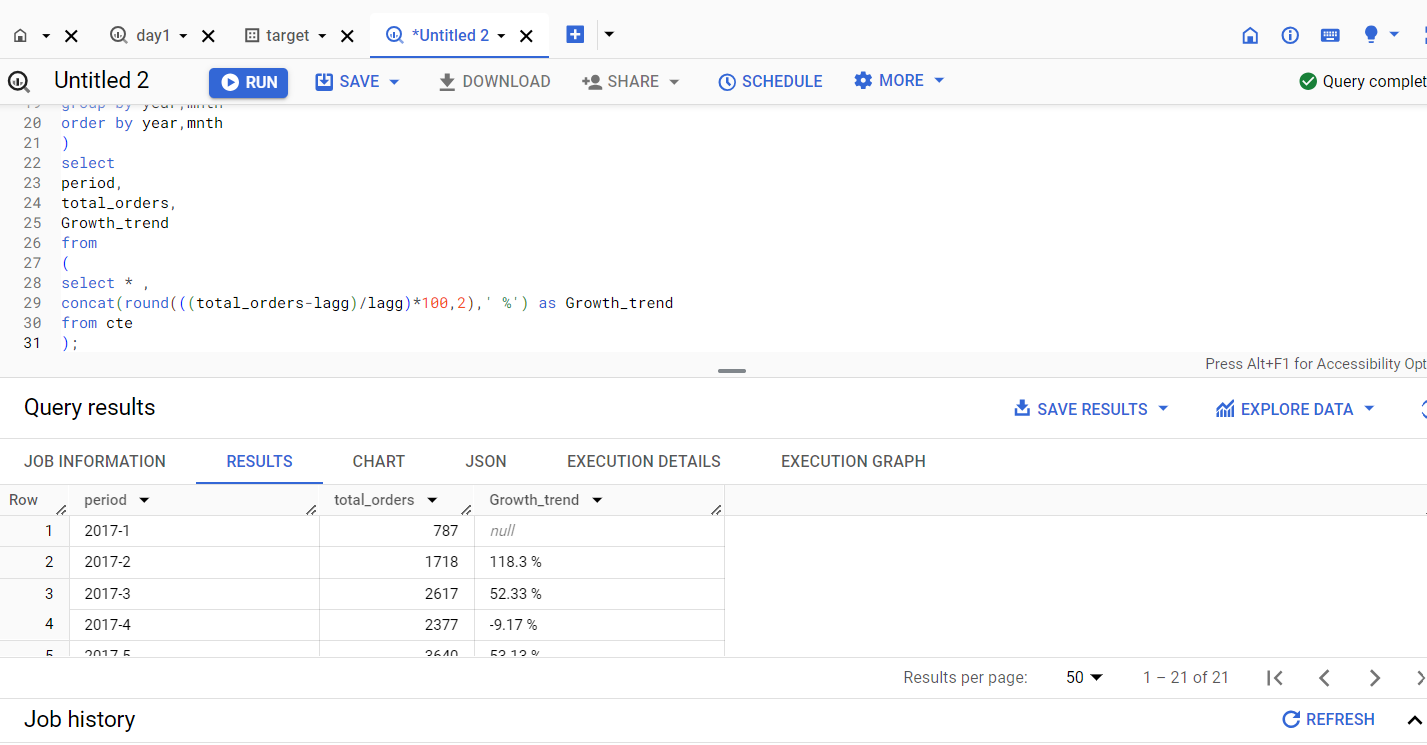
(

select \* ,

concat(round(((total\_orders-lagg)/lagg)\*100,2),' %') as Growth\_trend

from cte

);



**Insight**

Based on the query analysis, it is evident that there was substantial order growth between 2016 and 2017, followed by a decrease in growth rate in 2018. This analysis helps identify patterns and trends in order placements over time.

Overall, there is a consistent and significant upward trend in the number of orders. However, based on the revenue, there is a leap trend over the Brazil region, indicating positive business performance and increasing customer demand for Target's products. However, Target must work on revenue-based strategies to gain an uptrend.

Top of Form

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

select

concat(period,' -- is ',mth)as monthly\_season,

Max\_orders

from

(

select

concat(year,' - ',mnth)as period,

sum(no\_of\_orders)as Max\_orders,

case

when mnth=1 then 'January'

when mnth=2 then 'Feburary'

when mnth=3 then 'March'

when mnth=4 then 'April'

when mnth=5 then 'May'

when mnth=6 then 'June'

when mnth=7 then 'July'

when mnth=8 then 'August'

when mnth=9 then 'September'

when mnth=10 then 'October'

when mnth=11 then 'November'

else 'December'

end as mth

from

(

select

extract(year from order\_purchase\_timestamp)as year,

extract(month from order\_purchase\_timestamp)as mnth,

count(\*) as no\_of\_orders,

from `target.orders`

where order\_status <>'canceled'

group by order\_purchase\_timestamp,year,mnth

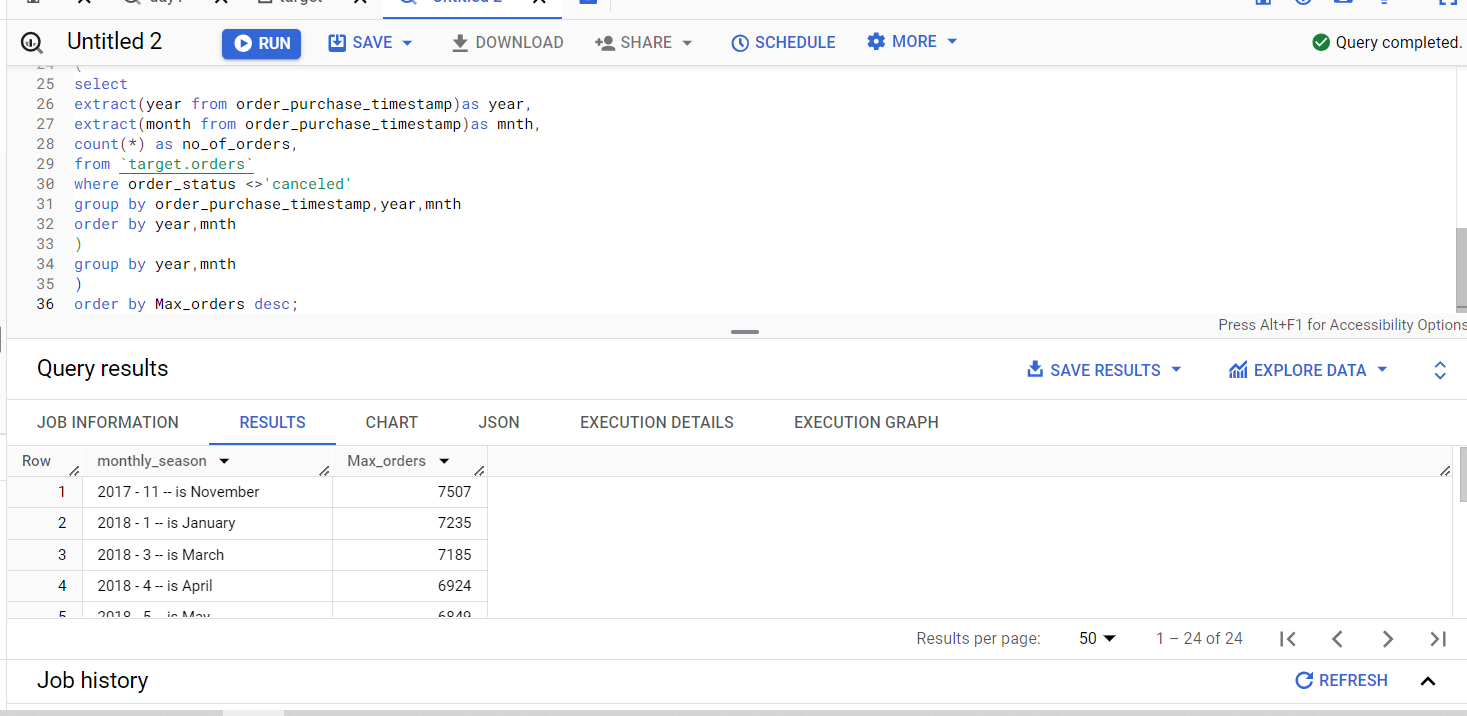
order by year,mnth

)

group by year,mnth

)

order by Max\_orders desc;



**Insight**

The objective of the query was to calculate the total count of orders for each month over the years to understand and manage the business's order patterns.

There is a clear monthly seasonal pattern in the number of orders, with increases and decreases throughout the year, perhaps influenced by holidays or specific shopping periods.The peak is in November 2017, showing the highest number of orders, while September 2016 has the lowest number of orders. Comparing with world events and Brazilian culture, the increase in orders during the year-end can be attributed to welcoming the new year and the materialization of home appliances and needs. November 2017 had the Black Friday sale, and during the FIFA World Cup 2018, there was an increase in orders related to soccer games. These events contributed to the peak in orders during these periods, demonstrating monthly seasonality hotspots.Comparing the number of orders across different years reveals growth and troughs in the number of orders. There was significant growth from 2016 to 2017, but a decrease in average orders in 2018 compared to the previous year.

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

with cte as

(

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 'MORNING'

when extract(hour from order\_purchase\_timestamp)

between 13 and 18 then 'AFTERNOON'

else 'NIGHT'

end as TIME\_OF\_DAY,

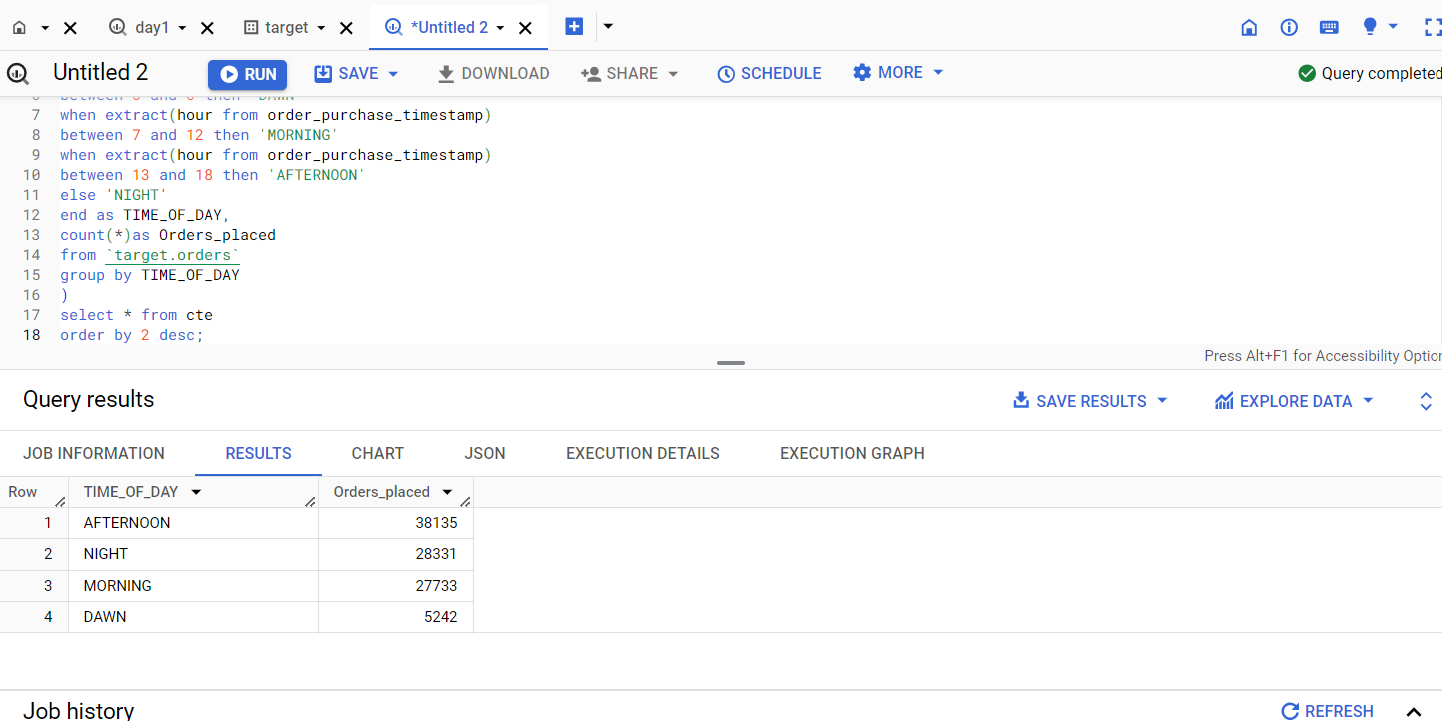
count(\*)as Orders\_placed

from `target.orders`

group by TIME\_OF\_DAY)

select \* from cte

order by 2 desc;



**Insight**

Analyzing the data allows us to determine the time of day when Brazilian clients often place their orders. This insight into their ordering habits and preferences can be valuable for scheduling customer assistance personnel or launching focused marketing efforts during the busiest ordering periods. For example, we find that Brazilian clients frequently order more in the afternoons, indicating that this is a period when people want to shop online. Additionally, customers tend to buy the least during dawn. This information can help us operate more efficiently and cater to customer needs more effectively.

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

* 1. Get the month on month no. of orders placed in each state.

select yr\_mnth,state,

min(monthly\_orders) over(partition by yr\_mnth)as Min\_orders\_recd,

max(monthly\_orders) over(partition by yr\_mnth) as Max\_orders\_recd,

round(avg(monthly\_orders) over(partition by state order by yr\_mnth),3)as Avg\_monthly\_order,

monthly\_orders as Total\_ordersPerMonth,

sum(monthly\_orders) over(partition by state)as Monthly\_state\_order

from

(

select customer\_state as state,

format\_timestamp("%Y-%m",order\_purchase\_timestamp) as yr\_mnth,

count(\*)as monthly\_orders

from `target.customers` c

join `target.orders` o

on c.customer\_id=o.customer\_id

where order\_status not in ('canceled','unavailable')

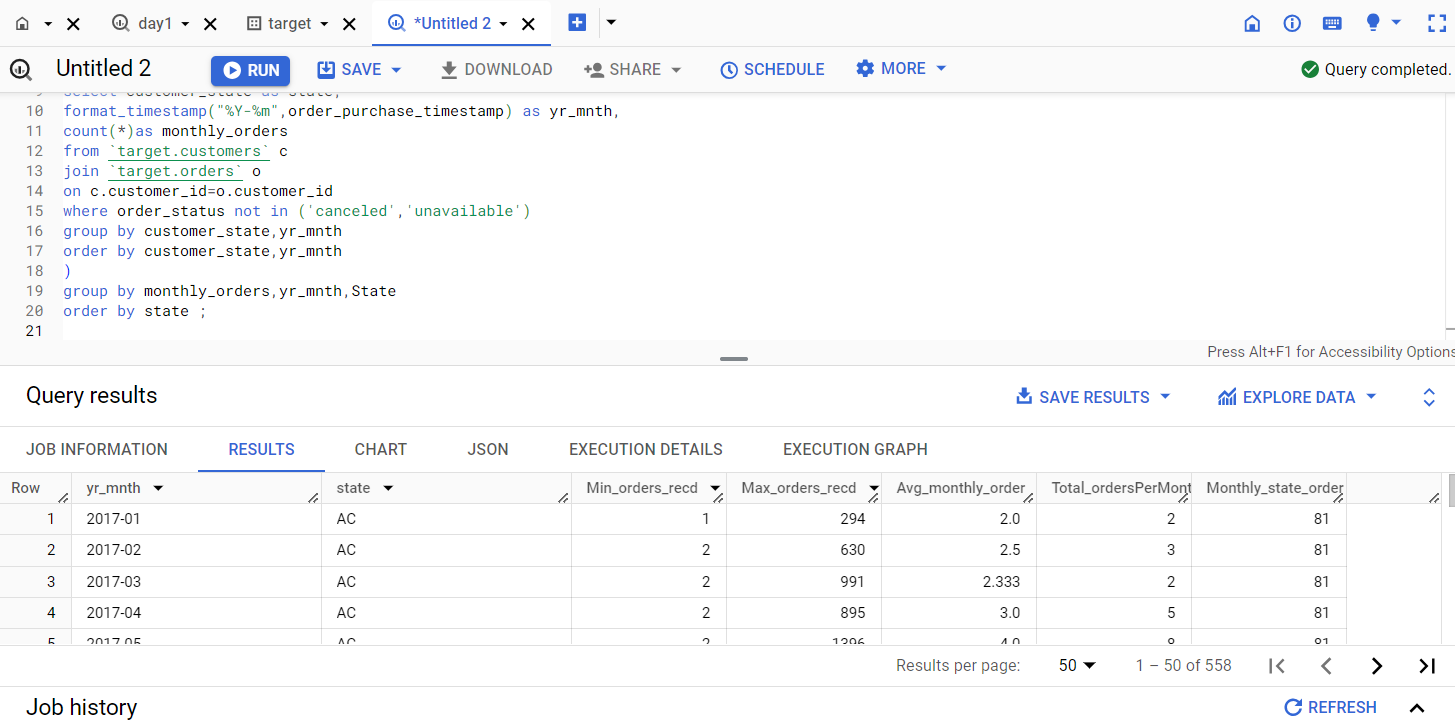
group by customer\_state,yr\_mnth

order by customer\_state,yr\_mnth

)

group by monthly\_orders,yr\_mnth,State

order by state ;



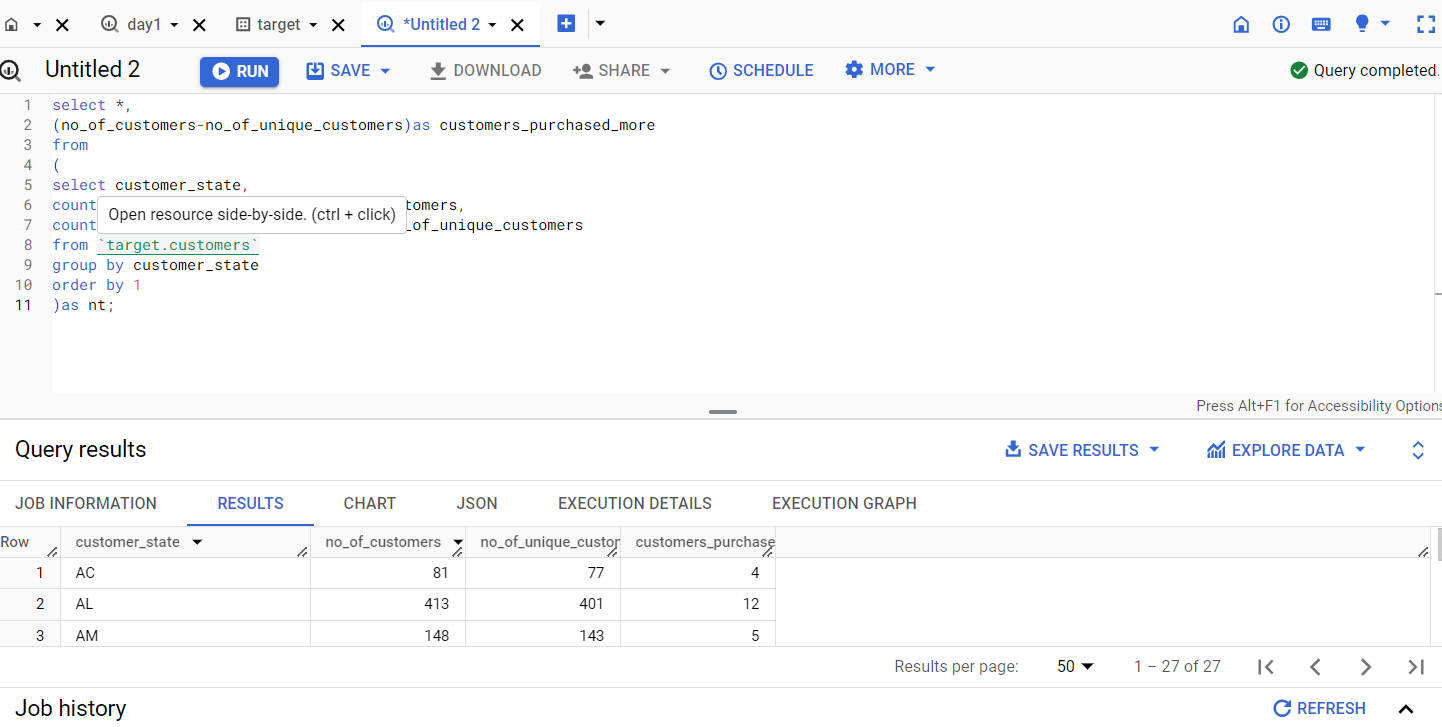
**Insight**

Analyzing the monthly order count for each state allows us to spot trends, patterns, or seasonality in order volume over time. We can identify which states consistently have high order volumes and pinpoint any months or states where order counts have significantly changed. In our data, we observe that the state called SP consistently has the highest number of orders every month.

These insights enable us to target marketing efforts in states with rising order volumes, identify potential operational issues in states with falling order volumes, and optimize inventory management based on order trends across different states.

* 1. How are the customers distributed across all the states?

1. select \*,
2. (no\_of\_customers-no\_of\_unique\_customers)as customers\_purchased\_more
3. from
4. (
5. select customer\_state,
6. count(distinct customer\_id)as no\_of\_customers,
7. count(distinct customer\_unique\_id)as no\_of\_unique\_customers
8. from `target.customers`
9. group by customer\_state
10. order by 1
11. )as nt;

****

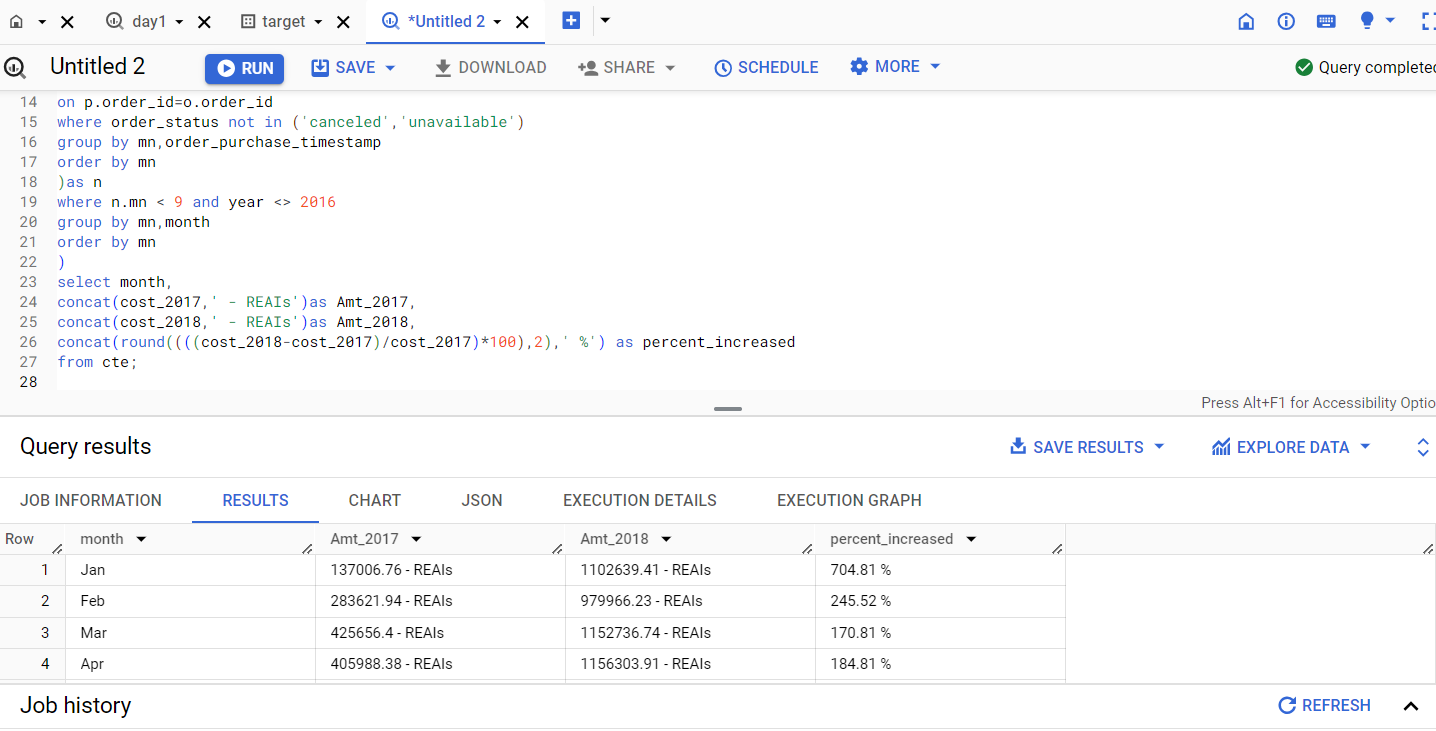
**Insight**

The query provides insights into the total number of customers, the number of customers who made more than one purchase, and the total orders placed state-wise. This information is valuable for geographical analysis.

**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.

1. with cte as (
2. select month,
3. round(sum(if(year=2018,amt,0)),2) as cost\_2018,
4. round(sum(if(year=2017,amt,0)),2) as cost\_2017
5. from
6. (
7. select
8. extract(year from order\_purchase\_timestamp)as year,
9. extract(month from order\_purchase\_timestamp)as mn,
10. format\_timestamp("%b",order\_purchase\_timestamp) as month,
11. sum(p.payment\_value)as amt
12. from `target.payments` p
13. join `target.orders` o
14. on p.order\_id=o.order\_id
15. where order\_status not in ('canceled','unavailable')
16. group by mn,order\_purchase\_timestamp
17. order by mn
18. )as n
19. where n.mn < 9 and year <> 2016
20. group by mn,month
21. order by mn
22. )
23. select month,
24. concat(cost\_2017,' - REAIs')as Amt\_2017,
25. concat(cost\_2018,' - REAIs')as Amt\_2018,
26. concat(round((((cost\_2018-cost\_2017)/cost\_2017)\*100),2),' %') as percent\_increased
27. from cte;



with cte as (

select

round(sum(if(year=2018,amt,0)),2) as cost\_2018,

round(sum(if(year=2017,amt,0)),2) as cost\_2017

from

(

select

extract(year from order\_purchase\_timestamp)as year,

extract(month from order\_purchase\_timestamp)as mn,

sum(p.payment\_value)as amt

from `target.payments` p

join `target.orders` o

on p.order\_id=o.order\_id

group by order\_purchase\_timestamp

order by year

)as n

where n.mn < 9 and year <> 2016

)

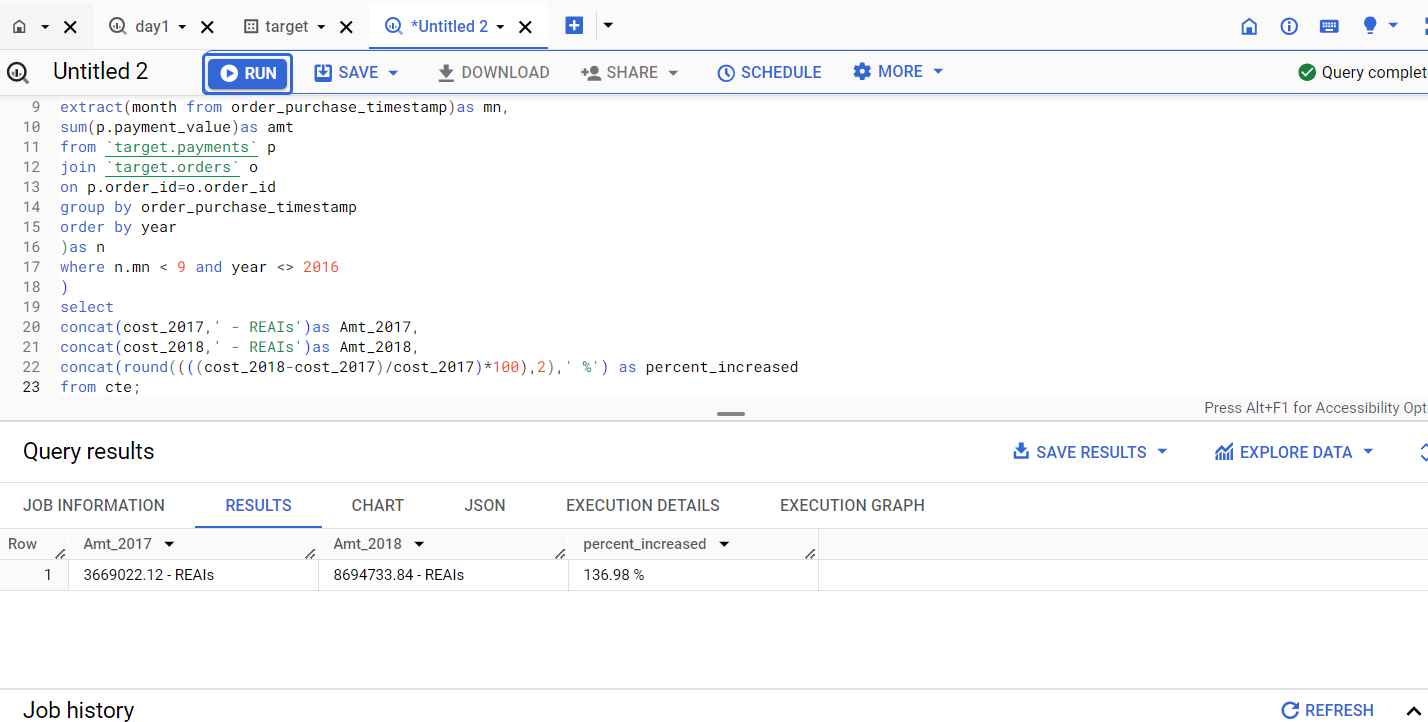
select

concat(cost\_2017,' - REAIs')as Amt\_2017,

concat(cost\_2018,' - REAIs')as Amt\_2018,

concat(round((((cost\_2018-cost\_2017)/cost\_2017)\*100),2),' %') as percent\_increased

from cte;



**Insight**

The data indicates a positive trend of increasing order amounts from 2017 to 2018, which suggests economic growth and potential success in the business across Brazilian states. The substantial decrease in the cost of products also contributes to this understanding.

Understanding these market dynamics is crucial for making changes and replicating successful strategies in the future. These insights help in understanding seasonal patterns and growth trends, supporting strategic planning and decision-making processes.

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

select customer\_state ,

concat(round(sum(payment\_value),2),' REAIs') as Total\_amount,

concat(round(avg(payment\_value),2),' REAIs') as Avg\_amt

from `target.customers` c

join `target.orders` o

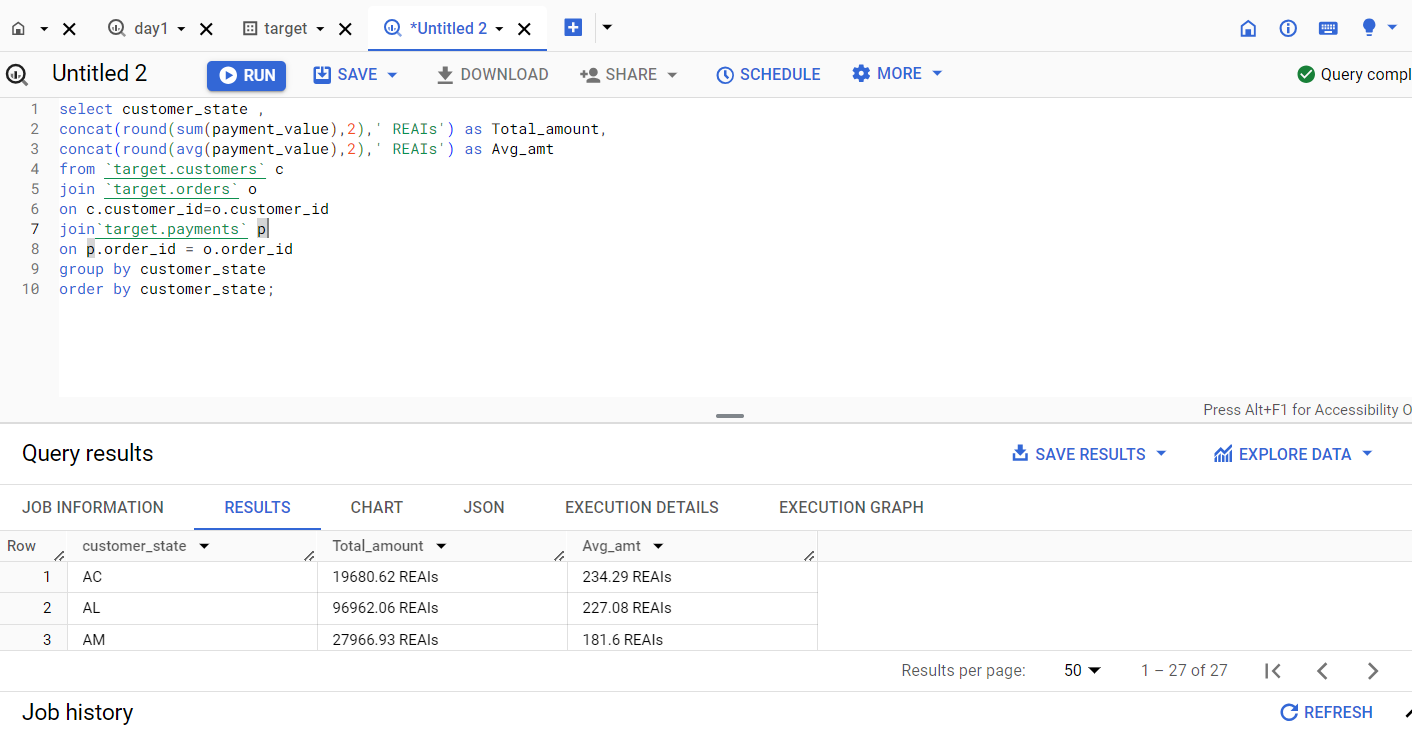
on c.customer\_id=o.customer\_id

join`target.payments` p

on p.order\_id = o.order\_id

group by customer\_state

order by customer\_state;



**Insight**

The analysis indicates significant variation in total and average order prices across different states. São Paulo (SP) and Rio de Janeiro (RJ) have higher total prices, suggesting larger revenue contributions, while Roraima (RR) and Amapá (AP) have relatively lower total prices.

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

select customer\_state ,

concat(round(sum(freight\_value),2),' REAIs') as Total\_freight\_value,

concat(round(avg(freight\_value),2),' REAIs') as Avg\_freight\_value

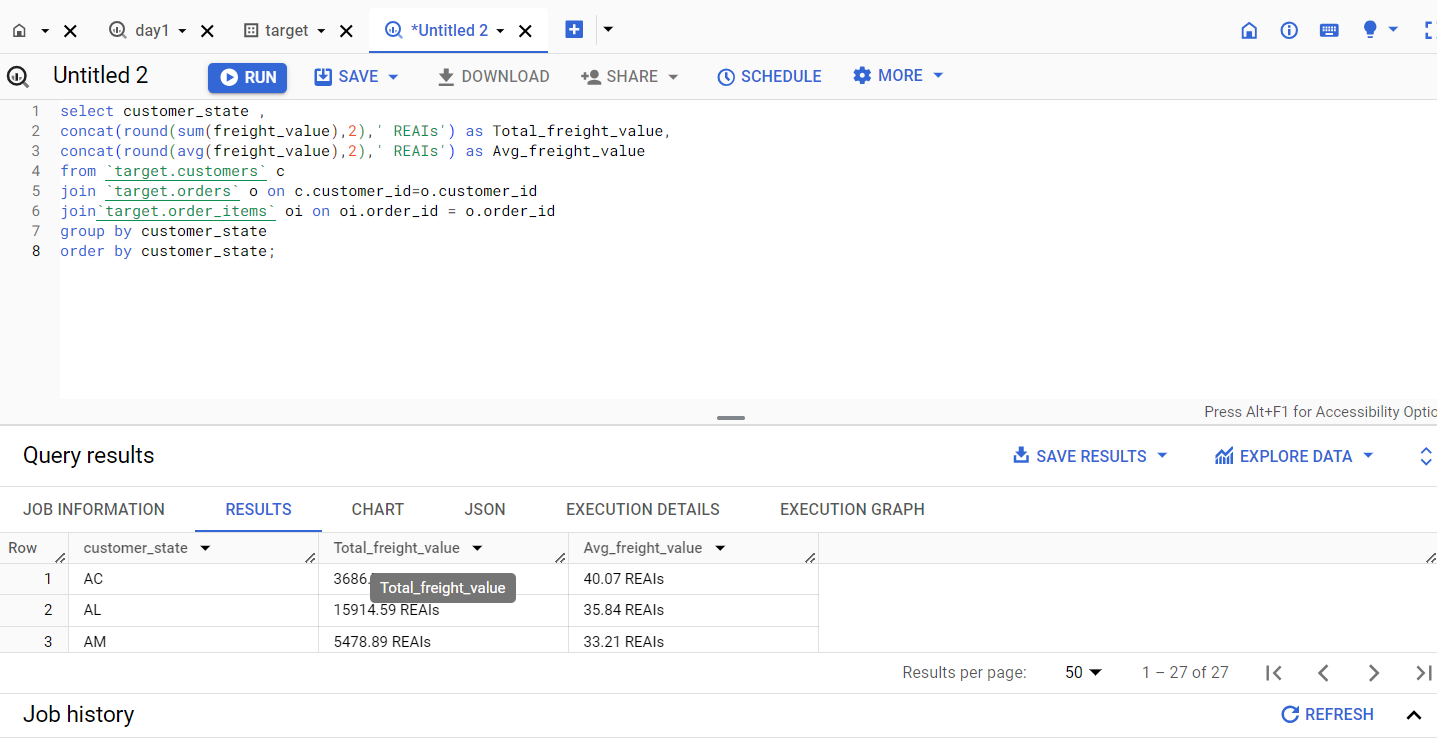
from `target.customers` c

join `target.orders` o on c.customer\_id=o.customer\_id

join`target.order\_items` oi on oi.order\_id = o.order\_id

group by customer\_state

order by customer\_state;

****

**Insight**

Average freight values reflect the average shipping cost per order, with states like São Paulo (SP) and Minas Gerais (MG) having lower average freight values, suggesting more efficient logistics networks or economies of scale. On the other hand, states like Paraíba (PB) and Piauí (PI) have higher average freight values, indicating relatively higher shipping costs per order. These insights can assist in optimizing logistics operations and understanding regional shipping trends.

**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\_delivered\_customer\_date - order\_estimated\_delivery\_date

select

order\_id,

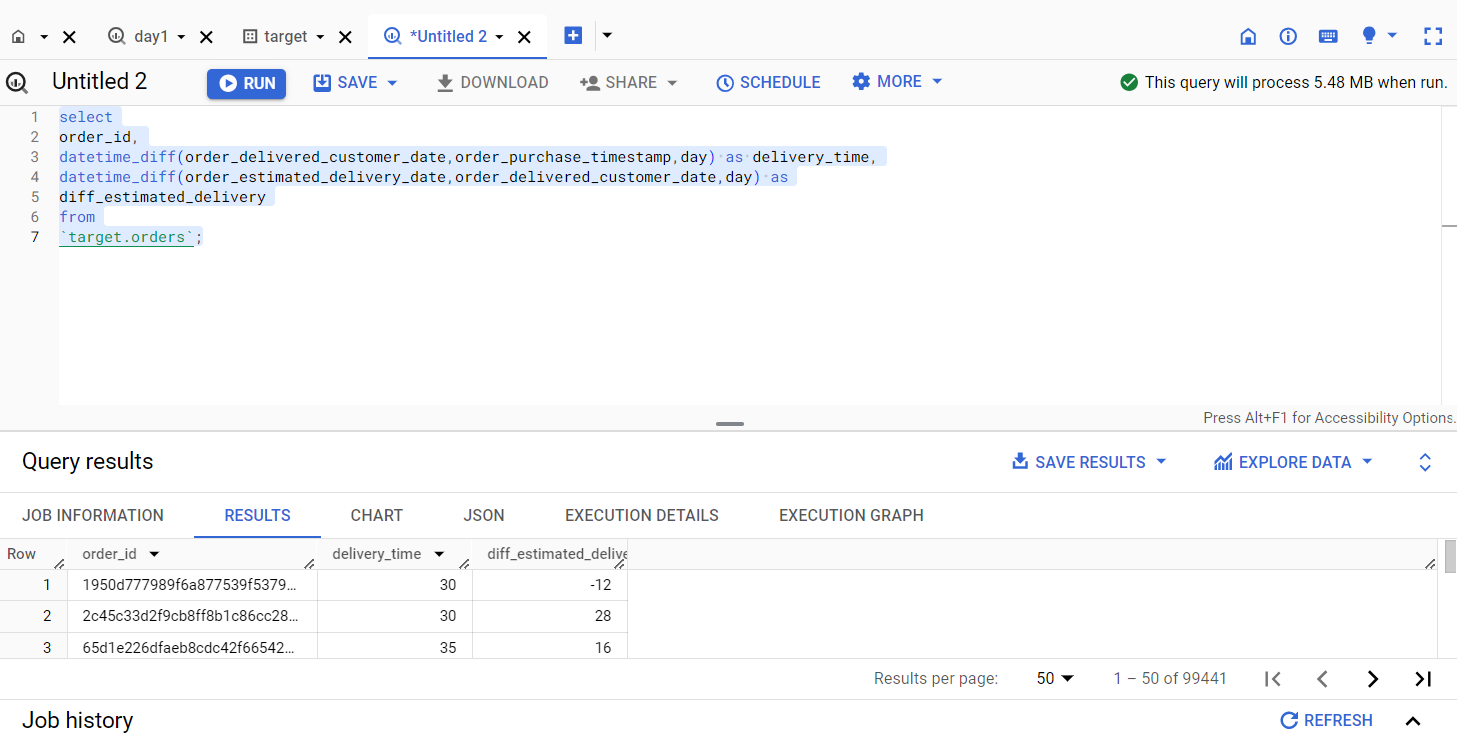
datetime\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) as delivery\_time,

datetime\_diff(order\_estimated\_delivery\_date,order\_delivered\_customer\_date,day) as

diff\_estimated\_delivery

from

`target.orders`;



**Insight**

From Query Analysis, The "Time\_to\_deliver" column represents the number of days taken to deliver an order to the customer from the purchase date, while the "diff\_estimated\_delivery" column indicates the difference between the estimated delivery date and the actual delivery date. By analyzing this data, we can identify orders that took longer to deliver and compare each delivery time with the average delivery timeline to assess delivery efficiency. Negative values in the "diff\_estimated\_delivery" column indicate delayed deliveries, while positive values indicate early deliveries. Digging deeper, the reasons for these variances can help improve delivery timelines and reduce the difference between estimated and actual delivery dates, leading to enhanced logistics and delivery processes.

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

select state ,avg\_freight\_value from

(

(

select concat('HIGH # ',customer\_state) as state,

max(freight\_value) as High\_freight\_value,

concat(round(avg(freight\_value),2),' REAIs') as avg\_freight\_value

from `target.customers` c

join `target.orders` o on c.customer\_id=o.customer\_id

join `target.order\_items` as p on o.order\_id = p.order\_id

group by customer\_state

order by 3 desc

limit 5

)

union all

(

select concat('LOW # ',customer\_state) as state,

min(freight\_value) as low\_freight\_value,

concat(round(avg(freight\_value),2),' REAIs') as avg\_freight\_value

from `target.customers` c

join `target.orders` o on c.customer\_id=o.customer\_id

join `target.order\_items` as p on o.order\_id = p.order\_id

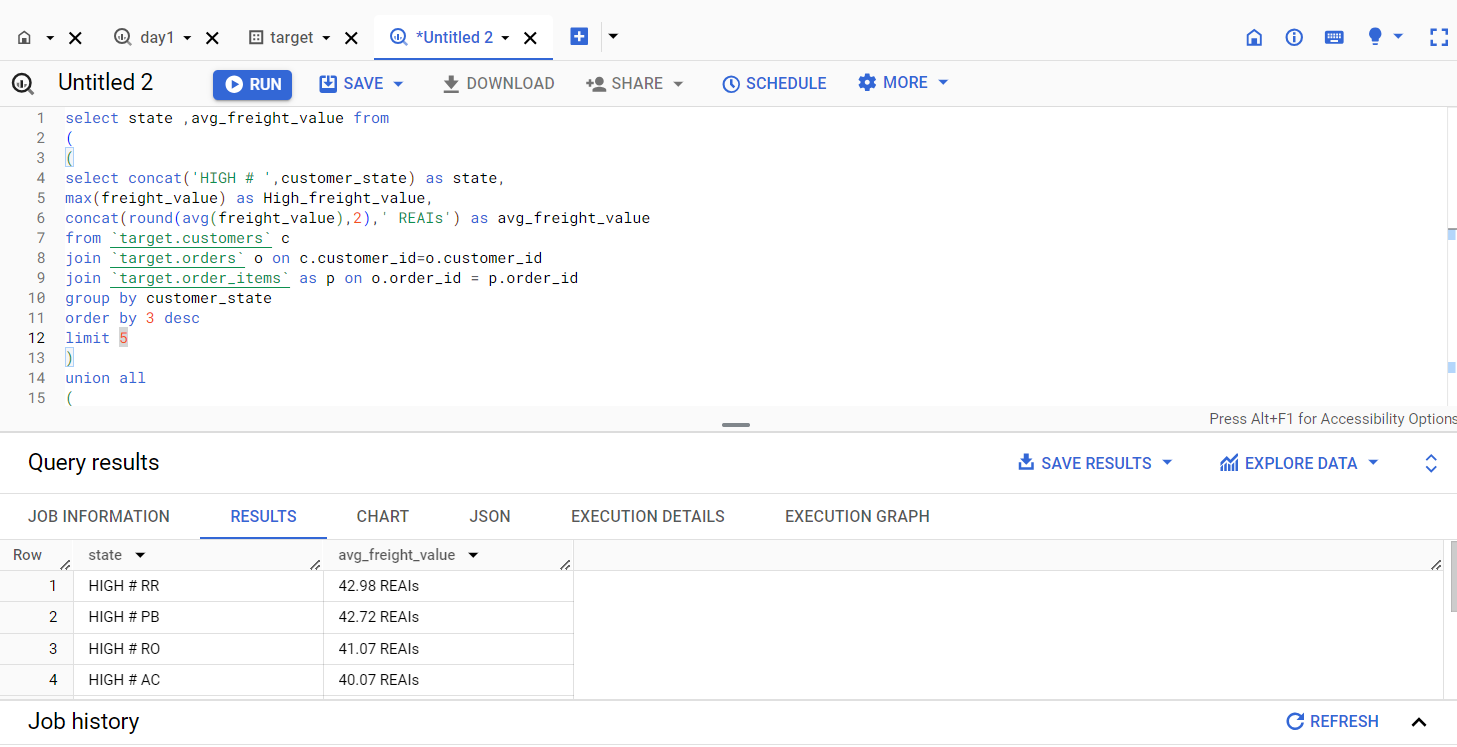
group by customer\_state

order by 3 asc

limit 5

)

)as t;



**Insight**

Certain states exhibit higher average freight values, indicating specific characteristics or logistical challenges that result in increased freight costs. Some states demonstrate lower average freight values, suggesting more favorable logistics infrastructure or other factors contributing to reduced freight costs. With average freight costs, we can that there is a notable difference (Significant Variation) in the average freight value among different states. These insights provide businesses with valuable information for optimizing their shipping and logistics operations. It enables them to make informed decisions regarding pricing strategies, supply chain optimization, and resource allocation.

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

with cte as

(

select state,'FAST'as val,avg(delivery\_time) as avg\_delivery\_time,

dense\_rank() over (order by avg(delivery\_time) desc) as rnk

from

(

select customer\_state as state,

datetime\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) as delivery\_time,

from `target.customers` as c

join `target.orders` as o on c.customer\_id = o.customer\_id

group by state,order\_delivered\_customer\_date,order\_purchase\_timestamp,delivery\_time

) nt1

group by state

union all

select state,'SLOW'as val,avg(delivery\_time) as avg\_delivery\_time,

dense\_rank() over (order by avg(delivery\_time) asc) as rnk

from

(

select customer\_state as state,

datetime\_diff(order\_delivered\_customer\_date,order\_purchase\_timestamp,day) as delivery\_time,

from `target.customers` as c

join `target.orders` as o on c.customer\_id = o.customer\_id

group by state,order\_delivered\_customer\_date,order\_purchase\_timestamp,delivery\_time

) nt2

group by state

)

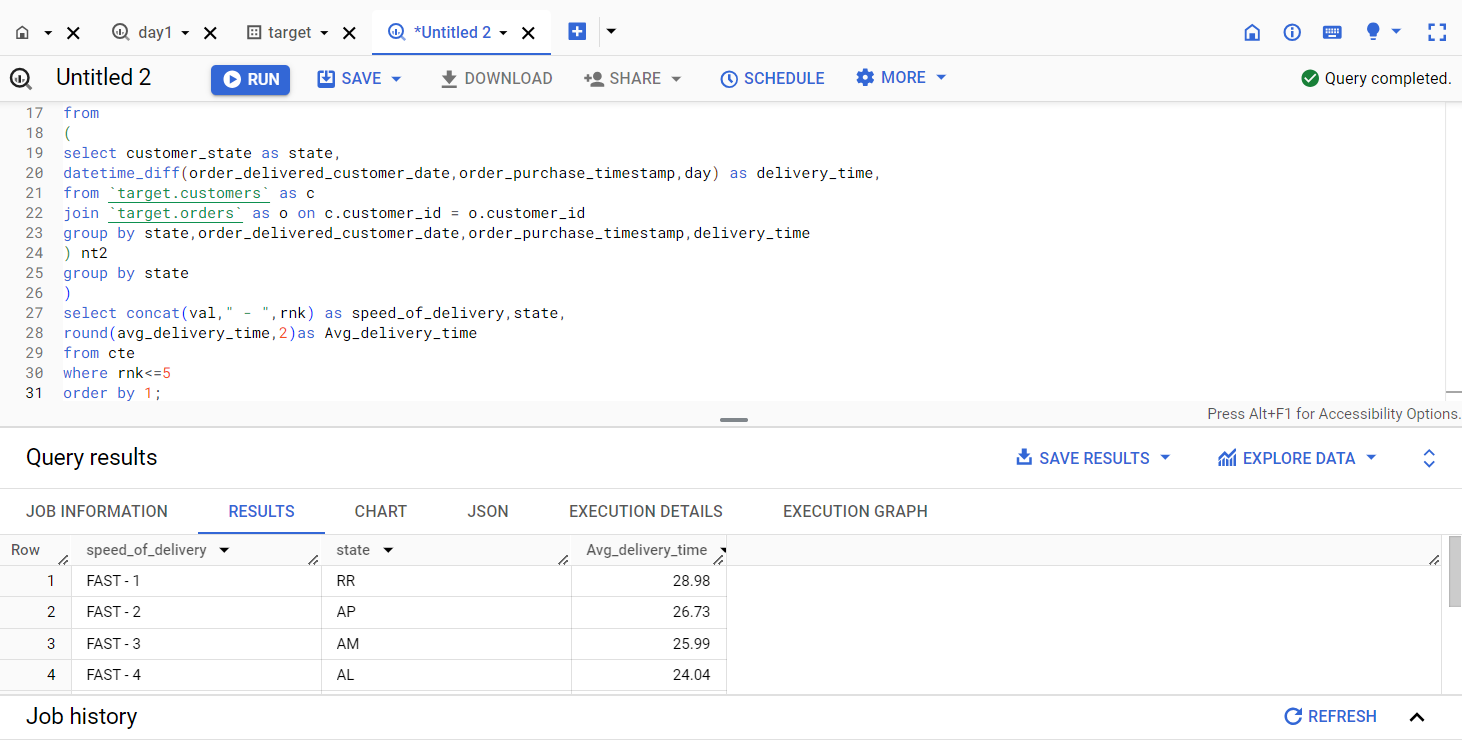
select concat(val," - ",rnk) as speed\_of\_delivery,state,

round(avg\_delivery\_time,2)as Avg\_delivery\_time

from cte

where rnk<=5

order by 1;



**Insight**

Open communication and collaborative relationships with shipping carriers and logistics partners are crucial for monitoring and addressing any issues promptly. By optimizing the last-mile delivery process and maintaining efficient relationships with partners, businesses can reduce delivery times, enhance customer experience, and improve overall satisfaction

4.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.

select customer\_state as state,

round(avg(date\_diff(o.order\_estimated\_delivery\_date,o.order\_delivered\_customer\_date,day)),2)as

avg\_speed\_delivery

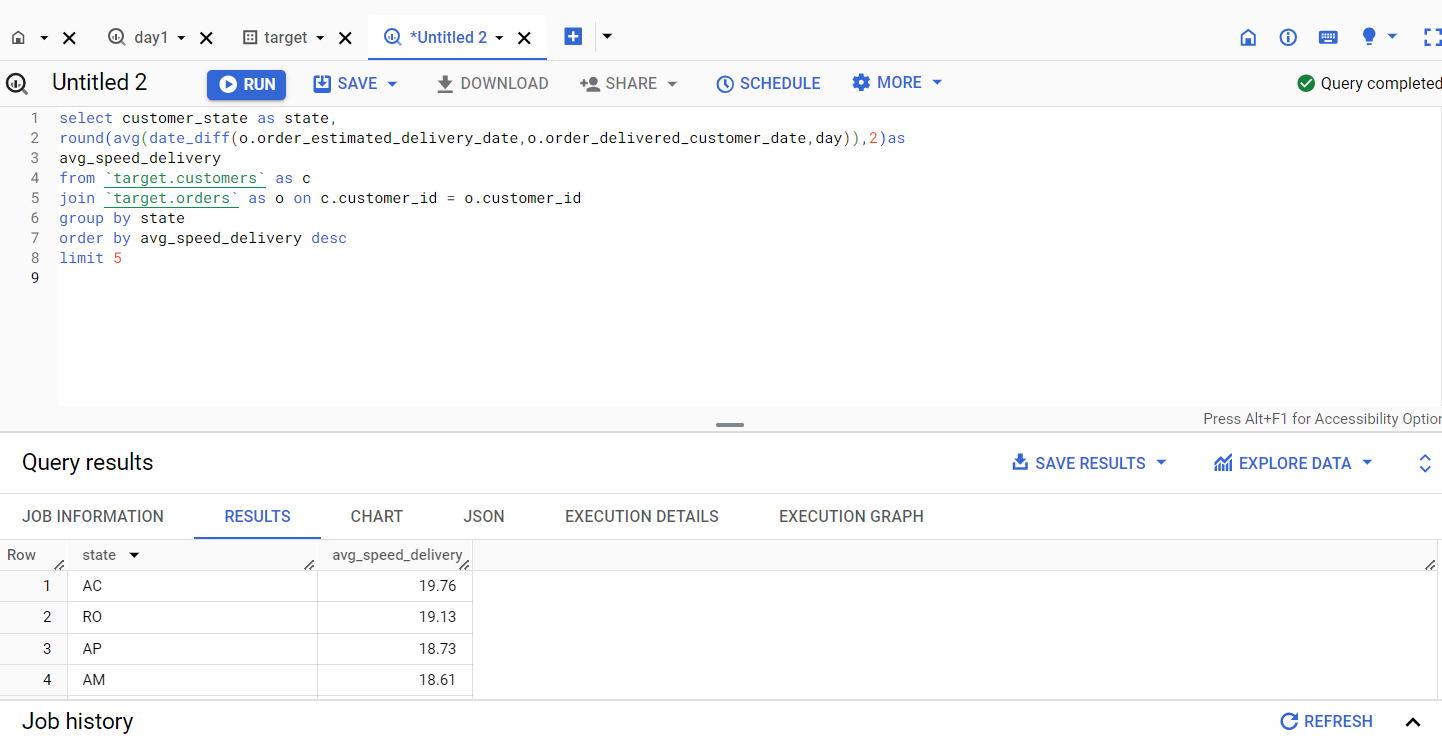
from `target.customers` as c

join `target.orders` as o on c.customer\_id = o.customer\_id

group by state

order by avg\_speed\_delivery desc

limit 5



**Insight**

Our company operating in these states called AC, RO, AP, and AM where average delivery speed is highest can take advantage of the quicker delivery times by highlighting their rapid and dependable service, thereby drawing more clients, and boosting client satisfaction.

These data can help us improve our operations, enhance customer experience, optimize logistics, or look for expansion prospects in areas with a track record of quick order delivery.

**6.Analysis based on the payments:**

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

select ym,

payment\_type,

sum(cnt)as Total\_orders

from

(

select

format\_timestamp("%Y-%m",order\_purchase\_timestamp) as ym,

payment\_type,

count(p.order\_id)as cnt

from `target.payments` p

join `target.orders` o

on p.order\_id=o.order\_id

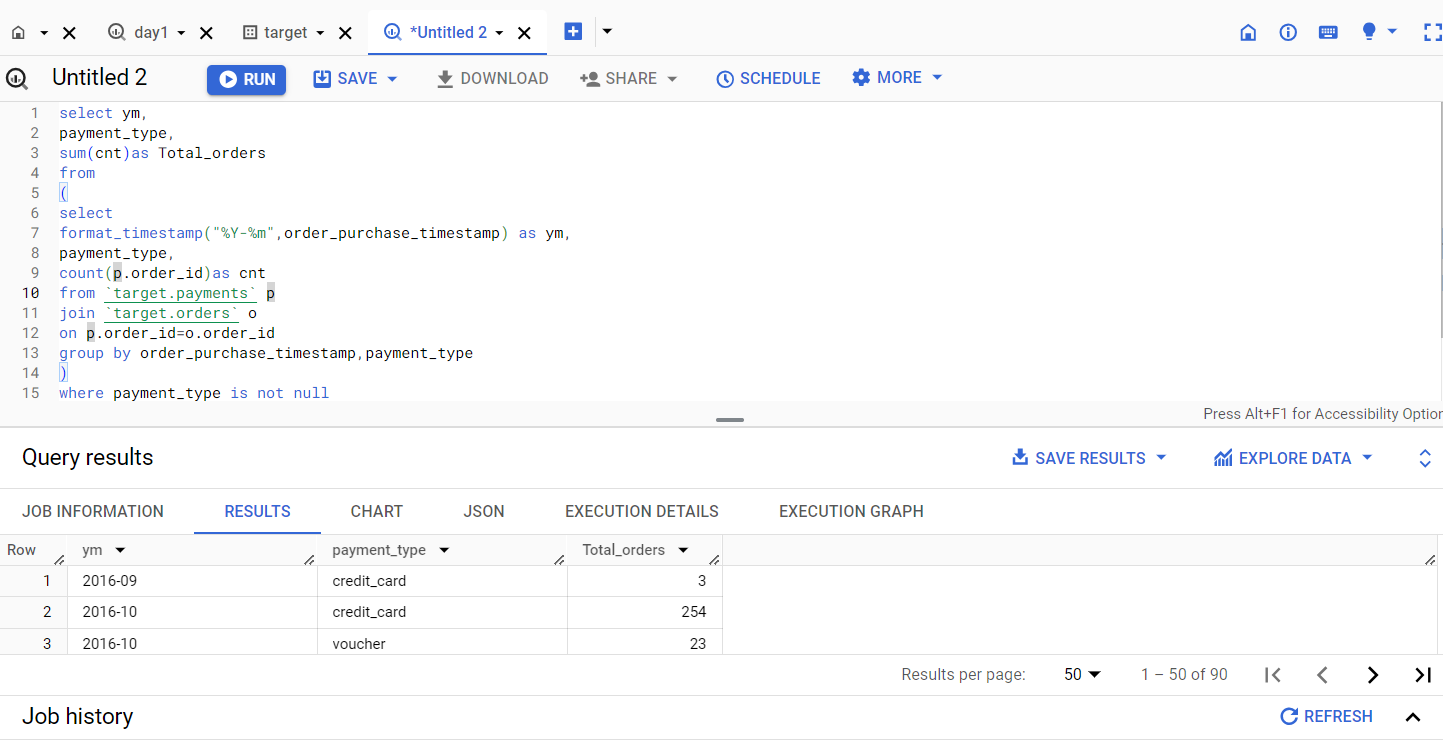
group by order\_purchase\_timestamp,payment\_type

)

where payment\_type is not null

group by payment\_type,ym

order by ym



**Insight**

We identify that credit card as a payment method was most used in November 2017.

To analyze seasonality, identify peak months, or evaluate the effects of marketing efforts or outside variables on consumer behavior, tracking the month-to-month trends in order counts can be helpful.

Based on the payment preferences noticed during various months, these insights might help firms optimize their payment procedures, customize marketing campaigns, or enhance customer experiences.

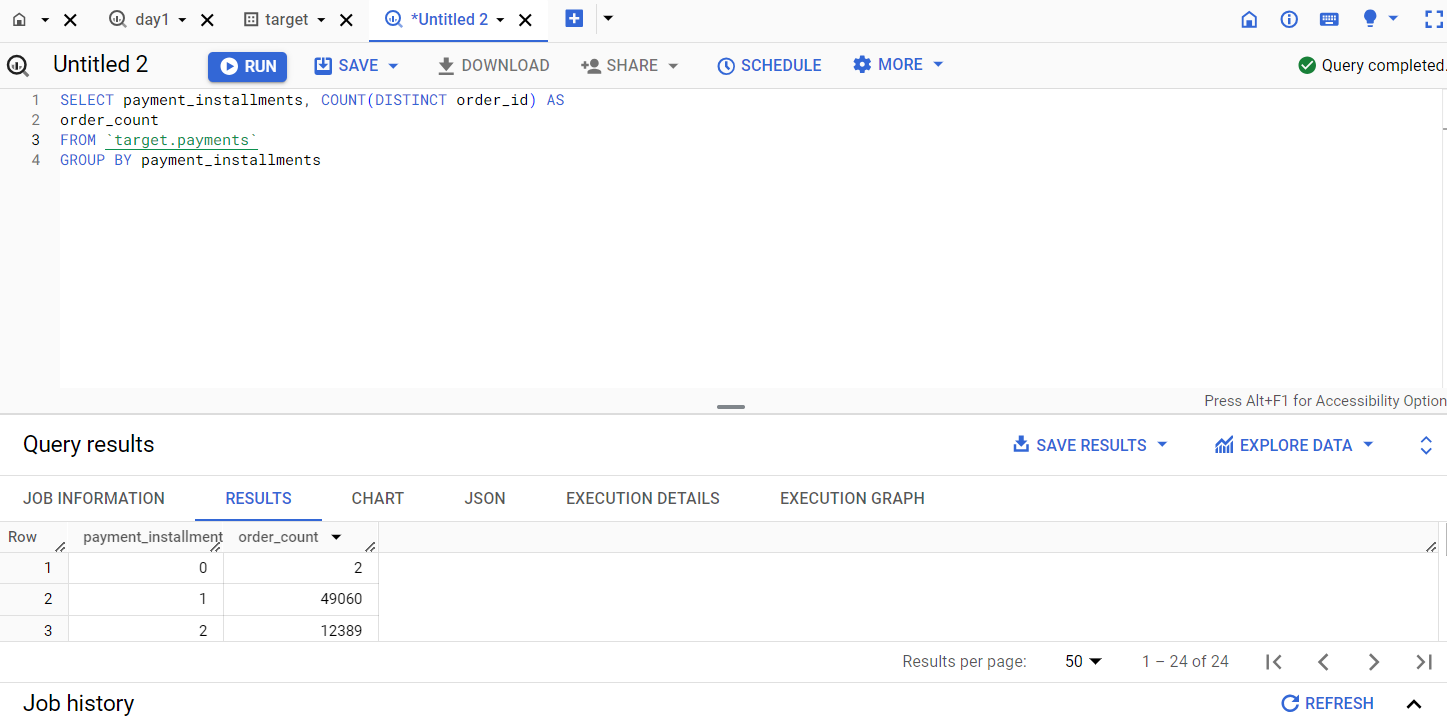
* 1. Find the no. of orders placed on the basis of the payment installments that have been paid.

SELECT payment\_installments, COUNT(DISTINCT order\_id) AS

order\_count

FROM `target.payments`

GROUP BY payment\_installments

****

**Insight**

49060 orders were placed where payment installment was 1.

This analysis can help determine whether payment installment alternatives are popular or preferred by clients.

Customers' preferences for budgeting or financing may be discerned by whether they tend to select a particular

number of payment installments. 4. Monitoring the distribution of orders according to payment installments might reveal information about the buying habits of clients and their preference for flexible payment methods.