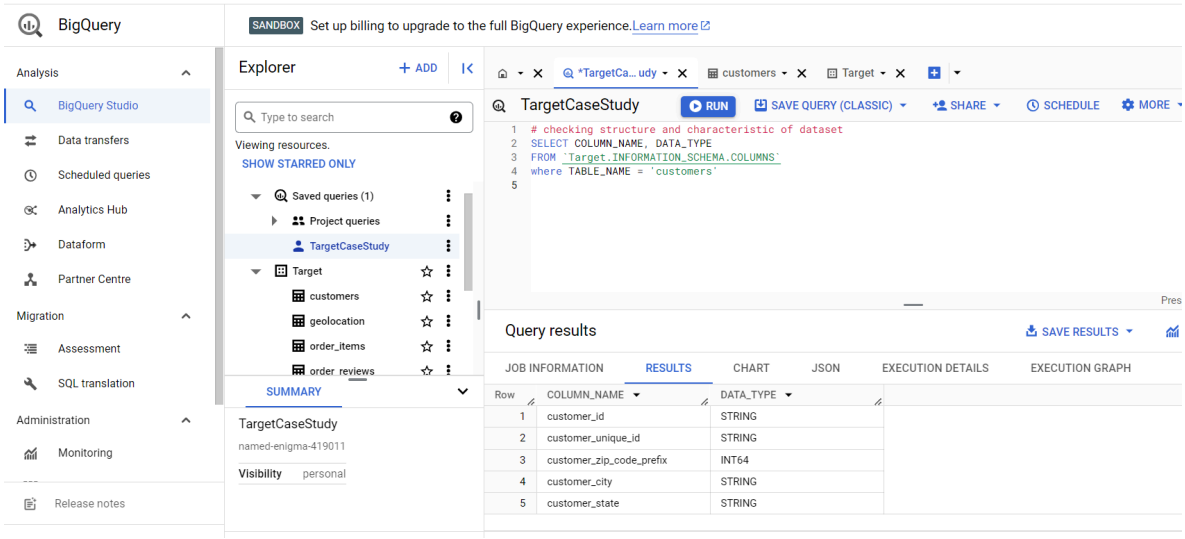
I.What does ‘good’ look like?

A. Data type of all columns in the “customers” table.

Query

```
# checking structure and characteristic of dataset
SELECT COLUMN_NAME, DATA_TYPE
FROM `Target.INFORMATION_SCHEMA.COLUMNS`
where TABLE_NAME = 'customers'
```

Screenshot



Insights and Action Items

1. All columns in this example are of STRING data type except customer_zip_code_prefix which is integer. Customer_id and customer_unique_id contain a unique identifier for each customer.
2. Customer_city column stores the name of the city where each customer resides.
3. This information could be used for demographic analysis, targeting customers based on location, or for address validation purposes.
4. In the customer_state column, the string data type indicates that state names could be alphanumeric and of varying lengths.
5. This information could be valuable for regional analysis, compliance purposes, or marketing campaigns targeting specific states.

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

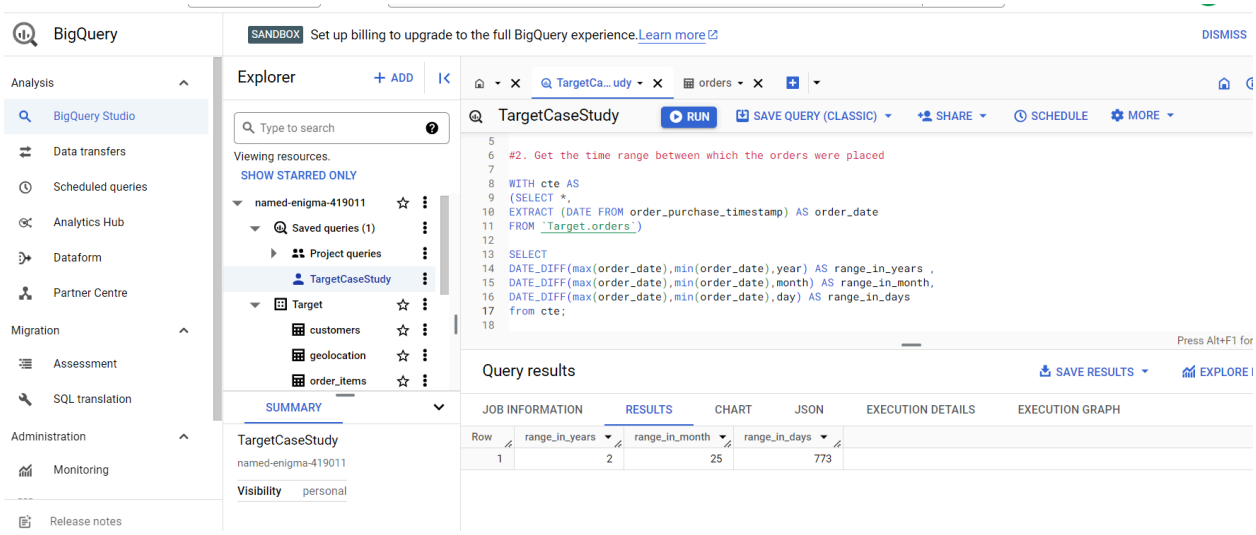
Query

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

WITH cte AS
(SELECT *,
EXTRACT (DATE FROM order_purchase_timestamp) AS order_date
FROM `Target.orders`)

SELECT
DATE_DIFF(max(order_date),min(order_date),year) AS range_in_years ,
DATE_DIFF(max(order_date),min(order_date),month) AS range_in_month,
DATE_DIFF(max(order_date),min(order_date),day) AS range_in_days
from cte;
```

Screenshot



Insights and Action Items

1. The orders in this case were placed between 2 years, 25 months, or 773 days.
2. To analyze trends, seasonality, and overall order patterns over a certain time, it can be helpful to know the time range of the orders.
3. By examining the range in years, you can identify whether the business is experiencing growth or decline over time. A positive trend suggests business expansion, while a negative trend may indicate a need for strategic adjustments.

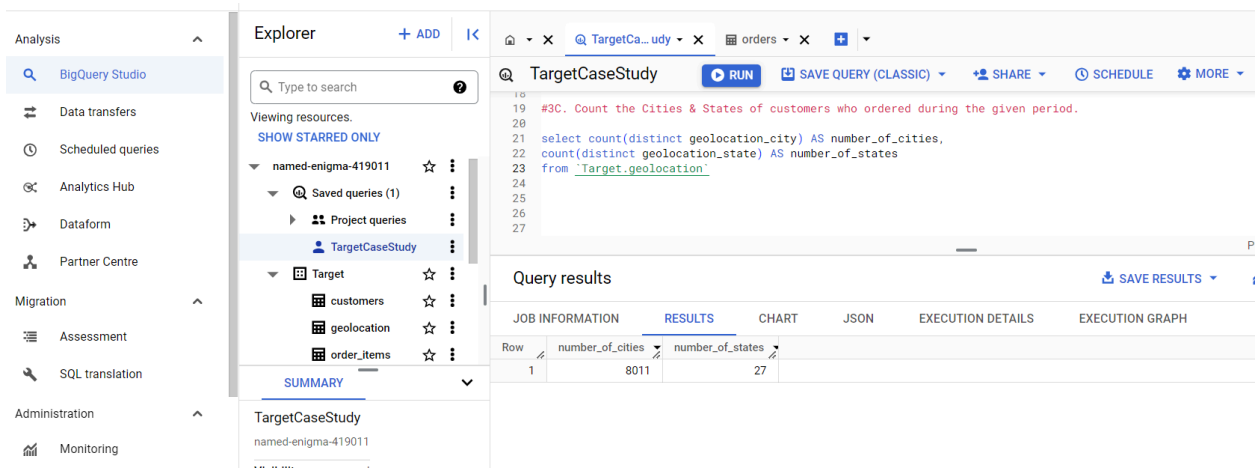
4. Use historical order data to forecast future demand and plan inventory, staffing, and resource allocation accordingly. The insights gained from the time span analysis can inform more accurate forecasts and strategic decision-making.

C. Count the Cities & States of customers who ordered during the given period.

Query

```
select count(distinct geolocation_city) AS number_of_cities,
count(distinct geolocation_state) AS number_of_states
from `Target.geolocation`
```

Screenshot



Insights and Action Items

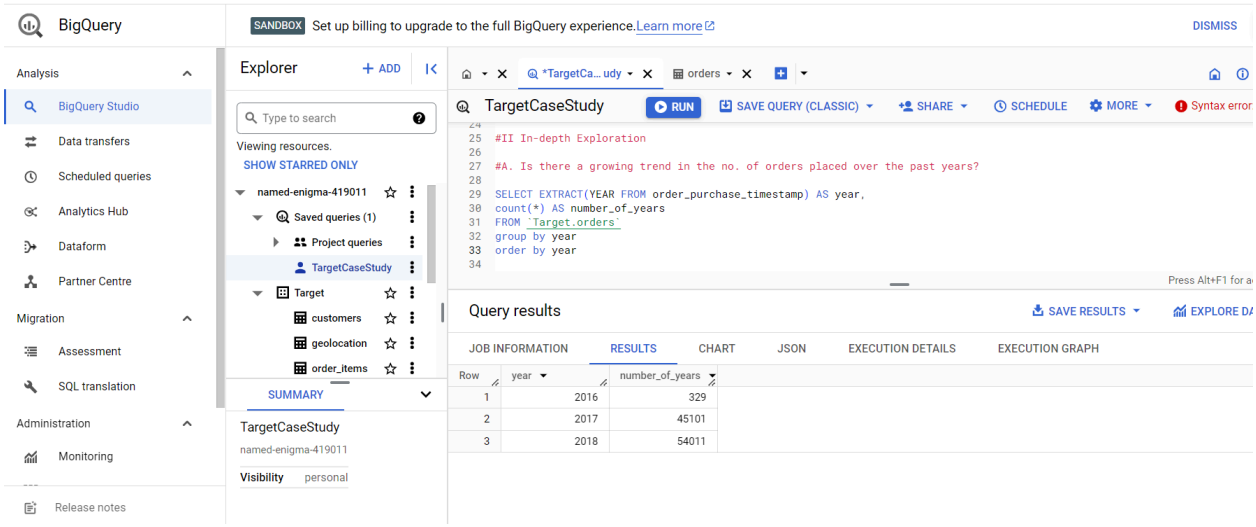
1. The dataset in this example has 27 distinct states and 8011 distinct cities. There is a diversity of cities and states from which customers are placing orders.
2. The Market Expansion Opportunities can be : Identify regions with a low number of unique cities or states to uncover potential market expansion opportunities. Targeting these areas with marketing campaigns or localized promotions could help increase brand awareness and customer acquisition.
3. Use geographical insights to optimize logistics and supply chain operations. Concentrating distribution centers or adjusting inventory levels based on the concentration of orders in specific cities or states can improve delivery efficiency and reduce shipping costs.
4. Customer experience can be enhanced using geographical insights by offering localized services, such as language-specific support or region-specific product recommendations. Tailoring the customer experience to local preferences can improve customer satisfaction and loyalty.

II. In-depth Exploration:

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

```
SELECT EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
count(*) AS number_of_years
FROM `Target.orders`
group by year
order by year
```

Screenshot



Insights and Action Items

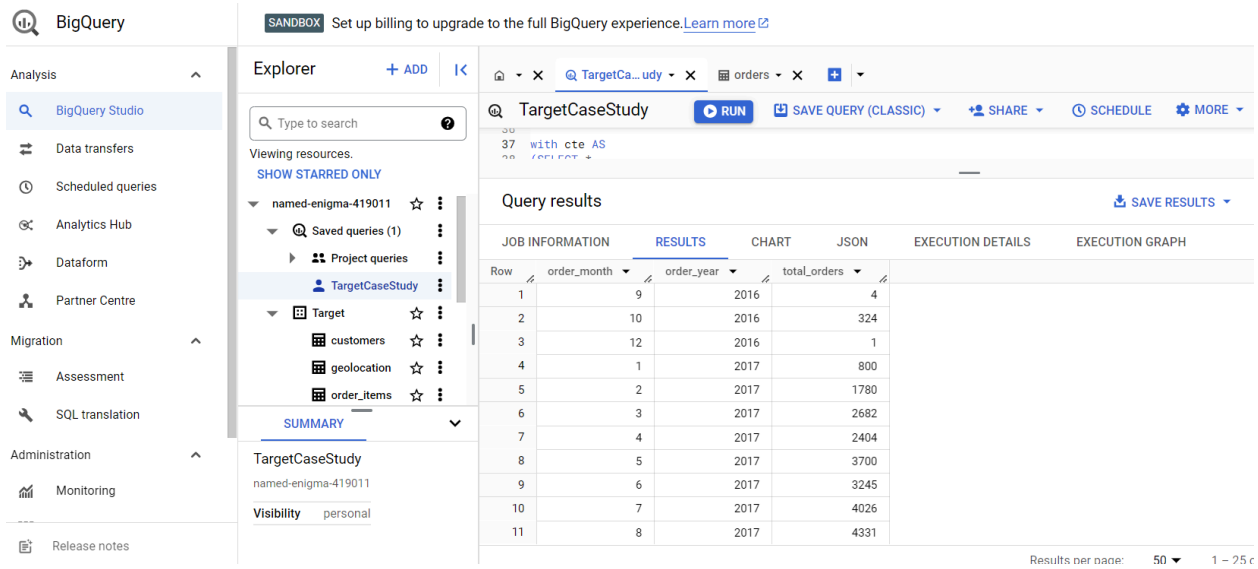
1. There has been an upward trend in the number of orders over the past few years after examining the results. A favorable trend can be seen if the order number regularly rises year over year.
2. There is opportunity for strategic expansion, such as increasing inventory, expanding product offerings, or entering new markets to capitalize on growing demand.
3. Understanding the trend in order volume can help in resource allocation planning. For instance, during peak order periods, additional resources may be needed in areas such as customer support, logistics, and fulfillment to ensure smooth operations and timely order processing.
4. Seasonal variations in order volume can inform the timing and content of marketing campaigns and promotions. Targeting customers during peak order periods with relevant offers and incentives can help maximize sales and customer engagement.

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

Query

```
with cte AS
(SELECT *,
EXTRACT(Date FROM order_purchase_timestamp)AS order_date,
EXTRACT(Year FROM order_purchase_timestamp)AS order_year,
EXTRACT(Month FROM order_purchase_timestamp)AS order_month,
FROM `Target.orders`
)
SELECT
order_month,
order_year,
count(order_id) AS total_orders
FROM cte
group by order_month,order_year
order by order_year,order_month
```

Screenshot



Insights and Action Items

1. We see a seasonal trend for Nov 2017 where there was Black Friday and there is a huge increase in the orders placed.
2. There is also a growth trend in Jan 2017 and Jan 2018 where New Years is experienced, and people may have preordered for the Carnival in Feb
3. Understanding monthly seasonality can help with operational planning, marketing tactics, and consumer behavior. It can aid in better planning of promotional activities, inventory management optimization, and peak period identification and resource allocation.

C. During what time of the day, do the Brazilian customers mostly place their

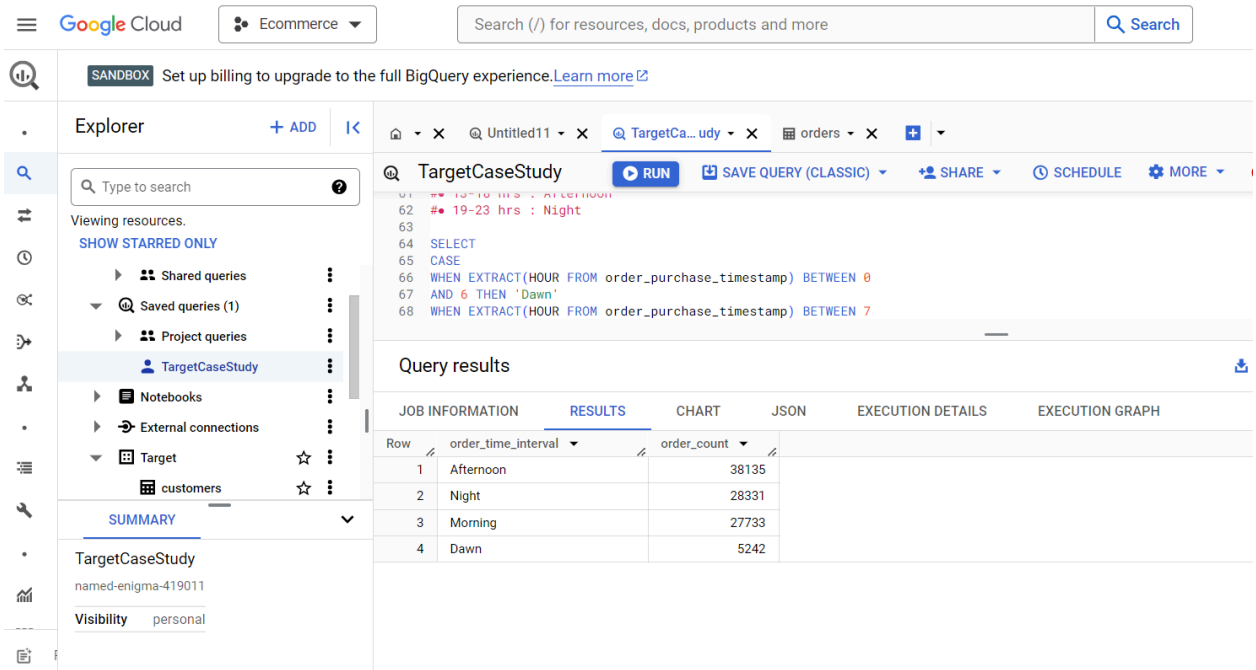
orders? (Dawn, Morning, Afternoon or Night)

- 0-6 hrs : Dawn
- 7-12 hrs : Mornings
- 13-18 hrs : Afternoon
- 19-23 hrs : Night

Query

```
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'
WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 19
AND 23 THEN 'Night'
END AS order_time_interval,
COUNT(*) AS order_count
FROM `Target.orders`
GROUP BY order_time_interval
ORDER BY order_count DESC
```

Screenshot



Insights and Action Items

1. The majority of orders from Brazilian customers are placed during the Afternoon time interval, followed by Night, Morning, and Dawn, in descending order of frequency. Also, customers are buying least during dawn.
2. Allocation of marketing resources and scheduling promotions to coincide with the Afternoon time interval can be done, when customer activity is highest. Tailor marketing messages and offers to maximize engagement during this peak period.
3. Supply chain operations and logistics can be streamlined to accommodate the higher order volume during the Afternoon time interval.Coordinating with suppliers and logistics partners will ensure timely delivery and inventory availability.
4. Considering regional variations in ordering patterns within Brazil can be used to tailor marketing efforts and promotions effectively. Analyze regional preferences and adjust strategies to optimize engagement and sales across different regions.

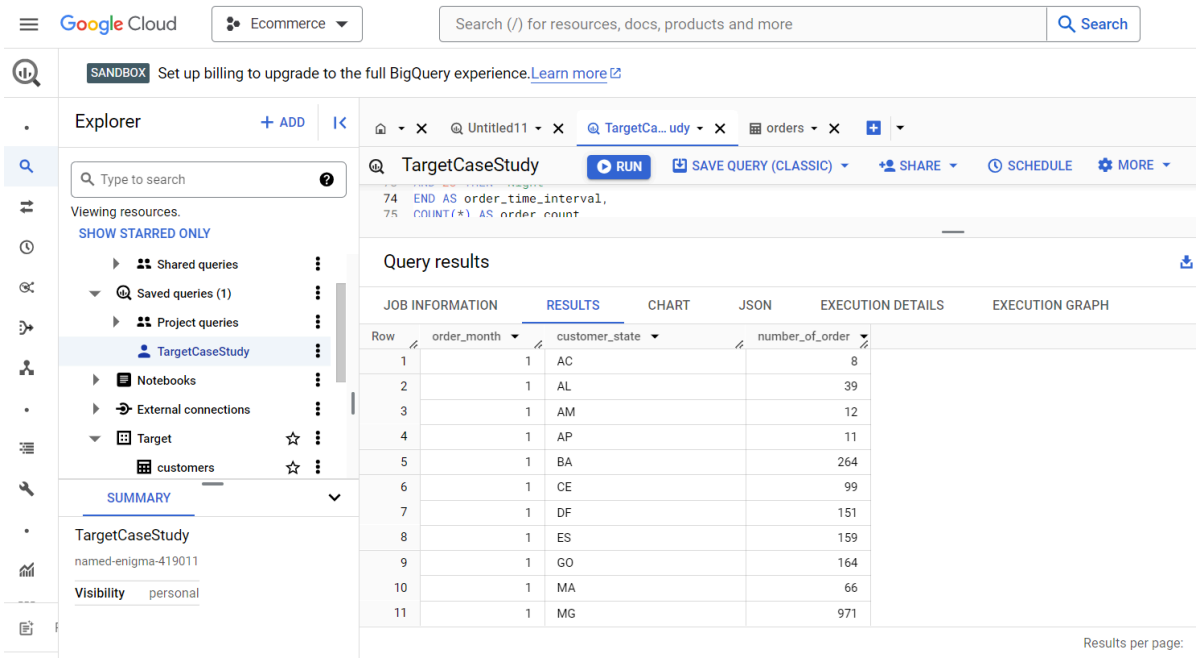
III. Evolution of E-commerce orders in the Brazil region:

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

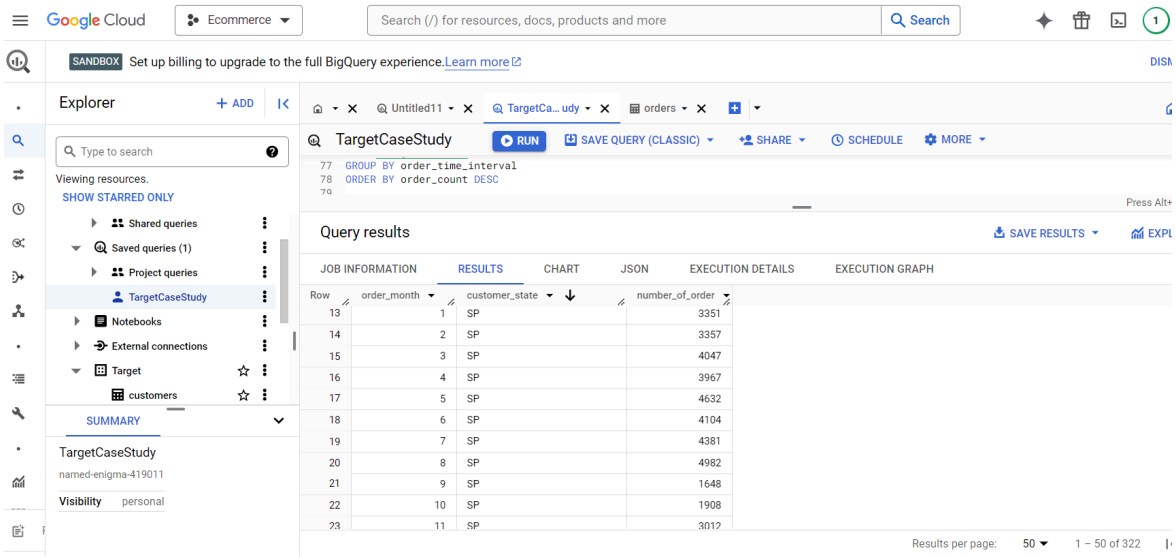
Query

```
SELECT EXTRACT(MONTH FROM o.order_purchase_timestamp) AS
order_month, c.customer_state, COUNT(*) AS number_of_order
FROM `Target.orders` AS o
JOIN `Target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY order_month, c.customer_state
ORDER BY order_month, c.customer_state;
```

Screenshot



Highest number of orders in state SP



Insights and Action Items

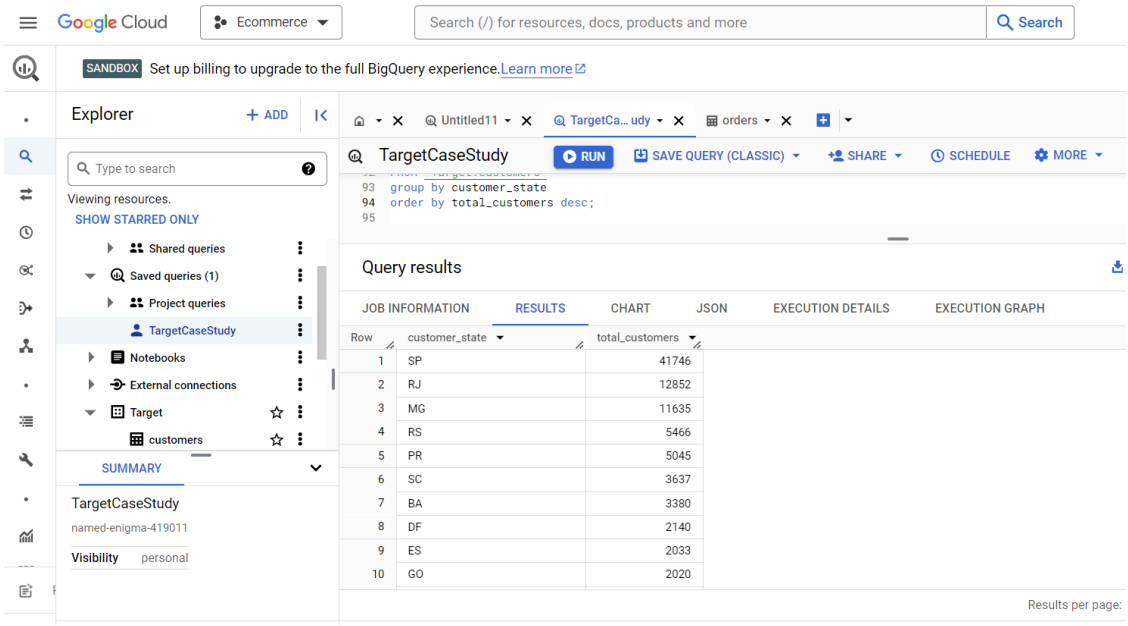
1. We can learn more about the monthly number of orders for each state by examining the query's results. Over time, we can spot trends, patterns, or seasonality in the order volume for various states. We can use it to determine which states have consistently high order volumes and to pinpoint any months or states where order counts have significantly changed. Here in our data, we can find that for every month the state called SP has the highest number of orders.
2. We can target marketing efforts in states with rising order volumes, spot potential operational issues in states with falling order volumes or optimize inventory management based on order trends across different states by analyzing these insights.

B. How are the customers distributed across all the states?

Query

```
select customer_state, count(distinct customer_id) AS total_customers
FROM `Target.customers`
group by customer_state
order by total_customers desc
```

Screenshot



Insights and Action Items

1. The query provides insights into the geographic distribution of customers, indicating the number of unique customers present in each state. This allows businesses to understand where their customer base is concentrated and identify areas with high or low customer density such as states called SP that have the highest clients and the state called RR that have the fewest clients.
2. There are several uses for this information, including: Market targeting, Expansion opportunities and Customer service.
3. We can learn more about the geographic distribution of our client base, spot prospective growth areas, and make wise decisions to optimize our company strategy by looking at the customer distribution between states.

IV. Impact on Economy: Analyze the money movement by e-commerce by looking at

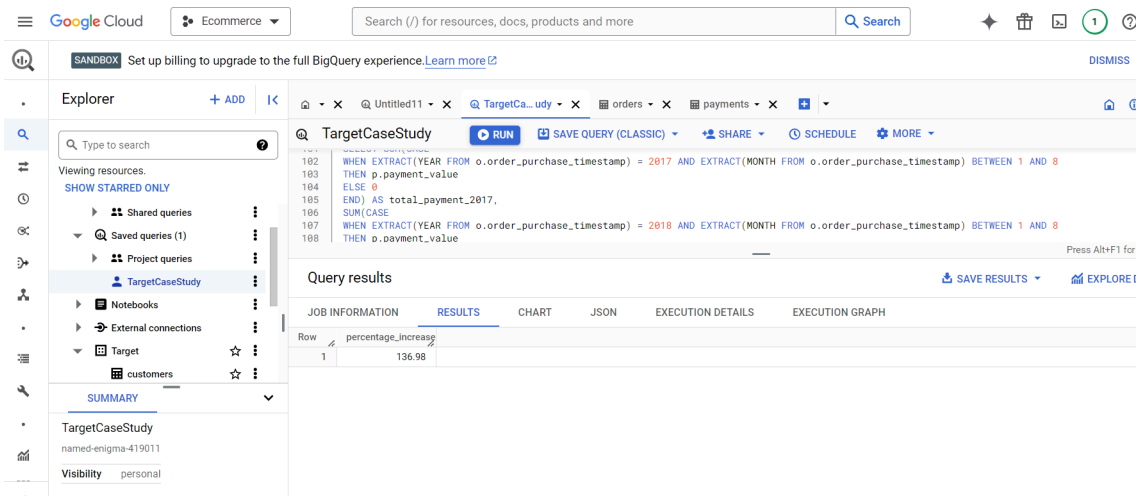
order prices, freight and others.

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

Query

```
select ROUND((((total_payment_2018 - total_payment_2017) / total_payment_2017) * 100),
2) AS percentage_increase
FROM (
  SELECT SUM(CASE
  WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 AND EXTRACT(MONTH FROM
o.order_purchase_timestamp) BETWEEN 1 AND 8
  THEN p.payment_value
  ELSE 0
  END) AS total_payment_2017,
  SUM(CASE
  WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 AND EXTRACT(MONTH FROM
o.order_purchase_timestamp) BETWEEN 1 AND 8
  THEN p.payment_value
  ELSE 0
  END) AS total_payment_2018
  from `Target.payments` p
  join `Target.orders` o
  on p.order_id = o.order_id
)
```

Screenshot



Insights and Action Items

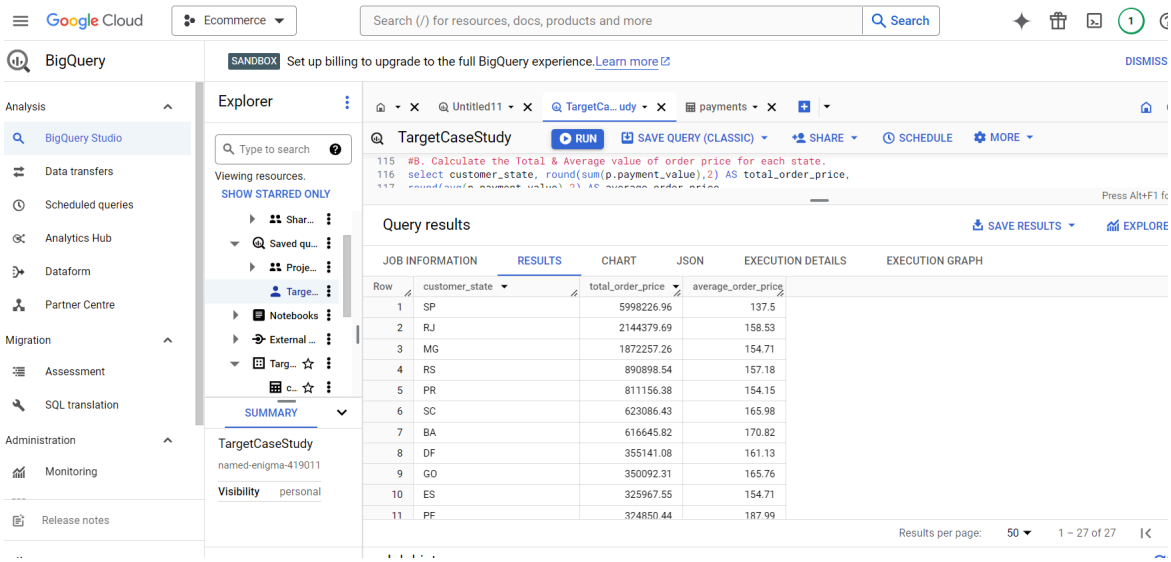
1. For both 2017 and 2018, only orders placed from January to August are considered.
2. To get the % increase, the query analyzes the monthly prices between 2017 and 2018.
3. The findings tell us a growth rate of approximately 137% from 2017 to 2018.

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

Query

```
select customer_state, round(sum(p.payment_value),2) AS total_order_price,
round(avg(p.payment_value),2) AS average_order_price
from `Target.payments` p
join `Target.orders` o
on o.order_id = p.order_id
join `Target.customers` c
on o.customer_id = c.customer_id
group by customer_state
order by total_order_price desc
```

Screenshot



Insights and Action Items

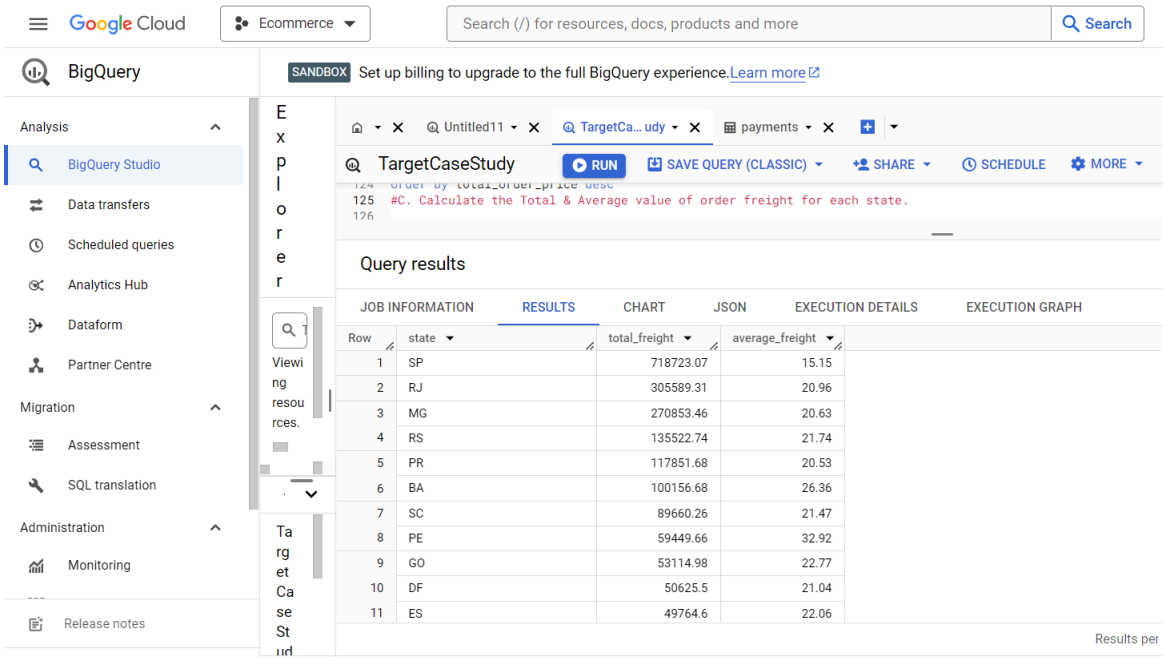
1. The sum of all order prices for each state is displayed in the "total_order_price" column, which represents the total amount of orders placed.
2. The "average_order_price" column shows the normal order value for each state together with the average order price for that state.
3. We can find states with large total order values, which point to potentially profitable marketplaces, by analyzing the results.By examining the total order prices, we can identify states that contribute the most to the overall revenue. This information is valuable for strategic planning and resource allocation.
4. Implement customer retention programs targeted at states with higher average order values. Offer loyalty rewards, exclusive discounts, or personalized incentives to encourage repeat purchases and foster long-term customer relationships.

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

Query

```
select customer_state as state,
round(sum(oi.freight_value),2) AS total_freight,
round(avg(oi.freight_value),2) AS average_freight
from `Target.orders` o
join `Target.order_items` oi
on o.order_id = oi.order_id
join `Target.customers` c
on c.customer_id = o.customer_id
group by state
order by total_freight desc
```

Screenshot



Insights and Action Items

1. We can find states with high total freight costs, here in our case a state called SP, by analyzing the results, which could point to regions with higher shipping prices or logistical difficulties.
2. When optimizing logistics operations or pricing strategies, it might be helpful to discover regions with higher or lower average shipping prices by comparing the average order freight costs across states.
3. Understanding the differences in order freight rates between states can offer information about local shipping habits, supplier locations, or client preferences that can be used to optimize processes and cut costs.

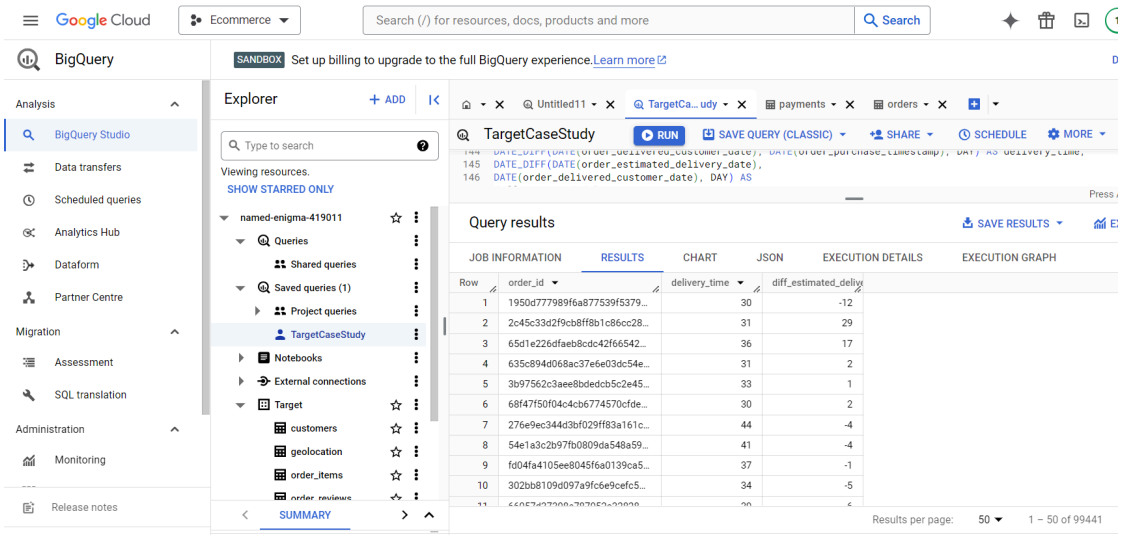
V. Analysis based on sales, freight and delivery time.

A. 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.

Query

```
select order_id,
DATE_DIFF(DATE(order_delivered_customer_date), DATE(order_purchase_timestamp), DAY) AS
delivery_time,
DATE_DIFF(DATE(order_estimated_delivery_date),
DATE(order_delivered_customer_date), DAY) AS
diff_estimated_delivery
from `Target.orders`
```

Screenshot



Insights and Action Items

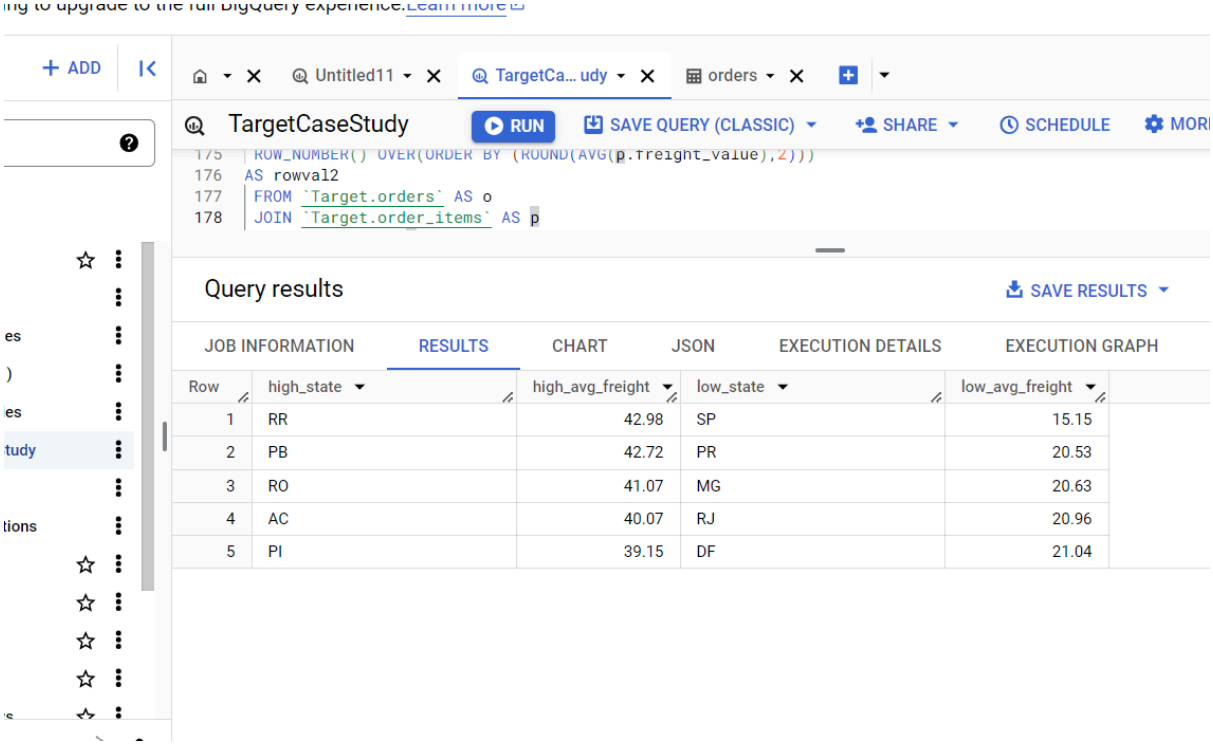
1. Insights into the effectiveness of the delivery process, including any delays or early deliveries compared to the projected time frame, can be gained by analyzing the delivery_time and diff_estimated_delivery columns.
2. Utilize to set benchmarks and targets for delivery times and minimize the difference between estimated and actual delivery dates. Implement strategies to improve delivery performance and enhance customer satisfaction by ensuring timely and reliable order fulfillment.
3. Continuously monitor delivery performance metrics and analyze trends over time to identify patterns and root causes of delays. Use data-driven insights to make informed decisions and implement targeted interventions for optimizing delivery operations.

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

```
SELECT high.customer_state AS high_state,
high.average_freight_value AS high_avg_freight,
low.customer_state AS low_state,
low.average_freight_value AS low_avg_freight
FROM
(
SELECT c.customer_state,
ROUND(AVG(p.freight_value),2) AS average_freight_value,
ROW_NUMBER() OVER(ORDER BY

(ROUND(AVG(p.freight_value),2))DESC) AS rowval1
FROM `Target.orders` AS o
join `Target.order_items` AS p
on o.order_id = p.order_id
join `Target.customers` AS c
on c.customer_id = o.customer_id
GROUP BY c.customer_state
order by average_freight_value desc
limit 5) AS high
JOIN
(
 SELECT
 c.customer_state,
 ROUND(AVG(p.freight_value),2) AS average_freight_value,
 ROW_NUMBER() OVER(ORDER BY (ROUND(AVG(p.freight_value),2)))
AS rowval2
 FROM `Target.orders` AS o
 JOIN `Target.order_items` AS p
 ON o.order_id = p.order_id
 JOIN `Target.customers`       AS c
 ON o.customer_id = c.customer_id
 GROUP BY
 c.customer_state
 ORDER BY
 average_freight_value
 LIMIT
 5
) AS low
ON high.rowval1 = low.rowval2
```

Screenshot



Insights and Action Items

1. The states with the highest average freight values like states called RR and PB may experience greater shipping prices due to reasons like remote locations, higher transportation costs, or supply chain difficulties.
2. It might be useful for Target to try to optimize logistics operations or save costs to locate places with relatively reduced shipping prices by looking at the states with the lowest average freight values like states such as SP and PR.
3. This data can help us develop focused initiatives, bargain freight costs, or spot possible opportunities to reduce costs in our supply chain operations.
4. When assessing the data and drawing conclusions from these insights, it is crucial to consider additional elements like distance, transportation infrastructure, carrier availability, or regional economic variations.

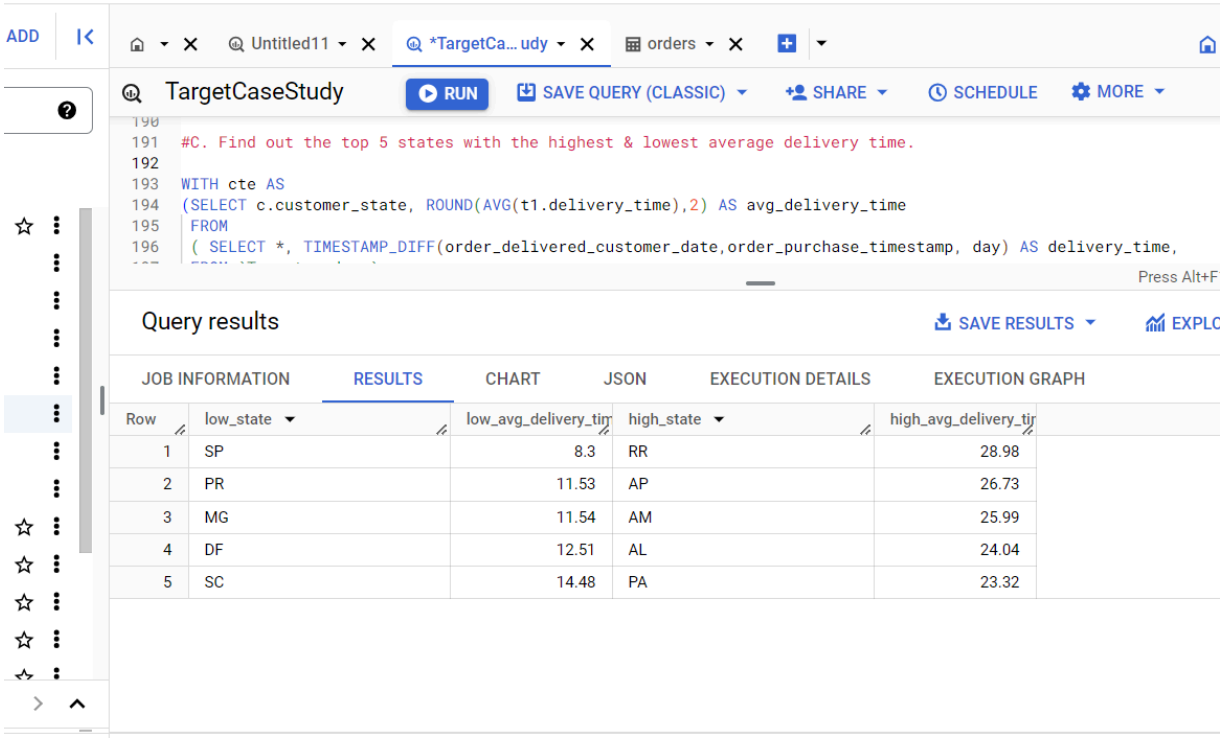
C. Find out the top 5 states with the highest & lowest average delivery time.

Query

```
WITH cte AS
(SELECT c.customer_state, ROUND(AVG(t1.delivery_time),2) AS avg_delivery_time
```

```
 FROM
 ( SELECT *, TIMESTAMP_DIFF(order_delivered_customer_date,order_purchase_timestamp,
day) AS delivery_time,
 FROM `Target.orders`
 WHERE order_status = 'delivered' AND order_delivered_customer_date IS NOT NULL
 ORDER BY order_purchase_timestamp) AS t1
JOIN `Target.customers` AS c
 ON t1.customer_id = c.customer_id
 GROUP BY c.customer_state
 ORDER BY avg_delivery_time
)
 SELECT c1.customer_state AS low_state, c1.avg_delivery_time AS low_avg_delivery_time,
 c2.customer_state AS high_state,c2.avg_delivery_time AS high_avg_delivery_time
FROM
(SELECT *,ROW_NUMBER() OVER (ORDER BY cte.avg_delivery_time DESC) AS rowval2
 FROM cte
 ORDER BY rowval2
) AS c2
JOIN
(SELECT *, ROW_NUMBER() OVER (ORDER BY cte.avg_delivery_time) AS rowval1
 FROM cte
 ORDER BY rowval1
) AS c1
ON c1.rowval1 = c2.rowval2
LIMIT 5
```

Screenshot



Insights and Action Items

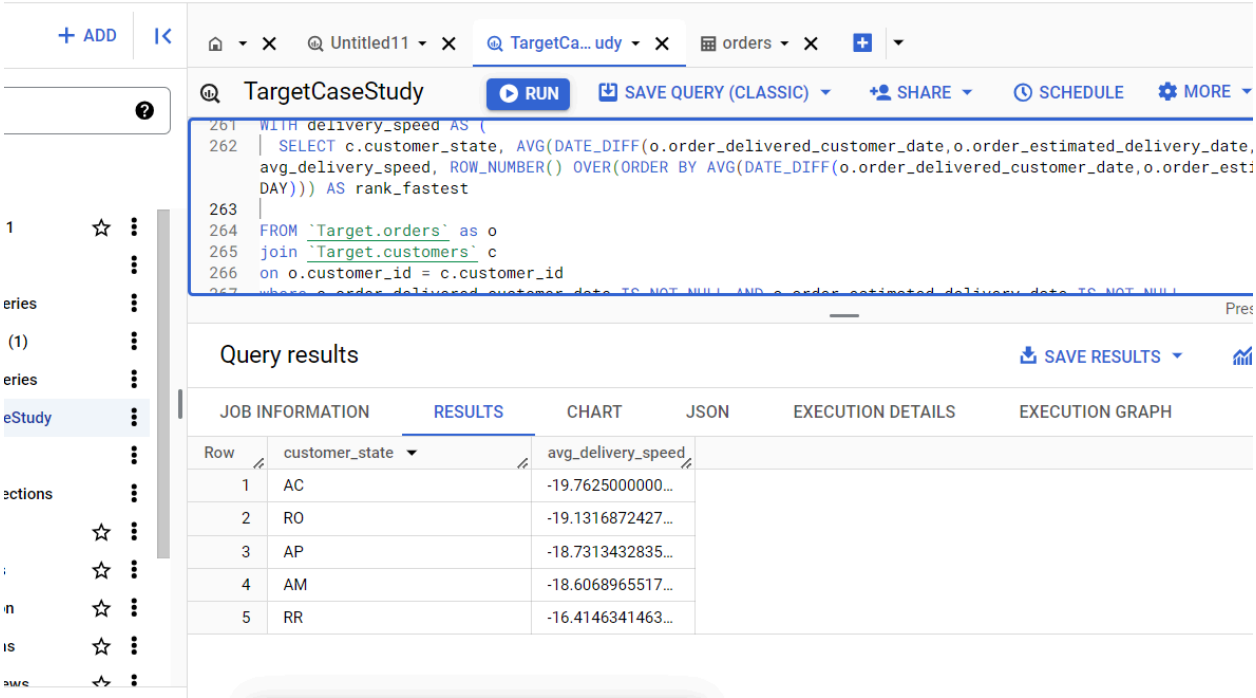
1. The states like SP and PR with the lowest average delivery times and states called RR and AP with highest average delivery times.
2. Differences in average delivery times between states may be influenced by geographic factors such as distance from distribution centers, population density, and transportation infrastructure. Understanding these factors can help pinpoint root causes of delivery delays or efficiencies in specific regions.
3. Longer delivery times may negatively impact customer satisfaction and perception of service quality, leading to potential churn or negative reviews. Conversely, shorter delivery times can enhance customer satisfaction and loyalty, driving repeat business and positive word-of-mouth.
4. Analyzing factors contributing to longer delivery times in states with high average delivery times can be done. Implementing operational improvements such as optimizing logistics routes, expanding fulfillment infrastructure, or enhancing last-mile delivery capabilities to reduce delivery times and improve efficiency can be important action items.

D. 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.

Query

```
WITH delivery_speed AS (
  SELECT c.customer_state,
AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_estimated_delivery_date,DAY)) AS
avg_delivery_speed, ROW_NUMBER() OVER(ORDER BY
AVG(DATE_DIFF(o.order_delivered_customer_date,o.order_estimated_delivery_date,DAY)))
AS rank_fastest
FROM `Target.orders` as o
join `Target.customers` c
on o.customer_id = c.customer_id
where o.order_delivered_customer_date IS NOT NULL AND o.order_estimated_delivery_date
IS NOT NULL
GROUP BY c.customer_state)
select customer_state,avg_delivery_speed
FROM delivery_speed
where rank_fastest <= 5
order by avg_delivery_speed
```

Screenshot



Insights and Action Items

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

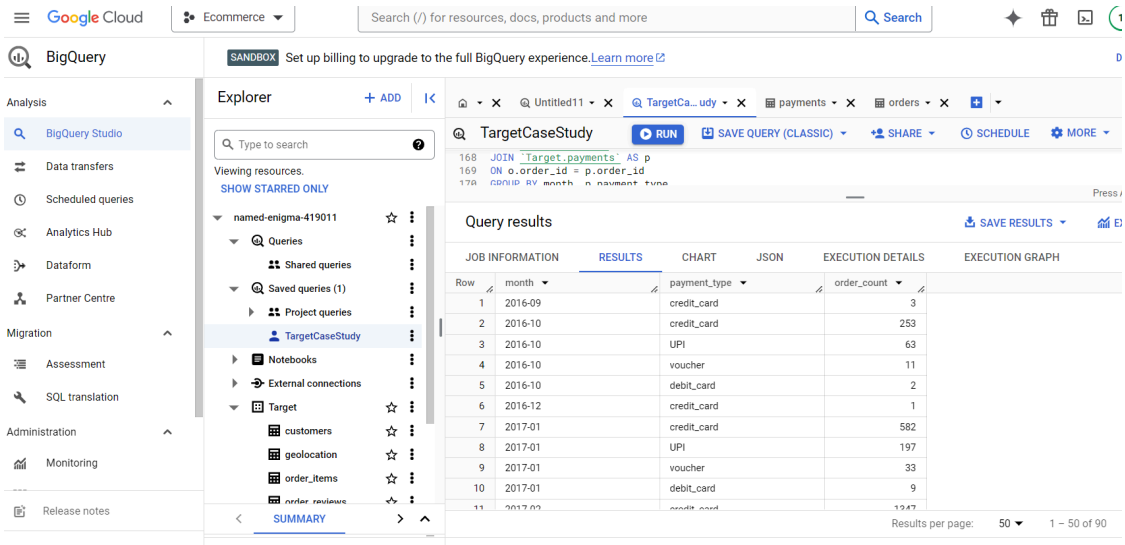
VI. Analysis based on the payments:

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

Query

```
SELECT
 FORMAT_TIMESTAMP('%Y-%m', o.order_purchase_timestamp) AS
month,
 p.payment_type,
 COUNT(DISTINCT o.order_id) AS order_count
FROM `Target.orders` AS o
JOIN `Target.payments` AS p
ON o.order_id = p.order_id
GROUP BY month, p.payment_type
ORDER BY month;
```

Screenshot



Insights and Action Items

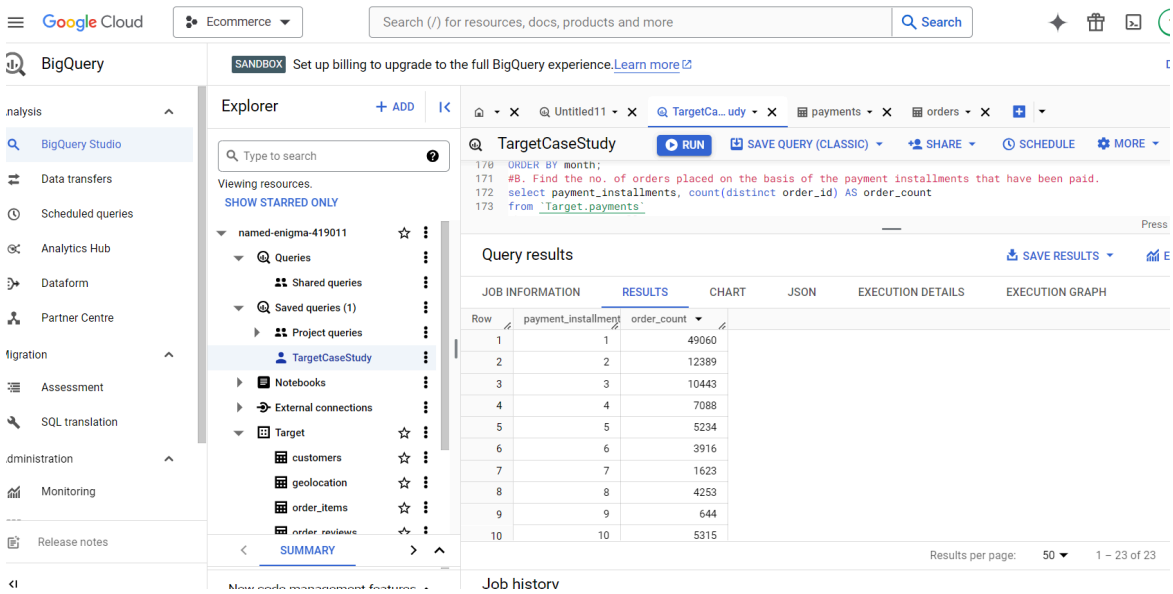
1. We identify that credit card as a payment method was most used in November 2017.
2. To analyze seasonality, identify peak months, or evaluate the effects of marketing efforts or outside variables on consumer behavior, tracking the month-to-month trends in order counts can be helpful.
3. Based on the payment preferences noticed during various months, these insights might help firms optimize their payment procedures, customize marketing campaigns, or enhance customer experiences

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

Query

```
select payment_installments, count(distinct order_id) AS order_count
from `Target.payments`
where payment_installments >= 1
group by payment_installments
order by payment_installments
```

Screenshot



Insights and Action Items

1. 49060 orders were placed where payment installment was 1.
2. This analysis can help determine whether payment installment alternatives are popular or preferred by clients.
3. 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.