TITLE: Business Case Study – TARGET SQL

Description: Target is a well-known brand on a global scale and a significant retailer in the US. By providing unmatched value, creativity, innovation, and an extraordinary customer experience that no other retailer can match, Target establishes itself as a preferred shopping destination.

The operations of Target in Brazil are the subject of this business case, which offers analytical data on 100,000 orders placed between 2016 and 2018. The dataset provides a detailed look at several variables, such as order status, pricing, payment and freight performance, customer geography, product features, and customer reviews.

This large dataset can be analyzed to reveal important information on Target's Brazilian operations. The data can provide insight into several business-related topics, including order fulfilment, pricing tactics, the effectiveness of payments and shipping, client demographics, product features, and customer satisfaction levels.

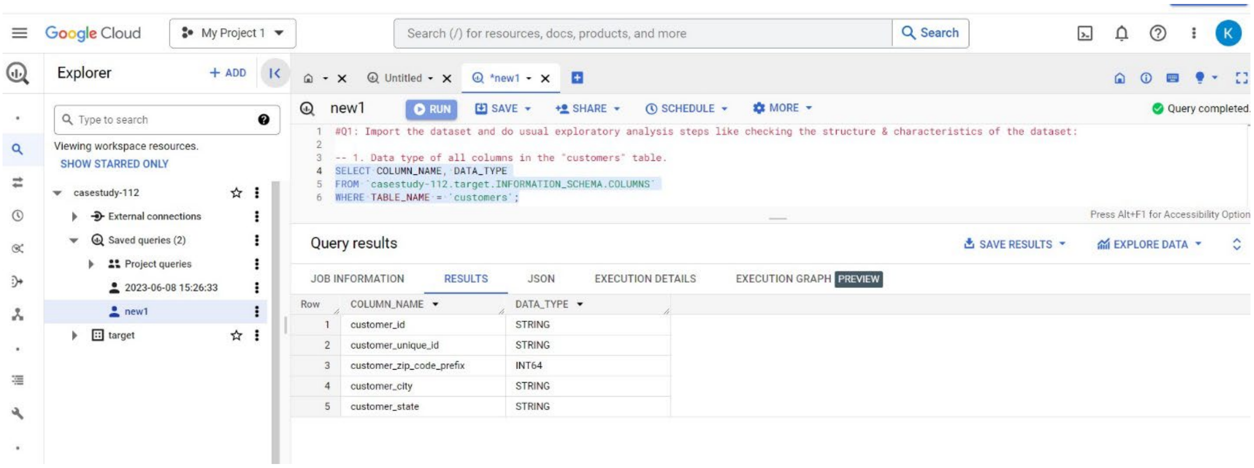
#Q1: Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

Answer Query:

```
SELECT COLUMN_NAME, DATA_TYPE
FROM `casestudy-112.target.INFORMATION_SCHEMA.COLUMNS`
WHERE TABLE_NAME = 'customers';
```

Output screenshot:



Insights:

1. **All columns in this example are saved as strings (VARCHAR), except customer_zip_code_prefix which is integer.** This implies that the data might mostly be textual in nature, and that any values relating to dates or numbers are probably saved as strings.

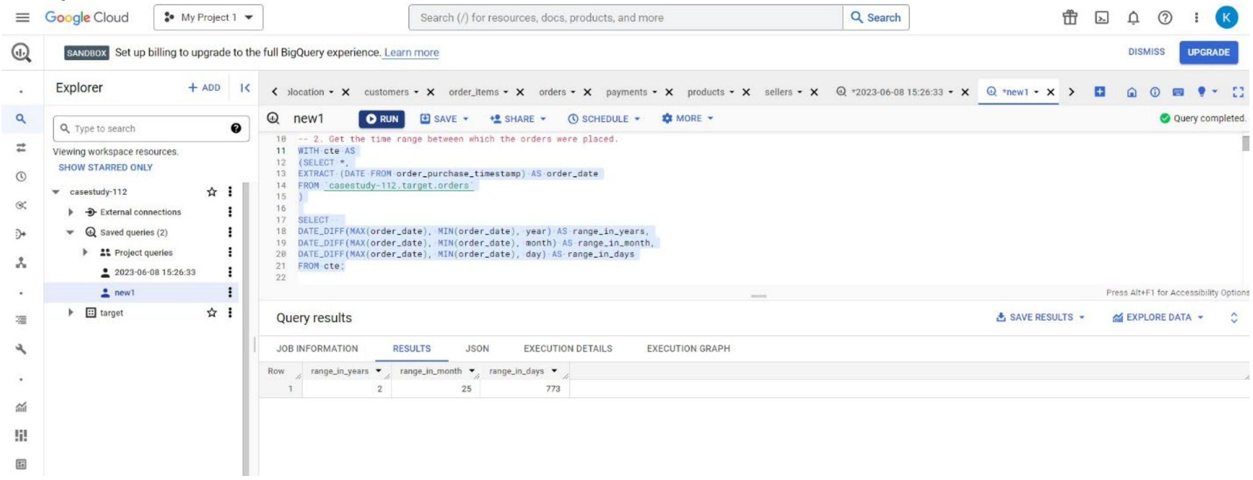
--2. Get the time range between which the orders were placed.

Answer Query:

```
WITH cte AS
(SELECT *,
EXTRACT (DATE FROM order_purchase_timestamp) AS order_date
FROM `casestudy-112.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;
```

Output screenshot:



Insights:

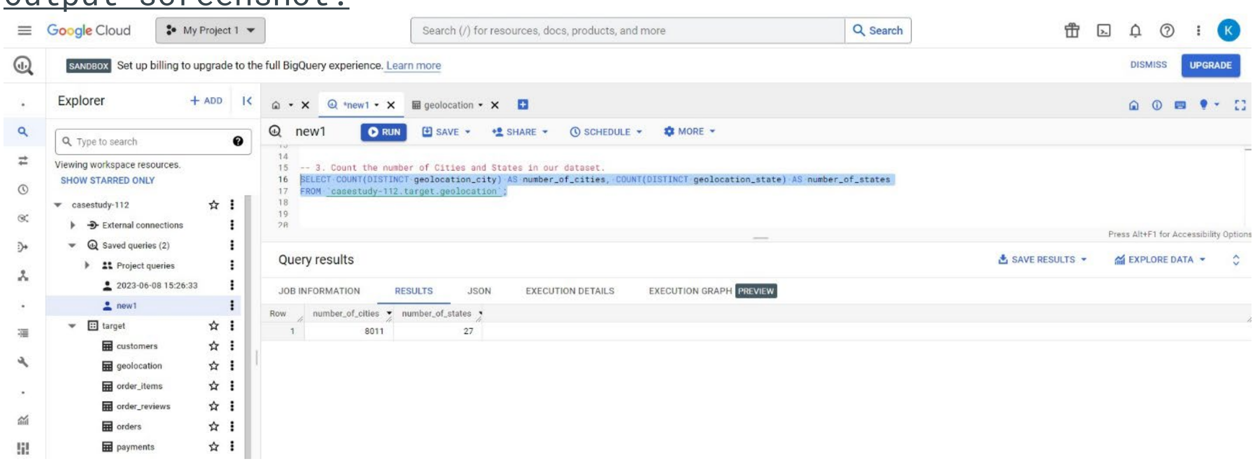
1. The orders in this case were placed between **2 years, 25 months, or 773 days.**
2. To analyze trends, seasonality, and overall order patterns over a certain time, it can be helpful to know the time range of the orders.

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

Answer Query:

```sql
SELECT COUNT(DISTINCT geolocation_city) AS number_of_cities,
COUNT(DISTINCT geolocation_state) AS number_of_states
FROM `casestudy-112.target.geolocation`;
```

Output screenshot:



Insights:

1. The dataset in this example has **27 distinct states and 8011 distinct cities.** These figures can aid in our comprehension of the geographic distribution of our clientele or the scope of our dataset.
2. Analyzing the distribution of cities and states might reveal information about the diversity or concentration of our clientele in various geographic areas. It might be helpful for regional analysis, figuring out hotspots, or figuring out how far our company has spread across the nation or the globe.

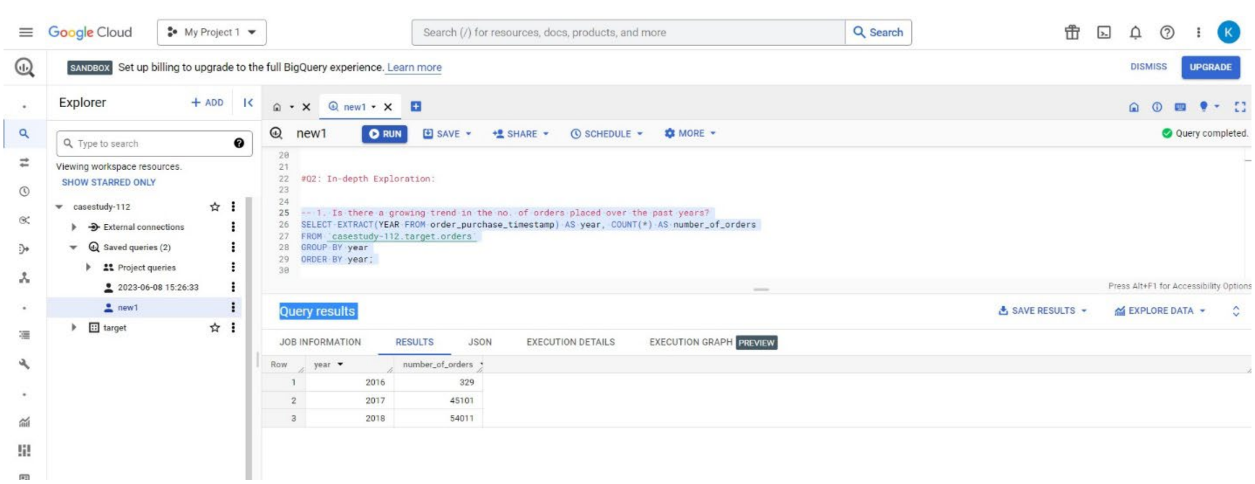
#Q2: In-depth Exploration:

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

Answer Query:

```
SELECT EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
COUNT(*) AS number_of_years
FROM `casestudy-112.target.orders`
GROUP BY year
ORDER BY year;
```

Output screenshot:



Insights:

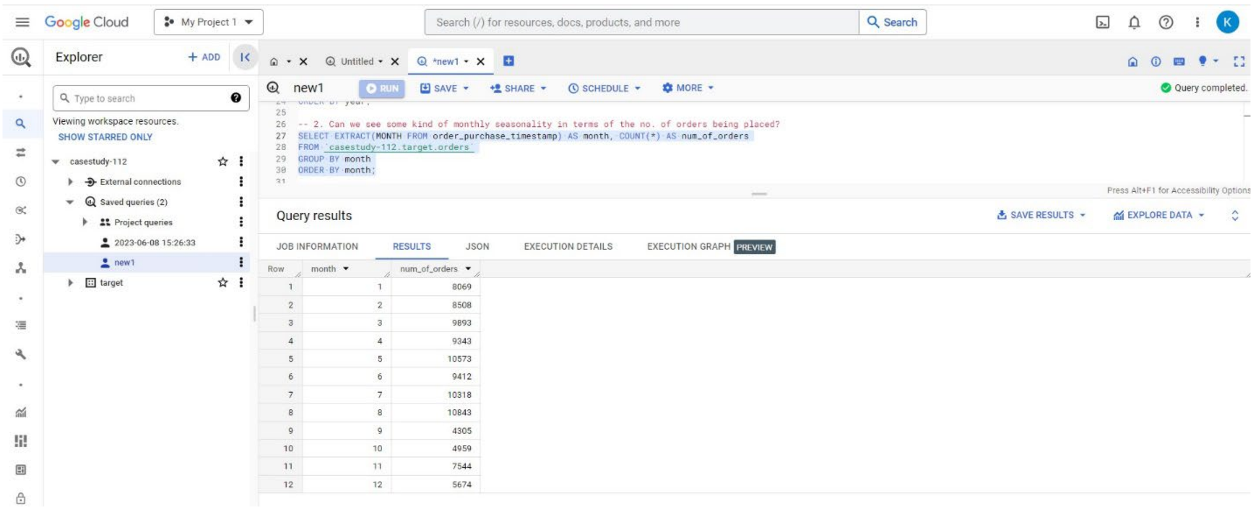
1. There has been **an upward trend in the number of orders over the past few years after examining the results.** A favorable trend can be seen if the order number regularly rises year over year. If there are variations or a negative tendency, however, it points to a different pattern.

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

Answer 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 `casestudy-112.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
```

Output screenshot:



Insights:

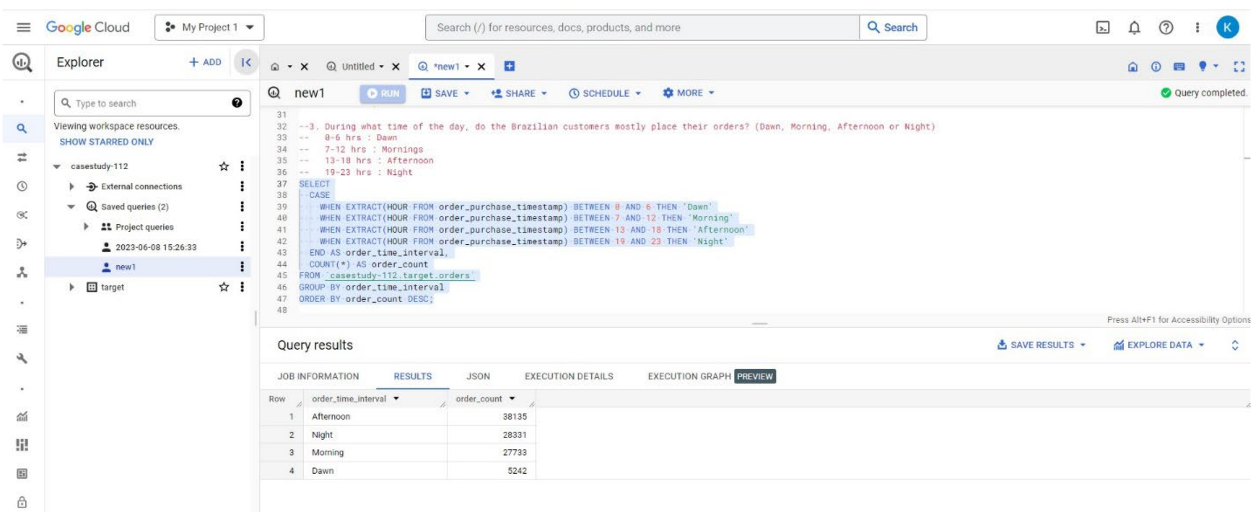
1. We see a seasonal trend for **Nov 2017 where there was Black Friday** and there is a huge increase in the orders placed.
2. There is also a **growth trend in Jan 2017 and Jan 2018 where New Years is experienced, and people may have preordered for the Carnival in Feb**
3. There is also an increase in orders in **Q1 of 2018 during which FIFA World Cup was scheduled.**
4. 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.

--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

Answer Query:

```sql
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 `casestudy-112.target.orders`
GROUP BY order_time_interval
ORDER BY order_count DESC;
```

Output screenshot:



Insights:

1. Based on the hour component of the timestamp, the query divides the order timestamps into various time groups (Dawn, Morning, Afternoon, Night). The results are then sorted based on how many orders fell inside each time frame.

2. We can ascertain the time of day when Brazilian clients often place their orders by analyzing the data. We will learn more about their ordering habits and preferences. For instance, we discover that **Brazilian clients frequently order more in the afternoons**, indicating that this is a period when people want to shop online. By scheduling customer assistance personnel or launching focused marketing efforts during the busiest ordering periods, for example, we may operate more efficient operations with the aid of this information. Also, **customers are buying least during dawn**.

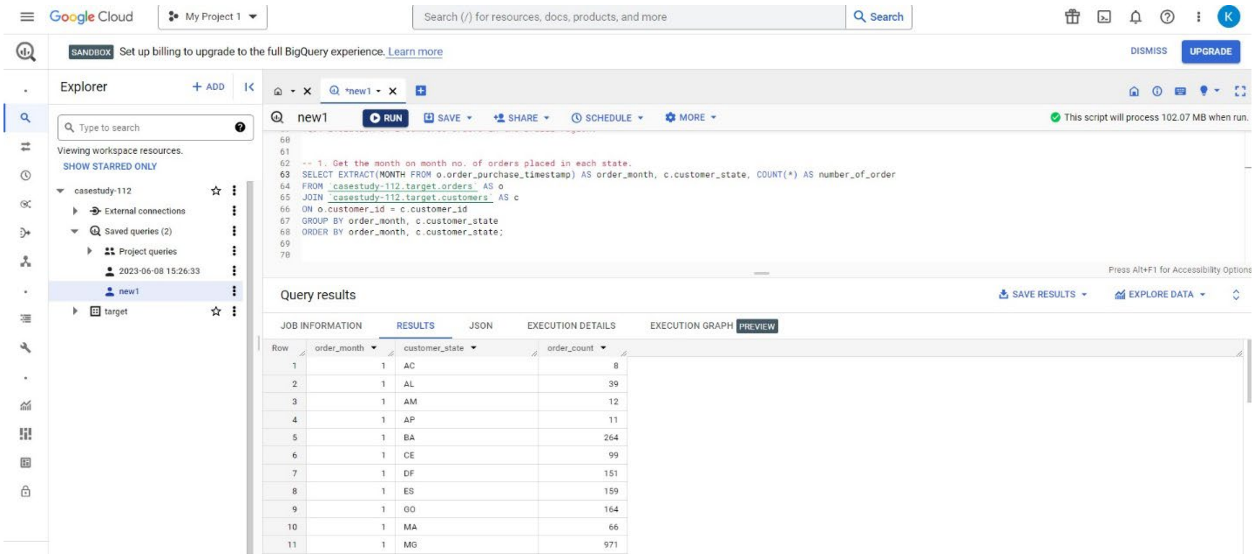
#Q3. Evolution of E-commerce orders in the Brazil region:

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

Answer Query:

```
SELECT EXTRACT(MONTH FROM o.order_purchase_timestamp) AS
order_month, c.customer_state, COUNT(*) AS number_of_order
FROM `casestudy-112.target.orders` AS o
JOIN `casestudy-112.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;
```

Output screenshot:



Insights:

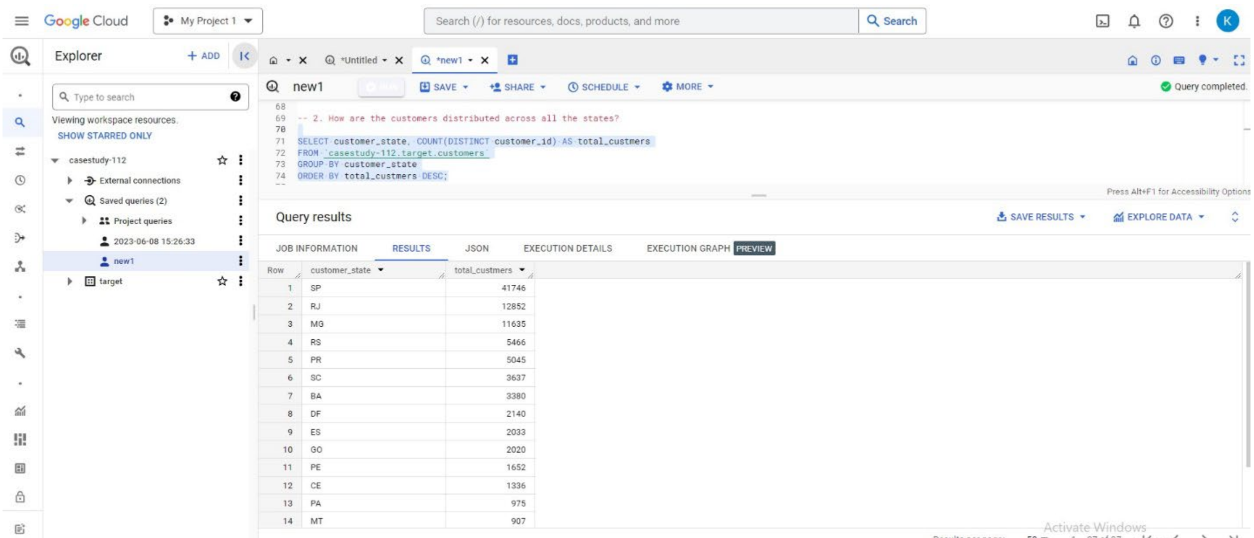
1. We can learn more about the monthly order count for each state by examining the query's results. Over time, we can spot trends, patterns, or seasonality in the order volume for various states. We can use it to determine which states have consistently high order volumes and to pinpoint any months or 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.

-- 2. How are the customers distributed across all the states?

Answer Query:

```
SELECT customer_state, COUNT(DISTINCT customer_id) AS
total_custmers
FROM `casestudy-112.target.customers`
GROUP BY customer_state
ORDER BY total_custmers DESC;
```

<u>Output screenshot:</u>



<u>Insights:</u>

1. The distribution of clients across states will be shown by analyzing the query's results. Which states have the most customers and which states have comparatively fewer consumers can be determined. Here the **<u>state called SP has the highest clients and the state called RR has the fewest clients.</u>** There are several uses for this information, including: Market targeting, Expansion opportunities and Customer service.
2. 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.

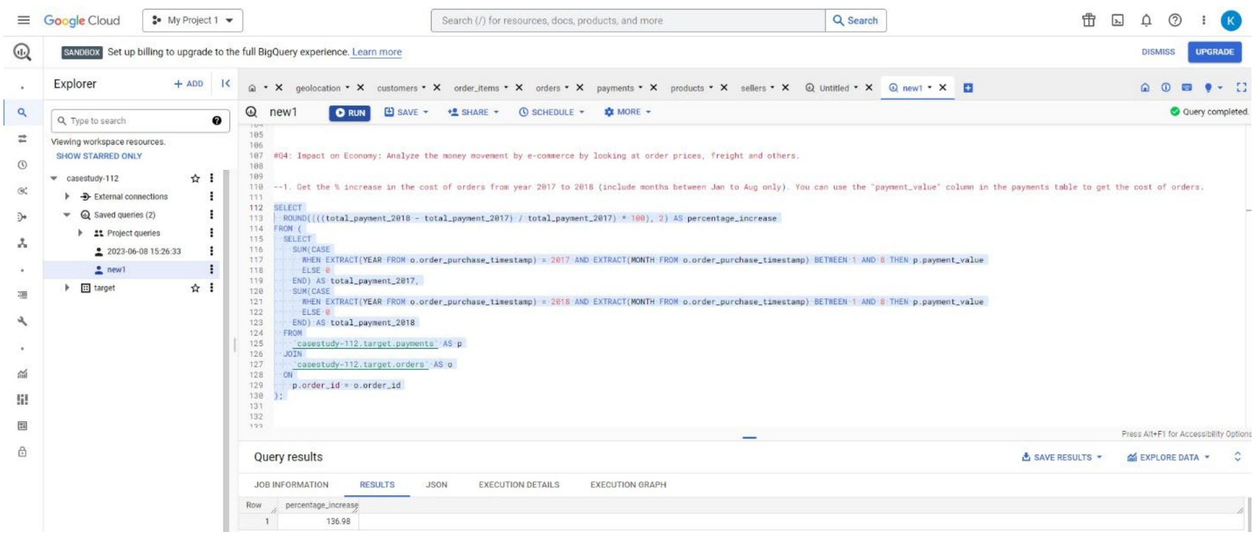
#Q4: Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight, and others.

--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.

Answer 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
    `casestudy-112.target.payments` AS p
  JOIN
    `casestudy-112.target.orders` AS o
  ON
    p.order_id = o.order_id
);
```

Output screenshot:



Insights:

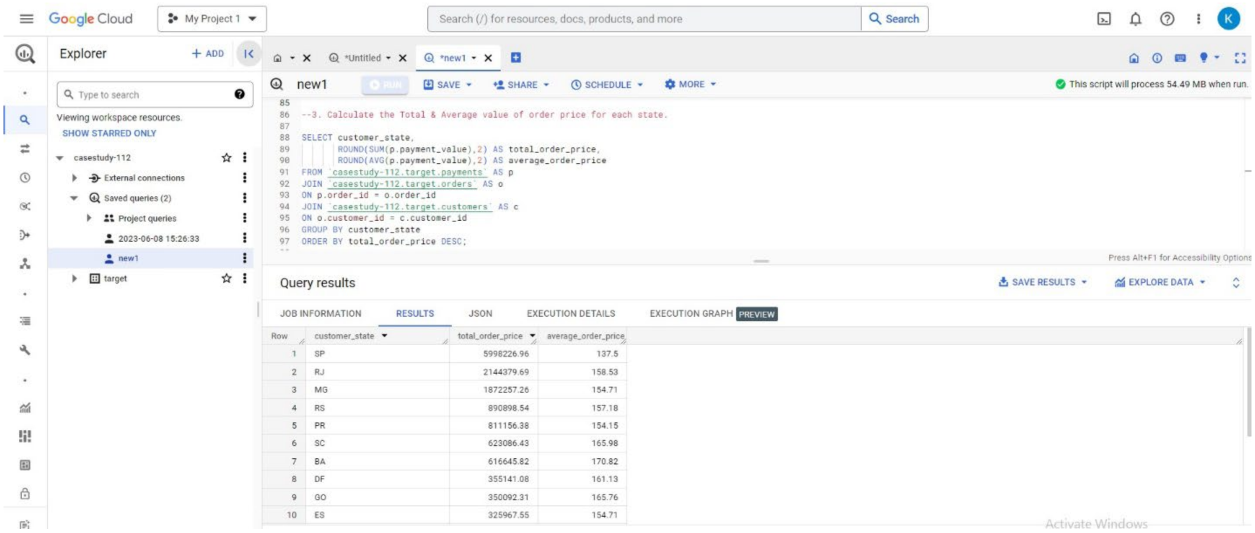
1. For both 2017 and 2018, only orders placed from January to August are considered.
2. To get the % increase, the query analyses the monthly prices between 2017 and 2018.
3. The findings tell us a **growth rate of approximately 137% from 2017 to 2018.**

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

Answer 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 `casestudy-112.target.payments` AS p
JOIN `casestudy-112.target.orders` AS o
ON p.order_id = o.order_id
JOIN `casestudy-112.target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state
ORDER BY total_order_price DESC;
```

Output screenshot:



Insights:

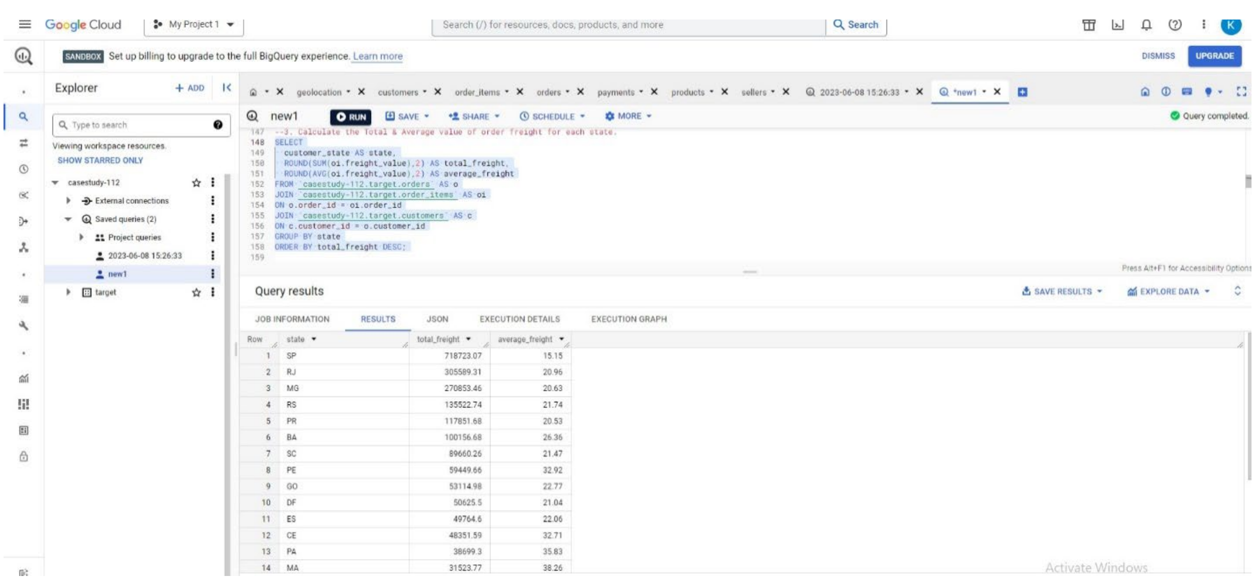
1. The sum of all order prices for each state is displayed in the "total_order_price" column, which represents the total amount of orders placed.
2. The "average_order_price" column shows the normal order value for each state together with the average order price for that state.
3. **We can find states with large total order values**, which point to potentially profitable marketplaces, by analyzing the results.
4. To develop focused marketing or pricing strategies, it can be helpful to compare the average order prices across states to find areas with higher or lower average spending.
5. To obtain more understanding and make wise judgements based on the data, it's critical to consider the context of each state, such as population, economic variables, or customer behavior.

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

Answer 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 `casestudy-112.target.orders` AS o
JOIN `casestudy-112.target.order_items` AS oi
ON o.order_id = oi.order_id
JOIN `casestudy-112.target.customers` AS c
ON c.customer_id = o.customer_id
GROUP BY state
ORDER BY total_freight DESC;
```

Output screenshot:



Insights:

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.

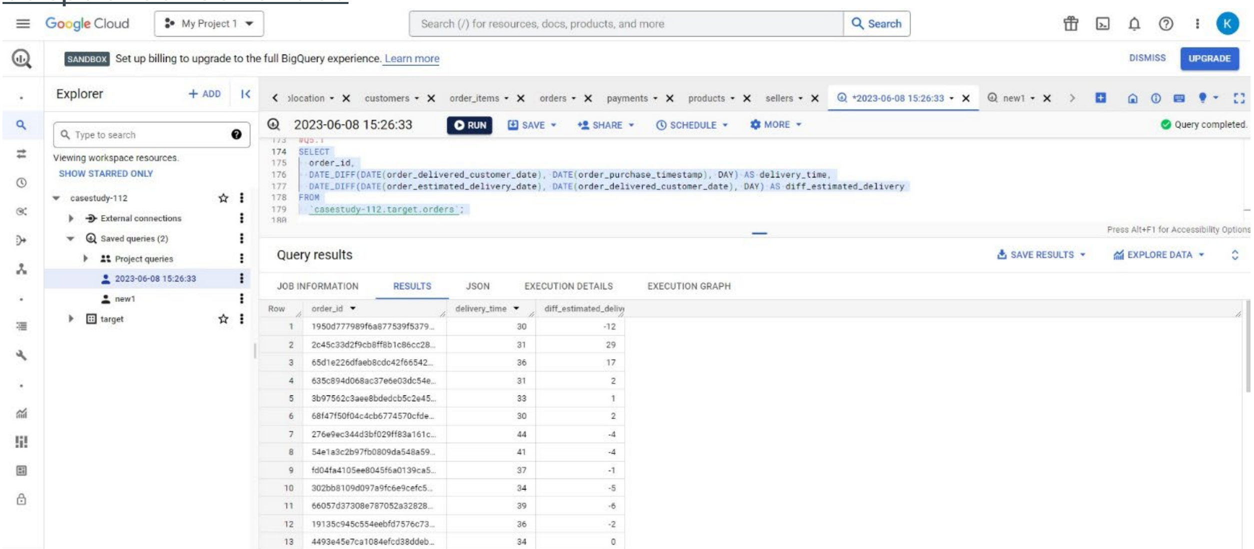
#Q5. Analysis based on sales, freight, and delivery time.

```
--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_estimated_delivery_date -
order_delivered_customer_date
```

Answer 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
  `casestudy-112.target.orders`;
```

Output screenshot:



Insights:

1. Insights into the effectiveness of the delivery process, including any delays or early deliveries compared to the projected timeframe, can be gained by analyzing the delivery_time and diff_estimated_delivery columns.
2. These columns can be further examined to find trends, outliers, or elements that affect delivery times or discrepancies between estimated and actual delivery dates.
3. These insights can be applied to manage customer expectations, enhance customer satisfaction, optimize the delivery process, and improve logistics operations.

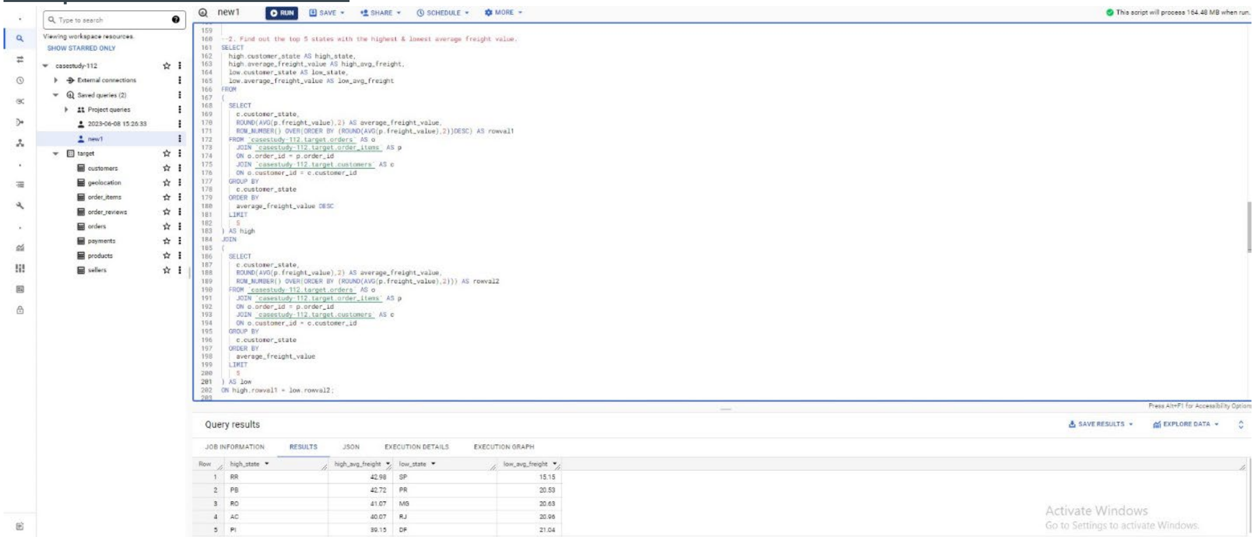
--2. Find out the top 5 states with the highest & lowest average freight value.

Answer 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 `casestudy-112.target.orders` AS o
    JOIN `casestudy-112.target.order_items` AS p
    ON o.order_id = p.order_id
    JOIN `casestudy-112.target.customers` AS c
    ON o.customer_id = c.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 `casestudy-112.target.orders` AS o
    JOIN `casestudy-112.target.order_items` AS p
    ON o.order_id = p.order_id
    JOIN `casestudy-112.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;
```

Output screenshot:



Insights:

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 our company 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.

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

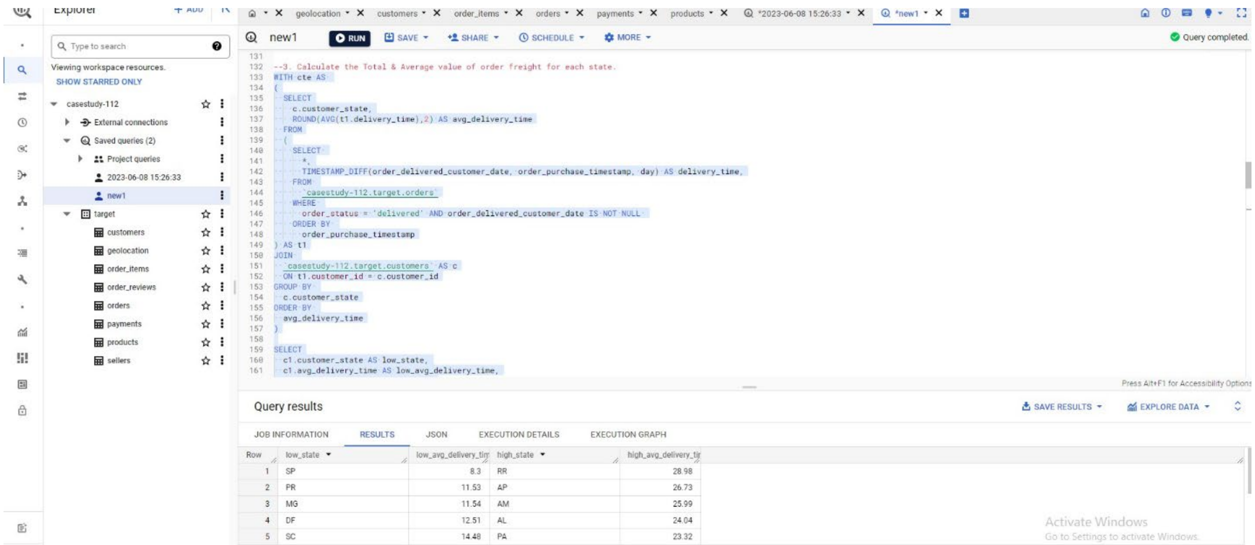
Answer Query:

```
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
      `casestudy-112.target.orders`
    WHERE
      order_status = 'delivered' AND
order_delivered_customer_date IS NOT NULL
    ORDER BY
      order_purchase_timestamp
) AS t1
JOIN
  `casestudy-112.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;
```

Output screenshot:



Insights:

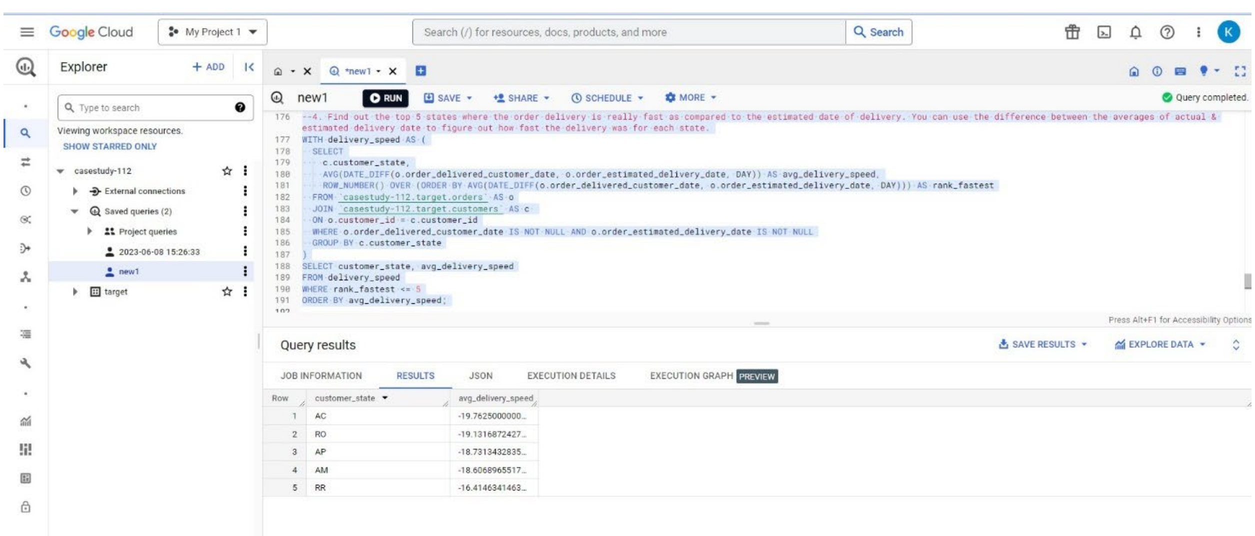
1. Finding areas with effective delivery operations, quicker transit times, or solid logistics networks can be done by looking at **the states like SP and PR with the lowest average delivery times and states called RR and AP with highest average delivery times.**
2. These insights can be helpful for our company looking to improve customer satisfaction, operational efficiency, delivery process optimization, and setting reasonable expectations for customers based on regional delivery time patterns.
3. When evaluating the data and drawing conclusions from these insights, it's crucial to take additional elements into account, such as population density, the distinction between urban and rural locations, customer expectations, or unique logistical restrictions.
4. Utilizing this information, our company can concentrate on areas where delivery efficiency improvements can be made, thereby improving customer experiences and operational efficiencies.

--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.

Answer 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 `casestudy-112.target.orders` AS o
  JOIN `casestudy-112.target.customers` AS 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;
```

Output screenshot:



Insights:

1. 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.
2. 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.

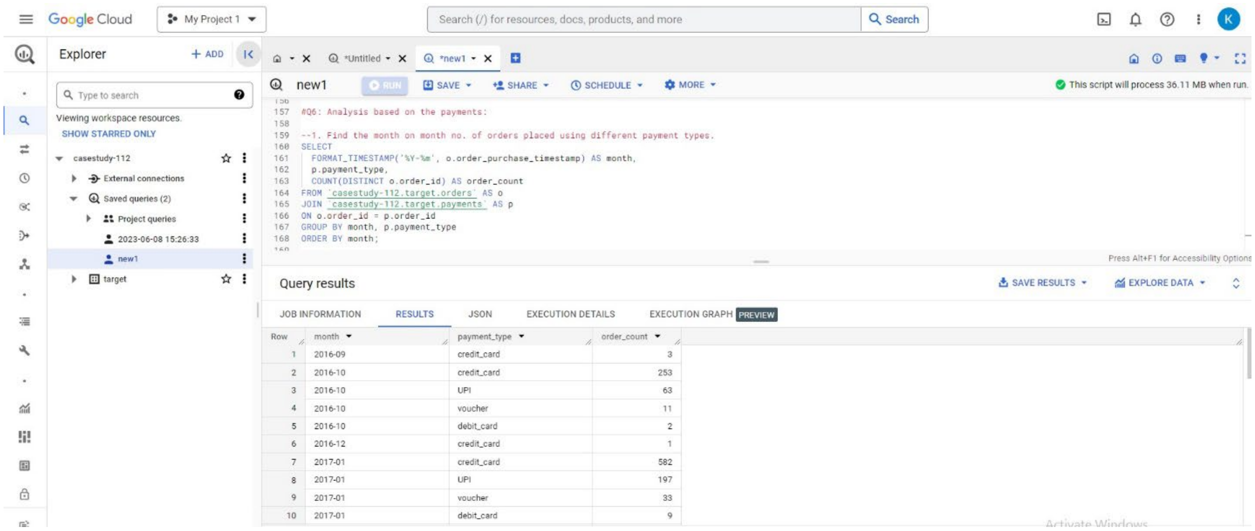
#Q6: Analysis based on the payments:

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

Answer Query:

```
SELECT
  FORMAT_TIMESTAMP('%Y-%m', o.order_purchase_timestamp) AS
month,
  p.payment_type,
  COUNT(DISTINCT o.order_id) AS order_count
FROM `casestudy-112.target.orders` AS o
JOIN `casestudy-112.target.payments` AS p
ON o.order_id = p.order_id
GROUP BY month, p.payment_type
ORDER BY month;
```

Output screenshot:



Insights:

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.

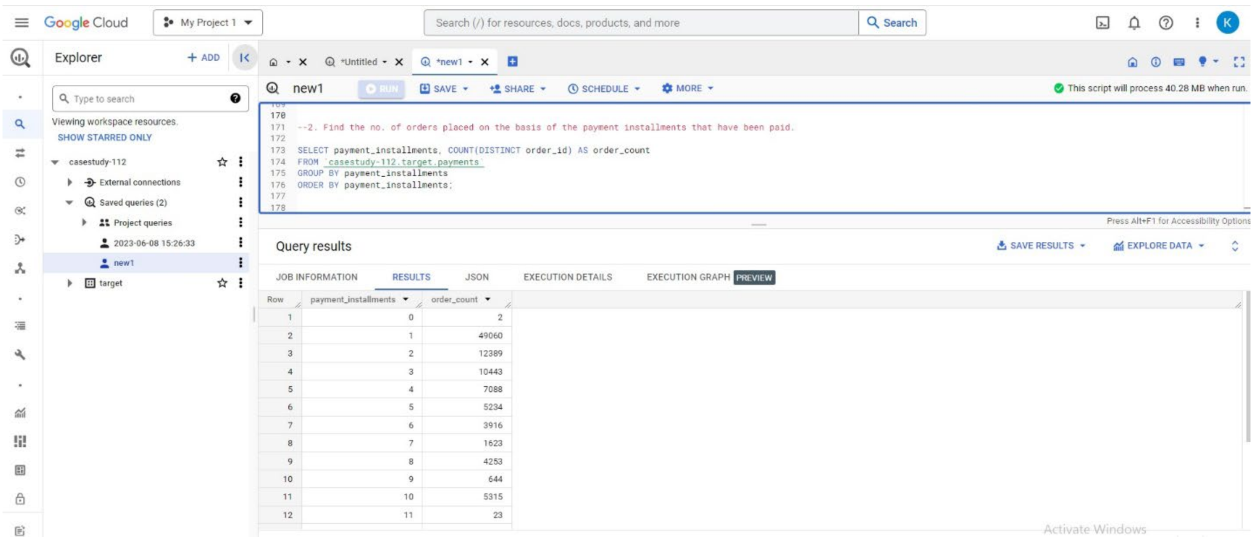
--2. Find the no. of orders placed on the basis of the payment installments that have been paid.

Answer Query:

```
SELECT payment_installments, COUNT(DISTINCT order_id) AS
order_count
FROM `casestudy-112.target.payments`
GROUP BY payment_installments
```

```sql
ORDER BY payment_installments;
```

<u>Output screenshot:</u>



<u>Insights:</u>

1. **<u>49060 orders were placed where payment installment was 1.</u>**
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.

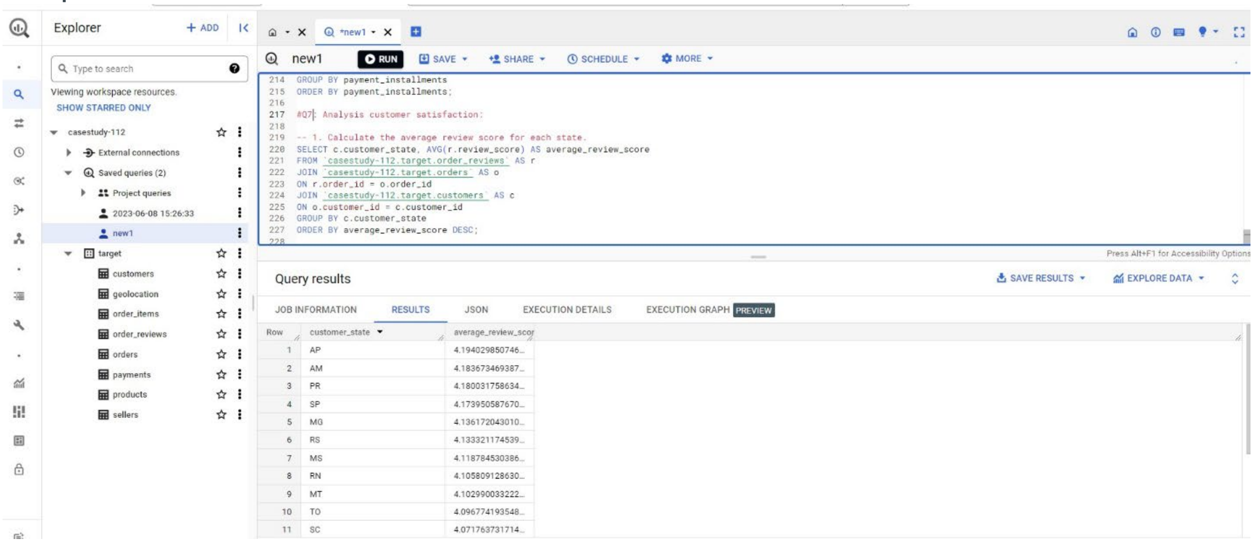
#Q7: Analysis customer satisfaction: (my own question)

```sql
-- 1. Calculate the average review score for each state.
```

Answer Query:

```
SELECT c.customer_state, AVG(r.review_score) AS
average_review_score
FROM `casestudy-112.target.order_reviews` AS r
JOIN `casestudy-112.target.orders` AS o
ON r.order_id = o.order_id
JOIN `casestudy-112.target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
ORDER BY average_review_score DESC;
```

Output screenshot:



Insights:

1. The average review score for each state gives a general idea of how satisfied customers are in various areas.

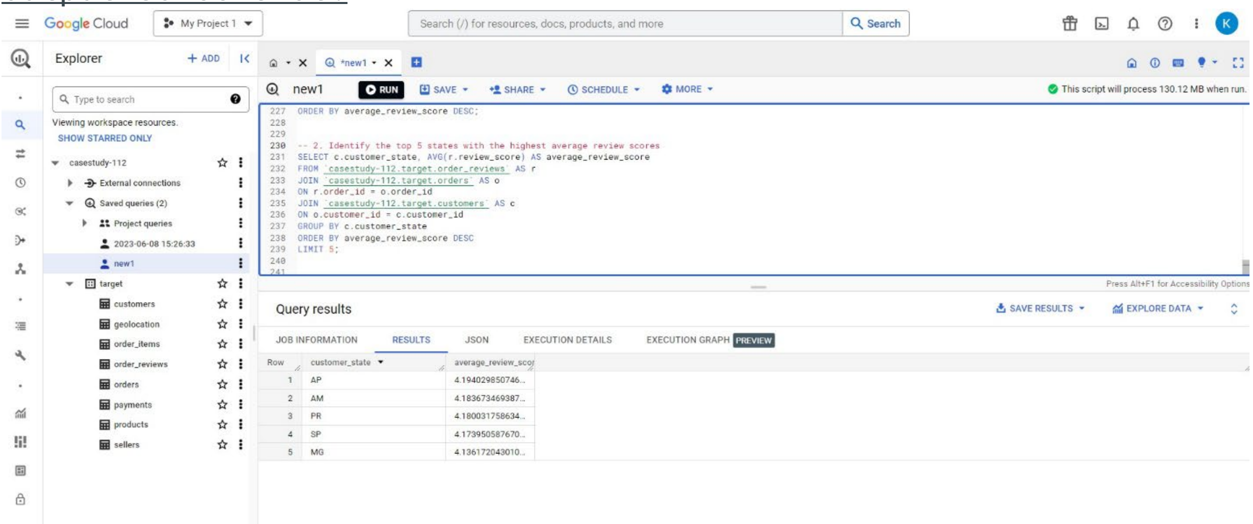
-- 2. Identify the top 5 states with the highest average review scores.

Answer Query:

```
SELECT c.customer_state, AVG(r.review_score) AS
average_review_score
FROM `casestudy-112.target.order_reviews` AS r
JOIN `casestudy-112.target.orders` AS o
```

```
ON r.order_id = o.order_id
JOIN `casestudy-112.target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY c.customer_state
ORDER BY average_review_score DESC
LIMIT 5;
```

Output screenshot:



Insights:

1. Customers have shown greater levels of satisfaction in the top 5 states with the highest average review scores.
2. Our company may concentrate on enhancing the client experience and resolving any issues by using these insights to pinpoint areas where customer satisfaction is high or low.

#Q8: Analysis on product popularity: (my own question)

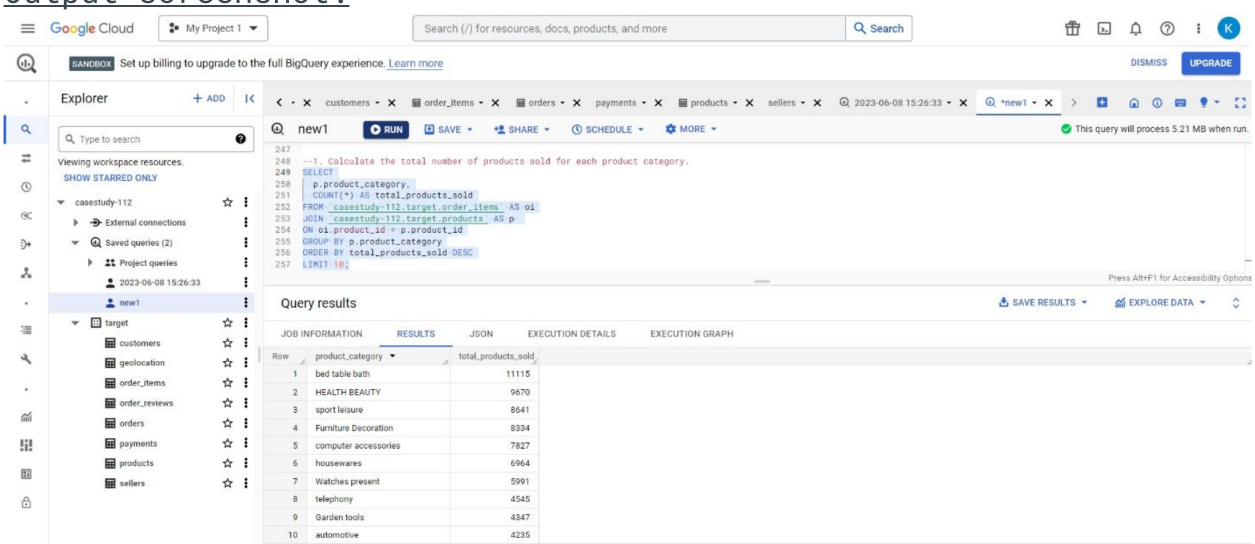
--1. Calculate the total number of products sold for each product category.

Answer Query:

```
SELECT
  p.product_category,
  COUNT(*) AS total_products_sold
```

```sql
FROM `casestudy-112.target.order_items` AS oi
JOIN `casestudy-112.target.products` AS p
ON oi.product_id = p.product_id
GROUP BY p.product_category
ORDER BY total_products_sold DESC
LIMIT 10;
```

Output screenshot:



Insights:

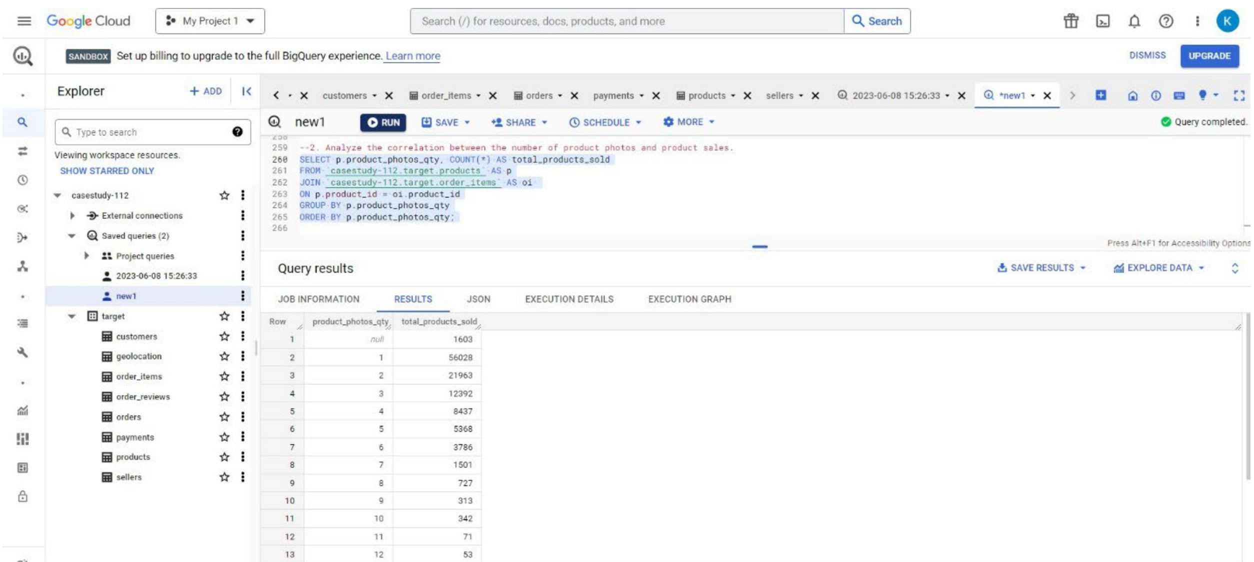
1. The popularity of various product categories can be determined by calculating the total number of products sold for each product category.
2. The analysis aids in identifying the product categories with larger sales volume, a sign of consumer preference.

```sql
--2. Analyze the correlation between the number of product photos and product sales.
```

Answer Query:

```sql
SELECT p.product_photos_qty, COUNT(*) AS total_products_sold
FROM `casestudy-112.target.products` AS p
JOIN `casestudy-112.target.order_items` AS oi
ON p.product_id = oi.product_id
GROUP BY p.product_photos_qty
ORDER BY p.product_photos_qty;
```

<u>Output screenshot:</u>



<u>Insights:</u>

1. Our company can ascertain the influence of visual material on customers' purchase selections by comprehending the relationship between the quantity of product photographs and product sales.
2. By concentrating on top-performing product categories and improving the visual presentation of products with more images, this information can be used to optimize product listings.