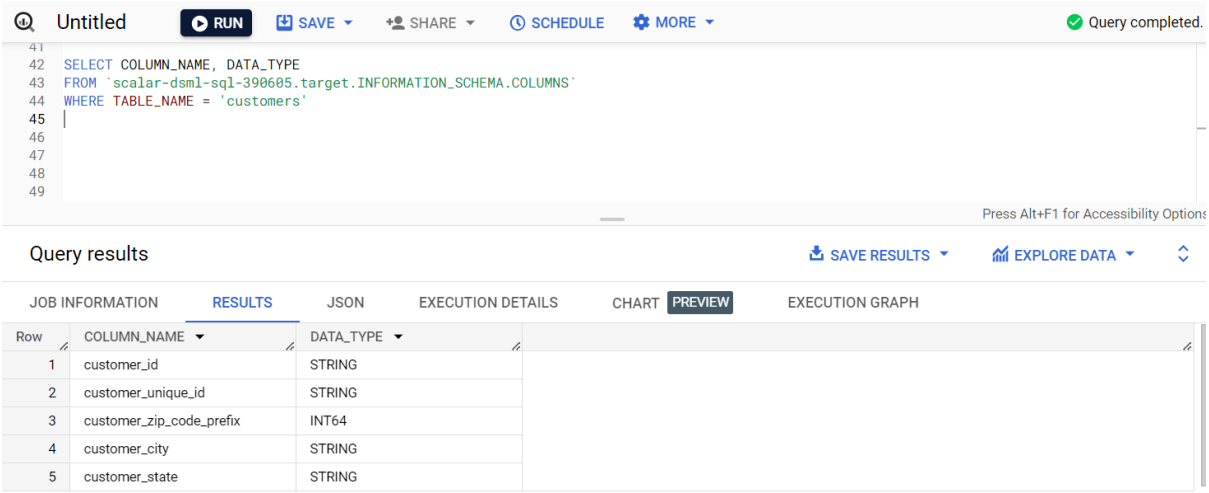
1. Import the dataset and do usual exploratory analysis steps like checking the structure & characteristics of the dataset:

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

Ans-

```
SELECT COLUMN_NAME, DATA_TYPE
FROM `scalar-dsml-sql-390605.target.INFORMATION_SCHEMA.COLUMNS`
WHERE TABLE_NAME = 'customers'
```



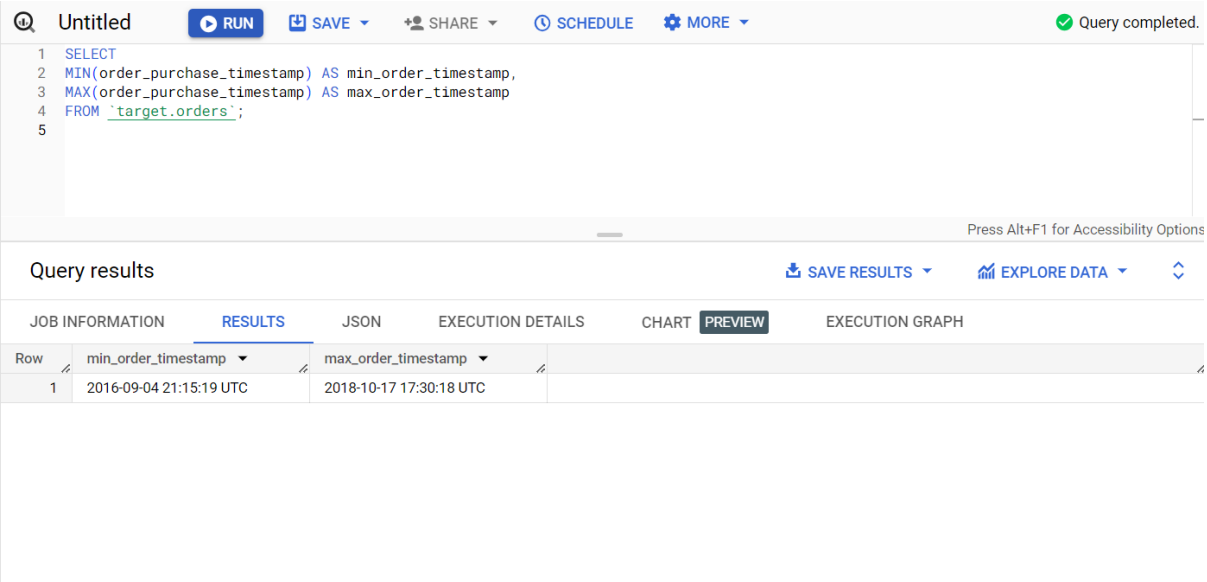
Row	COLUMN_NAME	DATA_TYPE
1	customer_id	STRING
2	customer_unique_id	STRING
3	customer_zip_code_prefix	INT64
4	customer_city	STRING
5	customer_state	STRING

- **Insights** - Understanding the data types is essential for correct interpretation and analysis. It helps ensure that we are working with the data in an appropriate manner.

- **Recommendations** - Ensure that the data types of the columns in the "customers" table match the nature of the data they contain. Incorrect data types can lead to errors and inaccurate analysis.

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

Ans-
```
SELECT
MIN(order_purchase_timestamp) AS min_order_timestamp,
MAX(order_purchase_timestamp) AS max_order_timestamp
FROM `target.orders`;
```

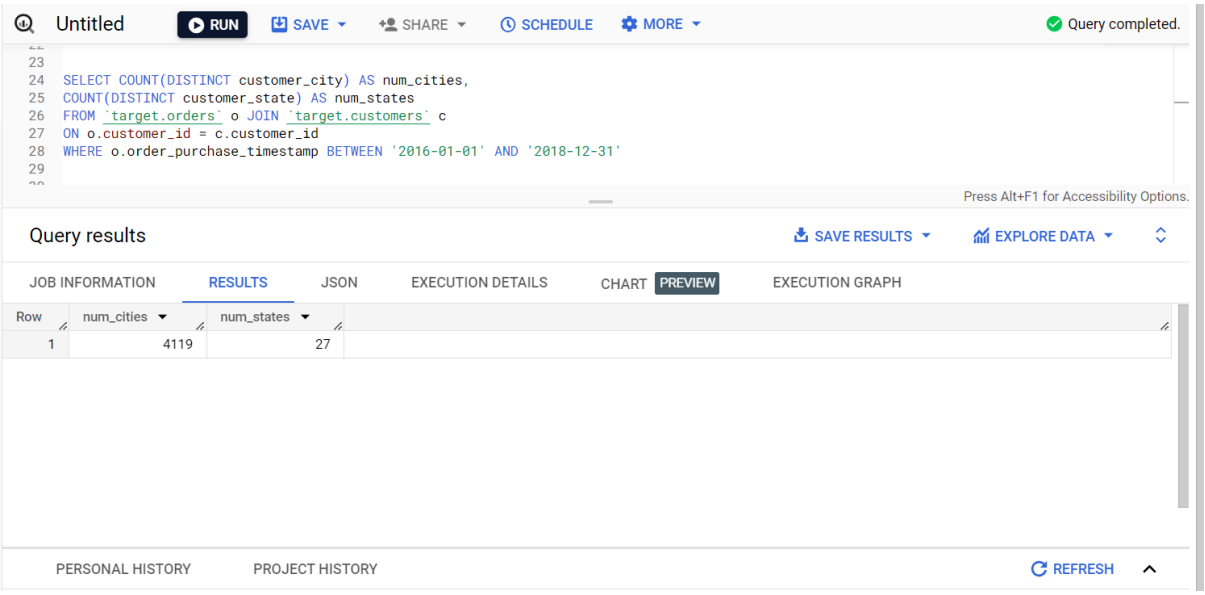


- **Insights -** The time range gives you a sense of the duration for which the dataset captures order information. This information is valuable for understanding historical trends and patterns.

- **Recommendations** - Identify the earliest and latest order purchase timestamps to establish the time range. This information is useful for understanding the period of order placements and for planning marketing campaigns and inventory management.

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

Ans-
```
SELECT COUNT(DISTINCT customer_city) AS num_cities,
COUNT(DISTINCT customer_state) AS num_states
FROM `target.orders` o JOIN `target.customers` c
ON o.customer_id = c.customer_id
WHERE o.order_purchase_timestamp BETWEEN '2016-01-01' AND '2018-12-31';
```



- **Insights** - This analysis helps identify the regions with the highest customer engagement during the specified time period. It can provide insights into where your customer base is concentrated.

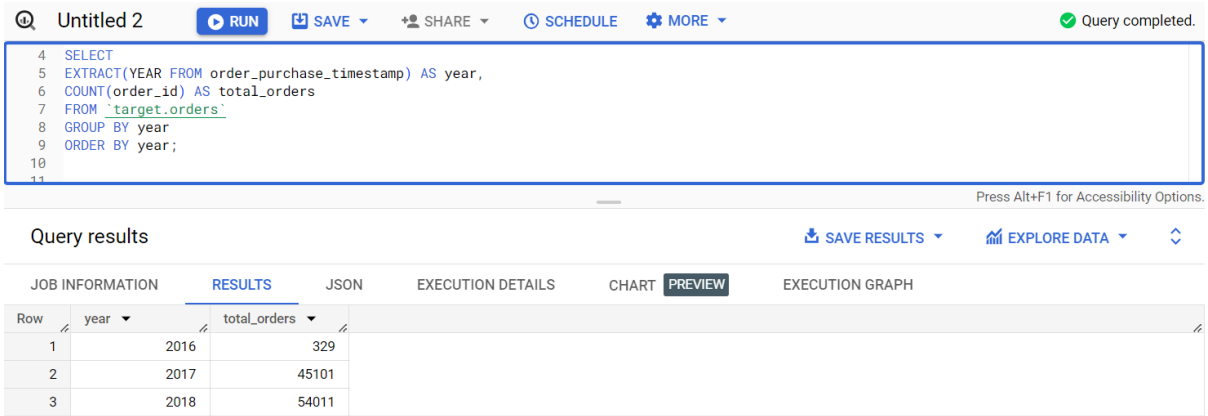
- **Recommendations** - Calculate the counts of customers in each city and state who placed orders within the specified time range. This will help in targeting specific regions for marketing efforts, optimizing delivery logistics, and tailoring services based on customer demographics.

2. In-depth Exploration:

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

Ans-

```
SELECT
EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
COUNT(order_id) AS total_orders
FROM `target.orders`
GROUP BY year
ORDER BY year;
```



Row	year	total_orders
1	2016	329
2	2017	45101
3	2018	54011

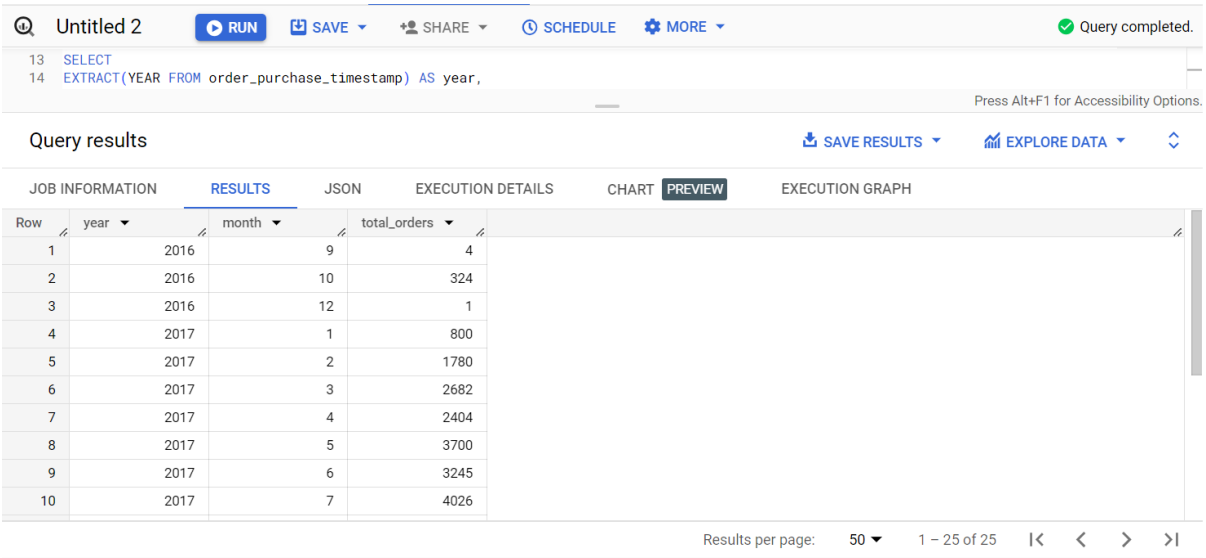
- **Insights** - Analyzing the order counts over the years can reveal there is a growing trend in the number of orders placed. A steady increase in orders over the years may suggest growth in Target's operations.

- **Recommendations** - Based on the analysis, There is a growing trend in the number of orders placed over the past years, Target may need to allocate resources, optimize logistics, and prepare for increased demand in the future years.

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

Ans-

```
SELECT
EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
COUNT(order_id) AS total_orders
FROM `target.orders`
GROUP BY year, month
ORDER BY year, month;
```



Row	year	month	total_orders
1	2016	9	4
2	2016	10	324
3	2016	12	1
4	2017	1	800
5	2017	2	1780
6	2017	3	2682
7	2017	4	2404
8	2017	5	3700
9	2017	6	3245
10	2017	7	4026

- **Insights** - Analysing the monthly order counts over the years can help identify any seasonality patterns. Peaks in certain months may indicate high-demand periods or specific shopping seasons.

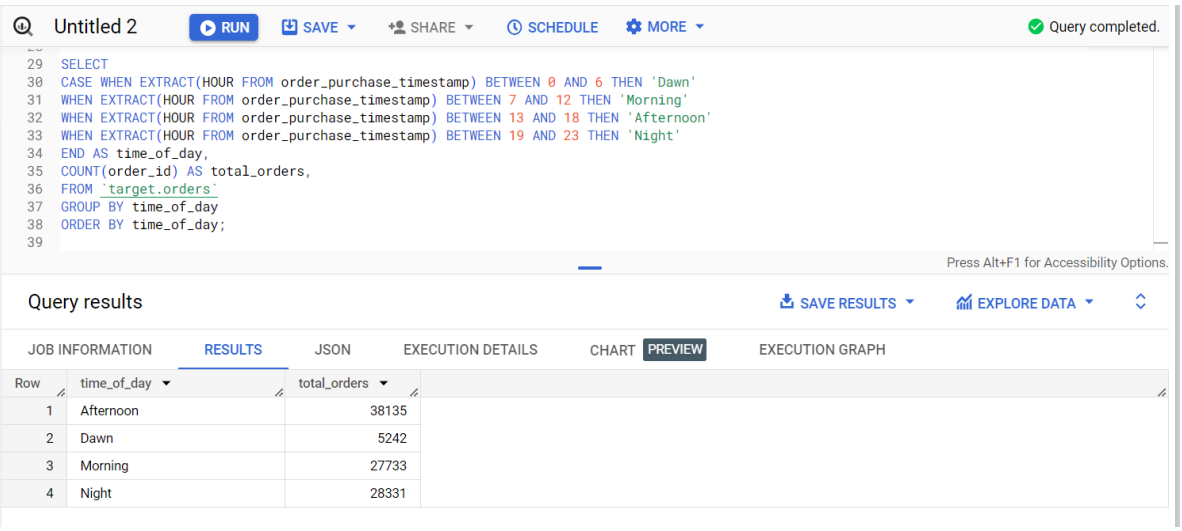
- **Recommendations** - There is evidence of monthly seasonality, Target can plan marketing campaigns, promotions, and inventory management accordingly to maximize sales during peak months.

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

Ans-

```
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 time_of_day,
COUNT(order_id) AS total_orders,
FROM `target.orders`
GROUP BY time_of_day
ORDER BY time_of_day;
```



Row	time_of_day	total_orders
1	Afternoon	38135
2	Dawn	5242
3	Morning	27733
4	Night	28331

- **Insights** - Understanding the time of day when customers mostly place orders can assist in optimizing marketing campaigns, promotions, and customer support services.

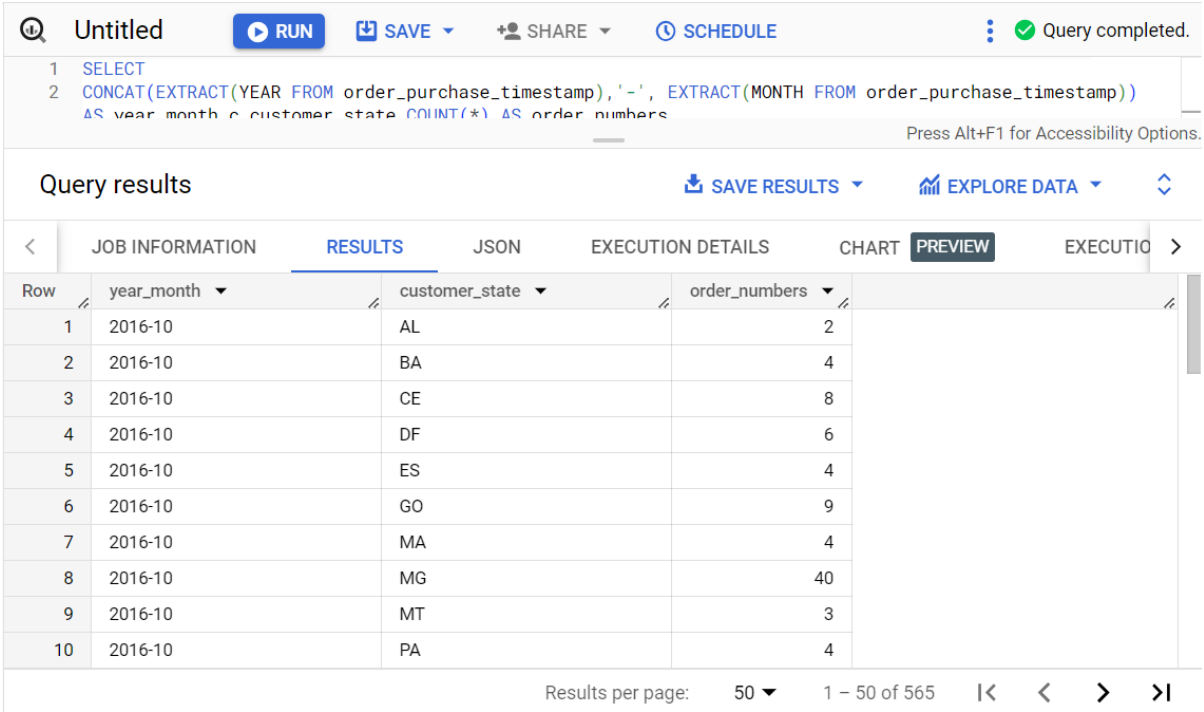
- **Recommendations -** Based on the analysis of when Brazilian customers mostly place their orders, Target can schedule marketing activities, customer support, and product releases to align with peak ordering times.

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

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

Ans-

```
SELECT CONCAT(EXTRACT(YEAR FROM order_purchase_timestamp),'-',
EXTRACT(MONTH FROM order_purchase_timestamp)) AS year_month,
c.customer_state,COUNT(*) AS order_numbers
FROM `target.orders` o JOIN `target.customers` c
ON o.customer_id = c.customer_id
GROUP BY year_month, c.customer_state
ORDER BY year_month, c.customer_state;
```



Row	year_month	customer_state	order_numbers
1	2016-10	AL	2
2	2016-10	BA	4
3	2016-10	CE	8
4	2016-10	DF	6
5	2016-10	ES	4
6	2016-10	GO	9
7	2016-10	MA	4
8	2016-10	MG	40
9	2016-10	MT	3
10	2016-10	PA	4

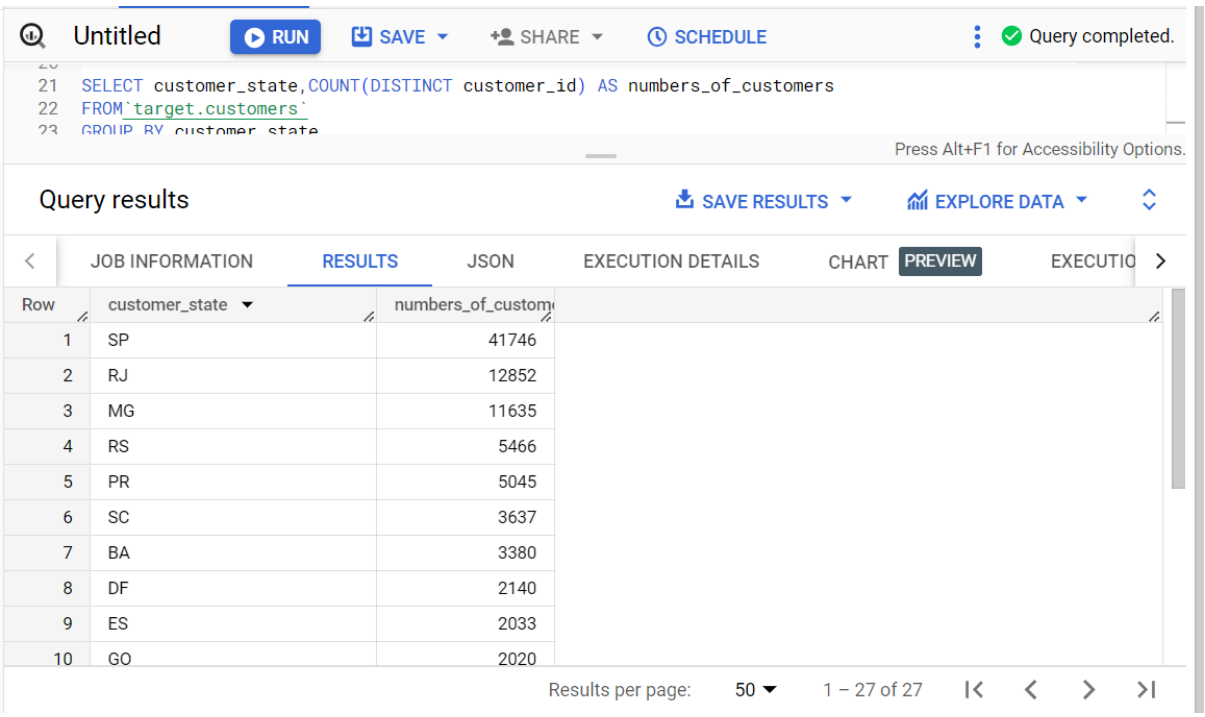
- **Insights** - The output of this query will allow to visualize the trends in e-commerce orders on a monthly basis for each state. By comparing the current month's order count with the previous month's order count, we can identify states that are experiencing significant growth or decline in e-commerce orders. This information can help Target allocate resources, plan promotions, and optimize logistics in different states.

- **Recommendations** - Analyze the patterns in the month-on-month order counts to identify seasonal trends. This information can be used to plan inventory, staffing, and marketing strategies accordingly.

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

Ans-

```
SELECT customer_state,
COUNT(DISTINCT customer_id) AS numbers_of_customers
FROM`target.customers`
GROUP BY customer_state
ORDER BY numbers_of_customers DESC;
```



Row	customer_state	numbers_of_custom
1	SP	41746
2	RJ	12852
3	MG	11635
4	RS	5466
5	PR	5045
6	SC	3637
7	BA	3380
8	DF	2140
9	ES	2033
10	GO	2020

- **Insights -** The result of this query will provide a clear view of the distribution of customers across all states. We can identify states with the highest number of unique customers, which can be valuable for targeting marketing campaigns, expanding operations, and tailoring services to specific regions. It also helps to understand the market reach and potential for growth in different states.

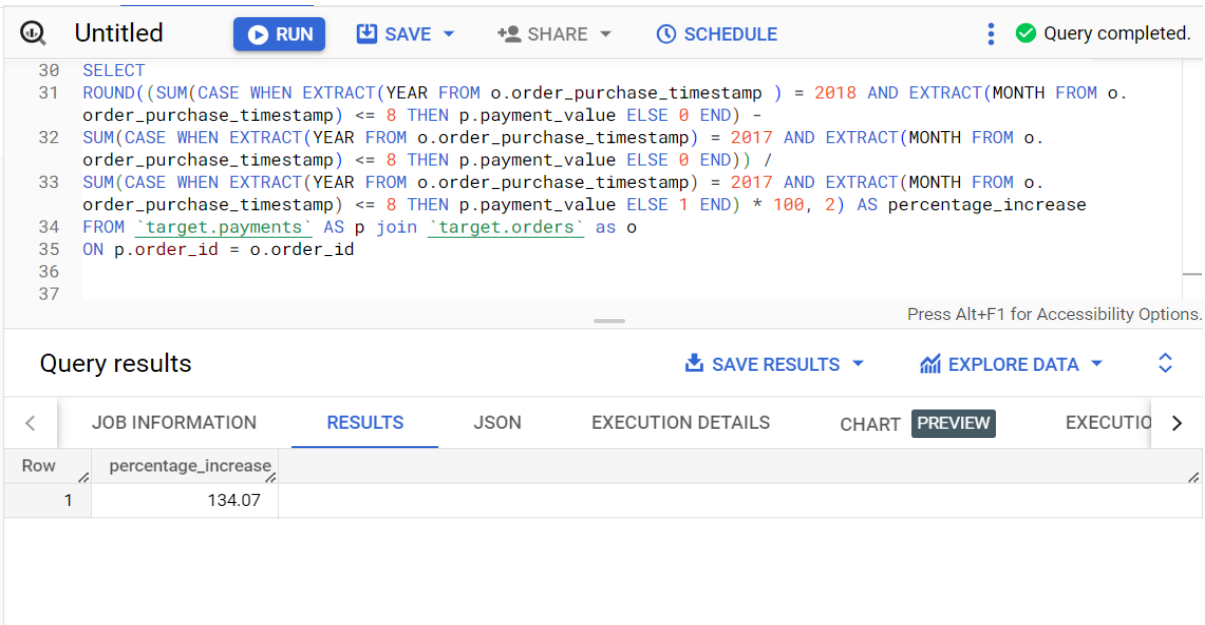
- **Recommendations -** Use the customer distribution data to tailor marketing campaigns for specific states or regions. Consider offering localized promotions and incentives to attract more customers from under represented states.

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).**

Ans-

```
SELECT
ROUND((SUM(CASE WHEN EXTRACT(YEAR FROM
o.order_purchase_timestamp ) = 2018 AND EXTRACT(MONTH FROM
o.order_purchase_timestamp) <= 8 THEN p.payment_value ELSE 0
END) -
SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
2017 AND EXTRACT(MONTH FROM o.order_purchase_timestamp) <= 8
THEN p.payment_value ELSE 0 END)) /
SUM(CASE WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) =
2017 AND EXTRACT(MONTH FROM o.order_purchase_timestamp) <= 8
THEN p.payment_value ELSE 1 END) * 100, 2) AS
percentage_increase
FROM `target.payments` AS p join `target.orders` as o
ON p.order_id = o.order_id
```



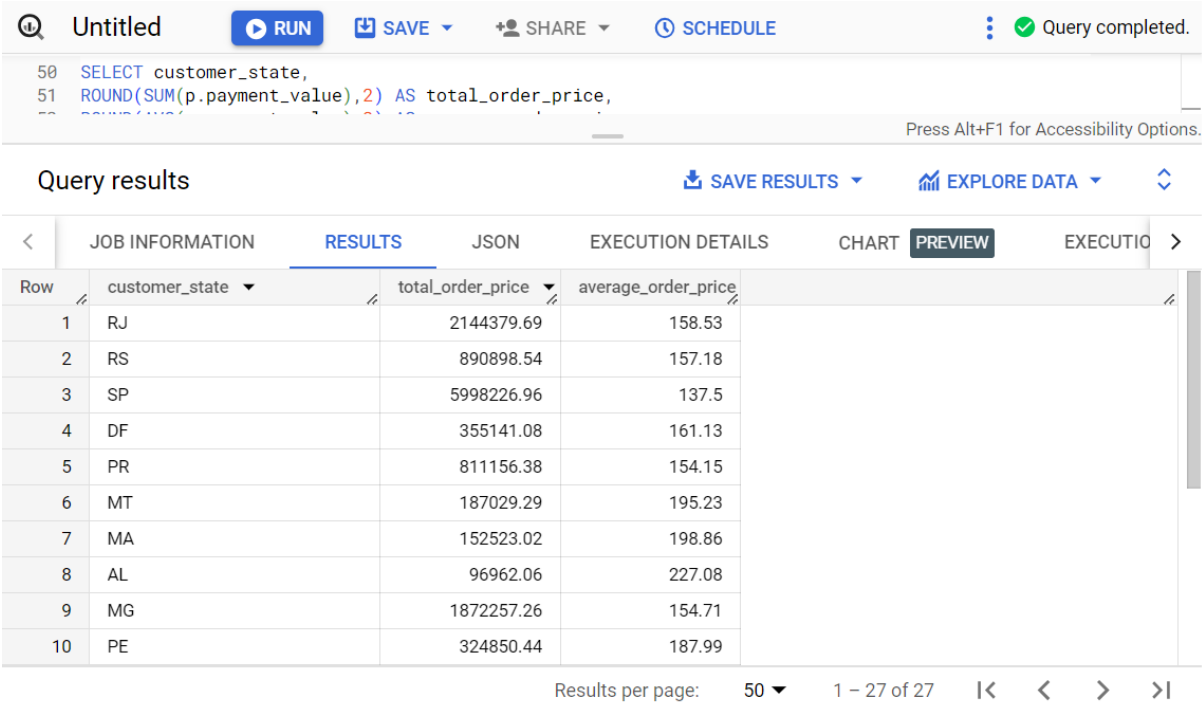
- **Insights** - Calculate the percentage increase in total payment from 2017 to 2018. This will provide insight into the growth of e-commerce in terms of order payments during this period.

- **Recommendations** - Analyze the reasons behind significant increases in payment values. It could be due to changes in customer behaviour, pricing strategies, or external factors. Adjust business strategies accordingly.

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

Ans-

```
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` AS p JOIN `target.orders` AS o
ON p.order_id = o.order_id
JOIN `target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state;
```



Row	customer_state	total_order_price	average_order_price
1	RJ	2144379.69	158.53
2	RS	890898.54	157.18
3	SP	5998226.96	137.5
4	DF	355141.08	161.13
5	PR	811156.38	154.15
6	MT	187029.29	195.23
7	MA	152523.02	198.86
8	AL	96962.06	227.08
9	MG	1872257.26	154.71
10	PE	324850.44	187.99

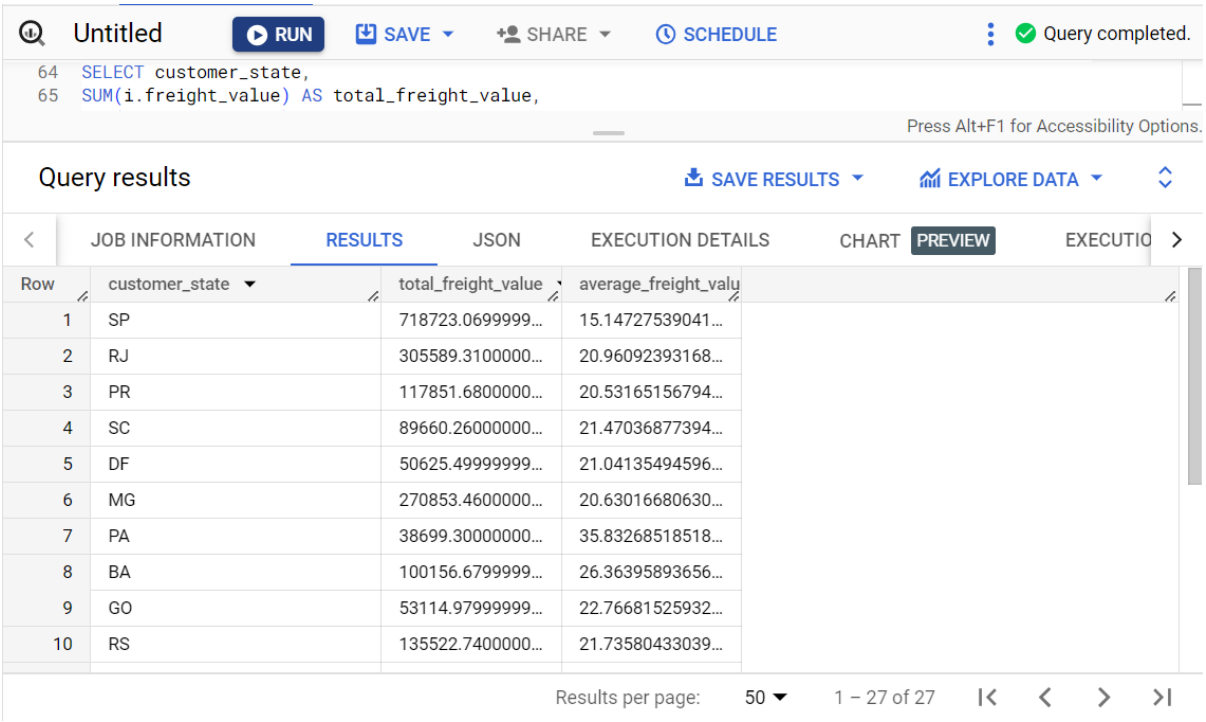
- **Insights** - Calculating the total and average order price for each state helps identify regions with higher overall spending and average transaction values. This information can guide targeted marketing efforts and inventory planning.

- **Recommendations** - Identify states with high average order values and consider targeted marketing efforts to further boost sales in those regions.

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

Ans-

```
SELECT customer_state,
SUM(i.freight_value) AS total_freight_value,
AVG(i.freight_value) AS average_freight_value
FROM `target.order_items` AS i JOIN `target.orders` AS o
ON i.order_id = o.order_id
JOIN `target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state;
```



Row	customer_state	total_freight_value	average_freight_valu
1	SP	718723.0699999...	15.14727539041...
2	RJ	305589.3100000...	20.96092393168...
3	PR	117851.6800000...	20.53165156794...
4	SC	89660.26000000...	21.47036877394...
5	DF	50625.49999999...	21.04135494596...
6	MG	270853.4600000...	20.63016680630...
7	PA	38699.30000000...	35.83268518518...
8	BA	100156.6799999...	26.36395893656...
9	GO	53114.97999999...	22.76681525932...
10	RS	135522.7400000...	21.73580433039...

Insights - Identify states with the highest total and average freight costs. This can help in optimizing logistics and delivery strategies.

Recommendations - Optimize logistics and shipping strategies in states with high average freight values to minimize costs and enhance customer satisfaction.

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.

Ans-

```
SELECT order_id,
TIMESTAMP_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY) AS time_to_deliver,
TIMESTAMP_DIFF(order_estimated_delivery_date,
order_delivered_customer_date, DAY) AS diff_estimated_delivery
FROM `target.orders`;
```

Untitled RUN MORE SAVE SHARE SCHEDULE Query completed.

```
67 SELECT order_id,
68 TIMESTAMP_DIFF(order_delivered_customer_date, order_purchase_timestamp, DAY) AS time_to_deliver,
69 TIMESTAMP_DIFF(order_estimated_delivery_date, order_delivered_customer_date, DAY) AS diff_estimated_delivery
```

Press Alt+F1 for Accessibility Options.

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	order_id	time_to_deliver	diff_estimated_deliv
1	1950d777989f6a877539f5379...	30	-12
2	2c45c33d2f9cb8ff8b1c86cc28...	30	28
3	65d1e226dfaeb8cdc42f66542...	35	16
4	635c894d068ac37e6e03dc54e...	30	1
5	3b97562c3aee8bdedcb5c2e45...	32	0
6	68f47f50f04c4cb6774570cfde...	29	1
7	276e9ec344d3bf029ff83a161c...	43	-4
8	54e1a3c2b97fb0809da548a59...	40	-4
9	fd04fa4105ee8045f6a0139ca5...	37	-1
10	302bb8109d097a9fc6e9cefc5...	33	-5

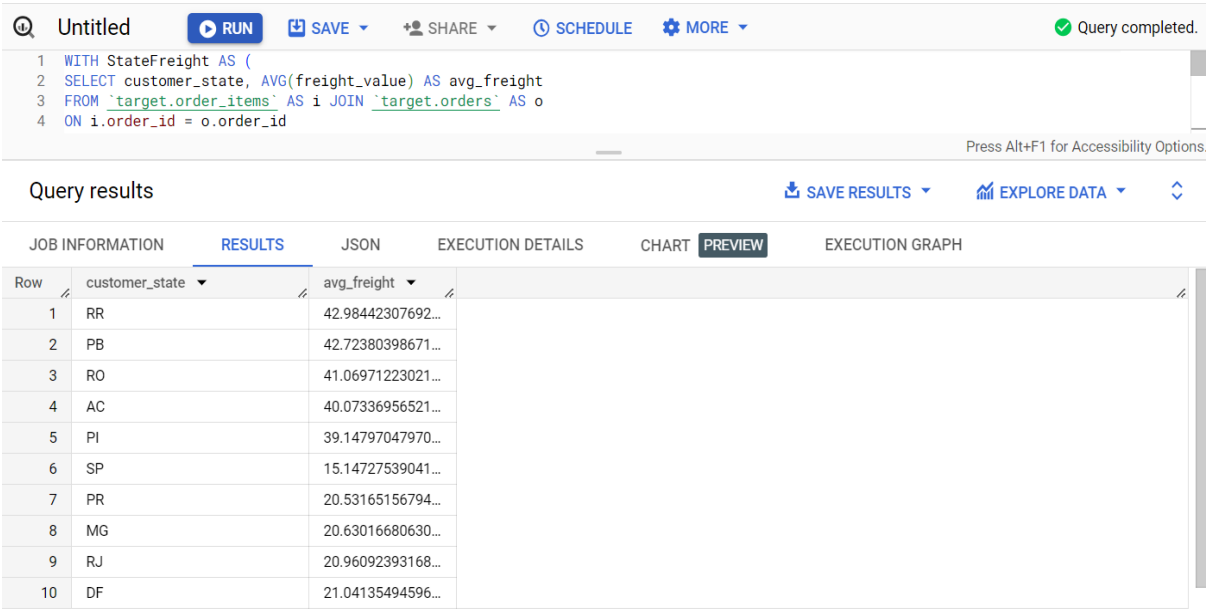
Results per page: 50 1 – 50 of 99441

- **Insights** - Delivery Time: This provides an understanding of the average time taken to deliver orders to customers after the purchase date. A shorter delivery time indicates better shipping efficiency**.**
  Difference in Estimated & Actual Delivery: This helps identify how well the estimated delivery dates align with the actual delivery dates. A negative value indicates orders were delivered earlier than expected, while a positive value indicates delays.

- **Recommendations** - Analyzing delivery times and the differences between estimated and actual delivery dates can help identify areas for improvement in Target's logistics and delivery processes. If the "diff_estimated_delivery" is consistently negative, it indicates that Target is delivering orders before the estimated delivery date, which can enhance customer satisfaction.

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

Ans-
```
WITH StateFreight AS (
SELECT customer_state, AVG(freight_value) AS avg_freight
FROM `target.order_items` AS i JOIN `target.orders` AS o
ON i.order_id = o.order_id
JOIN `target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state
)
(SELECT customer_state, avg_freight
FROM StateFreight
ORDER BY avg_freight DESC
LIMIT 5)
UNION ALL
(SELECT customer_state, avg_freight
FROM StateFreight
ORDER BY avg_freight ASC
LIMIT 5);
```



Row	customer_state	avg_freight
1	RR	42.98442307692...
2	PB	42.72380398671...
3	RO	41.06971223021...
4	AC	40.07336956521...
5	PI	39.14797047970...
6	SP	15.14727539041...
7	PR	20.53165156794...
8	MG	20.63016680630...
9	RJ	20.96092393168...
10	DF	21.04135494596...

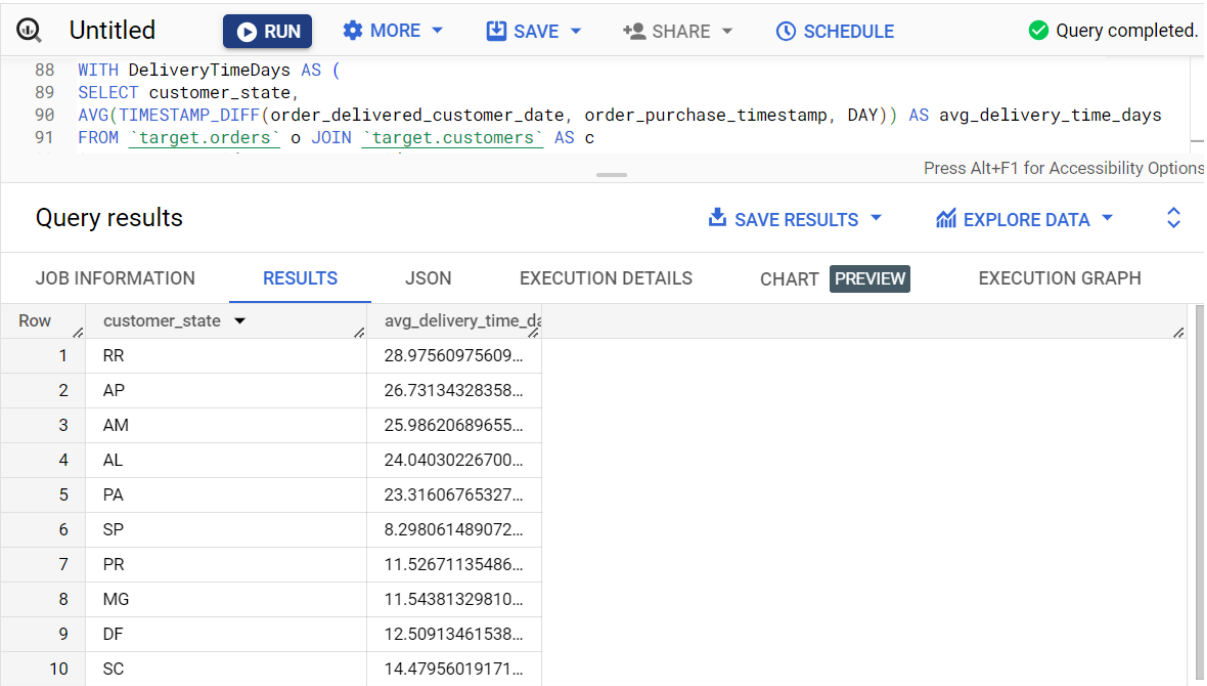
- **Insights** - High Freight States: Identify the states where customers pay higher freight costs on average. This information can be used to evaluate shipping strategies and pricing in those regions.
  Low Freight States: Discover states with lower average freight costs. This might help target marketing efforts or optimize shipping to improve cost efficiency.

- **Recommendations** - Identifying states with high and low average freight values can help in optimizing shipping costs. If the freight costs are consistently high in certain states, Target can explore options to negotiate better shipping rates or adjust pricing strategies.

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

Ans-

```
WITH DeliveryTimeDays AS (
SELECT customer_state,
AVG(TIMESTAMP_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)) AS avg_delivery_time_days
FROM `target.orders` o JOIN `target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state )
(SELECT customer_state,avg_delivery_time_days
FROM DeliveryTimeDays
ORDER BY avg_delivery_time_days DESC
LIMIT 5)
UNION ALL
(SELECT customer_state,avg_delivery_time_days
FROM DeliveryTimeDays
ORDER BY avg_delivery_time_days ASC
LIMIT 5);
```



Row	customer_state	avg_delivery_time_da
1	RR	28.97560975609...
2	AP	26.73134328358...
3	AM	25.98620689655...
4	AL	24.04030226700...
5	PA	23.31606765327...
6	SP	8.298061489072...
7	PR	11.52671135486...
8	MG	11.54381329810...
9	DF	12.50913461538...
10	SC	14.47956019171...

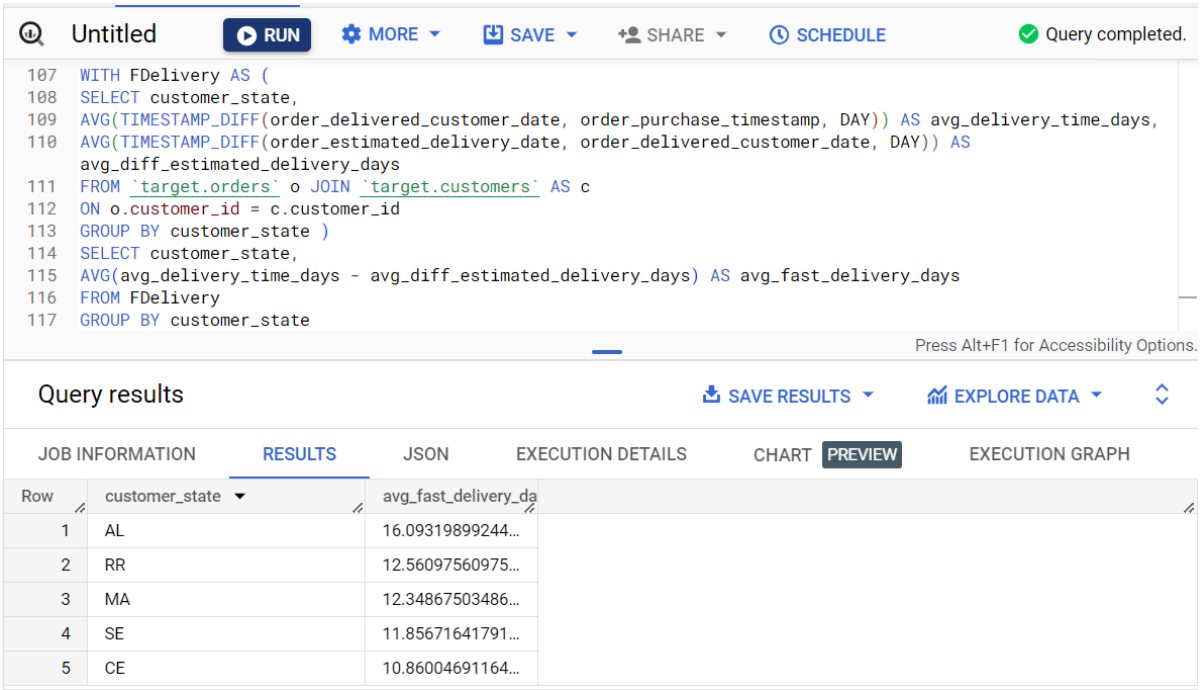
- **Insights** - High Delivery Time States: Identify states where customers experience longer average delivery times. This insight can help in pinpointing potential logistic issues or areas for improvement.
  Low Delivery Time States: Discover states with shorter average delivery times. These regions can serve as benchmarks for efficient delivery.

- **Recommendations** - Analyzing average delivery times across states can help Target identify areas where delivery processes need improvement. Addressing states with longer delivery times can lead to increased customer 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.

Ans-

```
WITH FDelivery AS (
SELECT customer_state,
AVG(TIMESTAMP_DIFF(order_delivered_customer_date,
order_purchase_timestamp, DAY)) AS avg_delivery_time_days,
AVG(TIMESTAMP_DIFF(order_estimated_delivery_date,
order_delivered_customer_date, DAY)) AS
avg_diff_estimated_delivery_days
FROM `target.orders` o JOIN `target.customers` AS c
ON o.customer_id = c.customer_id
GROUP BY customer_state )
SELECT customer_state,
AVG(avg_delivery_time_days - avg_diff_estimated_delivery_days)
AS avg_fast_delivery_days
FROM FDelivery
GROUP BY customer_state
ORDER BY avg_fast_delivery_days DESC
LIMIT 5;
```



Row	customer_state	avg_fast_delivery_da
1	AL	16.09319899244...
2	RR	12.56097560975...
3	MA	12.34867503486...
4	SE	11.85671641791...
5	CE	10.86004691164...

- **Insights** - Fast Delivery States: Identify states where actual delivery dates are consistently earlier than estimated delivery dates. This indicates efficient delivery operations and customer satisfaction.

- **Recommendations -** Identifying states where actual delivery is consistently faster than the estimated delivery date showcases efficient logistics operations. Target can study these states to learn from their practices and apply them to other regions.

6. Analysis based on the payments:

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

Ans-

```
SELECT
EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
payment_type,
COUNT(DISTINCT o.order_id) AS num_orders
FROM `target.orders` o join `target.payments` p
ON o.order_id = p.order_id
GROUP BY year, month, payment_type
ORDER BY year, month, payment_type;
```

Untitled RUN MORE SAVE SHARE SCHEDULE Query completed.

```
121 SELECT
122 EXTRACT(YEAR FROM order_purchase_timestamp) AS year,
123 EXTRACT(MONTH FROM order_purchase_timestamp) AS month,
```

Press Alt+F1 for Accessibility Options

Query results SAVE RESULTS EXPLORE DATA

JOB INFORMATION | RESULTS | JSON | EXECUTION DETAILS | CHART PREVIEW | EXECUTION GRAPH

Row	year	month	payment_type	num_orders
1	2016	9	credit_card	3
2	2016	10	UPI	63
3	2016	10	credit_card	253
4	2016	10	debit_card	2
5	2016	10	voucher	11
6	2016	12	credit_card	1
7	2017	1	UPI	197
8	2017	1	credit_card	582
9	2017	1	debit_card	9
10	2017	1	voucher	33

Results per page: 50 1 – 50 of 90

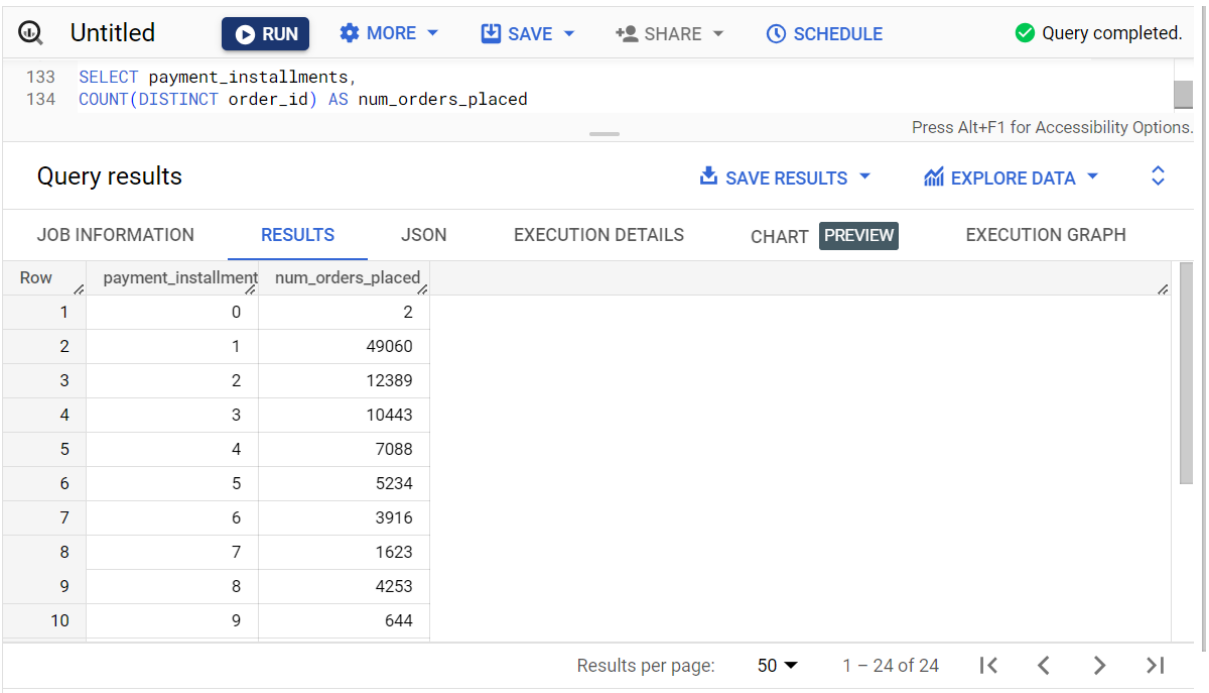
- **Insights** - We can identify trends in payment preferences over time. For instance, we might observe that credit card payments are more popular during certain months, while boleto payments are more common during others. This insight can help Target tailor its payment options and marketing strategies accordingly.

- **Recommendations** - Analyzing the month-on-month order count for different payment types can help Target understand which payment methods are preferred by customers. Target can tailor promotions or discounts to encourage the use of certain payment types.

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

Ans-

```
SELECT payment_installments,
COUNT(DISTINCT order_id) AS num_orders_placed
FROM `target.payments`
GROUP BY payment_installments
ORDER BY payment_installments;
```



Row	payment_installment	num_orders_placed
1	0	2
2	1	49060
3	2	12389
4	3	10443
5	4	7088
6	5	5234
7	6	3916
8	7	1623
9	8	4253
10	9	644

- **Insights** - We can understand how customers choice of payment installments affects their purchasing behavior. It might reveal that customers who choose higher installments tend to place larger orders or that customers prefer to pay in full for smaller orders. This insight can help optimize payment options and pricing strategies.

- **Recommendations** - Analyzing the distribution of orders based on payment installments can help Target understand how customers prefer to pay for their orders. This insight can guide marketing strategies or payment-related offers.