Business Case: Target SQL Solution

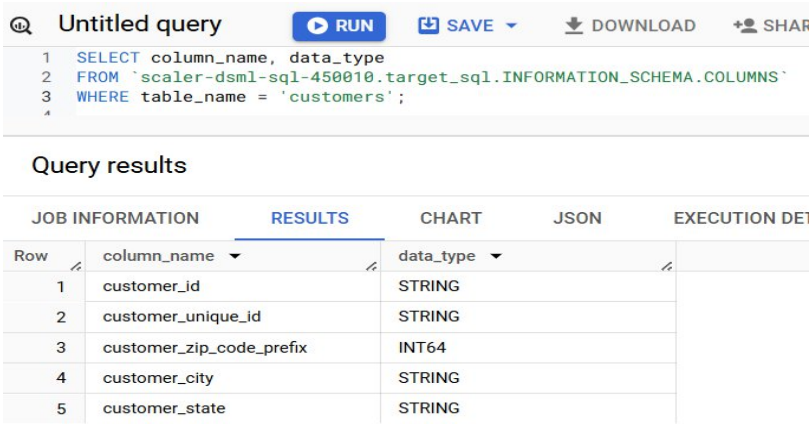
Here are the analysis outcomes.

1: Initial Exploration:

I have imported the dataset in BigQuery and usual exploratory analysis and preview the structure of each file. Started by checking the data structure of each file in the dataset.

1. Here is the schema of the '**customers'** table:

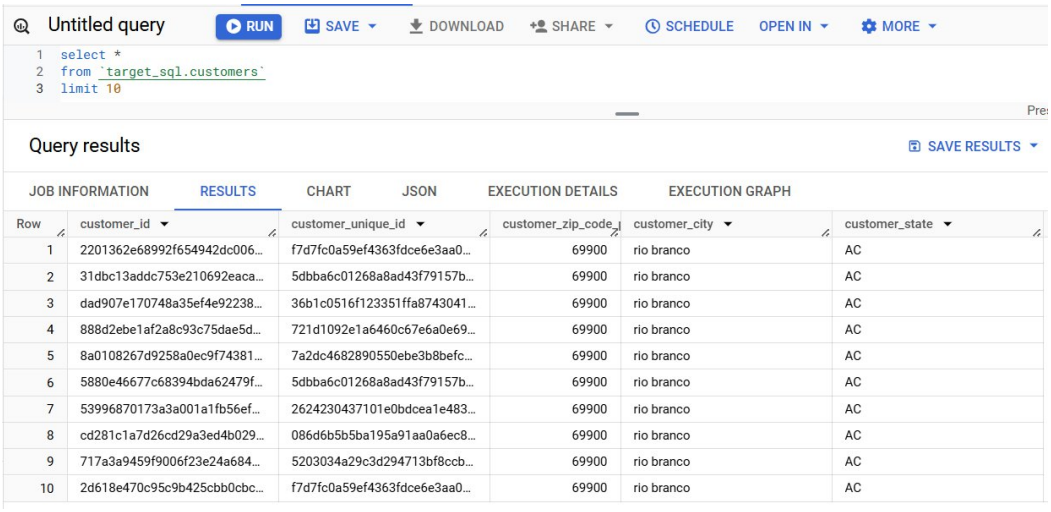
```
SELECT column_name, data_type
FROM `scaler-dsml-sql-450010.target_sql.INFORMATION_SCHEMA.COLUMNS`
WHERE table_name = 'customers';
```



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

Now, let's take a quick look of the first 10 rows of data from the '**customers'** table:

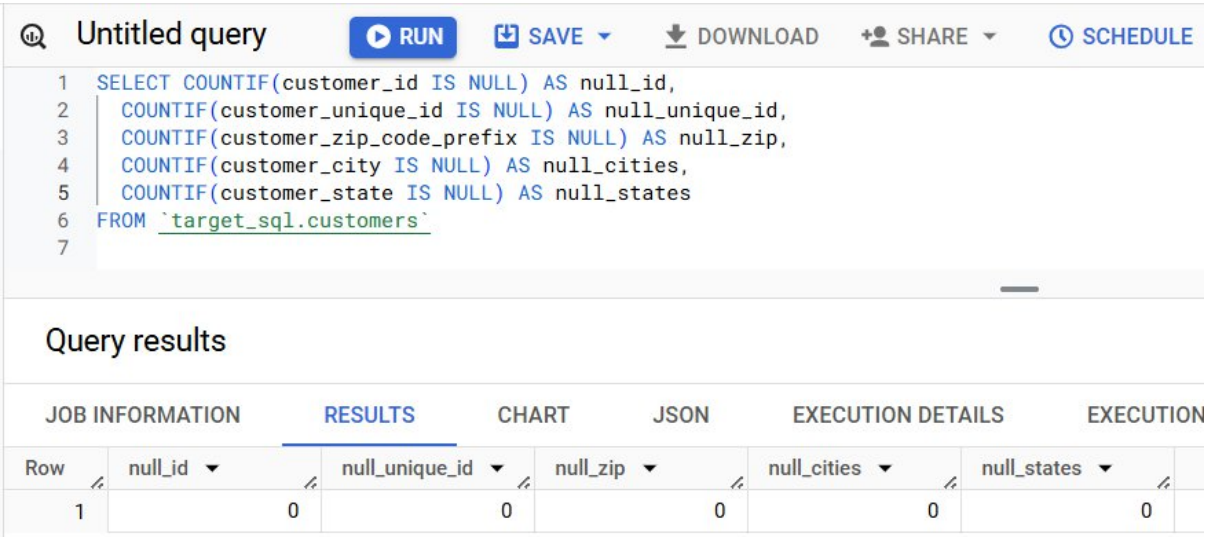
```
select *
from `target_sql.customers`
limit 10
```



Row	customer_id	customer_unique_id	customer_zip_code_p	customer_city	customer_state
1	2201362e68992f654942dc006...	f7d7fc0a59ef4363fdce6e3aa0...	69900	rio branco	AC
2	31dbc13addc753e210692eaca...	5dbba6c01268a8ad43f79157b...	69900	rio branco	AC
3	dad907e170748a35ef4e92238...	36b1c0516f123351ffa8743041...	69900	rio branco	AC
4	888d2ebe1af2a8c93c75dae5d...	721d1092e1a6460c67e6a0e69...	69900	rio branco	AC
5	8a0108267d9258a0ec9f74381...	7a2dc4682890550ebe3b8befc...	69900	rio branco	AC
6	5880e46677c68394bda62479f...	5dbba6c01268a8ad43f79157b...	69900	rio branco	AC
7	53996870173a3a001a1fb56ef...	2624230437101e0bdcea1e483...	69900	rio branco	AC
8	cd281c1a7d26cd29a3ed4b029...	086d6b5b5ba195a91aa0a6ec8...	69900	rio branco	AC
9	717a3a9459f9006f23e24a684...	5203034a29c3d294713bf8ccb...	69900	rio branco	AC
10	2d618e470c95c9b425cbb0cbc...	f7d7fc0a59ef4363fdce6e3aa0...	69900	rio branco	AC

This table captures information about the customers, such as their location (city and state), along with their unique identifiers.

Now, let's check for null values in 'customers' table through below sql query:



Here, we can see that no null values exists in the 'customers' table.

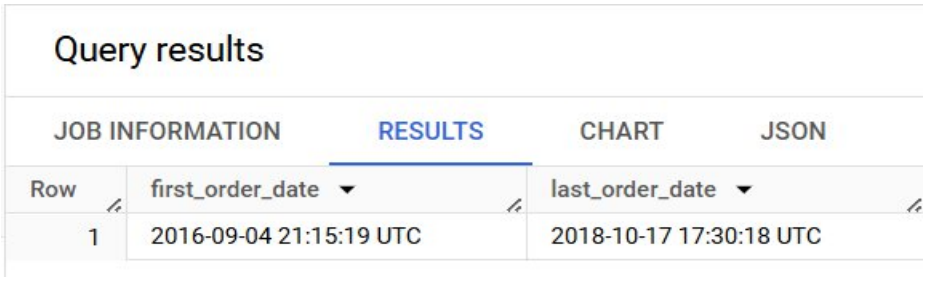
Similarly, I have checked the structure of all other tables '**geolocation**', '**order_items**', '**order_reviews**', '**orders**', '**payments**', '**products**', and '**sellers**'.

2. Let's find the range of dates between which the orders were placed.

Below is the query to find the earliest and latest order dates:

```
select min(order_purchase_timestamp) as first_order_date,
       max(order_purchase_timestamp) as last_order_date
from `target_sql.orders`
```

Here, is the result for above query:

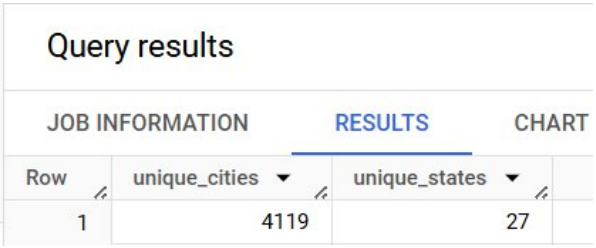


We can see that all orders are between 4th September 2016 to 17th October 2018.

3. Now, let's count the number of unique Cities & States of the customers who ordered during above time period. Below is the sql query.

```
select count(distinct customer_city) as unique_cities,
       count(distinct customer_state) as unique_states
from `target_sql.customers`
```

Here is the result:

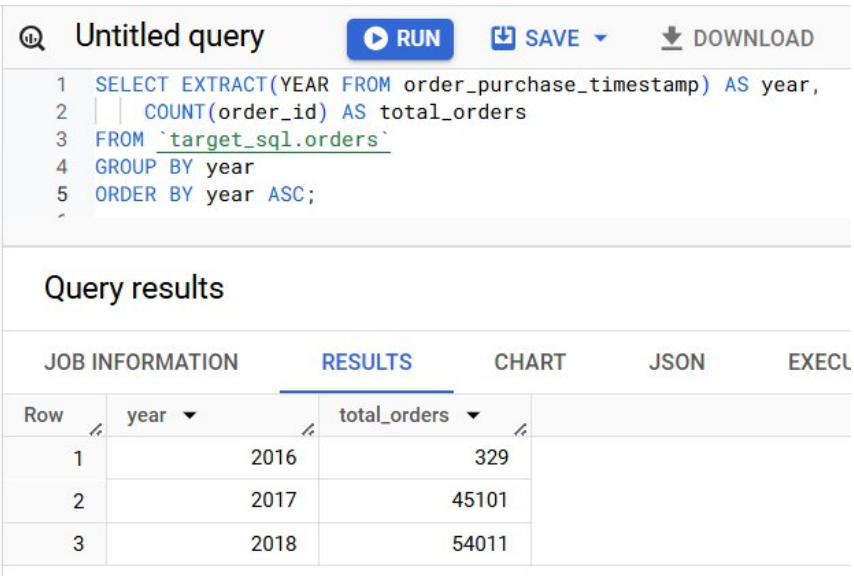


We can see that customers are from 4,119 unique cities and 27 unique states have placed orders.

2: In-depth Exploration:

Here are some in-depth explorations like growing yearly trends, monthly seasonality & time of day analysis.

1. **Growing Yearly Trends:** First, let's check the trend in number of orders placed over the past years. Below is the sql query and result.



Here's the summary for **Growing Yearly Trends** based on the result:

- **2016**: The dataset shows **329 orders** placed in 2016. This number is relatively low as data seems to start mid-year, around September, and may not represent the full year.
- **2017**: There is a significant surge in orders, reaching **45,101 total orders**. This indicates a strong adoption of Target's e-commerce services during this period.

- **2018**: Growth continues, with **54,011 total orders**, representing an increase compared to 2017. This suggests sustained momentum in Target's online operations in Brazil.

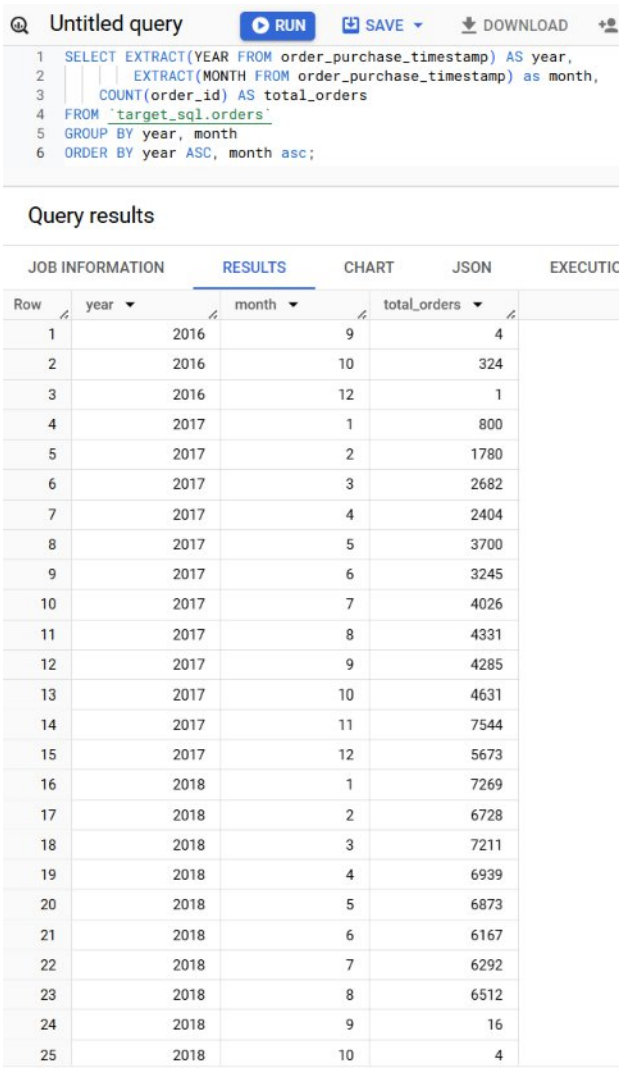
Observations:

- There is a **clear upward trend** in the number of orders year-over-year, signaling increasing customer engagement and adoption of online shopping.
- The dataset highlights Target's successful expansion or growing market presence in Brazil during 2016–2018.

Recommendations:

- **Marketing Strategies**: Capitalize on this growth by identifying high-demand periods and running targeted campaigns to sustain and further boost order volumes.
- **Infrastructure Scaling**: Ensure logistics and customer service teams are equipped to handle this growth in demand efficiently.

2. **Monthly Seasonality**:
   Now, let's check the monthly seasonality of the number of orders placed. Below is the query and result for the total monthly orders.



Untitled query RUN SAVE DOWNLOAD

```
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_sql.orders`
GROUP BY year, month
ORDER BY year ASC, month asc;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTIC

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
11	2017	8	4331
12	2017	9	4285
13	2017	10	4631
14	2017	11	7544
15	2017	12	5673
16	2018	1	7269
17	2018	2	6728
18	2018	3	7211
19	2018	4	6939
20	2018	5	6873
21	2018	6	6167
22	2018	7	6292
23	2018	8	6512
24	2018	9	16
25	2018	10	4

Here's a summary of the **Monthly Seasonality** analysis based on the above results:

2016 Orders:

Data is limited, starting from **September 2016**, with only a small number of orders recorded:

September: 4 orders.

October: 324 orders (a significant spike).

November and December: Just 1 order each, possibly due to incomplete data.

2017 Orders:

Orders grew **steadily throughout the year**, showing an upward trend overall:

Starting at 800 orders in January.

A noticeable peak in **November 2017** with **7544 orders**, potentially indicating a seasonal event like **Black Friday**.

December saw orders dip slightly to 6348.

2018 Orders:

Order volumes remained high, especially early in the year:

January 2018 recorded 7269 orders (the highest monthly orders in the dataset).

Following months show a gradual decline in orders, with 4 orders recorded in **October 2018** (likely due to incomplete or trailing data).

3. Time of Day Analysis:

Here are the count of orders placed in each time of day by Brazilian customers.

```
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 'Mornings'
    WHEN EXTRACT(HOUR FROM order_purchase_timestamp) BETWEEN 13 AND 18 THEN 'Afternoon'
    ELSE 'Night'
  END as time_of_day,
  COUNT(order_id) as total_orders
FROM `target_sql.orders`
GROUP BY time_of_day
ORDER BY total_orders desc
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTIO

w	time_of_day	total_orders
1	Afternoon	38135
2	Night	28331
3	Mornings	27733
4	Dawn	5242

Here’s the summary for **Time-of-Day Analysis** based on the result observed:

Key Insights:

1. **Afternoon Dominates**:
    - **38,135 orders** were placed during the **Afternoon** (13:00–18:00), making it the busiest time of day for e-commerce activity in Brazil.
2. **Night Comes Next**:
    - **28,331 orders** were placed at **Night** (19:00–23:00). This shows that many customers prefer shopping after traditional work hours.
3. **Morning Activity**:
    - **27,733 orders** were placed during the **Morning** (07:00–12:00), indicating a steady flow of orders during this period.
4. **Least Activity at Dawn**:
    - **5,242 orders** were placed during the **Dawn** hours (00:00–06:00), likely reflecting reduced customer engagement during late-night and early-morning hours.

Observations:

- **Afternoon and Night Together**:
    - These two periods account for the majority of the orders, with customers primarily active in the post-lunch and evening hours.
- **Low Engagement at Dawn**:
    - Minimal order activity during the early morning hours suggests a natural drop-off due to customers' sleeping patterns.

Recommendations:

1. **Targeted Marketing**:
    - Schedule promotions, discounts, or ads during **Afternoon** and **Night**, as these are the peak hours for customer activity.
2. **Operational Efficiency**:
    - Align logistics, customer support, and inventory planning to handle higher demand during these periods.
3. **Boost Dawn Engagement**:
    - Consider offering special “early bird” discounts or flash sales to encourage shopping during the quieter Dawn hours.

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

Now, let's analyse the region-wise e-commerce growth in Brazil.

We will first check month-on-month orders per state.

1. Month-on-Month Orders per State:

To get the number of orders per state we will use below sql query. The results are displayed chronologically (by year and month) and alphabetically (by state).

```
SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) as Year,
      EXTRACT(MONTH FROM o.order_purchase_timestamp) as Month,
      c.customer_state, COUNT(o.order_id) as total_orders
FROM `target_sql.orders` o join `target_sql.customers` c
on o.customer_id = c.customer_id
GROUP BY Year, Month, customer_state
ORDER BY Year asc, Month asc, c.customer_state asc
```

Attached sample output for reference.

Year	Month	customer_state	total_orders
2016	9	RR	1
2016	9	RS	1
2016	9	SP	2
2016	10	AL	2
2016	10	BA	4
2016	10	CE	8
2016	10	DF	6
2016	10	ES	4
2016	10	GO	9
2016	10	MA	4
2016	10	MG	40
2016	10	MT	3
2016	10	PA	4
2016	10	PB	1
2016	10	PE	7
2016	10	PI	1
2016	10	PR	19
2016	10	RJ	56
2016	10	RN	4
2016	10	RR	1

2. Customer Distribution Across States:

To find the customers distribution across all the states we can use below sql query. It will calculate unique customers per state.

```
SELECT customer_state,
      COUNT(DISTINCT customer_unique_id) as unique_customers
FROM `target_sql.customers`
GROUP BY customer_state
ORDER BY unique_customers desc
```

Attaching the result of above query, here is the number of unique customers across all state.

Query results

JOB INFORMATION | RESULTS | CHART | JS

Row	customer_state	unique_customers
1	SP	40302
2	RJ	12384
3	MG	11259
4	RS	5277
5	PR	4882
6	SC	3534
7	BA	3277
8	DF	2075
9	ES	1964
10	GO	1952
11	PE	1609
12	CE	1313
13	PA	949
14	MT	876
15	MA	726
16	MS	694
17	PB	519
18	PI	482
19	RN	474
20	AL	401
21	SE	342
22	TO	273
23	RO	240
24	AM	143
25	AC	77
26	AP	67
27	RR	45

Key Insights:

1. **State with the Highest Customers**:
    - The state of **São Paulo (SP)** stands out significantly, with **40,302 unique customers**, representing the largest share of the customer base.
2. **Other High-Performing States**:
    - **Rio de Janeiro (RJ):** 12,384 unique customers.

- **Minas Gerais (MG):** 11,259 unique customers.
- These states, along with São Paulo, dominate the customer distribution, indicating high engagement in southeastern Brazil.

3. **Mid-Tier States**:
    - States like **Rio Grande do Sul (RS):** 5,277 customers, **Paraná (PR):** 4,882 customers, and **Santa Catarina (SC):** 3,534 customers show moderate customer engagement.
4. **Low Engagement States**:
    - Some states have minimal customer representation, such as **Amapá (AP):** 67 customers and **Roraima (RR):** 45 customers, indicating low e-commerce penetration in northern Brazil.

Observations:

- There is a clear concentration of e-commerce activity in the southeastern states (SP, RJ, MG), which are likely urban hubs with better digital infrastructure and larger populations.
- Northern and northeastern states show much lower participation, suggesting either logistical challenges or fewer customer touchpoints in these regions.

Recommendations:

1. **Expand Marketing Efforts**:
    - Increase targeted marketing and awareness campaigns in underperforming states like **AP** and **RR** to boost customer engagement.
2. **Logistics Optimization**:
    - Assess logistical challenges in low-engagement states to improve delivery services and attract more customers.
3. **Leverage Strength in High-Performing States**:
    - Consolidate customer loyalty in high-performing states (SP, RJ, MG) through promotions, better delivery options, and enhanced customer service.

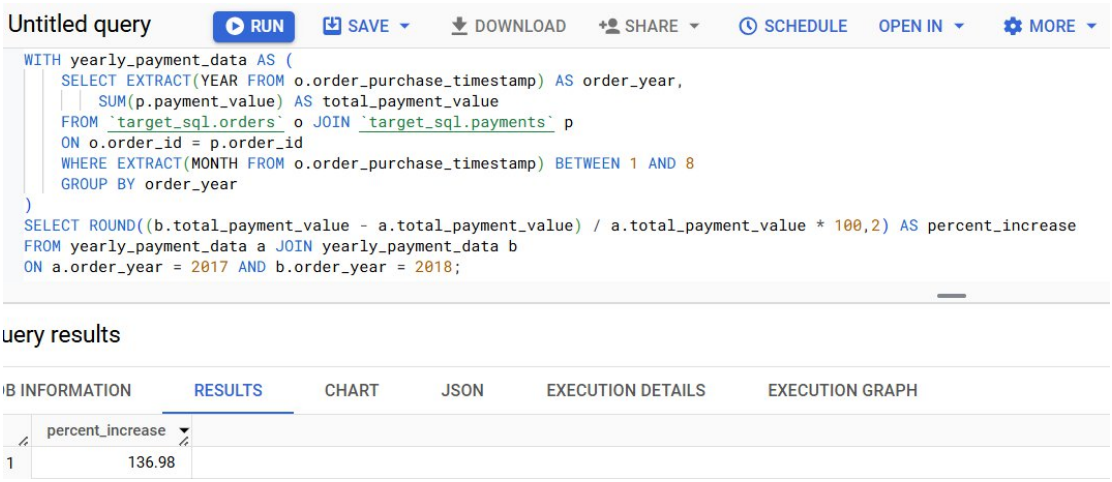
4: Impact on Economy: Analyze the money movement by e-commerce by looking at order prices, freight and others.

Now, let's find the economic impact of the business.

1. **Calculate the year-on-year percentage increase in the total payment value (payment_value) from 2017 to 2018 (Jan–Aug).**

    To calculate the percentage increase in total payment value from January to August in 2017 to the same period in 2018 below query is used.

```
WITH yearly_payment_data AS (
   SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) AS order_year,
      SUM(p.payment_value) AS total_payment_value
   FROM `target_sql.orders` o JOIN `target_sql.payments` p
   ON o.order_id = p.order_id
   WHERE EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
   GROUP BY order_year
)
SELECT ROUND((b.total_payment_value - a.total_payment_value) /
a.total_payment_value * 100,2) AS percent_increase
FROM yearly_payment_data a JOIN yearly_payment_data b
ON a.order_year = 2017 AND b.order_year = 2018;
```



From January to August, the **total payment value** for e-commerce orders saw a significant growth between **2017** and **2018 (Jan–Aug)**. The percentage increase in the total payment value during this period is **136.98%**. This reflects a remarkable surge in economic activity and suggests:

- A growing customer base.
- Increased spending by customers.
- Potential enhancements in Target's e-commerce strategy in Brazil.

2. **Calculate the Total & Average value of order price for each state.**
   To calculate the total and average payment values for each state based on orders, below query is used.

```
SELECT c.customer_state,
    ROUND(SUM(p.payment_value),2) AS total_order_price,
    ROUND(AVG(p.payment_value),2) AS avg_order_price
FROM `target_sql.customers` c JOIN `target_sql.orders` o
ON c.customer_id = o.customer_id
JOIN `target_sql.payments` p ON o.order_id = p.order_id
GROUP BY c.customer_state
ORDER BY total_order_price DESC;
```

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTI

Row	customer_state	total_order_price	avg_order_price
1	SP	5998226.96	137.5
2	RJ	2144379.69	158.53
3	MG	1872257.26	154.71
4	RS	890898.54	157.18
5	PR	811156.38	154.15
6	SC	623086.43	165.98
7	BA	616645.82	170.82
8	DF	355141.08	161.13
9	GO	350092.31	165.76
10	ES	325967.55	154.71
11	PE	324850.44	187.99
12	CE	279464.03	199.9
13	PA	218295.85	215.92
14	MT	187029.29	195.23
15	MA	152523.02	198.86
16	PB	141545.72	248.33
17	MS	137534.84	186.87
18	PI	108523.97	207.11
19	RN	102718.13	196.78
20	AL	96962.06	227.08
21	SE	75246.25	208.44
22	TO	61485.33	204.27
23	RO	60866.2	233.2
24	AM	27966.93	181.6
25	AC	19680.62	234.29
26	AP	16262.8	232.33
27	RR	10064.62	218.8

Key Insights:

Order Prices:

1. **High-Performing States:**

- **São Paulo (SP)** leads with a total order price of **₹5,998,226.96**, reflecting its role as the hub of e-commerce activity in Brazil. However, its average order price of **₹137.50** is relatively low due to high transaction volume with smaller orders.
- **Rio de Janeiro (RJ)** and **Minas Gerais (MG)** follow with total order prices of **₹2,144,379.69** and **₹1,872,257.26**, respectively, with average order prices slightly higher at **₹158.53 (RJ)** and **₹154.10 (MG)**.

2. Low-Performing States:

- **States like Roraima (RR)** and **Amapá (AP)** have much lower total order prices at **₹10,064.62** and **₹19,680.62**, respectively. However, their average order prices are high, indicating less frequent but higher-value transactions.

3. **Calculate the Total & Average value of order freight for each state.**
   To calculate total and average value of order freight for each state below query is used.

```
SELECT c.customer_state,
    ROUND(SUM(oi.freight_value),2) AS total_freight_price,
    ROUND(AVG(oi.freight_value),2) AS avg_freight_price
FROM `target_sql.customers` c JOIN `target_sql.orders` o
ON c.customer_id = o.customer_id
JOIN `target_sql.order_items` oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY total_freight_price DESC;
```

Untitled query RUN SAVE DOWNLOAD

```
SELECT c.customer_state,
      ROUND(SUM(oi.freight_value),2) AS total_freight_price,
      ROUND(AVG(oi.freight_value),2) AS avg_freight_price
FROM `target_sql.customers` c JOIN `target_sql.orders` o
ON c.customer_id = o.customer_id
JOIN `target_sql.order_items` oi
ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY total_freight_price DESC;
```

Processing location: US

Query results

JOB INFORMATION | RESULTS | CHART | JSON | EXECUTIO

Row	customer_state	total_freight_price	avg_freight_price
1	SP	718723.07	15.15
2	RJ	305589.31	20.96
3	MG	270853.46	20.63
4	RS	135522.74	21.74
5	PR	117851.68	20.53
6	BA	100156.68	26.36
7	SC	89660.26	21.47
8	PE	59449.66	32.92
9	GO	53114.98	22.77
10	DF	50625.5	21.04
11	ES	49764.6	22.06
12	CE	48351.59	32.71
13	PA	38699.3	35.83
14	MA	31523.77	38.26
15	MT	29715.43	28.17
16	PB	25719.73	42.72
17	PI	21218.2	39.15
18	MS	19144.03	23.37
19	RN	18860.1	35.65
20	AL	15914.59	35.84
21	SE	14111.47	36.65
22	TO	11732.68	37.25
23	RO	11417.38	41.07
24	AM	5478.89	33.21
25	AC	3686.75	40.07
26	AP	2788.5	34.01
27	RR	2235.19	42.98

Key Insights:

Freight Costs:

1. **Total Freight Costs:**

- **São Paulo (SP**) also leads in total freight costs at **₹718,732.07**, consistent with its high transaction volume.
- **Rio de Janeiro (RJ)** and **Minas Gerais (MG)** follow with ₹305,893.31 and **₹270,858.46**, respectively.

2. **Average Freight Costs:**

- States like **Acre (AC)** and **Roraima (RR)** have some of the **highest average freight costs** at **₹40.07 (AC)** and **₹42.98 (RR)**, compared to **São Paulo's** much lower **₹15.15**. This likely reflects the logistical challenges of delivering to remote regions.

3. **Regional Freight Dynamics:**

- Northern and smaller states tend to have much higher average freight costs, highlighting challenges in last-mile delivery and transportation.

Overall Observations:

- **Southeastern Brazil (SP, RJ, MG)** dominates both in terms of total order and freight values, demonstrating its status as the economic powerhouse of Brazil.
- **Smaller northern states** like **AC**, **RR**, and **AP** have fewer total transactions but higher average costs for both orders and freight, indicating a need for infrastructure improvements and e-commerce expansion.

Recommendations:

1. **Target Remote Areas**:
    - Optimize delivery networks and reduce freight costs in northern states to encourage more frequent purchases and improve accessibility.
2. **Leverage High-Volume States**:
    - Continue strengthening operations in high-performing states like SP, RJ, and MG to maintain their dominance.
3. **Tailored Promotions**:
    - Offer promotions or discounts on freight charges in high-cost regions to drive order growth.

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.**

```
SELECT order_id,
        DATE_DIFF(order_delivered_customer_date,order_purchase_timestamp, DAY) AS
        time_to_deliver ,
        DATE_DIFF(order_delivered_customer_date,order_estimated_delivery_date,DAY) AS
        diff_estimated_delivery
FROM `target_sql.orders`
ORDER BY order_id
```

Here are the sample output of the above query with ‘time_to_deliver’ and ‘diff_estimated_delivery’ values in Days for each order.

Row	order_id	time_to_deliver	diff_estimated_delivery
1	00010242fe8c5a6d1ba2dd792...	7	-8
2	00018f77f2f0320c557190d7a1...	16	-2
3	000229ec398224ef6ca0657da...	7	-13
4	00024acbcdf0a6daa1e931b03...	6	-5
5	00042b26cf59d7ce69dfabb4e...	25	-15
6	00048cc3ae777c65dbb7d2a06...	6	-14
7	00054e8431b9d7675808bcb8...	8	-16
8	000576fe39319847cbb9d288c...	5	-15
9	0005a1a1728c9d785b8e2b08...	9	0
10	0005f50442cb953dcd1d21e1f...	2	-18
11	00061f2a7bc09da83e415a52d...	4	-10
12	00063b381e2406b52ad42947...	10	0

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

We can find highest & lowest average freight value using UNION ALL.

```
SELECT customer_state, avg_freight_value
FROM (
    (SELECT c.customer_state, ROUND(AVG(oi.freight_value),2) AS avg_freight_value
FROM `target_sql.customers` c join `target_sql.orders` o
ON c.customer_id = o.customer_id
JOIN `target_sql.order_items` oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY avg_freight_value DESC
LIMIT 5)

UNION ALL

(SELECT c.customer_state, ROUND(AVG(oi.freight_value),2) AS avg_freight_value
FROM `target_sql.customers` c join `target_sql.orders` o
ON c.customer_id = o.customer_id
JOIN `target_sql.order_items` oi ON o.order_id = oi.order_id
GROUP BY c.customer_state
ORDER BY avg_freight_value ASC
LIMIT 5)) AS Combined_result
```

Below is the result for above query which have highest and lowest average freight value. Highlighted states are lowest average freight value.

Row	customer_state	avg_freight_value
1	RR	42.98
2	PB	42.72
3	RO	41.07
4	AC	40.07
5	PI	39.15
6	SP	15.15
7	PR	20.53
8	MG	20.63
9	RJ	20.96
10	DF	21.04

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

We can use same approach as above to calculate the average delivery time for top 5 states.

```
SELECT customer_state, avg_delivery_time
FROM (
    (SELECT c.customer_state,
    ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) AS avg_delivery_time
FROM `target_sql.customers` c join `target_sql.orders` o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_delivery_time DESC
LIMIT 5)

UNION ALL

(SELECT c.customer_state,
    ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_purchase_timestamp, DAY)),2) AS avg_delivery_time
FROM `target_sql.customers` c join `target_sql.orders` o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY avg_delivery_time ASC
LIMIT 5)) AS Combined_result
```

Here is the result for above query, average delivery time is calculated in Days. Lowest average time states are highlighted.

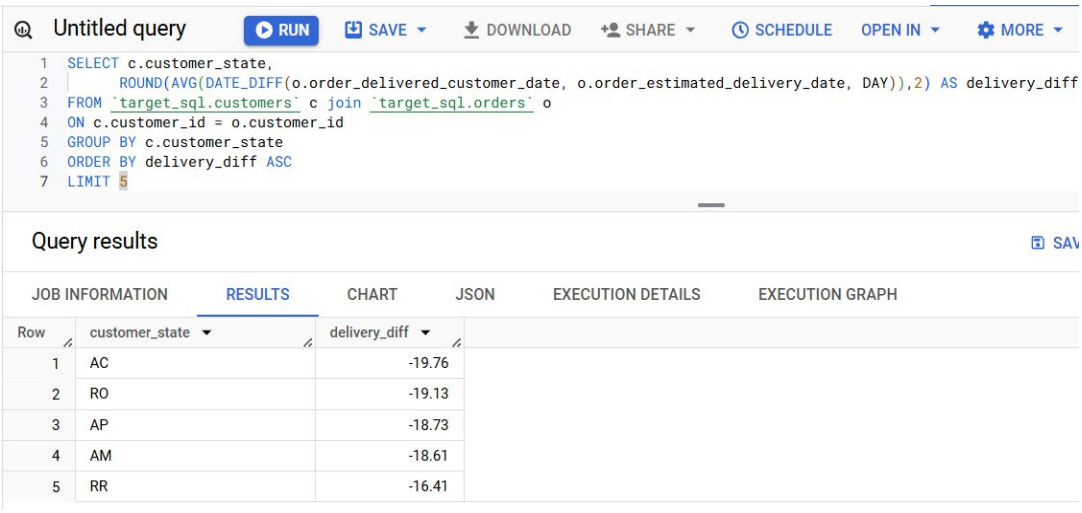
Row	customer_state	avg_delivery_time
1	RR	28.98
2	AP	26.73
3	AM	25.99
4	AL	24.04
5	PA	23.32
6	SP	8.3
7	PR	11.53
8	MG	11.54
9	DF	12.51
10	SC	14.48

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.**

   Here is query to find the top 5 states having really fast deliveries.

```
SELECT c.customer_state,
    ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date,
o.order_estimated_delivery_date, DAY)),2) AS delivery_diff
FROM `target_sql.customers` c join `target_sql.orders` o
ON c.customer_id = o.customer_id
GROUP BY c.customer_state
ORDER BY delivery_diff ASC
LIMIT 5
```

   The top 5 states are below:



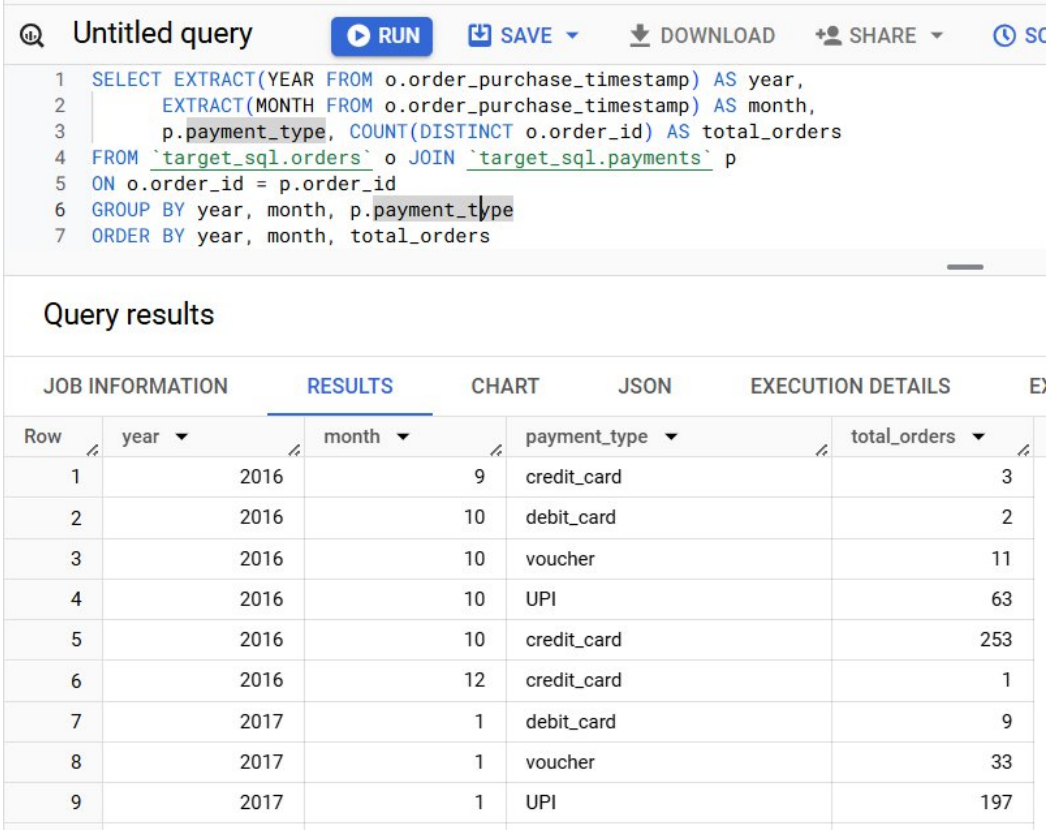
6: Analysis based on the payments:

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

Here is the query to calculate the number of orders placed each month using different payment types.

```
SELECT EXTRACT(YEAR FROM o.order_purchase_timestamp) AS year,
    EXTRACT(MONTH FROM o.order_purchase_timestamp) AS month,
    p.payment_type, COUNT(DISTINCT o.order_id) AS total_orders
FROM `target_sql.orders` o JOIN `target_sql.payments` p
ON o.order_id = p.order_id
GROUP BY year, month, p.payment_type
ORDER BY year, month, total_orders
```

Below is the sample result displayed using above query.



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

Key Insights:

1. **Dominance of Credit Card Payments:**
   - Credit cards are consistently the most preferred payment type across all months, with the highest number of orders (e.g., 253 orders in October 2016).

2. **UPI Usage Rises:**
    - UPI payments show notable adoption, especially in January 2017, with 197 orders, indicating growing customer preference for digital payments.
3. **Minimal Use of Debit Cards and Vouchers:**
    - Debit card and voucher payments have significantly lower transaction counts. For example, only 2 orders with a debit card in October 2016.
4. **Seasonality Impact:**
    - October and January exhibit higher order activity across all payment types, potentially reflecting seasonal sales or holiday periods.

Observations:

- Credit card remains the most widely used payment type, while UPI emerges as a strong alternative.
- Payment preferences align with broader digital adoption trends in Brazil during the period.

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

```
SELECT payment_installments,
    COUNT(DISTINCT order_id) AS total_orders
FROM `target_sql.payments`
GROUP BY payment_installments
```

Here are the result for above query, the number of orders grouped by payment installments.

Row	payment_installments	total_orders
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
11	10	5315
12	11	23
13	12	133
14	13	16
15	14	15
16	15	74
17	16	5
18	17	8
19	18	27
20	20	17
21	21	3
22	22	1
23	23	1
24	24	18

7: Actionable Insights & Recommendations

1. Growing Yearly Trends

Insight:

- Significant growth in e-commerce orders from 2016 to 2018, with a **136.98% increase in payment value** from January to August (2017 vs. 2018).
- A clear upward trend in customer engagement and online shopping adoption.

Recommendation:

- **Expand Marketing Efforts**: Invest in targeted advertising during peak months (e.g., November and January) to capitalize on seasonal demand.
- **Infrastructure Scaling**: Strengthen logistics and customer support to accommodate sustained growth and maintain customer satisfaction.

2. Monthly Seasonality & Time of Day

Insight:

- **Afternoon (13:00–18:00)** and **Night (19:00–23:00)** dominate customer activity, contributing to the majority of orders.
- Peak months are **November and January**, likely influenced by Black Friday and New Year sales.
- Minimal activity during **Dawn (00:00–06:00)** hours.

Recommendation:

- **Target Promotions During High Activity Periods**: Schedule discounts and flash sales during Afternoon and Night to maximize engagement.
- **Boost Dawn Engagement**: Launch early morning "Early Bird Discounts" to increase order volume during quieter periods.
- **Seasonal Campaigns**: Prepare for high demand in November and January by optimizing inventory, delivery, and marketing strategies.

3. Regional Performance

Insight:

- **São Paulo (SP), Rio de Janeiro (RJ), and Minas Gerais (MG)** dominate, with high customer distribution, order volumes, and payment values.
- Northern states like **Roraima (RR)** and **Amapá (AP)** have low customer engagement but high average order values and freight costs, reflecting logistical challenges.

Recommendation:

- **Strengthen Operations in High-Performing States**: Consolidate presence in SP, RJ, and MG with improved customer experience initiatives.

- **Expand to Underserved Areas**: Improve infrastructure and reduce freight costs in northern regions like RR and AP to encourage e-commerce adoption.
- **Localized Marketing**: Tailor campaigns to urban hubs in southeastern states while increasing awareness in less engaged northern regions.

4. Payment Trends

Insight:

- **Credit Card** remains the most preferred payment method.
- Growing adoption of digital methods like UPI (Unified Payments Interface).
- Higher payment installments are less frequent, indicating a preference for fewer installments or full payments.

Recommendation:

- **Incentivize Digital Payments**: Encourage UPI and digital methods through cashback offers and discounts.
- **Flexibility in Payment Plans**: Offer installment-friendly plans to attract high-value customers who may prefer spreading out payments.

5. Freight and Delivery

Insight:

- States like **São Paulo (SP)** have the lowest average freight costs, but northern states such as **Acre (AC)** and **Roraima (RR)** have the highest due to logistical challenges.
- Delivery times vary widely, with certain states performing better than others. The fastest states are outperforming delivery expectations significantly.

Recommendation:

- **Optimize Logistics**: Invest in better distribution networks for high-cost regions (e.g., AC and RR) to lower freight expenses and improve last-mile delivery.
- **Leverage Fast-Delivery States**: Analyze and replicate practices from top-performing states to improve delivery efficiency across the board.
- **Customer-Centric Policies**: Offer free or discounted shipping for high-value orders to boost satisfaction and loyalty in high-cost regions.

6. Insights on Delivery Performance

Insight:

- Some states deliver much faster than estimated delivery dates, showcasing high operational efficiency. However, others face delays, particularly in remote regions.

Recommendation:

- **Benchmark Best Practices**: Identify and implement strategies from states with the most efficient deliveries.

- **Proactively Communicate Delays**: Keep customers informed about delays and offer compensations like coupons or discounts to maintain trust.