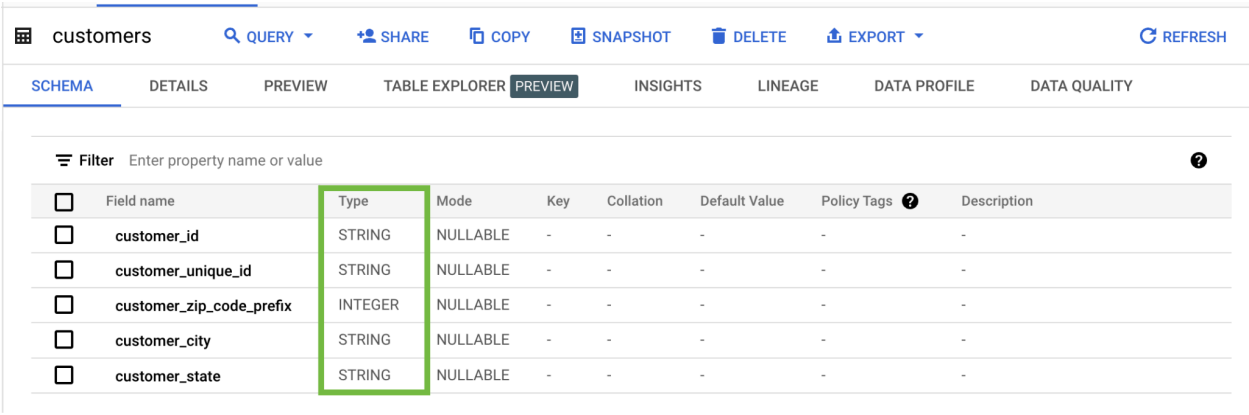
Business Case: Target SQL

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

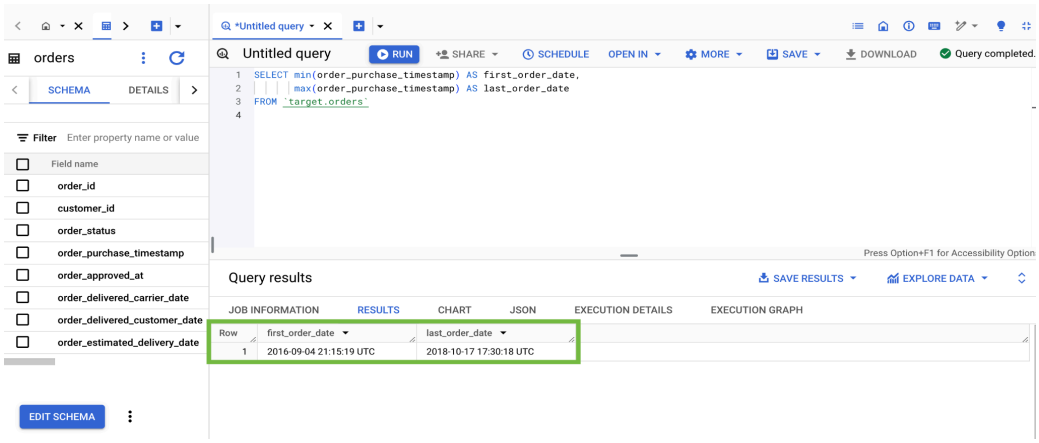
1.1 - Data type of all columns in the "customers" table.



Insight

- In the `customers_table`, all columns have a `STRING` data type except for `customer_zip_code_prefix`, which is of type `INT`.

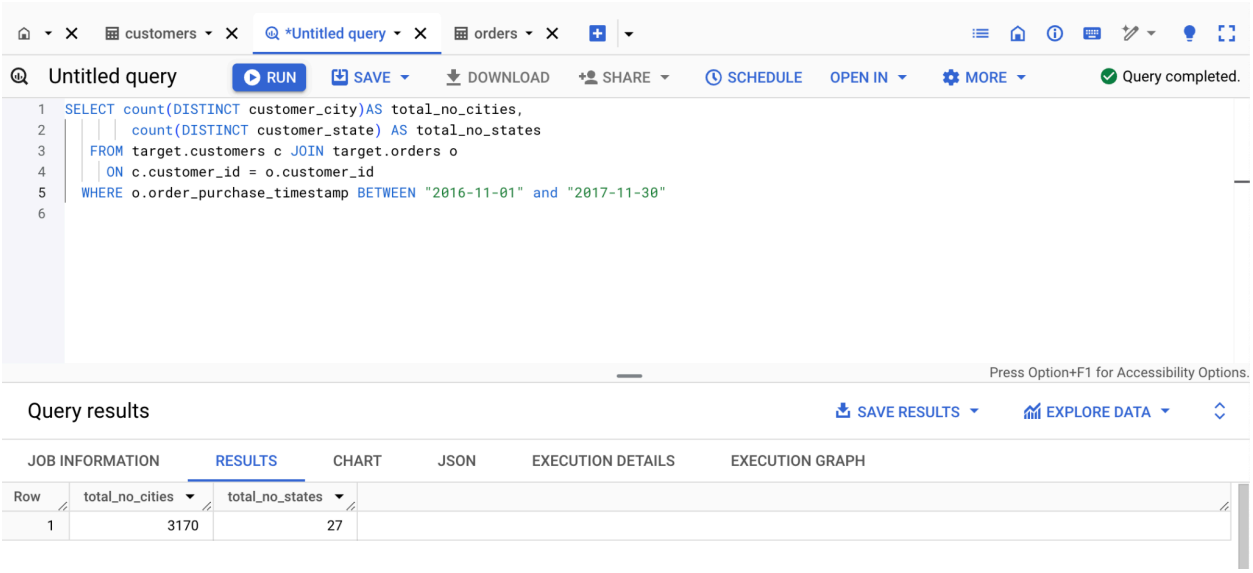
1.2 - Get the time range between which the orders were placed.



Insight

- The dataset spans approximately two years, from September 2016 to October 2018, providing a comprehensive view of this period.

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



Insight

- Between November 2016 and November 2017, customers placed orders from 3,170 unique cities across 27 states. This highlights the wide geographic reach of the customer base during this period.

2. In-depth Exploration:

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

customers | *Untitled query | orders

Untitled query — RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE — Query completed.

```
1  SELECT
2      EXTRACT(YEAR FROM order_purchase_timestamp) AS order_year,
3      COUNT(DISTINCT order_id) AS total_orders
4  FROM
5      target.orders
6  GROUP BY
7      order_year
8  ORDER BY
9      order_year;
10
11
```

Press Option+F1 for Accessibility Options.

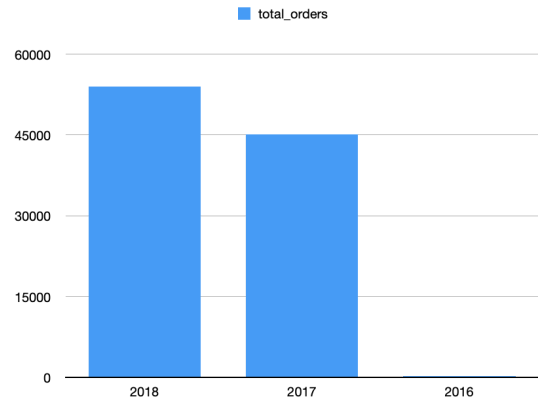
Query results — SAVE RESULTS | EXPLORE DATA

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

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

Insight

- There is a growing trend in the number of orders placed, with 2018 surpassing 2017 in total orders, indicating an increase in customer activity over the past year.



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

customers | *Untitled query | orders

Untitled query | RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE | This query will process 3.98 MB when run.

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

Press Option+F1 for Accessibility Options.

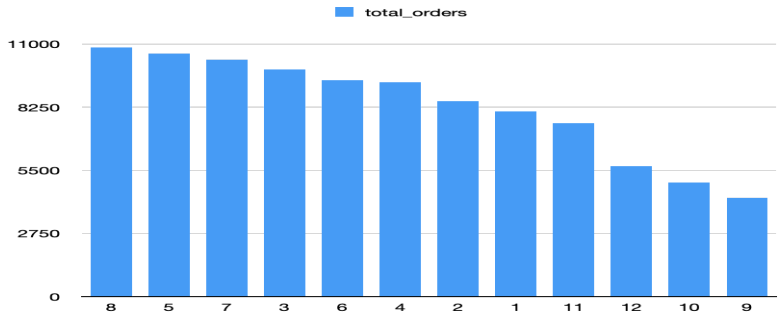
Query results | SAVE RESULTS | EXPLORE DATA

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

Row	order_year	order_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

Insight 1:

- The highest number of orders were placed in August, followed by May and July, when combining the sales from 2017 and 2018. These months indicate peak sales periods, suggesting that marketing or seasonal factors during these times could be driving higher order volumes.



Insight 2:

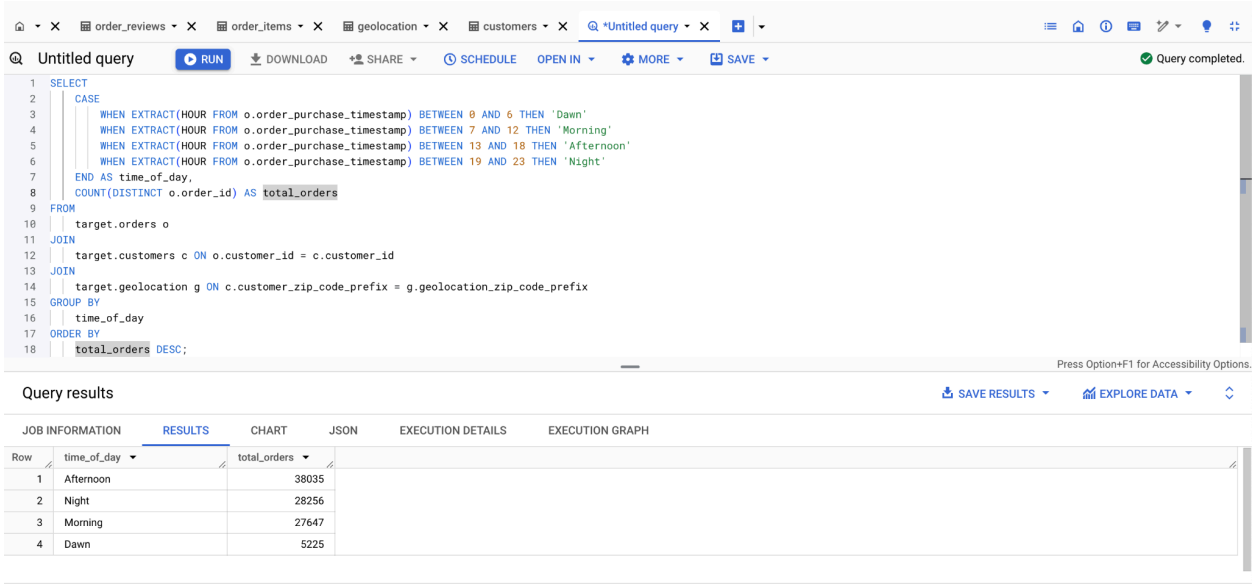
- In 2017, November had the highest number of orders with 7,544 orders.
- In 2018, January recorded the highest number of orders with 7,269 orders.

Insight 3:

- The number of orders showed a consistent increase month-over-month, peaking in November 2017.

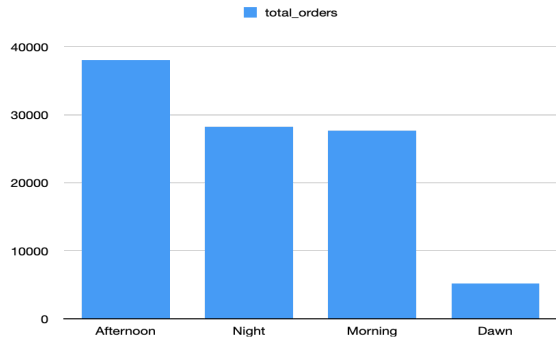
- After that, there was a slight decline in December 2017, followed by a small jump in January 2018.
- Post-January 2018, the number of orders remained constant, maintaining a minimum of 6,000 orders per month throughout the year.

2.3 - During what time of the day, do the Brazilian customers mostly place their orders? (Dawn, Morning, Afternoon or Night)



Insight

- **Afternoon**: The majority of orders are placed during the afternoon, indicating this is the peak time for Brazilian customers.
- **Night**: Orders placed at night come in second, showing a significant level of activity during this time.
- **Morning**: The number of orders placed in the morning is almost the same as at night, showing a moderate level of activity.
- **Dawn**: The least number of orders are placed during dawn, indicating minimal customer activity at this time.



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

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

orders | *Untitled query

Untitled query | RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE | Query completed.

```
SELECT count(DISTINCT o.order_id) total_orders,
       EXTRACT(MONTH FROM o.order_purchase_timestamp ) as month,
       c.customer_state AS customer_state
FROM `target.orders` o JOIN `target.customers` c
    ON o.customer_id = c.customer_id
    JOIN `target.geolocation` g
    ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
GROUP BY customer_state,month
ORDER BY month,total_orders desc,customer_state
```

Press Option+F1 for Accessibility Options.

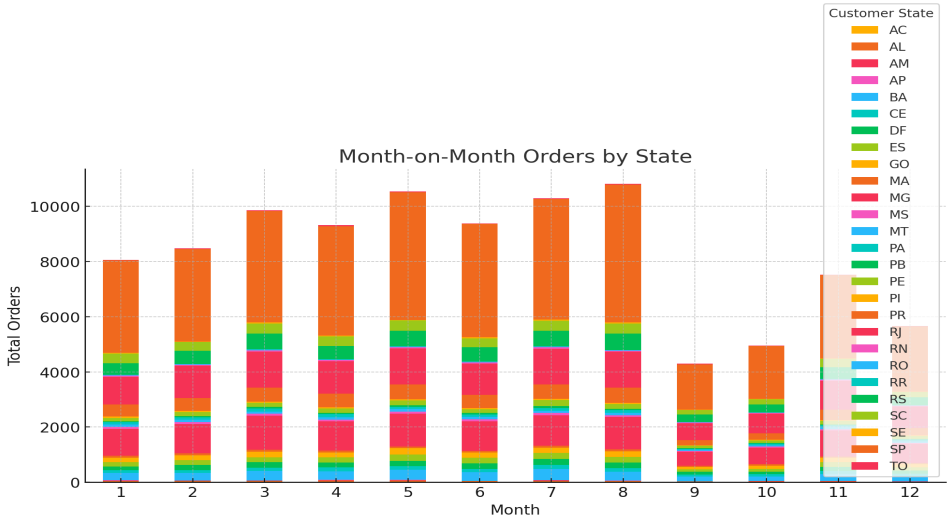
Query results | SAVE RESULTS | EXPLORE DATA

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

Row	total_orders	month	customer_state
1	3351	1	SP
2	990	1	RJ
3	971	1	MG
4	443	1	PR
5	427	1	RS
6	345	1	SC
7	264	1	BA
8	164	1	GO
9	157	1	ES
10	139	1	DF

Insight:

- **State "SP"** has the highest number of month-on-month orders across all months.
- **August** is the month with the highest number of orders, followed by **May** and **July**.



3.2 - How are the customers distributed across all the states?

customers | *Untitled query

Untitled query | RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE | Query completed.

```
1  SELECT customer_state,
2         COUNT(DISTINCT customer_id) AS no_of_customers,
3         (COUNT(*) * 100.0 / SUM(COUNT(*)) OVER ()) AS percentage
4  FROM `target.customers`
5  GROUP BY customer_state
6  ORDER BY no_of_customers DESC
```

Press Option+F1 for Accessibility Options.

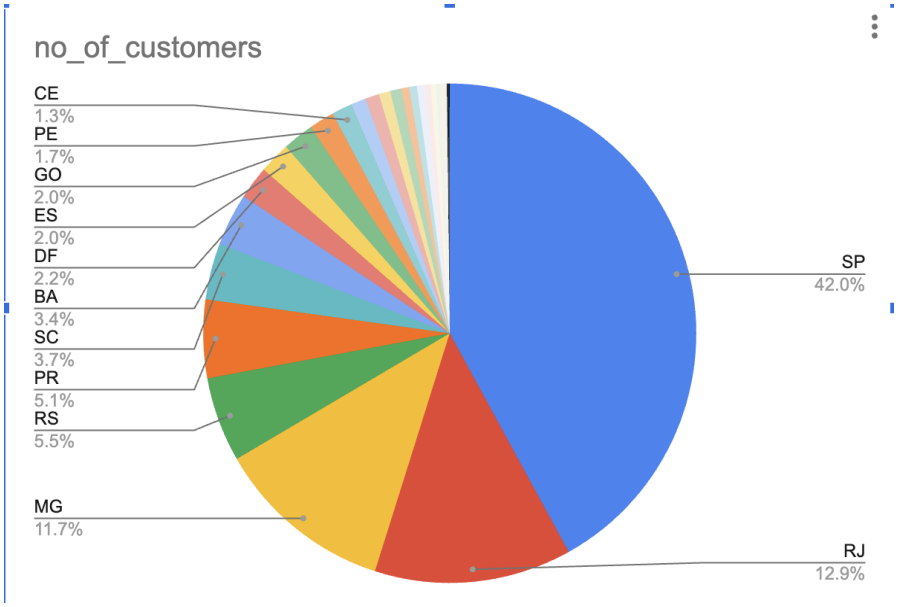
Query results | SAVE RESULTS | EXPLORE DATA

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

Row	customer_state	no_of_customers	percentage
1	SP	41746	41.98067195623...
2	RJ	12852	12.92424653814...
3	MG	11635	11.70040526543...
4	RS	5466	5.496726702265...
5	PR	5045	5.073360082863...
6	SC	3637	3.657445118210...
7	BA	3380	3.399000412304...
8	DF	2140	2.152029846843...
9	ES	2033	2.044428354501...
10	GO	2020	2.031355275992...

Insight

- São Paulo (SP) boasts the largest customer base across all of Brazil, highlighting its prominence as a key market for our business.
- followed by Rio de Janeiro (RJ) and Minas Gerais (MG), emphasizing these states as our most significant markets.



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

4.1 - Get the % increase in the cost of orders from 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.

Untitled query | RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE | This query will process 8.14 MB when run.

```sql
SELECT
    ROUND(SUM(CASE
        WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 THEN p.payment_value
        ELSE 0
    END),2) AS total_2017,

    ROUND(SUM(CASE
        WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 THEN p.payment_value
        ELSE 0
    END),2) AS total_2018,

    ROUND(((SUM(CASE
        WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2018 THEN p.payment_value
        ELSE 0
    END) -
    SUM(CASE
        WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 THEN p.payment_value
        ELSE 0
    END)) /
    SUM(CASE
        WHEN EXTRACT(YEAR FROM o.order_purchase_timestamp) = 2017 THEN p.payment_value
        ELSE 0
    END)) * 100, 2) AS percentage_increase

FROM `target.orders` o
JOIN `target.payments` p
    ON o.order_id = p.order_id
WHERE EXTRACT(MONTH FROM o.order_purchase_timestamp) BETWEEN 1 AND 8
  AND EXTRACT(YEAR FROM o.order_purchase_timestamp) IN (2017, 2018);
```

Press Option+F1 for Accessibility Options.

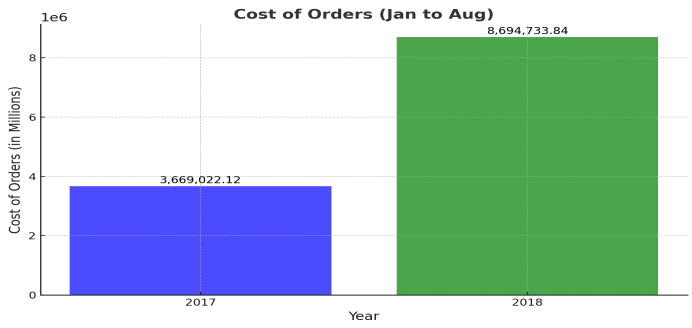
Query results | SAVE RESULTS | EXPLORE DATA

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

Row	total_2017	total_2018	percentage_increase
1	3669022.12	8694733.84	136.98

Insight

- **Significant Growth**: The cost of orders from January to August increased by 136.9% from 2017 to 2018, indicating either a rise in order volume or higher order values.



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

customers | *Untitled query | payments | geolocation | orders

Untitled query — RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE — Query completed.

```
SELECT g.geolocation_state,
       SUM(p.payment_value) AS total_value,
       AVG(p.payment_value) AS average_value
FROM `target.customers` c JOIN `target.geolocation` g
       ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
       JOIN `target.orders` o
       ON  c.customer_id = o.customer_id
       JOIN `target.payments` p
       ON o.order_id = p.order_id
GROUP BY g.geolocation_state
ORDER BY total_value DESC,average_value DESC
```

Press Option+F1 for Accessibility Options

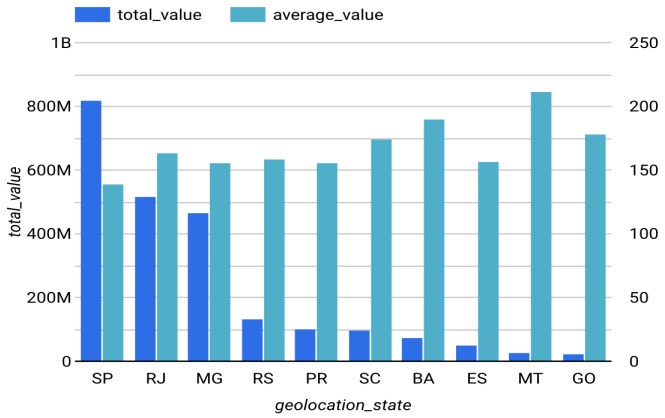
Query results — SAVE RESULTS | EXPLORE DATA

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

Row	geolocation_state	total_value	average_value
1	SP	819324973.1272...	139.1800016902...
2	RJ	516214891.0892...	163.3892353243...
3	MG	468418138.0098...	155.9394130293...
4	RS	131886278.0300...	158.3569008647...
5	PR	101367928.9499...	155.6463115200...
6	SC	96577779.98002...	174.3123905423...
7	BA	74143589.03002...	190.0976312787...
8	ES	51507204.52001...	157.0576229985...
9	MT	27075810.91000...	211.8707522262...
10	GO	24572389.12999...	178.1046716582...

Insight

- São Paulo (SP) leads by a large margin in terms of total order value, showing it's the biggest market, while Paraíba (PB) and Alagoas (AL) stand out for having the highest average order values, meaning customers in these states tend to make more expensive purchases.
- Minas Gerais (MG) and Rio de Janeiro (RJ) follow closely in total value, highlighting their strong customer base, but most states show similar average order values, indicating fairly consistent spending habits across the country.



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

customers | *Untitled query | order_items

Untitled query — RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE — Query completed.

```
SELECT g.geolocation_state,
       ROUND(SUM(oi.freight_value),2) AS total_freight_value,
       ROUND(AVG(oi.freight_value),2) AS average_freight_value
FROM `target.customers`c JOIN `target.geolocation` g
       ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
       JOIN `target.orders` o
       ON  c.customer_id = o.customer_id
       JOIN `target.order_items` oi
       ON o.order_id = oi.order_id
GROUP BY g.geolocation_state
ORDER BY total_freight_value DESC,average_freight_value DESC
```

Press Option+F1 for Accessibility Options

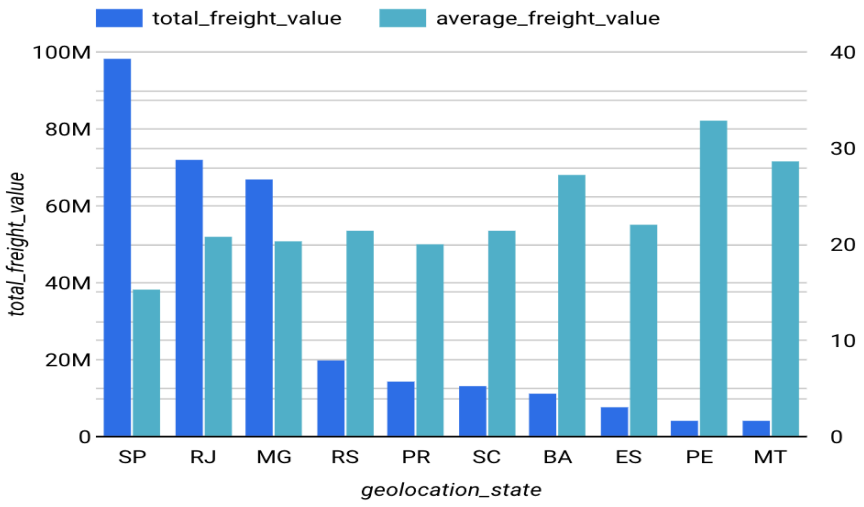
Query results — SAVE RESULTS | EXPLORE DATA

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

Row	geolocation_state	total_freight_value	average_freight_valu
1	SP	98574572.43	15.41
2	RJ	71966793.75	20.9
3	MG	67058347.09	20.46
4	RS	19910834.35	21.52
5	PR	14432159.77	20.15
6	SC	13472314.62	21.55
7	BA	11345094.0	27.22
8	ES	7799979.09	22.05
9	PE	4195977.72	32.87
10	MT	4177068.03	28.72

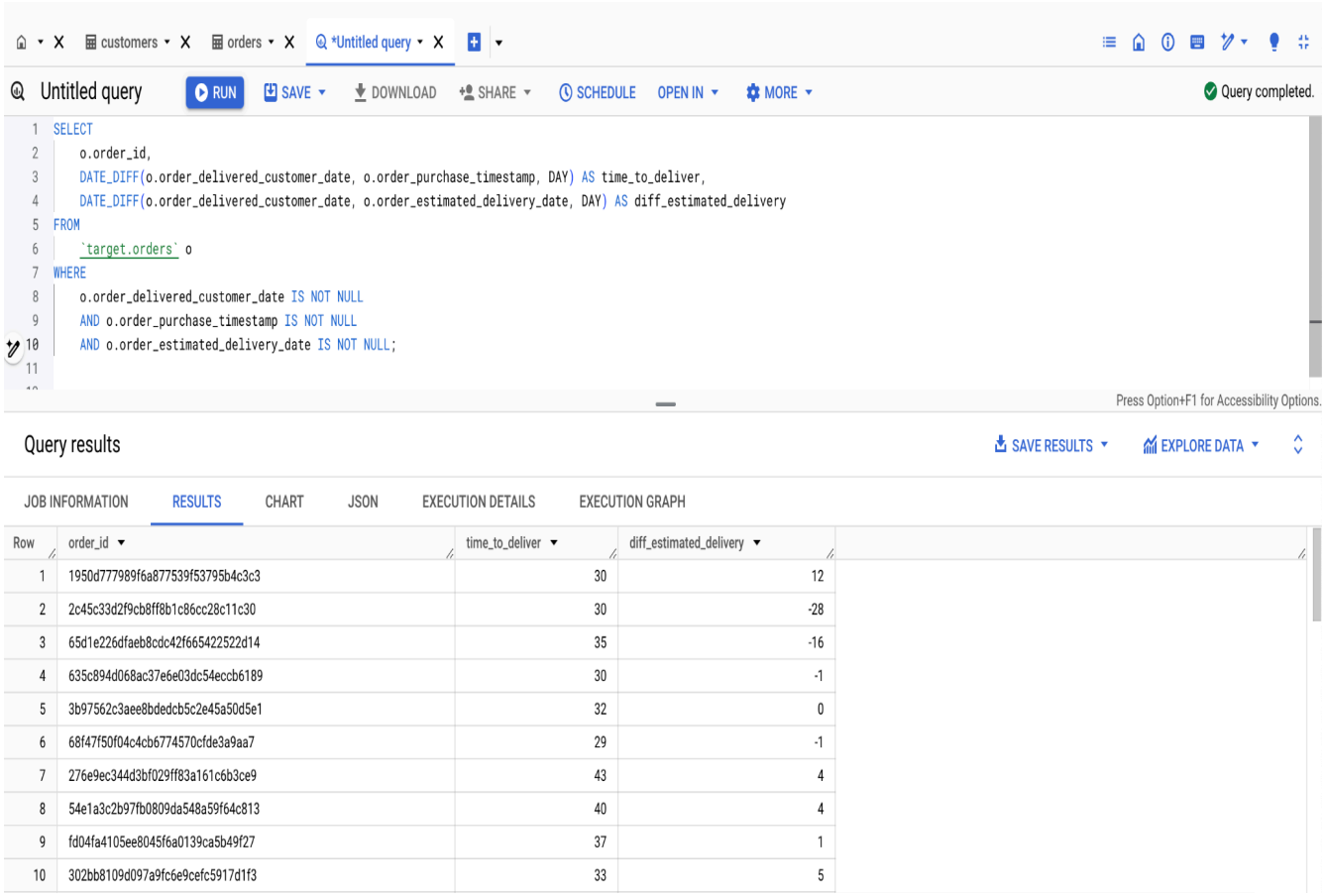
Insight

- Distrito Federal (DF) leads with the highest total freight value, followed by states like MS (Mato Grosso do Sul) and PB (Paraíba), indicating that these regions are handling a significant share of the freight.
- States such as PB (Paraíba), RR (Roraima), and AC (Acre) have the highest average freight values, suggesting that while the overall freight volume may be lower in these states, each shipment tends to be more expensive.



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

5.1 - 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.



Row	order_id	time_to_deliver	diff_estimated_delivery
1	1950d777989f6a877539f53795b4c3c3	30	12
2	2c45c33d2f9cb8ff8b1c86cc28c11c30	30	-28
3	65d1e226dfaeb8cdc42f665422522d14	35	-16
4	635c894d068ac37e6e03dc54eccb6189	30	-1
5	3b97562c3aee8bdedcb5c2e45a50d5e1	32	0
6	68f47f50f04c4cb6774570cfde3a9aa7	29	-1
7	276e9ec344d3bf029ff83a161c6b3ce9	43	4
8	54e1a3c2b97fb0809da548a59f64c813	40	4
9	fd04fa4105ee8045f6a0139ca5b49f27	37	1
10	302bb8109d097a9fc6e9cefc5917d1f3	33	5

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

Untitled query RUN SCHEDULE OPEN IN MORE SAVE DOWNLOAD SHARE Query completed.

```sql
WITH RankedFreightValues AS (
    SELECT
        g.geolocation_state,
        ROUND(AVG(oi.freight_value), 2) AS average_freight_value,
        RANK() OVER (ORDER BY AVG(oi.freight_value) DESC) AS rank_highest,
        RANK() OVER (ORDER BY AVG(oi.freight_value) ASC) AS rank_lowest
    FROM
        `target.customers` c
    JOIN
        `target.geolocation` g ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
    JOIN
        `target.orders` o ON c.customer_id = o.customer_id
    JOIN
        `target.order_items` oi ON o.order_id = oi.order_id
    GROUP BY
        g.geolocation_state
)
SELECT
    geolocation_state,
    average_freight_value,
    'Highest' AS category
FROM RankedFreightValues
WHERE rank_highest <= 5

UNION ALL

SELECT
    geolocation_state,
    average_freight_value,
    'Lowest' AS category
FROM RankedFreightValues
WHERE rank_lowest <= 5
ORDER BY category DESC, average_freight_value DESC;
```

Query results SAVE RESULTS EXPLORE DATA

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

Row	geolocation_state	average_freight_valu	category
1	DF	21.01	Lowest
2	RJ	20.9	Lowest
3	MG	20.46	Lowest
4	PR	20.15	Lowest
5	SP	15.41	Lowest
6	PB	42.77	Highest
7	RR	42.47	Highest
8	PI	39.48	Highest
9	AC	39.1	Highest
10	MA	38.08	Highest

Insight

- High Freight Values: States with the highest average freight values may have larger geographic areas, more complex logistics, or higher transportation costs, which can impact the overall freight pricing for goods.

- Low Freight Values: States with the lowest average freight values could indicate more efficient logistics, smaller geographical regions, or potentially lower demand for freight services, leading to more competitive pricing in those areas.

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

customers orders *Untitled query *Untitled query

Untitled query RUN SAVE DOWNLOAD SHARE SCHEDULE OPEN IN MORE Query completed.

```sql
WITH RankedDeliveryTimes AS (
    SELECT
        g.geolocation_state,
        ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY)), 2) AS average_delivery_time,
        RANK() OVER (ORDER BY AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY)) DESC) AS rank_highest,
        RANK() OVER (ORDER BY AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_purchase_timestamp, DAY)) ASC) AS rank_lowest
    FROM
        `target.customers` c
    JOIN
        `target.geolocation` g ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
    JOIN
        `target.orders` o ON c.customer_id = o.customer_id
    WHERE
        o.order_delivered_customer_date IS NOT NULL
    GROUP BY
        g.geolocation_state
)
SELECT
    geolocation_state,
    average_delivery_time,
    'Highest' AS category
FROM RankedDeliveryTimes
WHERE rank_highest <= 5

UNION ALL

SELECT
    geolocation_state,
    average_delivery_time,
    'Lowest' AS category
FROM RankedDeliveryTimes
WHERE rank_lowest <= 5
ORDER BY category DESC, average_delivery_time DESC;
```

Query results SAVE RESULTS EXPLORE DATA

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

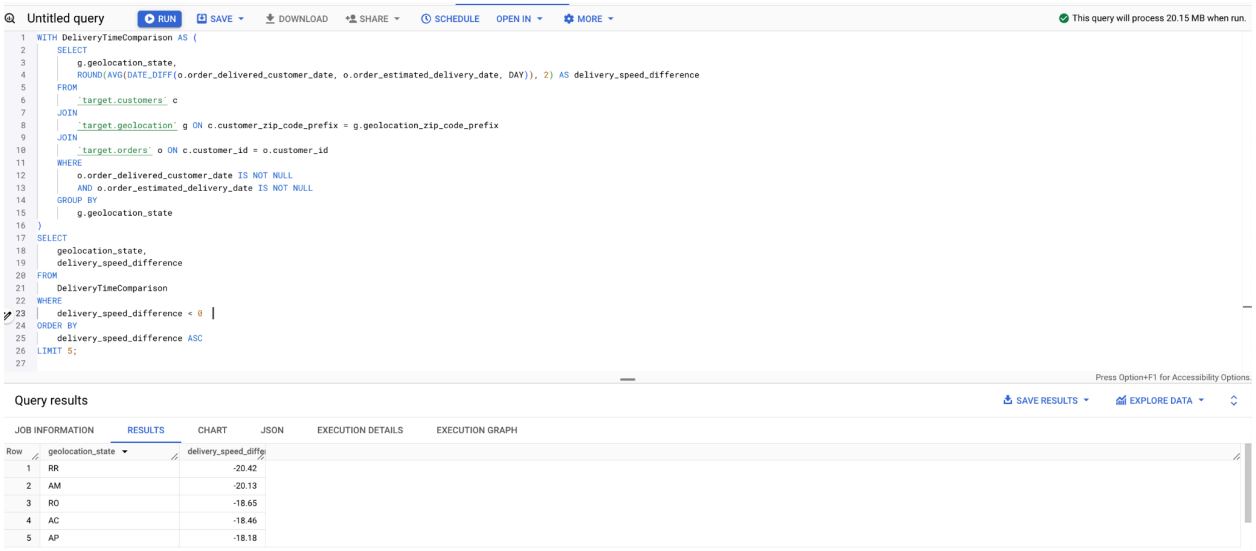
Row	geolocation_state	average_delivery_time	category
1	SC	14.49	Lowest
2	DF	12.5	Lowest
3	MG	11.42	Lowest
4	PR	11.04	Lowest
5	SP	8.47	Lowest
6	AP	27.99	Highest
7	AM	24.65	Highest
8	RR	24.52	Highest
9	AL	23.14	Highest
10	PA	22.55	Highest

Insight

- States with Fastest Delivery Times: States like SP, PR, MG, DF, and SC have the lowest average delivery times, indicating efficient logistics and quicker order fulfillment processes in these regions.

- States with Slowest Delivery Times: States like AP, AM, RR, AL, and PA have the highest average delivery times, which might be due to challenges such as geographic isolation, infrastructure limitations, or higher volumes of deliveries.

5.4 - 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.



Untitled query — RUN | SAVE | DOWNLOAD | SHARE | SCHEDULE | OPEN IN | MORE — This query will process 20.15 MB when run.

```
WITH DeliveryTimeComparison AS (
    SELECT
        g.geolocation_state,
        ROUND(AVG(DATE_DIFF(o.order_delivered_customer_date, o.order_estimated_delivery_date, DAY)), 2) AS delivery_speed_difference
    FROM
        `target.customers` c
    JOIN
        `target.geolocation` g ON c.customer_zip_code_prefix = g.geolocation_zip_code_prefix
    JOIN
        `target.orders` o ON c.customer_id = o.customer_id
    WHERE
        o.order_delivered_customer_date IS NOT NULL
        AND o.order_estimated_delivery_date IS NOT NULL
    GROUP BY
        g.geolocation_state
)
SELECT
    geolocation_state,
    delivery_speed_difference
FROM
    DeliveryTimeComparison
WHERE
    delivery_speed_difference < 0
ORDER BY
    delivery_speed_difference ASC
LIMIT 5;
```

Press Option+F1 for Accessibility Options.

Query results — SAVE RESULTS | EXPLORE DATA

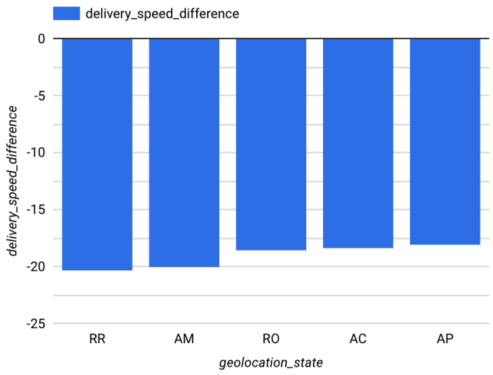
JOB INFORMATION | RESULTS | CHART | JSON | EXECUTION DETAILS | EXECUTION GRAPH

Row	geolocation_state	delivery_speed_diffe
1	RR	-20.42
2	AM	-20.13
3	RO	-18.65
4	AC	-18.46
5	AP	-18.18

Insight

- Fastest Deliveries Compared to Estimates: States like RR, AM, RO, AC, and AP have significantly negative delivery speed differences, indicating that orders were delivered much faster than the estimated delivery date. This suggests highly efficient delivery systems or favorable logistical conditions in these regions.

- Highly Efficient Logistics: These states demonstrate a strong ability to fulfill orders ahead of schedule, which may be attributed to factors such as optimized delivery networks, better infrastructure, or smaller geographic areas that allow for quicker shipping times.



6. Analysis based on the payments:

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

*Untitled query

Untitled query RUN SAVE DOWNLOAD SHARE SCHEDULE OPEN IN MORE

```
SELECT
    FORMAT_DATE('%Y-%m', o.order_purchase_timestamp) AS order_month,
    p.payment_type,
    COUNT(DISTINCT o.order_id) AS number_of_orders
FROM
    `target.orders` o
JOIN
    `target.payments` p ON o.order_id = p.order_id
WHERE
    o.order_purchase_timestamp IS NOT NULL
GROUP BY
    order_month, p.payment_type
ORDER BY
    order_month ASC, number_of_orders DESC;
```

Press Option+F1 for Accessibility Options.

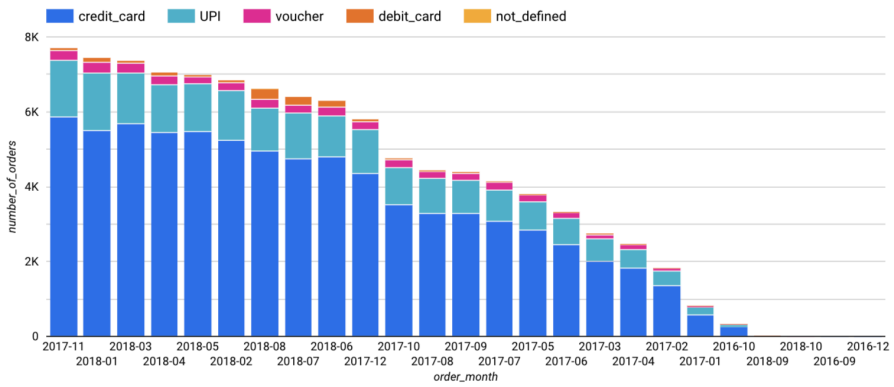
Query results SAVE RESULTS EXPLORE DATA

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

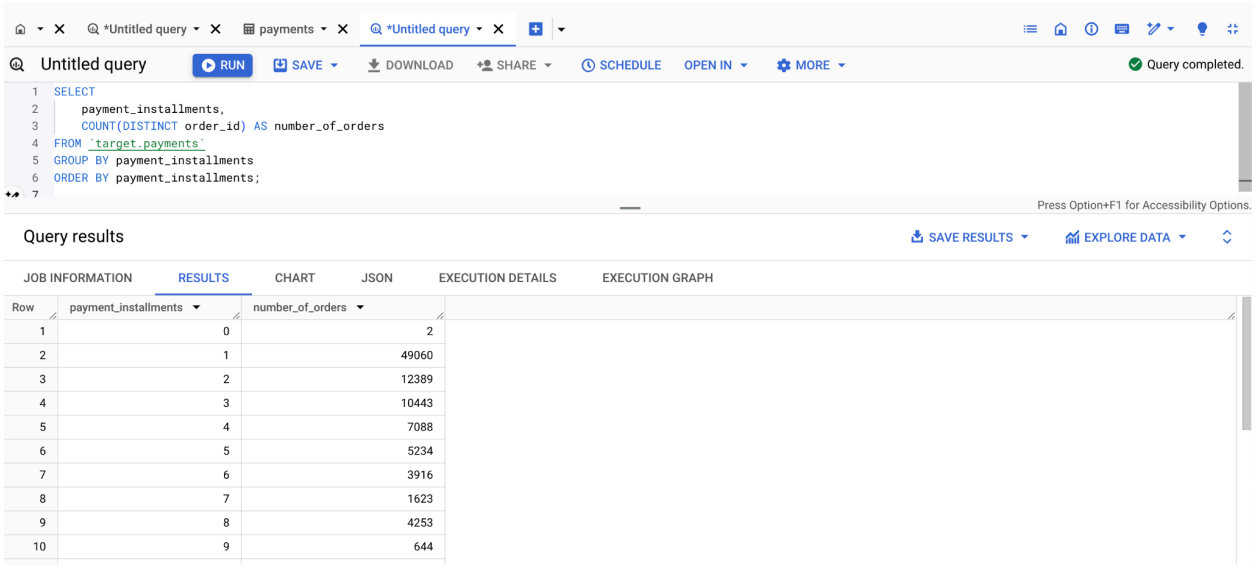
Row	order_month	payment_type	number_of_orders
1	2016-09	credit_card	3
2	2016-10	credit_card	253
3	2016-10	UPI	63
4	2016-10	voucher	11
5	2016-10	debit_card	2
6	2016-12	credit_card	1
7	2017-01	credit_card	582
8	2017-01	UPI	197
9	2017-01	voucher	33
10	2017-01	debit_card	9

Insight

- **Credit Cards Dominate Payments**: The majority of orders are consistently paid using credit cards, with a significant margin over other payment types across all months. This highlights the preference for credit card transactions among customers.
- **UPI Usage Shows Growth**: UPI (Unified Payments Interface) demonstrates a noticeable contribution, particularly in the later months, indicating an increasing shift towards digital and instant payment methods.



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



Row	payment_installments	number_of_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

Insight

- Strong Preference for Fewer Installments: The highest number of orders (49,060) were made with a single installment, showing a clear preference for straightforward, one-time payments.

- Low Adoption of High Installment Plans: Orders with more than three installments drop sharply in number, with many installment counts (e.g., 21-24 installments) having only a handful of orders, highlighting a minimal demand for extended payment plans.